18th Global Symposium for Regulators (Geneva, 2018)

New Regulatory Frontiers

Reports, Papers, and Series for Discussion during GSR-18

Reports, papers and series for discussion during GSR-18

AI for Development Series

The AI for development series was developed to deepen the understanding, and promote further discussion and collaboration, among policy makers and regulators of the significance of artificial intelligence (**AI**) and the policy and regulatory issues that are beginning to emerge from the development of AI.

The series includes four modules:

1. **The Introductory module**

   This module introduces some of the key aspects of AI and the important policy and regulatory issues that arise and that are discussed elsewhere in the AI Series.

2. **The module on Setting the Stage for AI Governance: Interfaces, Infrastructures, and Institutions for Policymakers and Regulators**

   This module examines governance strategies for AI to limit the risks arising from these innovative applications and helping to unlock their opportunities.

3. **The module on AI, ethics and society**

   This module examines the ethical and societal issues arising from AI.

4. **The module on AI, IoT and security aspects**

   This module examines the relevance of AI in the current and future development of the Internet of Things (IoT) and how security should be addressed, including in relation to data protection and privacy.

Setting the stage for 5G: Opportunities and challenges

This report has been prepared as part of the overall framework of AI series of reports to help governments, information communications and technology (ICT) regulators or national regulatory authorities (NRAs) prepare for AI and 5G digital transformation. It reviews 5G expectations and examines the infrastructure and investment requirements on the private and public sectors to prepare for 5G, to support emerging use cases and services, it provides a high-level cost model to estimate the potential capital investment and identifies key issues for consideration by policy makers and regulators to facilitate 5G deployment.

Release of the 2017 edition of the ICT Regulatory Tracker

The ICT Regulatory Tracker is an evidence-based tool to help decision-makers and regulators make sense of the rapid evolution of ICT regulation. The Tracker enables various analytical features to pinpoint the changes taking place in the CT regulatory environment. Using both quantitative and qualitative data, the Tracker makes possible benchmarking and the identification of trends in ICT legal and regulatory frameworks. It likewise helps identify the gaps in existing regulatory frameworks, making the case for further regulatory reform towards achieving a vibrant and inclusive ICT sector. The Tracker covers up to 190 ITU Member States over the period 2007 – 2017.

Artificial Intelligence (AI) for Development Series

Introductory module

July 2018

Work in progress, for discussion purposes
Comments are welcome!
Please send your comments on this paper at: gsr@itu.int by 30 July 2018

This Introductory module was prepared by Malcom Webb, M Webb Ltd, under the direction of the ITU/BDT Regulatory and Market Environment Division, in collaboration with the ITU/BDT Telecommunication Division and under close coordination with the Chief of the ITU/BDT Infrastructure, Enabling Environment, and E-Applications Department. We would like to thank the ITU General Secretariat and the ITU Standardization Bureau for their contributions.

The author would like to thank Callum Webb for his research and valuable insights that substantially contributed to this paper. .

1. The AI Series

The Telecommunication Development Bureau (**BDT**) of the International Telecommunications Union are promoting an initiative to deepen the understanding, and promote further discussion and collaboration, among policy makers and regulators of the significance of artificial intelligence (**AI**) and the policy and regulatory issues that are beginning to emerge from the development of AI.

The AI Series is a part of this initiative. The AI Series includes:

- This introductory module which introduces some of the key aspects of AI and the important policy and regulatory issues that arise and that are discussed elsewhere in the AI Series;
- A module on AI governance examining governance strategies for AI to limit the risks arising from these innovative applications and helping to unlock their opportunities;
- A module on the ethical and societal issues arising from AI; and
- A module on the relevance of AI in the current and future development of the Internet of Things (IoT) and how security should be addressed, including in relation to data protection and privacy.

2. Introduction to AI

There is no accepted definition of artificial intelligence (**AI**). Professor Nilsson, from Stanford University, describes AI, and intelligence, as follows:

> "*Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment*"[1]

At the ITU's AI for Good Summit 2017, AI was described as:

> *"… a set of associated technologies and techniques that can be used to complement traditional approaches, human intelligence and analytics and/or other techniques".*

AI comprises a broad range of computational technologies, some of which are developments of existing technologies and some brand new.

One of the themes of this AI Series is that policy makers and regulators need to increase their understanding of AI technologies and the policy implications of this technology, exchange experiences and discuss possible governance and regulatory frameworks to capture the benefits of AI and address its challenges. It will be important for policy makers and regulators to develop a cross-sectoral and interdisciplinary approach to facilitate AI.

Although the term AI has only recently come into widespread public consciousness, AI itself is not new. AI traces its roots back over 50 years. However, AI systems and their use in many fields have developed significantly in the last few years, revealing the true potential of this technology. It may still be debated whether AI is a revolutionary technology. Many experts in the field believe so. Whether AI is revolutionary, or an evolutionary technology, it shines light on a wide range of policy

[1] Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge, UK: Cambridge University Press, 2010).

and regulatory issues from a different angle and with a more intense focus than we have seen with other technologies.

As illustrated in the module on AI in society "These two factors – a growing global ubiquity and an emerging set of risks and rewards – is why AI presents such a wide array of increasingly sticky ethical and societal concerns".

AI is popularly used today in a narrow and relatively basic form, including image and voice recognition systems; Siri and Alexa; Amazon and Netflix recommendations; subtitles on over a billion YouTube videos; fraud detection by credit card companies, etc. Today, AI systems recognise images and words as well as, if not better than, most humans.

Robots and AI are different. Robots are automated, but they are not usually autonomous. Automated means being able to do physical or mental work that could have been done by a person. It generally involves repetitive tasks. Autonomous systems are designed to operate in changing circumstances without human control. They look for patterns and learn from their experience, without following a programmed set of instructions. Normally, automated systems do not use AI, but increasingly robots and other automated systems will use AI in performing manual or cognitive tasks.

In this AI Series, when we refer to AI, we do not usually mean "artificial general intelligence", which may be defined as the ability to do any intellectual task that a human is capable of. It is too early in the development of the technology to consider general AI in any substantive way.

Rather, this AI Series generally refers to "narrow AI", which are narrower applications of human-like intelligence[2].

> "*AI products tend to evolve from laughably weak to interesting but feeble, then to artificial but useful, and finally to transcendent and superior to humans*"[3]

3. Applications of AI

There are **three key things** that are propelling momentum for AI today: the availability of far greater quantities of data, increased computer processing power (particularly cloud computing) and algorithmic advances. We may also add to this list the increasing ubiquity of high speed broadband networks.

a) Current and potential applications

Current and potential applications of AI across the digital ecosystem include:

Healthcare	Education
More accurate diagnoses and treatment; personalised medicine; improved medical decision-making; forecasting health risks and improving preventative responses; virtual agents to guide patients; remote patient monitoring and consultations	Automating teacher tasks; virtual teaching assistants; automating assessments; programming assignments; personalised or customised learning; students learning at their own pace; remote teaching and assessments; personalisation at scale

[2] When chapters do discuss "artificial general intelligence", it is highlighted and dealt with specifically

[3] Garry Kasparov, *Deep Thinking: Where machine intelligence ends and human creativity begins* (John Murray, 2017)

Public services	Utilities
Better forecasting; more efficient and targeted provision of public services	Optimising management and use of utility infrastructure; better predictions of demand and supply; condition-based maintenance, rather than scheduled maintenance; increasing capital productivity
Meteorology	**Climate change**
Analysis of weather patterns; predicting adverse weather-related events	More accurate climate models
Transport	
More efficient transport systems; as well as autonomous vehicles; making public and private transport safer	

b) Sustainable Development Goals

AI is expected to be a key enabler for countries to achieve the Sustainable Development Goals.

At the ITU's AI for Good Summit 2017, the significance, and implications, of AI for developing countries was discussed:

> *"Developing countries may have the most to gain from AI, but unless we are vigilant, they may also have the most to lose. In order to reap the benefits of AI, vast amounts of data are needed, which are only available through mass digitization – an area where developing countries lag far behind. There can be no mass digitization without universal and affordable access to broadband, which is central to ITU's mission. We need to avoid a deepening of the digital divide, so the benefits of AI can be distributed equitably"*[4].

This highlights the challenges for emerging countries around digitisation, and broadband access, to the successful application of AI technologies, which we discuss further in this module and in this AI Series.

A critical point was also made at the AI for Good Summit 2017 that:

> *"… it is vital that the needs of a diverse range of people, including the most vulnerable, guide the design and development of AI systems. Those who are furthest behind in terms of social and economic development are at the centre of the SDGs and need to be at the centre of design and application of technologies such as AI".*

We discuss the potential of AI, and some of these challenges, further in this module and elsewhere in this AI Series.

4. Status of AI development and availability around the world

AI development is currently mainly concentrated in large wealthy countries or regions (in particular, the United States, China and the European Union). AI development is also concentrated in sectors

[4] ITU, AI for Good Summit 2017 report: https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf

which are early adopters in digital technologies (the high-technology sector, telecommunications, financial services, etc)[5]. A key characteristic of each of these sectors is that industry participants have access to large volumes of structured data.

According to the McKinsey Global Institute:

> *"AI investment is growing fast, dominated by digital giants such as Google and Baidu. Globally, we estimate tech giants spent $20 billion to $30 billion on AI in 2016, with 90 percent of this spent on R&D and deployment, and 10 percent on AI acquisitions. VC and PE financing, grants, and seed investments also grew rapidly, albeit from a small base, to a combined total of $6 billion to $9 billion. Machine learning, as an enabling technology, received the largest share of both internal and external investment".*

Much of the AI investment today relates to machine learning (almost 60% of investment according to McKinsey), which is an enabling technology for other AI developments. Autonomous vehicles, for example, is a relatively small investment class currently, but experts predict it is likely to emerge quickly. Autonomous vehicles is a high public profile technology and will be a bellwether of AI and its acceptance, or resistance, by the public.

Adoption of AI in health and education is growing, from a low base. Both sectors face the challenge of having to build the trust of professionals in that field, the public and regulators.

India case study: In February 2018, India's finance minister Arun Jaitley informed Parliament, during the 2018-2019 budget speech, that Niti Aayog, the premier policy think-tank for the government, will oversee a National Programme on AI, focussing on research and development of AI and its application.

The Digital India programme, the government's initiative for the promotion of AI, machine learning, and other related fields, intends to emphasise promotion of AI in 2018 and has set up four committees to encourage research related to AI. These committees are focused on researching and working on development of AI, creating a data platform, skilling, re-skilling, research and development, legal regulatory, ethical and cyber security issues. Digital India's funding nearly doubled to US$477 million for 2018-2019.

5. The technologies used in AI

a) Machine learning

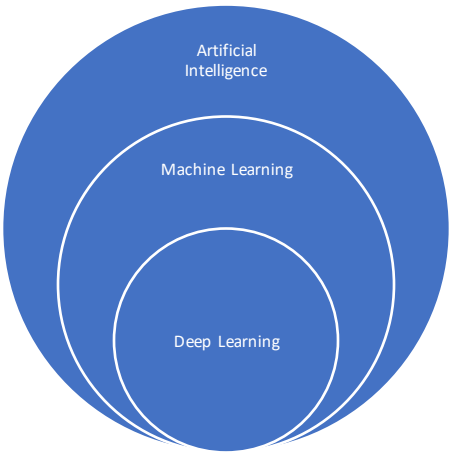
Machine learning is a subset of AI. It's what most people tend to think of when they imagine AI. Machine learning allows systems to learn directly from data, without being explicitly programmed. Structured and unstructured data provide raw material for algorithms, using training data to identify statistical rules and correlating inputs with successful outputs, learning to make predictions and recommendations. Machine learning algorithms tend to emphasise outcomes over processes. They use induction and decision-tree techniques, building context by analysing new data.

As noted in the module on AI and IoT in security aspects, "In a nutshell, machine learning is all about automatically learning a highly accurate predictive or classifier model, or finding unknown patterns in data, by leveraging learning algorithms and optimization techniques".

[5] McKinsey Global Institute

b) Deep learning

Deep learning is in turn a subset of machine learning that uses algorithms to gain more abstract insights from data. There are other forms of machine learning (such as search, symbolic and logical reasoning, etc), but deep learning is the most prominent. There have been some highly successful developments in machine learning and deep learning in recent times. Artificial neural networks are trained on enormous data sets powered by high performance computing.



c) Neural networks

Neural networks, with layers of "neurons", each receiving an input, detect patterns and then provide an input to the next layer of neurons. A neural network generally contains many layers and many neurons in each layer, with intricate webs of connections between the layers. The technology is very loosely inspired by how the human brain and nervous system works.

d) Computer vision and natural language processing

Computer vision and natural language processing are important AI technologies. Computer vision has accelerated with the developments in deep learning. Computer vision takes advantage of powerful graphics processing units (GPUs), which are used in machine learning, which can, e.g., quickly and accurately process images or video to identify objects and position.

Natural language processing enables AI systems to understand what is said or written and its meaning. The technology is now moving beyond responding to simple text enquiries into being able to engage in more complex interactions with people.

Computer vision is an important technology for healthcare and natural language processing for education.

e) Supervised and unsupervised learning

Supervised learning algorithms are trained on datasets that include labels estimated by a data scientist, indicating the importance of features within the problem. Backpropagation applies an algorithm which makes it possible for machines to predict an outcome based on input information provided. The algorithm runs many trials, learning from each trial by analysing the difference between the assigned expected outcome and the outcome reached. The algorithm then adapts its previous guess and attempts the process again. This is repeated until the algorithm has ran all its cycles (or epochs), resulting in the "actual" outcome based on the initial values given.

Unsupervised learning algorithms must determine the importance of features within the problem on its own, by analysing inherent patterns in the data. An amalgamation of these methods are semi-supervised or reinforcement learning algorithms.

The module on AI and IoT in security aspects includes a detailed description of the key technologies underlying the development in AI.

6. Investment and ICT infrastructure requirements

AI is supported by ICT infrastructure. This includes cloud-based computers with high processing powers, but also IoT networks of sensors and devices that can feed vast quantities of real-world, real-time data in to AI systems.

To support AI, ICT infrastructure will need to be flexible, very low latency, reliable, secure and adaptable to different use cases.

a) Communications infrastructure

AI will require "smarter" communications networks, which involve softwarisation, cloud infrastructure, virtualisation and more complex network structures.

a. Mobile telecommunications networks

In mobile telecommunications networks, the foundation communications infrastructure for the foreseeable future will be IMT-2020 or 5G, which is expected to be commercially widely available in the early 2020's. IMT 2020 will facilitate increasing "softwarisation" of the network – greater virtualisation and centralisation of operations (reducing cost, increasing flexibility in meeting customer and network requirements). These technologies benefit network providers by reducing their costs, but also AI users will benefit from the scalability and customisation of these technologies.

Some key IMT-2020 technologies:

- *Software defined networking (SDN)* – allows greater flexibility, agility and control in large networks; the foundation of many emerging network technologies[6] [7]
- *Network function virtualisation (NVF)* - allows operators to use commercial servers for base station hardware; decreases complexity of hardware needs
- *Network slicing* – allows operators to provide isolated sub-networks, each optimised for specific types of traffic characteristics

5G networks, with far greater capacity requirements, will require "densification" of the networks, with more base stations and access points, at both the macro and small cell layers.

[6] See the ITU's SDN portal here: https://www.itu.int/en/ITU-T/sdn/Pages/default.aspx

[7] See: Recommendation ITU-T Y.3150 "High level technical characteristics of network softwarization for IMT-2020".

5G is optimised for Internet of Things (**IoT**) capabilities, where an enormous range of devices will connect to the network. The World Economic Forum estimates there will be as many as 30 billion IoT devices in the next ten years[8].

Smart cities are a use case that goes beyond IoT, but IoT is integral to smart cities.

Although 5G will in time be the foundation communications infrastructure to support AI, networks operating 4G and possibly 3G can still provide a reliable infrastructure for some applications. Sensors, for example, that feed AI applications and are only required to communicate occasionally with small amounts of data may indeed operate over 3G or even 2G.

b. Fibre networks

Increasingly, fibre infrastructure will be necessary to support the more advanced mobile telecommunications networks. Fibre backhaul will be required to connect to the base stations and access points to provide low latency and high capacity.

Although fibre infrastructure is available in main centres of large developed countries, and also in central urban areas of many emerging countries, it remains a huge challenge for governments to facilitate the expansion of the range of fibre networks with the sort of density that will be required for advanced mobile telecommunications networks, but even more so outside of urban areas and into rural areas.

c. Investment and market structures

The investment requirements for new fibre-rich, high-speed mobile broadband networks to support the full realisation of an AI future will be considerable in most countries. In many countries, new duct or pole infrastructure will be required to push fibre deeper into the networks. Environmental and health concerns around the world may create real consenting obstacles for densified high-speed mobile networks.

This pressure may result in calls for single networks, at least at the passive layer, which will be shared by retail service providers and others providing IoT and AI applications. Whether existing market structures built around infrastructure competition will still be fit for purpose in this new era may be a valid question in some countries.

Governments and regulators will need to consider how this new infrastructure will be financed and how access will be provided. Co-investment models may be appropriate in some jurisdictions, or new infrastructure could be owned by non-traditional telecoms investors, such as infrastructure funds. Governments may be investors, supplemented by donor funds in some cases.

There are many issues for governments and regulators to consider to drive the deployment of new broadband infrastructure. AI is a use case for high speed broadband and there are many other applications for broadband networks. However, we emphasise the importance of high quality communications infrastructure as a key enabler of an AI future.

b) Cloud infrastructure

AI relies on robust cloud infrastructure to provide the computing power to run the algorithms and massive data sets that are required.

[8] https://www.accenture.com/t20170411T115809Z__w__/us-en/_acnmedia/Accenture/Conversion-Assets/WEF/PDF/Accenture-Telecommunications-Industry.pdf

Some key cloud technologies:

- *Infrastructure as a Service (IaaS)* – infrastructure elements are hosted by a third party, which may include hardware, software, storage, with associated services
- *Platform as a Service (PaaS)* – a service that allows users to develop, run and manage their own applications, using common infrastructure
- *Software as a Service (SaaS)* – a service that provides access to software over cloud infrastructure and platforms

This also requires physical data centre infrastructure to run these cloud applications. There are hundreds if not thousands of data centres around the world and they are essential for centralised cloud computing. They also consume a large amount of energy and so access to low cost and high quality electricity systems is important for their development. Indeed, as with all ICT technologies, a certain base level of electricity infrastructure will be required to realise AI's full potential.

7. Socio-economic impact of AI

AI will allow certain functions to be performed more accurately and efficiently than humans are capable of. The implications of this are wide ranging and will impact on socio-economic matters such as employment, training and the future of work. The same can be said for automation. Indeed, it is helpful to consider the socio-economic impact of both AI and automation, as they both are emerging as potent technologies and will both have wide ranging effects.

This is not a new issue. Society has been dealing with the impact of technology and mechanisation on jobs for hundreds of years. As with previous generations of technology, the growth of AI and automation is expected to adversely impact on employment in some areas, but also create new employment in other areas (e.g., data science fields).

All of this is uncertain. We may be able to anticipate jobs that are likely to be lost because of AI and automation, but we don't know when this is likely to occur. We expect it won't occur evenly around the world. Some countries, and some sectors, will be affected earlier than others. Some countries may be insulated from its effects for some time.

We also don't know what new jobs will be in demand in an AI future and whether there will be a net gain or net loss of employment.

Nevertheless, we can anticipate that many people around the world will eventually be affected to some degree by the impact of AI and automation. The broad implications of these effects may require considerations of development of social safety nets and ideas such as universal basic income.

Governments need to develop a sense for where and when the benefits and risks of AI will be experienced, how those benefits and risks are likely to be realised (broadly or narrowly) and where the opportunities are for broadly shared benefits.

a) Jobs that will be affected by AI

Over time, we can anticipate that the impact of AI and automation on employment may be profound. In one prominent 2013 study by Frey and Osborne[9], the authors estimate "*… around 47*

[9] https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

percent of total US employment is … at risk – i.e. jobs we expect could be automated relatively soon, perhaps over the next decade or two".

In emerging countries, the impact of AI on employment could be more significant, with greater proportions of low skilled workers performing manual or repetitive tasks. These jobs are potentially most at risk of being replaced by automation.

In a World Bank study[10], the authors found that two-thirds of all jobs are susceptible to automation in the developing world. However, the impact of AI and automation in emerging countries is likely to be cushioned for a period by slower technology adoption than in developed countries and lower wages. Lower wages in emerging countries may attract jobs that cannot be efficiently undertaken in developed countries impacted by AI and automation. On the other hand, the United Nations believes that the inevitable increased usage of robots in developed countries will erode the labour-cost advantage which emerging countries have enjoyed[11].

Jobs will increasingly require people to work collaboratively with AI, just as we do today with new technologies. AI and robotics will tackle manual or repetitive tasks, while humans will undertake more creative or strategic tasks, which complement the respective strengths of machines and humans.

b) Preparing people for the age of AI

It is clear that policy makers should begin the process of adapting their education and training systems to prepare their people for the age of AI.

Throughout formal education, there has been a primary focus on literacy and numeracy, which have been important skills for many jobs in today's workforce. However, recent studies show that current AI techniques are close to performing literacy and numeracy tasks at or above the proficiency of 89% of adults in OECD countries (Elliott, 2017)[12].

This suggests that policy makers should consider preparing students beyond literacy and numeracy to include training and skills in such areas as problem solving, data and statistical literacy, computational thinking and digital technology.

If the future of work is likely to involve humans working in complementary areas alongside AI, then education and training should prepare people in those complementary areas. This training will be required from an early age, through primary and secondary school and on to tertiary education.

Just as importantly, it will also be necessary to consider the needs of those already in the workforce and those of working age, who will require training in new skills. Continuous learning itself will be a core skill going forward. People need to be prepared and adaptable to meet the needs of a changing work environment.

SkillsFuture initiative case study: In January 2016, the Singapore government created the SkillsFuture initiative. This initiative provides guidance on expected areas of employment growth and training subsidies. To enable Singaporeans to take time out of full employment, a credit for workforce

[10] https://openknowledge.worldbank.org/handle/10986/23347

[11] http://unctad.org/en/PublicationsLibrary/presspb2016d6_en.pdf

[12] See http://www.keepeek.com/Digital-Asset-Management/oecd/education/computers-and-the-future-of-skill-demand_9789264284395-en#page90 pg 88-90, 96.

retraining has been given to everyone aged over 25, with further training subsidies available to those over 40 (The Economist).

If countries aspire to be AI hubs, then serious training in AI development is required at graduate and post-graduate level. Considerable investment in tertiary level capabilities and facilities will be necessary.

c) Impact on taxation revenue

On its face, if a robot or AI process displaces a human for the same job, this will not necessarily impact on the income tax revenue from that person, so long as that person is able to find another job for a similar income. However, if another job for a similar income is not available, or not available immediately to replace the displaced job, then income tax revenue for the government will diminish.

This has led policy-makers in some countries to consider how to manage any shortfall in income tax revenue that may arise as a result of robots or AI processes replacing jobs.

Robot tax case study: As part of EU-wide legislative talks on regulation of automation, a robot tax was proposed, and rejected, in February 2017. This tax would have been levied on robot owners, to pay for retraining of workers who lost their job. Robot tax is a colourful description of a tax on automation. EU Commissioner Andus Ansip described such a tax as a "tax on progress", which would result in Europe falling behind others in AI development.

On the other hand, South Korea has begun limiting tax incentives for investments in automated machines. South Korea's "robot tax" involves reducing currently available tax deductions for automation investments. While not directly taxing the employment disruption caused by robots, it is intended to provide comparable results. The reform would reduce the current deductions of three to seven percent for automation investment by up to two percent.

In 2016, the United Nations Conference on Trade and Development (UNCTAD) remarked that:

"Clearly, without the introduction of a major tax on robots as capital equipment, robot-based manufacturing cannot boost the fiscal revenues needed to finance both social transfers, to support workers made redundant by robots, and minimum wages, to stem a decline in the living standards of low-skilled and medium-skilled workers."[13]

d) Safety nets

Universal Basic Income (**UBI**) has been proposed by some experts as a solution to address the social consequences of the expected displacement of jobs by automation (and AI). Under UBI, all citizens would receive a reasonable amount of money to ensure at least a minimum standard of living. Top economists, such as the chief economic advisor to the Government of India, Arvind Subramanian, and economics Nobel prize winner Sir Chris Pissarides, among others, have shown their support for UBI.

In emerging countries such as India, Subramanian argues that a safety net from UBI would support people impacted by poverty due to droughts, declining agricultural opportunities etc. In a similar light, a decline in manual or repetitive tasks due to automation may also require a safety net to

[13] http://unctad.org/en/PublicationsLibrary/presspb2016d6_en.pdf p.3

catch those whose work becomes redundant. Sir Chris Pissarides advocates for UBI as a solution to inequality, which may be expected to rise because of automation.

The idea of a universal basic income has existed since the industrial revolution. In 1849, John Stuart Mill famously proposed that a "certain minimum" should be assigned by the government for the subsistence of every member of the community, whether capable or not of labour.

e) Other policy proposals

The ITU paper on the social and economic impact of digital transformation on the economy for GSR17[14] examined many of these issues in depth. A number of proposals are put forward for policies aimed at promoting innovation in advanced technologies while mitigating workforce disruption in developed economies, including:

- "*Increase public expenditures in education to increase the skills (including digital skills) acquired through formal training;*
- *Implement labor policies focused on workers being able to retain their current jobs or move to new areas of demand (job placement services, special labor market programs, apprenticeship programs);*
- *Put in place subsidies to lessen job disruption of low-skilled workers (tuition-free education, temporary cut in payroll taxes, basic income guarantees);*
- *Implement policies aimed at increasing geographic mobility (reduction of relocation costs, subsidized housing; and*
- *Promote demand for skilled workers by accelerating the rate of innovation in areas likely to be affected by job disruption effects*".

8. Significance of a strong foundation in data

AI requires a strong foundation in data. Access to data is needed to train AI systems, to allow them to identify patterns, which in turn enables those systems to make predictions and recommendations. In comparison to data, computing power has almost become a commodity and so perhaps less important to the development of AI as access to data.

a) Open data and open standards

Open data and open standards for public data are likely to be an important enabler of AI in many countries. Open data improves the quality of public services, through learnings from the data made available and providing new insights, which will also be valuable for AI. Open standards will assist AI systems in making sense of the complexity of data.

Governments can promote open standards to build a robust data ecosystem in their country, particularly for public data, making systems and data interoperable. These include common standards for metadata, which will allow the provenance of data to be traced as data is used and reused for different purposes[15].

[14] https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2017/Soc_Eco_impact_Digital_transformation_finalGSR.pdf

[15] For further discussion about open data in developing countries, see Verhulst and Young, "Open Data in Developing Economies: towards building an evidence base on what works and how", July 2017

This may be one of the most significant steps that Governments can take to prepare themselves for the age of AI. A quality government data environment, with open standards, will be foundational to maximise the achievement of the Sustainable Development Goals through the use of AI. Conversely, a poor government data environment, with inconsistent and incoherent standards, will impede the potential for AI.

Governments can play an important role in developing and adopting effective anonymisation or de-identification techniques which can be implemented through these open standards, providing an appropriate balance between re-identification risks and the public benefit in using this information. We discuss anonymisation or de-identification techniques further below.

The Open Data Charter is a collaborative effort between data experts and over 70 governments operating with the objective of opening up public data. The Open Data Barometer (ODB) is an initiative of the World Wide Web Foundation[16]. In its most recent survey (2016), they found that 79 of 115 countries studied had operational open data initiatives.

Open data charter case studies: The principles of the Open Data Charter are in summary: open by default, timely and comprehensive, accessible and usable, comparable and interoperable, for improved governance and citizen engagement and for inclusive development and innovation[17].

The G8 have an open data charter[18]. Under the open data charter, all government data is expected to be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is published.

b) International data exchange

Governments can promote the international and regional exchange of data and global collaborative efforts. Medical research is an area where there are expected to be particular benefits from international exchange of appropriately anonymised clinical data.

The European Parliament is currently discussing the final stages of a bill to allow free flow of non-personal data between EU countries. A draft bill allowing nearly unrestricted flow of non-personal data (aside from where there may be concerns for public security) was passed in the Council of the EU on 20 December 2017. The legislative work is expected to be finished by June 2018.

The OECD Privacy Framework encourages transborder flows of personal data between countries where safeguards and effective enforcement exists consistent with the OECD guidelines. Any restrictions which are made to transborder data flows should be proportionate to the risks presented. The OECD framework intends to avoid the creation of unjustified obstacles to economic and social development. They use the example of excessive protection of personal data, exceeding the requirements necessary. The OECD recognises the benefits towards efficiency gains and convenience of increased transborder flows of personal data, however they are concerned with respect to the protection and enforcement of privacy.

c) Data for public good

The public good applications of AI are likely to be considerable for all countries. However, the public good applications in emerging countries may be even more significant than in developed countries,

[16] https://opendatabarometer.org
[17] See here for more details: https://opendatacharter.net/principles/
[18] https://opendatacharter.net/resource/g8-open-data-charter/

where commercial applications of AI will likely, at least initially, receive most of the investment and attention of the major players in the field.

However, much of the data that will be valuable from a public good perspective will be either personal data or commercial and proprietary data. This creates a tension between public good on the one hand and personal privacy and commercial strategic value on the other hand.

Personal data, which is provided and may only be used for certain purposes, may have substantial public benefits if it was accessible, for example for research purposes or for providing improved public services on an anonymised or de-identified basis using AI technologies.

Healthcare data case study: More health-related data is being collected than ever before, including by mobile apps, Fitbit, etc. Access to anonymised patient data may be highly beneficial for medical professionals and researchers.

Governments can develop rules for who can access this sensitive data, what it can be used for and how it is stored, protected from cyber risks and how it should be anonymised or de-identified.

Consideration will need to be given to providing incentives and mechanisms to share health data for these public benefit purposes. The same likely applies for education data (information about student performance, etc).

A different set of issues arises with commercial or proprietary data. Some commercial or proprietary data will be derived from personal data provided in return for digital services. Other commercial or proprietary data, with potential public good applications, will be developed for the purposes of providing commercial services (such as mapping data collected by various companies).

There is likely to be strategic value in that data which weighs against use outside of the business concerned. Costs and legal risk will also be a consideration. For example, there may be costs associated with the anonymisation or de-identification of any personal information, which would not have needed to be incurred if that information was not released. There may be legal risks for the business, particularly around re-identification or third-party confidentiality rights.

Nevertheless, public benefit may be realised in accessing appropriately protected, or aggregated, commercial or proprietary data for new AI based public services. Questions of incentives, and safeguards (e.g., protections from liability), for holders of proprietary data to share that data for public services need to be considered.

These are critical issues for governments preparing for the AI age. Where the public benefit from AI technologies and its reliance on high quality data is growing, these tensions between personal privacy and commercial considerations need to be deliberated by governments, and the public and the major holders of personal data. While these issues are present in the current era, they are likely to become more prominent in the AI age.

9. Ethical, legal and regulatory issues

a) Personal data

Data protection laws protect personal data, which is information about an individual. Access to personal data, and its protection, will be critical to the future evolution of AI. AI will not succeed if people lose confidence in the ability of AI to protect their personal data.

The European Union has a comprehensive regulatory framework for the protection of personal data. The approach in the European Union is to treat personal data as information about an identified or identifiable individual. This has been broadly followed in the OECD guidelines and the Privacy Framework of the Asia-Pacific Economic Cooperation[19], among others. By contrast, for example, the United States has pursued more of a sector-specific approach for personal data.

Privacy laws do not apply to non-personal data (information that does not relate to an identified or identifiable person), or to data where the person's identity has been sufficiently anonymised or de-identified[20].

Existing data protection laws were usually established at a time of limited collection and limited usage of personal data. What has changed in the intervening period is that more data is now collected about individuals, in new ways, at far greater scale. Personal data is used (and re-used) for a much wider range of purposes than ever before, often far beyond the original purpose. An increasing number of entities are involved in the collection and in the processing of data, often without explicit knowledge of the individual. There is very limited public awareness of these activities.

Collection case study: Data is being collected through sensors, social networks, vehicles, etc. It is captured as a by-product of interaction with devices, services, etc. Data is collected directly, e.g. through use of device, and indirectly, e.g., through sensors, Wi-Fi hotspots, or just being in places, including in the home. Data may still be private, even if it is captured in public places.

Big data and analytics allows for greater insights to be obtained from collected data, beyond what had been the original purpose of collection. Sometimes those insights may be apparent much later than the time of collection.

The EU approach, and in many other countries, is to protect information that relates to an "identified" individual, but also information that relates to a person that is "identifiable" (that is, they could be identified). In the EU[21], there is a test of reasonable likelihood of identification, but the test is dynamic, in that information may not be "identifiable" at the time of collection, but it may become identifiable as a reasonable likelihood through the progress of technology change.

De-identification case study: The anonymisation or de-identification of data can remove immediate privacy concerns. However, developments in advanced analytics over recent years has meant that

[19] OECD guidelines define personal data as "any information relating to an identified or identifiable individual (data subject)". The Privacy Framework of the Asia-Pacific Economic Cooperation 2004 defines PII as "any information about an identified or identifiable individual."

[20] e.g., in Recital 26 of the GDPR, personal data that is "rendered anonymous in such a manner that the data subject is not or no longer identifiable" is excluded

[21] Recital 29, GDPR

personal data may increasingly be inferred from de-identified data. Professor Paul Ohm has highlighted that re-identification risks, arising out of modern analytics technologies, render data increasingly identifiable[22]:

"Easy reidentification represents a sea change not only in technology but in our understanding of privacy. It undermines decades of assumptions about robust anonymization, assumptions that have charted the course for business relationships, individual choices, and government regulations."

These developments, and these concerns, have in turn propelled the search for advanced new technologies that can substantially reduce re-identification risks. These new technologies include differential privacy and homomorphic encryption.

Differential privacy is a method of data collection which applies random noise to the dataset on collection, where an individual's true information is distorted and will not be recognisable in the dataset.

Homomorphic encryption allows for the computational use of encrypted data, without knowledge of the true (decrypted) data. By never needing to decrypt the data, the privacy of users is uncompromised during the computational process. Although homomorphic encryption has existed as an idea for nearly 40 years, full homomorphic encryption is not expected to be usable for several decades due to the intensive computing power required.

The tensions between data protection and the realisation of IoT will create new grey areas with space to circumvent legislative boundaries. These are addressed in the module on AI and IoT in security aspects. The module also examines how data protection may be threatened in an IoT environment and further discusses AI-based privacy enhancing techniques and mechanisms in greater detail.

Some experts have called into question whether the traditional data protection law models are suitable in the AI and big data age, where information is increasingly identifiable. Some academics argue that the distinction between "identified" and "identifiable" information is becoming meaningless[23]:

Koop argues that[24]:

> *"Current data protection law … might be considerably more productive if, instead of trying fitfully to establish where the border lies between personal and non-personal data, we would allow for categories of data that have certain effects on people when they are processed, regardless of whether or not they relate to identifiable individuals."*[25]

This brings up the issue of context. Attitudes of the public to data protection (the benefits and the risks) depend on the context. For example, people often disclose personal information to receive digital services and feel no strong need to protect themselves. But that same information, used in another context, or when combined with other data, may be very concerning for the individual. Therefore, another possible approach to data protection law is to be more context-specific and

[22] Ohm, P. (2010) "Broken Promises of Privacy," 57 UCLA L. REV. 1701 (2010)

[23] For example, see Schwartz, P. and Solove, D. (2011) "The PII Problem: Privacy and a New Concept of Personally Identifiable Information" 86 N.Y.U. L. Rev. 1814

[24] BJ Koops, 'The trouble with European data protection law' International Data Privacy Law, Volume 4, Issue 4, 1 November 2014

[25] http://www.isaca.org/Groups/Professional-English/privacy-data-protection/GroupDocuments/2014-08-24%20%20The%20Trouble%20with%20European%20Data%20Protection%20Law.pdf

provide different levels of protection of data depending on the sensitivity and proposed use of the data.

These issues extend far beyond AI, but have a relevance in the AI context due to the use of personal data, or anonymised or de-identified data, in AI processes.

b) Notice and consent

While most countries, in most circumstances, require notification of the individual of the purpose of collection, use and disclosure of their personal data, there are different approaches taken in relation to obtaining consent. Some jurisdictions (e.g., the EU countries) adopt a notice and consent approach, while others are notification based, with consent required in limited circumstances.

People have often "consented" to the collection of their data, e.g., to receive the benefit of a digital service. Individuals are presented with detailed terms and conditions, which few are likely to have read or understood before accepting. The purpose of collection is often broadly described and individuals in many cases have little knowledge of information that is collected about them and what it's used for.

And things can change. The use of a service may vary from the time of initial notification and consent. With AI and big data, new insights can be gained from old data, including through combining data, and questions arise whether the original notification or consent was sufficient.

The solution to these difficulties is likely to involve greater transparency and public awareness of the benefits, and the risks, of intensive data usage.

Some experts have argued that data protection law should recognise the trade-offs involved when public good may be derived from personal data. Tene and Polonetsky suggest that: "*Where prospective data uses are highly beneficial and privacy risks minimal, the legitimacy of processing should be assumed even if individuals decline (or are not asked) to consent*".

Case study Singapore data consultation: The Personal Data Protection Commission of Singapore (PDPC) recently conducted a public consultation on approaches to managing personal data in the digital economy[26]. Singapore's data protection legislation primarily provides for consent as the basis for collection, use and disclosure of personal data. The PDPC noted that, in today's analytics-driven world, it may not always be possible to anticipate the purposes for use and disclosure at the outset. Also, it may not be possible always to obtain consent from individuals when their data is collected or attempt to identify the individuals to seek their consent for every new purpose.

The PDPC proposed an approach where notifying individuals of the purpose can be appropriate, where there is no foreseeable adverse impact on the individuals arising out of the collection, etc, of the personal data. The PDPC also proposed that there be a "legitimate interest" in collection, etc, of personal data without consent. This is a more limited ground, intending to apply in situations such as prevention of fraud.

[26] https://www.pdpc.gov.sg/Legislation-and-Guidelines/Public-Consultations#ACTR1

c) Bias or fairness issues

AI systems should be developed to ensure the equal and fair treatment of people that are affected by decisions made by that system.

This issue can arise as a result of bias inherent in the data on which the algorithms are trained (which may be derived from human biases at the time of collection). For example, AI systems may have been trained on limited data with under-representation of certain demographics, or systems which are selectively used for marginalised populations.

As suggested in the module on AI, Ethics and Society, "the risks for bias in AI is probably greater due to the qualities of its datasets than for any "hand coded" biases of its algorithms".

Heat map case study: Jessica Saunders et al.,[27] illustrate the results of bias through police "heat maps", which attempt to predict where best to patrol. Through increased patrolling, more criminals are caught in those areas, leading to the system being trained to increase patrolling further. Saunders et al., discovered that the use of "heat maps" by police has led to disproportionate harassment of African Americans.

The lack of diversity of those involved in AI research (a "sea of dudes" (Mitchell) and a "white guy problem" (Crawford, 2016)) is another issue. This lack of diversity may in turn create certain types of biases in AI systems, created through the lens of white male AI developers.

At a practical and technical level, it is very difficult for developers to ensure data or algorithms are free from bias.

But AI may also be the solution to this problem. AI systems are likely to produce more impartial results than humans as they are not susceptible to conscious or unconscious biases if they are designed properly. They can be used to detect and eliminate biases.

This issue was recognised in the Korea Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society[28]: "*As the massive quantities of data involved and high complexity of AI algorithms will make it nearly impossible for humans to rid these systems of biases once they begin operating and evolving, policymakers may well need to develop and establish refined methods for applying and testing ethical standards for their development at every stage (e.g., requirements for testing the fairness and reliability of data, enforcing the fiduciary duty of developers, preventing reverse choices, etc.)*"

d) Interpretation and transparency

AI systems today are rarely set up to be transparent and provide reasons for a decision that it makes. As a result, AI systems can be difficult to interpret. However, in certain circumstances (where the outputs are consequential for people, e.g., granting a mortgage, insuring a home, etc.), reasons may need to be provided for an AI decision.

[27] Jessica Saunders et al., Predictions Put into Practice: A Quasi-Experimental Evaluation of Chicago's Predictive Policing Pilot, 12 J. EXPERIMENTAL CRIMINOLOGY, 347, 350-51 (2016).

[28] http://www.msip.go.kr/dynamic/file/afieldfile/msse56/1352869/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf

In part, this relates to the bias and fairness issues discussed above. Where decisions are consequential, and where bias or fairness (or simply errors) can be a concern, then issues of interpretation and transparency become increasingly important.

Human decisions, of course, are also not necessarily interpretable or transparent. AI systems are more easily audited than humans.

European law makers have introduced a "right to explanation" in the GDPR, which requires "meaningful information about the logic involved"[29]. Questions arise over what an explanation is and whether disclosure of the program is sufficient. There is also a right for the person concerned "… not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her"[30].

Black box case study: There are inherent difficulties with transparency and interpretation of "black box" deep learning or neural networks. These systems are high performance, but they are also opaque and less transparent. Concerns have been raised by AI engineers over whether transparency requirements may compromise performance.

Specialised tests may be required that look for bad outcomes. Artificial inputs could be used to test for unusual situations that can produce unexpected outcomes. Some details of the system's design may be published, enabling analysis, without revealing proprietary or private information.

Despite the difficulties in "black box" systems, transparency is likely to be essential in building public trust in AI, at least in those circumstances where outputs are consequential for people. Failure to build and maintain public trust will likely lead to underuse of this important technology.

Ethical guidelines will be required. In some consequential areas, people will expect the right to understand the decision-making process, etc. In certain highly sensitive areas, there may be a need to limit development of AI to areas where human explanation is possible.

The AI, Ethics and Society module explains in greater depth the issues involved in interpretability and transparency in AI systems. Dr Best refers to the Statement on Algorithmic Transparency and Accountability released by the ACM US Public Policy Council (2017), that advises transparency of data used to train AI systems, as well as explainability of their decisions. The statement also suggests auditing systems in case of harm, redressing groups adversely affected by algorithms, and holding accountable the entity producing the algorithm.

e) Accountability and liability

What happens if AI goes wrong? Accidents and even crimes can happen due to AI decision making.

AI is in many respects no different to other technologies. The designers and manufacturers of AI systems, and the users of those systems, are potentially accountable and liable for how those systems operate, depending on the civil or criminal law of the country concerned.

[29] Articles 13 and 14, GDPR
[30] Article 22, GDPR

In most countries, a victim of harm can sue the wrongdoer under civil law rules for negligence, failure of statutory duty, etc. But who is liable where it is the AI system that is doing the wrong?[31]

While there is discussion in academic circles of the concept of legal personhood for AI systems, it is currently too abstract to be given serious practical consideration. So, the AI system itself, lacking legal personhood, would not be liable for the harm that it causes.

We can imagine the owner or user of the AI system to be potentially liable, notwithstanding that they did not cause the harm. Owners and users of technology commonly have legal responsibility, when the harm is caused by technology that they control (say, industrial machinery). This may be under tort laws or strict liability laws.

What's different with AI is that, being to some degree "intelligent", it operates autonomously and potentially in ways that are not expected by the owner or user of the AI system. To the extent that some fault is required on the part of the owner or user, then it may be difficult to prove this with autonomous systems.

Another option for a victim of harm is legal action against the manufacturer of the AI system, under product liability laws in some countries (usually without having to prove fault by the manufacturer), or under civil law where some fault would have to be demonstrated.

Governments need to consider whether liability for AI systems should be based on fault (like negligence) or strict liability (where no fault needs to be shown). Should AI systems be treated like domestic animals, unpredictable, but where public policy approaches to risk allocation make the owner of the animal liable for its actions? The answer may be different for different types of systems. And if the owner or user of an AI system is strictly liable, then they would need to claim against the manufacturer when the issue arose out of a defect, which may not be straightforward.

This also gives rise to the question of whether compulsory insurance should be acquired by an owner or user of AI, where they are strictly liable, such as occurs with vehicles in a number of countries. Compulsory insurance means that the victim can claim against the insurer.

Compulsory car insurance case study: The UK government has proposed a system where compulsory car insurance will be required to provide cover for motorists when they hand over control to an autonomous vehicle. Motorists, or their insurers, will then rely on existing rules of product liability and negligence to ascertain who's responsible.

In the criminal law domain, there may be a question over whether a person intends to commit a crime when it is committed by an AI-enabled machine that the person owns or uses.

f) Appropriate standards

If regulation is to be applied to AI, what standards should algorithms be required to meet? This is a new and evolving area and policy makers in some countries have begun to consider regulation in the context of autonomous vehicles.

[31] For further discussion on these issues, see European Commission staff working document "Liability for emerging digital technologies" SWD(2018) 137 24 April 2018; also, Petit, "Law and Regulation of Artificial Intelligence and Robots: Conceptual Framework and Normative Implications" 9 March 2017 and the European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))

Autonomous vehicles case study: The Germany Ethics Commission published a report on automated driving guidelines for the programming of automated driving systems in August 2017. The report consisted of 20 proposals, such as that automated driving is an ethical imperative if the systems cause fewer accidents than human drivers, in every driving situation it must be clearly regulated and apparent who is responsible for the driving task – the human or the computer.[32]

During a recent hearing before the United States Congress subcommittee on Digital Commerce & Consumer Protection in the United States[33], the chairman of the committee proposed that AI, such as autonomous vehicles, should be implemented under the condition that they are safer drivers than humans.

Various questions arise when considering the extent of safety required before acceptance of an autonomous system being "safer than humans". For example, is an autonomous vehicle expected to have fewer accidents on average than a human, or is it expected to outperform a human with access to the best safety features currently available? How much safer than humans do we expect an autonomous vehicle to be before they are accepted by policymakers? Further, who undertakes the certification process, testing the safety levels before public implementation, and to what extent?

In many respects, the high-profile area of autonomous vehicles will be the bellwether for these sorts of issues going forward.

g) Verification and validation

For critical systems, AI companies will be expected, or required, to be able to verify whether the technology is operating as intended under actual operating conditions, with no unwanted or unpredictable behaviours. This will require manufacturers to prove, test, measure and evaluate systems before they are deployed.

However, AI machine behaviour can change as algorithms evolve. This creates complications when it comes to verification. How long would a verification be expected to be effective for, before needing to be re-verified? Traditional software verification may not be adequate. In safety critical systems and infrastructure, like planes and bridges, there are robust and accepted processes for addressing verification and validation to ensure safety and reliability. Manufacturers of AI systems will need to address how to manage the risk and building a safety case for the technology.

h) Security threats

AI systems will give rise to cyber-security threats. Hackers will look to access AI machines or datasets used by machines or IoT sensor networks in ways that may negatively impact on AI behaviour.

While AI presents another attack vector for cyber-criminals, AI can also be used to improve cybersecurity by anticipating attacks, identifying vulnerabilities and taking steps to prevent attacks.

The main subject matter of the module on AI and IoT in security aspects is on cyber-risks in IoT environments. It describes in detail the overall features and technical issues that arise in these

[32] See https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

[33] Self-Driving Vehicle Legislation: Hearing Before the Subcomm. on Digital Commerce & Consumer Prot. of the H. Comm. on Energy & Commerce, 115th Cong. (2017) (opening statement of Representative Greg Walden, Chairman, Subcommittee on Digital Commerce and Consumer Protection).

environments. A framework for the adoption of AI and how it can be used to enforce the security of IoT devices and networks is discussed.

The point is also made that "In security, [available] of big data means AI techniques can be exploited to analyse and recognize patterns of security vulnerabilities to prevent such attacks. Thus, the ability of IoT based platform to learn from data to analyse, identify and mitigate security threats is an important feature that every IoT system should incorporate".

i) *Market structure issues*

A relatively small number of firms in the private sector are currently at the forefront of pioneering AI development and they are deepening their expertise.

Research facilities within private sector organisations are moving towards becoming larger than universities or public facilities in AI, attracting leading practitioners in the AI industry. These firms also possess enormous troves of data, gained as a result of the digital services that they provide. It is difficult for smaller firms to compete in the market, given the concentration of in-depth analysis within these companies and their access to massive data sets.

There may be questions whether this leads to market power, including consideration of issues over barriers to entry in AI-related markets. For example, the data, although deep, may not be unique to that firm, which would reduce barriers to entry. These issues may arise in a mergers and acquisitions context, where one of these firms seeks to acquire another firm with AI capabilities.

More broadly, stresses are likely to emerge from a situation where the key inputs to this important technology (data, algorithms, know-how and IP) are held by the private sector, often outside the jurisdiction concerned, but where the public good benefits for a country are so great.

We expect this is an area where ICT regulators, and competition authorities, will need to examine closely in years to come.

10. Institutional framework and cross-sectoral and interdisciplinary approaches

a) Establishment of an oversight body

In most countries, there is a case for a government body or committee to be responsible for oversight over AI activities. It would not be premature to create such a body now.

This government body or committee may be newly established or it may be an existing body or committee, or indeed an existing regulator or government department, that perhaps has oversight over emerging technologies and their implications for policy-making. It may include people from outside of government, including academics, people from industry, consumer representatives and so on.

However it is constituted, an oversight body would be charged with providing advice to government more broadly. Its tasks may include:

- promoting public knowledge and meaningful public dialogue about AI and its benefits;

- research and analysis of regulatory and policy issues, as well as future technological developments;
- providing support for, and coordination with, sector-specific regulators;
- establishing standards, codes, ethical guidelines reflecting community values; and
- coordinating with other similar bodies internationally.

We see this oversight body having recommendatory powers, rather than enforcement powers.

International case studies: The Advisory Board on Artificial Intelligence and Human Society[34] was established in May 2016 under the Japanese Minister of State for Science and Technology Policy to advance research and development and use of AI technologies.

The French Digital Council[35] was established as an independent advisory commission that issues independent opinions and recommendations on questions relating to the impact of digital technologies on the economy and society and consults on new legislation or draft regulation.

Similarly, the UK Parliament has recommended a standing Commission on Artificial Intelligence.

Governments generally will need to up-skill in AI, to understand its policy implications. AI technical and policy capability should in due course be spread throughout government, providing more diverse perspectives on AI technology within the public sector.

The authors of the module on AI governance, discuss the importance of reducing information asymmetries in government when it comes to AI. As well as building internal capacity within government, they propose various ideas for government to interact with experts in the private sector, including through "tours of duty" and positions that operate outside of traditional bureaucratic structures. They suggest establishing ongoing interfaces with experts, that can supplement or replace the need to hire experts.

b) Sector-specific policy

Sector-specific regulators are likely to lead policy developments in their respective areas. For example, issues around enabling infrastructure would appropriately be dealt with by ICT regulators and ministries, issues around transport by transport regulators, medical applications by health authorities, financial markets by financial regulators, consumer protection by consumer protection authorities, etc.

c) Cross-sectoral policy

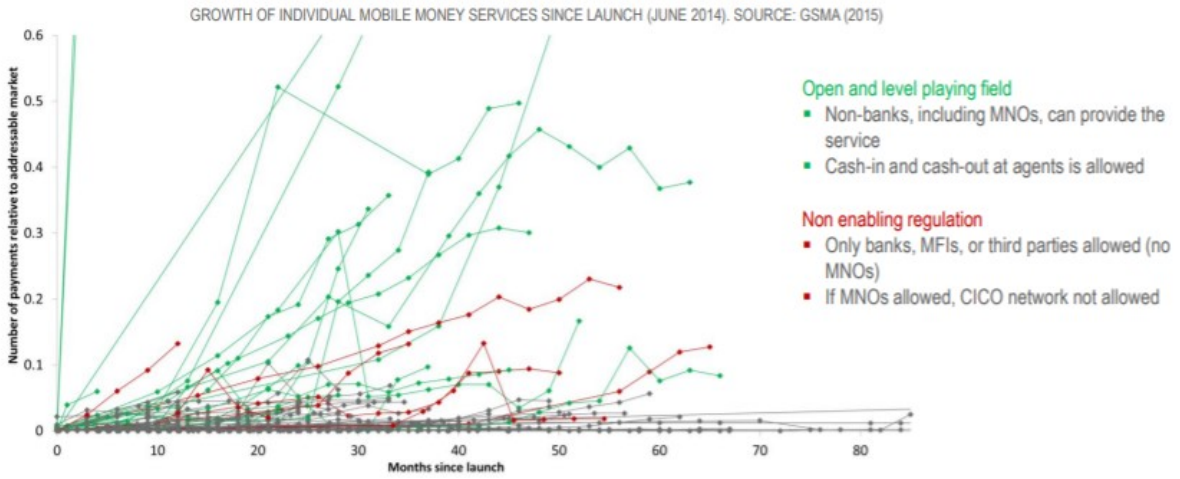
However, we can also envisage that a new technology, such as AI, will require increasing cross-sectoral approaches, which will require collaboration between different sectoral regulators. The ITU has emphasised the benefits of collaborative "G5" regulation, with the need to define the

[34] http://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_en.pdf
[35] https://cnnumerique.fr/en/french-digital-council/

foundation, platforms and mechanisms for working with other sector regulators to help achieve the Sustainable Development Goals[36]. This will also be pertinent in the context of AI policy.

Models for addressing cross-sectoral issues case study: Although not an example directly related to AI, the field of mobile money has required cross-sectoral regulatory approaches, in this context through the financial sector regulators alongside ICT regulators. The financial sector regulators tend to focus on increasing competition and efficiency, while the ICT regulators tend to focus on providing broad policy guidance on data protection, consumer protection etc.

Mobile money is also an example where, by implementing enabling regulation, growth and market penetration increased much faster relative to non-enabling regulation.



In some areas, governments may promote the sharing of incident and safety data related to AI among different sectoral regulators, such as what occurs with civil aviation with incident or near miss data.

d) Multistakeholder governance generally

One of the issues discussed in the module on AI governance is building effective multistakeholder governance groups. They propose a set of principles to guide the establishment of these groups and a range of tools that policy makers and regulators can deploy to engage with diverse stakeholders in advancing AI governance.

e) Data protection regulation

Because of the importance of data, and personal data, to the emergence of AI, the regulator with responsible for maintaining data protection laws will play a prominent role.

This may be an area of responsibility for the ICT regulator, or the privacy or data protection regulator (if a general data protection regulator has been established).

[36] See, for example: https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/Events/2016/Regulatory%20Conference/ITU_RegulatoryTrends%20Sept%202016_J_Ponder.pdf and https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/Events/2017/Regulatory%20Conference/Session%202%20Rosheen%20Collaborative%20Regulation_MontenegroITU.pdf

Whichever authority has responsibility for data protection, it is clear that the privacy implications of AI will be a critical area of focus for that regulator going forward. They will need the appropriate resource and powers to undertake this role.

f) Exploratory regulatory approaches

While we consider it is premature to implement specific AI regulation, we see merit in beginning to put in place structures and methodologies for exploring the potential regulatory implications of AI. These may include regulatory "sandboxes".

Sandboxes case study: Regulatory sandboxes allow for the piloting of new AI technologies in safe environments. The objective would be to promote innovative investment in AI for local application, starting with a contained, low risk, rule that permits something that would otherwise have been limited or prevented by regulation. This allows developers and regulators to observe, experiment, test and adapt further from there.

In the AI governance module, it is suggested that "*policymakers and regulators can create spaces that allow them to experiment in an iterative fashion with policies and regulatory approaches, that still allow for the development of new AI technologies, while still advancing core values of public safety, privacy, consumer protection, and due process*".

11. A roadmap for regulators

There will clearly be a need for policies and regulation that promote and facilitate the use of AI technologies, while at the same time addressing the potential challenges that these technologies present. These challenges may be different between developed and emerging countries.

a) Risk of over-regulation in growth phase

As discussed earlier, there is no clear definition of AI. AI developments are likely to occur gradually and incrementally, but they may experience a rapid acceleration (S-curve model).

There is a risk of over-regulation in the incremental growth phase that we are currently in. Overarching regulation appears to be inappropriate right now. Indeed, existing regulatory frameworks may be fit for purpose or may require relatively minor change. There will be different considerations in different contexts. Policy makers should consider how AI can reduce risks, as well as the risks that it creates.

b) Public awareness and trust

AI has received considerable media attention, but much of it has been superficial. There is little public awareness of the problems that could be solved by AI, with more public attention on extreme situations where AI might go wrong (e.g., accidents caused by autonomous vehicles).

This is an area where governments play an important role in helping society to prepare and adapt to AI. Governments can help to develop public trust and understanding in AI technologies and what the implications of these technologies will be for people.

This stage is critical and will be one of the first things that governments should be doing. If there is a lack of trust and understanding among the public, or if there is excessive fear of the consequences of AI (e.g., in employment), the potential benefits of AI may well not be realised. People may be reluctant to allow their data to be used in the development of these systems and may not be prepared to use them or allow them to be used.

c) Addressing the digital divide

The module on AI governance also highlights supporting local ecosystems of entrepreneurship and start-ups, as well as supporting capacity development at universities. It considers government programmes to facilitate the growth of entrepreneurial ecosystems, technology business incubators and other methods.

d) An AI national plan

Governments should consider developing an AI national plan. This would be a document that outlines the key strategies for preparing the country for AI. It should address the opportunities and risks, many of which are outlined in this AI Series.

For major economies, with research and investment ambitions, the AI national plan will be a comprehensive document.

AI national plan case studies: The United States[37] and China[38] have both produced substantial national AI plans, with India planning to release their own shortly. These plans focus on the research needs, regulatory requirements, data sharing and preparing an AI savvy workforce.

Emerging countries may have different objectives, but planning at a national level is still necessary. While it may be premature to regulate for AI, it is not too early to lay the groundwork for the emergence of AI. As discussed elsewhere in this introductory module, there are some key things that all governments can do now in anticipation of AI, including:

- Beginning a public dialogue to raise awareness of AI as a technology, of its benefits and potential consequences and how the government is preparing for it;
- Developing a quality government data environment, with open standards;
- Engaging with businesses operating in the country that are investing in AI internationally, and that may hold large amounts of personal data, and discussing the development of AI applications with public good purposes and that may fulfil Sustainable Development Goals objectives;
- Considering potential infrastructure roadblocks that could limit the potential for AI (e.g., access to 5G spectrum and deployment of fibre infrastructure); and
- Laying the groundwork for resolving some of the important macro issues that we discuss, including the future of work and how to prepare people for a changing work environment and the fiscal, and safety net, issues that may arise with AI and how to prepare for these impacts.

We also suggest an oversight body, which we discussed in the previous section.

37 https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

38 https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-government-outlines-ai-ambitions-through-2020/

This is an area where potential international or regional collaboration can be highly productive. Many countries are encountering these issues at almost the same time. Also, some of the actors involved will include global international businesses, that will be viewing these issues on an international or regional basis.

We raise many other issues in this introductory module, and in this AI Series, but it is not necessary for individual governments to resolve all of these issues up front, at least in the first iteration of the AI national plan. Some of these issues will take time to emerge and show their true contours and, apart from maintaining awareness of them, will probably not need to be addressed immediately. Other issues are likely to be substantially resolved at a global level and may not require resolution locally.

Bibliography

Artificial Intelligence Index, 2017 annual report, November 2017

British Academy and Royal Society, Data management and use: governance in the 21st century, June 2017

British Parliament Report on AI (https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/14504.htm)

Calo, Artificial Intelligence Policy: A Primer and Roadmap, August 2017

Elliot, 2017. Computers and the Future of Skill Demand. (http://www.oecd-ilibrary.org/education/computers-and-the-future-of-skill-demand_9789264284395-en)

European Parliament, Civil law rules on robotics, February 2017

Government of the Republic of Korea, Mid to long term master plan in preparation for the intelligent information society, [date?]

House of Lords, Select Committee on artificial intelligence, AI in the UK: ready, willing and able?, April 2018

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html)

Levendowski, How copyright law can fix artificial intelligence's implicit bias problem, July 2017

McKinsey Global Institute, Artificial intelligence the next digital frontier, June 2017

McKinsey Global Institute, Harnessing automation for a future that works, January 2017

Obama White House: Office of Science and Technology Policy, Preparing for the Future of Artificial Intelligence, May 2016

Obama White House: National Science and Technology Council, The national artificial intelligence research and development strategic plan, October 2016

Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop. 239 (2013). http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

Royal Society; Machine learning: the power and promise of computers that learn by example, April 2017

Stanford University, Artificial intelligence and life in 2030, September 2016

The Economist, 2017. Retraining low-skilled workers. (https://www.economist.com/news/special-report/21714175-systems-continuous-reskilling-threaten-buttress-inequality-retraining-low-skilled)

World Economic Forum (https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/)

World Economic Forum, White Paper, Digital Transformation Initiative, Telecommunications Industry, January 2017 (https://www.accenture.com/t20170411T115809Z__w__/us-en/_acnmedia/Accenture/Conversion-Assets/WEF/PDF/Accenture-Telecommunications-Industry.pdf)

Artificial Intelligence (AI) for Development Series

Module on Setting the Stage for AI Governance: Interfaces, Infrastructures, and Institutions for Policymakers and Regulators

July 2018

Work in progress, for discussion purposes
Comments are welcome!
Please send your comments on this paper at: gsr@itu.int by 30 July 2018

This module was prepared by Urs Gasser, Ryan Budish, Amar Ashar, members of the Ethics and Governance of AI initiative by the Berkman Klein Center for Internet & Society at Harvard University and the MIT Media Lab [1], under the direction of the ITU/BDT Regulatory and Market Environment Division, in collaboration with the ITU/BDT Telecommunication Division and under close coordination with the Chief of the ITU/BDT Infrastructure, Enabling Environment, and E-Applications Department. We would like to thank the ITU General Secretariat and the ITU Standardization Bureau for their contributions.

The authors would like to thank Wendy Chu, Elena Goldstein, Aida Joaquin Acosta, Levin Kim, Sean Rail, and Jenna Sherman for their research and writing assistance. Deep appreciation is also expressed to the many individuals who were gracious with their time through participation in the Berkman Klein Center's Global AI Dialogue Series and individual interviews, and whose inputs were incredibly influential to our thinking. In particular we thank Paul Nemitz (Principal Adviser, Directorate-General for Justice and Consumers, European Commission), Gabriele Mazzini (Policy Officer, European Commission), Danil Kerimi (Head of Information Technology and Electronics Industries, World Economic Forum), and Terah Lyons (Executive Director, Partnership on AI).

I. Introduction

This module is one part of a four-part series on AI for Development. The series covers a range of issues relevant to policymakers and regulators as they seek to understand and address the challenges and opportunities of AI technologies. The series covers AI, its potential societal impacts, governance questions, and cybersecurity and Internet of Things issues.

The other modules of the AI for development series make it clear that "AI" is not one thing—it refers to a range of different technologies and applications used in many different ways. The other briefing papers also highlight that we are at different places in our collective empirical and normative understanding of these technologies, their impact on humans and society, and the best ways to deal with the changes ahead. This state of uncertainty is mirrored in the contemporary debates about the governance and ethics of AI, where both public and private sector leaders and experts have many different ideas about how to best limit the risks of these innovative AI applications and help unlock their opportunities.[2] For example, some experts have called for the formation of new regulatory agencies that specialize in AI or robotics.[3] Governments and international organizations have started to create non-binding standards to govern the creation and use of AI.[4] Individual companies are releasing their own ethical guidelines constraining their own use of AI.[5] Multistakeholder partnerships are currently formulating their own best practices for the development and deployment of AI.[6] And academics have created frameworks to ensure that AI training data and outputs are used "for good."[7]

Although there are many proposals for addressing AI's challenges, these efforts do not represent a holistic governance framework ready for application in the real world. Instead, existing proposals are possible building blocks and elements towards a more comprehensive approach. That said, there have been efforts to sketch at a conceptual level what a holistic governance framework for AI might look like. For example, one of the authors of this paper has proposed a layered model for AI governance that describes three layers of different governance approaches: (1) the technical layer – focusing on technical standards

---

[2] Throughout this paper we collectively refer to these opportunities and risks as the "challenges" of AI.

[3] See Ryan Calo, "The Case for a Federal Robotics Commission," *Brookings Institution Center for Technology Innovation,* September 1, 2014, https://ssrn.com/abstract=2529151; Bruce Schenier, "Click Here to Kill Everyone," New York Magazine, January 27, 2017, https://www.schneier.com/essays/archives/2017/01/click_here_to_kill_e.html; European Parliament, "Motion for a European Parliament Resolution: Recommendations to the Commission on Civil Law Rules on Robotics," 2015/2103(INL), January 27, 2017, http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&mode=XML&reference=A8-2017-0005&language=EN#title1.

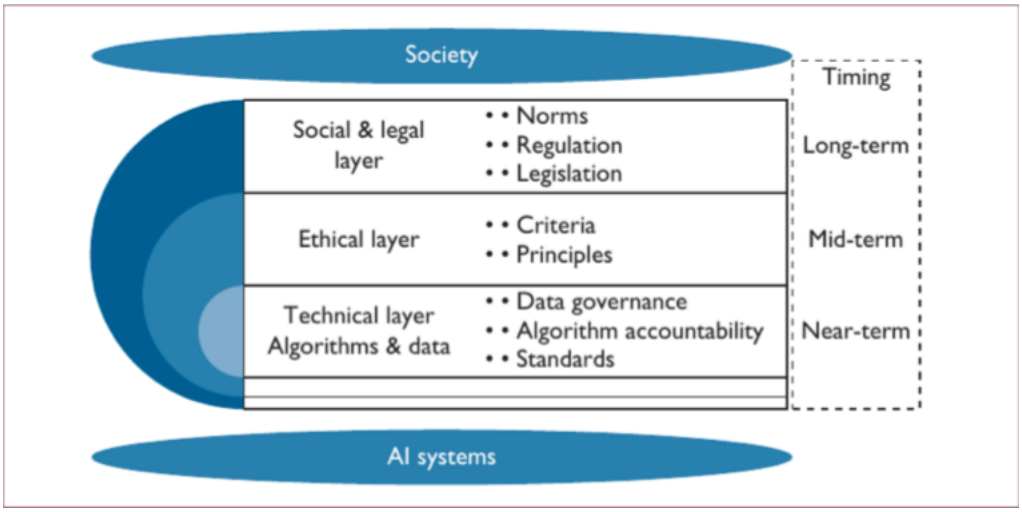
[4] See, for example, "Discussion Paper on Artificial Intelligence (AI) and Personal Data – Fostering Responsible Development and Adoption of AI," Personal Data Protection Commission: Singapore, June 5, 2018, https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data; Japan's Institute for Information and Communications Policy (IICP), the Ministry of Internal Affairs and Communications (MIC), "Draft AI R & D Guidelines," July 28, 2017, http://www.soumu.go.jp/main_content/000507517.pdf; IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design Version 2, 2017, http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

[5] Sundar Pichai, "AI at Google: Our Principles," June 7, 2018, https://blog.google/topics/ai/ai-principles/.

[6] See Partnership on AI, https://www.partnershiponai.org.

[7] ITU News, "'Roadmap Zero' to AI and Data Commons," ITU News, May 25, 2018, http://news.itu.int/roadmap-zero-to-ai-and-data-commons/.

and constraints on the collection, use, and management of data by AI algorithms; (2) the ethical layer – focusing on ethical and human rights principles; and (3) the social and legal layer – focusing on creating institutions for regulating autonomous systems.[8]



Credit: Urs Gasser and Virgilio A.F. Almeida (2017)

From this model one can see how existing efforts like those described at the outset, such as Singapore's accountability-based framework, or the IEEE's Ethically Aligned Design Principles, can each play a role in a larger governance framework. However, the envisioned framework is like an evolving library; we know where the books would go, but many of the books are still being written. In addition, a series of structural challenges make it currently difficult (and in our view inadvisable) to build an all-encompassing single AI governance framework:

1. **Unknown societal impact:** Governance frameworks aim to address particular societal problems based on evidence, and yet for most aspects of AI we currently lack a solid empirical understanding of the short- and long-term consequences of the technologies. In many cases, reliable metrics to track societal impacts beyond unemployment and GDP are not readily available—a task that is particularly difficult given the dual-use nature of AI, with a variety of positive and negative impacts.
2. **Undefined questions:** In many areas of application, researchers are still defining some of the "right" questions to be asked. For example, when we look at issues like disinformation or hate speech online and how AI might be used to help counter those challenges, we do not yet even fully understand the scope of the problem. We struggle to answer many foundational questions, such as: How do we define harmful speech or "fake news"? What is the role of platforms in addressing these issues? What level of error are we willing to accept from automated systems that police content online? And what interventions—technological, legal, social, normative or otherwise—will help us address it?

[8] Urs Gasser and Virgilio A.F. Almeida, "A Layered Model for AI Governance," IEEE Internet Computing 21, November, 2017, http://nrs.harvard.edu/urn-3:HUL.InstRepos:34390353, 58-62.

3. **Diversity of frameworks:** AI is not emerging in a total vacuum and is shaped in important ways by existing norms and governance frameworks. For example, medicine, automobiles, and data collection are spaces with complex, local, national, and international governance structures in place that interact with the development and deployment of next generation technology. The EU's recently enacted GDPR, for instance, is not an AI-specific governance framework, but already constrains how data can be collected and used for machine learning.

Given these structural challenges, the quest for a comprehensive and detailed governance framework for AI seems unrealistic—at least for the time being. Instead, we believe that a more productive approach would be to focus on the development of governance elements and strategies and the interplay and interoperability between these building blocks in a layered model. It is in this spirit that this paper offers a set of approaches aimed at assisting policymakers and regulators in building capacity so that they can engage in the shaping and creation of the evolving governance frameworks for AI. This is no small task. For many policymakers and regulators, addressing the challenges of AI may seem daunting in large part because they are complex, fast-moving, and dynamic, and as a result a handful of broad areas of concerns have emerged. Based on formal and informal dialogues at international forums such as the ITU's AI for Good Summit, the Internet Governance Forum, and the Global Summit on AI & Inclusion, as well as the Berkman Klein Center's Global AI Dialogue Series[9], original interviews with leaders and experts in the field, and a review of AI policy materials, four broad areas of concern have crystallized:

1. Increasing information asymmetries: Knowledge about AI is increasingly held within a handful of companies within the private sector, creating knowledge gaps and information asymmetries that challenge policymakers and regulators, who often struggle to keep pace with the latest societal developments and their societal implications, and lack the technological depth to understand the full range of possible approaches and the tradeoffs that they might entail.
2. Inadequacy of unilateral public-sector action: Just as most of the knowledge about AI is held within the private sector, so too is much of the control of the technical development of AI technologies. Even if information asymmetries can be fully bridged, many of the most effective approaches may require private sector participation and support.
3. Exacerbating the digital divide: The digital divide has long been a concern for decisionmakers, and AI is making these concerns ever more pressing, both from an impact and development perspective. From an

Global AI Dialogue Series

Over the last year, the Berkman Klein Center for Internet & Society has convened several workshops with policymakers and a diverse range of stakeholders from around the world as part of the Global AI Dialogue Series. The inclusive series is aimed at identifying opportunities as well as challenges related to AI that need to be addressed from an international perspective through evidence-based dialogue. The Series works to build an institutional knowledge base, foster human capacity, and strengthen interfaces with industry and policy-makers at an international scale. Initial meetings have taken place in the US, Seoul, China, Hong Kong, Switzerland, and Italy, with projected meetings set at this stage for Singapore and Thailand.

[9] See sidebar and Berkman Klein Center for Internet & Society, "Ethics and Governance of Artificial Intelligence Initiative: Research Sprints and Pilots," cyber.harvard.edu, https://cyber.harvard.edu/research/ai/research (accessed May 2, 2018)

impact perspective, AI technologies often require substantial digital infrastructure and data to be effective. This means that areas with the most data and the most robust digital infrastructure will be the first to reap the benefits of these technologies, leaving underresouced, less connected communities even further behind than they are now. And from a development perspective, areas without strong technical capacities (both human and digital) may find it challenging to participate in the global governance dialogue, and to compete with more established market competitors from places like Silicon Valley and China.

4. Creating and maintaining a competitive environment: Because AI is so dependent on data, existing privacy and intellectual property regulations, as well as legal interoperability across jurisdictions, can have outsized impacts on the development of AI technologies. Moreover, existing market incumbents with large amounts of data can leverage those datasets to create lock-in and network effects. The sum effect is that decision makers may struggle to support local entrepreneurial efforts in AI technologies.

These concerns are important and real. Some of these concerns are familiar issues that are exacerbated by new AI technologies, and others are novel concerns emerging from new applications. Either way, there are tools that policymakers and regulators can deploy to address these concerns and ensure fair, accountable, and transparent outcomes.[10] In the remainder of this paper, we go through each of these four areas of concerns and identify some of the many approaches and tools[11] that policymakers and regulators can and should experiment with. This "toolbox" builds upon previous experiences when dealing with disruptive technologies, including the Internet, and the governance challenges those technologies have created.[12] In particular, we offer tools that address each of the following challenges that policymakers and regulators face when addressing AI governance issues:

1. Addressing Information Asymmetries
2. Building Public-Private Partnerships
3. Bridging the Digital Divide
4. Sustaining a Competitive Environment

---

[10] "Discussion Paper on Artificial Intelligence (AI) and Personal Data – Fostering Responsible Development and Adoption of AI," Personal Data Protection Commission: Singapore, June 5, 2018, https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data, 5.

[11] Throughout this paper we refer to these as "tools." We recognize, however, that much of what we describe are governance approaches. Many of AIs challenges that policymakers seek to address are closely connected to, or are proxies for, difficult social, economic, and political challenges. As such, there are neither simple fixes nor external solutions that can simply be dropped in.

[12] To be clear, we do not suggest that AI governance will look like Internet governance, but we recognize that because this space is still in flux and because not everything is new, Internet governance and other areas of governance can serve as a source of inspiration when exploring how different tools might be used to address certain challenges.

II. Tools for Addressing Information Asymmetries

Emerging technologies such as AI has the potential for tremendous societal benefits, so long as risks can be managed and mitigated through informed, evidence-based decision making. However, the more complex the technology, the harder it can be for regulators and policymakers to understand the potential impacts of the technology and the potential ramifications (both intended and not) of any proposed policy intervention. This is particularly true with AI, where even technical experts struggle to fully explain the inner workings of their systems.[13] [14] Moreover, much of the expertise that exists about AI has been consolidated into the hands of a small group of companies, further exacerbating the difficulty that policymakers face when trying to craft informed policies. This challenge is one of information asymmetries -- a growing imbalance of foundational knowledge between the private and public sectors that is particularly acute for AI technologies.

Although information asymmetries are stark with respect to AI, such asymmetries are a fundamental challenge for regulators and policymakers seeking to address the challenges of any complex technology. For example, policymakers have frequently struggled to overcome information asymmetries regarding cybersecurity and encryption. Cryptography is a complex mathematical field, and encryption is increasingly important in a range of technologies.[15] For that reason, even well-informed policy makers can struggle to predict the impacts of encryption policies.[16] For example, two months after the 2016 San Bernardino shooting, two US Senators proposed legislation that would have effectively banned the distribution of secure web browsers.[17]

Just as policymakers must address information asymmetries to craft effective encryption policies, so too must they bridge information asymmetries in order to craft effective artificial intelligence policies. For example, policymakers with a limited understanding of how machine learning applications can reinforce existing societal biases may disproportionately rely on AI to solve difficult societal issues, such as the

---

[13] Will Knight, "The Dark Secret at the Heart of AI," *MIT Technology Review*, April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[14] David Weinberger, "Our Machines Now Have Knowledge We'll Never Understand," *Wired Magazine,* April 18, 2017, https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/.
[15] OECD, "OECD Guidelines for Cryptography Policy," OECD.com, https://www.oecd.org/sti/ieconomy/guidelinesforcryptographypolicy.htm (Accessed May 2, 2018)
[16] Ryan Budish, Herbert Burkert , Urs Gasser, "Encryption Policy And Its International Impacts: A Framework For Understanding Extraterritorial Ripple Effects," Hoover Institution, March 2, 2018, https://www.hoover.org/research/encryption-policy-and-its-international-impacts.
[17] Andy Greenburg, "The Senate's Draft Encryption Bill Is 'Ludicrous, Dangerous, Technically Illiterate,"' *Wired Magazine*, April 8, 2016, https://www.wired.com/2016/04/senates-draft-encryption-bill-privacy-nightmare/.

prevalence of pre-trial detention in the criminal justice system,[18] [19] and unintentionally make things worse, not better.

Guiding Principles

Policymakers and regulators need not become technical experts in AI, but it is important to work to reduce information asymmetries. In that process there are a few high level principles that can guide policymakers and regulators:

- **Create compelling opportunities for experts to join government.** AI expertise is incredibly valuable within the private sector, making it cost prohibitive for governments to compete outright for hiring such experts. Governments are unable to compete with the salaries or earning potential available from large technology companies or innovative startups. However, policymakers and regulators can offer short-term appointments or "tours of duty" that appeal to civic responsibility, rather than competing outright with the private sector.
- **Reduce participatory friction for experts.** Technical experts are often unfamiliar with bureaucratic and government processes, and short-term appointments do not afford the opportunity to learn how to effectively advocate within these systems. Instead of asking AI experts to operate within existing, complex bureaucratic structures, policymakers and regulators can create positions that operate outside of traditional bureaucratic structures, such as Chief Innovation Officer roles that report directly to the agency heads, so that technical experts can influence AI governance without needing to first learn how to navigate governmental institutions.
- **Obtain hands-on experiences with AI technologies.** There is no substitute for hands-on experience. Through building AI technologies in-house, visiting AI labs and businesses (both local and in other countries), and testing products, policymakers and regulators will learn about the way AI technologies are being developed and applied, and how they might evolve in the future.

In putting these principles into action, there are a range of tools that policymakers and regulators can deploy to narrow the information gap--some more general to emerging technologies,[20] and some specific to AI.[21] In particular, policymakers and regulators can bridge the knowledge gaps between governments and the private sector by (1) building internal capacity and (2) developing knowledge exchange interfaces between regulators and experts. Below, we explore some specific approaches for both.

(1) Building Internal Capacity

---

[18] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[19] Megan Stevenson, Assessing Risk Assessment in Action, George Mason Law & Econ. Research Paper №17–36, 2017, https://ssrn.com/abstract=3016088.

[20] For example, given the close relationship between cloud computing and big data, on the one hand, and AI on the other, some of the same approaches that policymakers have used to close the information gaps with respect to big data and cloud computing are equally applicable for AI. See Urs Gasser and David O'Brien, "Governments and Cloud Computing: Roles, Approaches, and Policy Considerations," Berkman Klein Center for Internet & Society, March 17, 2014, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2410270.

[21] For example, the challenges of AI interpretability may necessitate some new approaches to bridging information asymmetries. See Finale Doshi-Velez, Mason Kortz, etc, "Accountability of AI Under the Law: The Role of Explanation," Berkman Klein Center for Internet & Society, April 13, 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064761.

One way for policymakers and regulators to bridge information gaps is to bolster the technical expertise within government. This brings technical expertise in-house to policymakers and regulators to work side-by-side with policymakers and regulators in crafting policies and interventions, and can help reduce bureaucratic frictions. This kind of capacity development can occur at the individual layer, through special hiring and recruitment processes, or at the institutional layer, through creating and utilizing expertise across government agencies and departments.

(a) **Recruiting Individual Expertise**. At the individual layer, policymakers and regulators can attract AI experts to government by creating and funding more job positions, supporting residency programs that temporarily place technical experts within the government[22], and partnering with universities to create public sector employment pipelines for aspiring experts.[23] It is difficult to compete with the private sector, given the high salaries that AI experts can expect, but creating special temporary positions can appeal to policy-minded technical experts. For example, several United States regulatory agencies have experimented with creating positions like "Chief Technology Officer"[24] or "Chief Innovation Officer"[25] in order to attract experts with diverse backgrounds in security, telecommunications, privacy, and Internet governance. And programs like the Presidential Innovation Fellows Program and the U.S. Digital Service are designed to attract software engineers, designers, and product managers to perform "tours of duty" within government agencies.[26]

In some cases individual expertise may already exist internally, and simply needs to be activated through the creation of systems that identify employees with relevant technical skills and empower them to more fully engage in the governance process. For example, the World Bank created the SkillFinder network to help employees find technical and other experts from within its 27,000 employees, consultants, and alumni.[27] The United States Department of Health and Human Services took a similar approach in order to better utilize employees' technical expertise for medical device safety review panels in order to speed up approval processes.[28]

(b) **Building Institutional Expertise.** At the institutional level, capacity building can involve better connecting policymakers and regulators with the expertise and knowledge that exists in silos

---

[22] Examples include the Presidential Innovation Fellows Program, etc., see IEEE, "Artificial Intelligence Research,Development and Regulation", IEEEUSA.org, https://ieeeusa.org/wp-content/uploads/2017/10/AI0217.pdf, p. 5 (Accessed May 2, 2018)

[23] Examples include the Data Science for Good Program at the University of Chicago.

[24] See Neil Chilson's biography, Lorrie Faith Cranor's biography, and Latanya Sweeny's biography.

[25] US Commodity Futures Trading Commission, "The CFTC Announces Appointment of John Rogers as Chief Information Officer," CFTC.gov, September 6, 2011, https://www.cftc.gov/PressRoom/PressReleases/pr6106-11 and U.S. Department of Transportation. "J. Christian Gerdes biography," Volpe.dot.gov, https://www.volpe.dot.gov/events/chris-gerdes-biography (Accessed May 2, 2018).

[26] Presidential Innovation Fellows, "About," Presidential Innovation Fellows, https://presidentialinnovationfellows.gov/about/ (Accessed May 2, 2018); U.S. Digital Service, "Join," USDS.gov, https://www.usds.gov/join#who (Accessed May 2, 2018).

[27] GovLab, "Smarter State Case Studies: The World Bank: Skillfinder," thegovlab.org, February 10, 2016, http://www.thegovlab.org/static/files/smarterstate/skillfinder.pdf.

[28] GovLab, "Health And Human Services: HHS Profiles," thegovlab.org, February 10, 2016, http://www.thegovlab.org/static/files/smarterstate/HHS.pdf.

within other parts of government. For example, the United Kingdom Cabinet Office created open standard principles for common and secure information technology infrastructure[29] in order to promote data standardization practices to make data transfers more efficient and useful, enabling policymakers and regulators to take advantage of knowledge from across government and reducing frictions for data exchange.[30] Institutional expertise can also exist in peer agencies in other countries. For example, the International Development Research Centre recently held a research capacity mentorship and workshop program to facilitate ICT scholarship in the Global South.[31]

In some cases the demand for expertise may be so great that policymakers and regulators need to pool together expertise through establishing institutions to house specialized professionals. This can counter the fragmentation that can occur with overlapping agency jurisdiction, preempt political battles for turf, and attract new specialized talent.[32] For example, the European Data Protection Supervisor is an independent supervisory authority bringing together lawyers, IT specialists and administrators to advance privacy and data protection in the EU.[33]

(2) Developing Knowledge Exchange Interfaces With Experts

Policymakers and regulators can also address information asymmetries by establishing ongoing interfaces with experts, which can supplement or replace the need to hire experts. By engaging academic and industry experts, policymakers and regulators can leverage external expertise and gain increased experience with AI technologies.

(a) Leveraging academic expertise. Through the support of academic research centers and research projects, policymakers and regulators can address specific knowledge gaps while building a pipeline of expertise. For example, the April 25, 2018 Communication on Artificial Intelligence, the EU announced a EUR 2.7 billion investment in "market-creating innovation such as AI"

---

[29] See UK Government, "UK open standard principles," UK.gov, April 5, 2018 https://www.gov.uk/government/publications/open-standards-principles.

[30] Interagency data-sharing allows employees with different subject-matter expertise to collaborate on overlapping jurisdictions. In order to facilitate these partnerships, governments should develop a mechanism for agencies to dictate the reach and limits of their partnerships. For example, data-sharing can be facilitated by Memoranda of Understanding; the FTC and the Federal Communications Commission used a Memorandum of Understanding to share Internet Freedom consumer complaints subject to the agencies' confidentiality policies. See FTC, "FTC, FCC Outline Agreement to Coordinate Online Consumer Protection Efforts Following Adoption of The Restoring Internet Freedom Order," FTC.gov, December 11, 2017, https://www.ftc.gov/news-events/press-releases/2017/12/ftc-fcc-outline-agreement-coordinate-online-consumer-protection.

[31] International Development Research Centre, "Research Capacity Peer Mentorship Program: Case Study on the ICT and Development Conference," IDRC.ca, https://www.idrc.ca/en/project/research-capacity-peer-mentorship-program-case-study-ict-and-development-conference (Accessed May 2, 2018).

[32] Because of the effect of AI on a myriad fields, there is more potential for fragmented jurisdiction. In the United States, for example; at least sixteen federal agencies govern sectors related to AI.

[33] EU, "About: European Data Protection Supervisor," Europa.eu, https://edps.europa.eu/about-edps_en (Accessed May 2, 2018). In a similar manner, the Brookings Institute has called for a U.S. Federal Robotics Commission to combine the efforts of the National Highway Traffic Safety Administration, the National Aeronautics and Space Administration, the Food and Drug Administration, and various other bodies. See Ryan Calo, "The case for a federal robotics commission," Brookings Institute, September 15, 2014, https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission/.

through the European Innovation Council.[34] Similarly, the EU is supporting AI excellence centers to facilitate academic collaboration on AI. In the cybersecurity space, the US has leveraged academic expertise through national security-focused university research projects like START and CERT,[35] which have advanced research in the areas of cybersecurity and terrorism while creating "a pipeline of approximately 30,000 data experts."[36]

(b) Activating stakeholder expertise. Expertise may also exist outside of academia, in civil society organizations, the private sector and others. Through both on-going and more time-limited engagements, policymakers and regulators can also leverage the expertise throughout these other organizations. For example, the European Parliament launched a three-month public consultation process about the future of robotics and artificial intelligence; 298 individuals, organizations, and corporations answered general[37] and specialized[38] questionnaires about rules on ethics, liability rules, safety and security, and institutional coordination and oversight.[39] Similarly, the White House AI Workshop Series invited experts to contemplate safety and control, legal issues, and social good in events all over the United States. The series informed the preparation of the National Science and Technology Council's report on the future of AI.[40]

Ongoing collaborations between policymakers and external experts are also possible; although they require greater commitments from the participants, they reservoir of trust established over time can help address more complex questions. For example, from 2010 to 2013 Germany formed an Enquete Commission, consisting of politicians and technical experts to inform its Internet policy decisions.[41] Similarly, the European Commission's GEAR 2030 high-level group convened member states, industry groups, civil society organizations, and various observers to set

---

[34] "Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines," European Commission, April 25, 2018, http://europa.eu/rapid/press-release_IP-18-3362_en.htm.
[35] National Consortium for the Study of Terrorism and Responses to Terrorism, "START," University of Maryland, http://www.start.umd.edu/about/about-start; "The CERT Division," Carnegie Mellon University, https://www.sei.cmu.edu/about/divisions/cert/index.cfm.
[36] Joel Tito, "Destination unknown:Exploring the impact of Artificial Intelligence on Government," Center for Public Impact, September 2017, https://publicimpact.blob.core.windows.net/production/2017/09/Destination-Unknown-AI-and-government.pdf, p.58.
[37] See example of a general questionnaire: European Parliamentary Research Service, European Added Value Unit, "General Questionnaire Civil Law: Rules on Robotics," Committee on Legal Affairs of the European Parliament, April 2017, http://www.europarl.europa.eu/cmsdata/committees/juri-public-consultation/civil-law-rules-on-robotics/general_questionnaire.pdf.
[38] See example of specialized questionnaire: European Parliamentary Research Service, European Added Value Unit, "General Questionnaire Civil Law Rules on Robotics," Committee on Legal Affairs of the European Parliament, April 2017, http://www.europarl.europa.eu/cmsdata/committees/juri-public-consultation/civil-law-rules-on-robotics/specialised_questionnaire.pdf.
[39] "Public consultation – Future of Robotics and Artificial Intelligence," European Parliament Committees, March 22, 2017, http://www.europarl.europa.eu/committees/en/juri/robotics.html?tab=Introduction.
[40] Ed Felton, "Preparing for the Future of Artificial Intelligence," Obama White House Archives, May 3, 2016, https://obamawhitehouse.archives.gov/blog/2016/05/03/preparing-future-artificial-intelligence.
[41] Urs Gasser, Ryan Budish, and Sarah Myers West, "Multistakeholder as Governance Groups: Observations from Case Studies," Berkman Klein Center for Internet & Society, January 15, 2015, https://cyber.harvard.edu/publications/2014/internet_governance.

objectives, specify milestones, and assign responsibilities to "reinforce the competitiveness of the European automotive value chain."[42]

1. Addressing Information Asymmetries
2. **Building Public-Private Partnerships**
3. Bridging the Digital Divide
4. Sustaining a Competitive Environment

III. Tools for Building Public-Private Partnerships

Even where policymakers and regulators can bridge information asymmetries, developing effective governance approaches for AI's challenges will often require private sector and civil society participation and support. AI technologies are increasingly embedded within many sectors of society, and its impacts are widespread and deep, posing difficulties for policymakers and regulators seeking to intervene. Consider for example, autonomous vehicles. The potential impacts of autonomous vehicles include: passenger safety, public safety, traffic, automobile manufacture and sales, the environment, labor markets, and more. The sheer variety of AI's potential impacts require solutions that include span a variety of perspectives and expertise: technical, social, political, and economic, among others. Moreover, even if policymakers and regulators could navigate these diverse perspectives, addressing some of these impacts may be most effectively addressed by the private sector or others outside of government.

Building effective multistakeholder governance groups is not easy. Bringing together these diverse perspectives is important, but utilizing these perspectives to create effective governance strategies requires more than just placing people into a room. Bridging the gap between different sectors, experiences, norms, and cultures is complex because stakeholders may come to the table without a shared, common language about AI technologies or governance, with different levels of comfort with AI technologies, with different levels of resources available to invest in governance processes, or with a lack of trust.

The design of effective multistakeholder systems is challenging and is a rich and deep field of study.[43] AI is particularly challenging, because as an academic field AI has existed for over 50 years, which can make it difficult to bring new perspectives and disciplines to the AI conversation, because trust and intellectual depth may be missing. That said, there a variety of tools that policymakers can use to strengthen the connections between the public and private sectors.

Guiding Principles

- **Develop a terminology, shared across all stakeholders.** A major challenge for both describing the challenges of AI and for developing shared solutions is that there's no common language across stakeholders. World like "Rules" and "function," for example, mean very different things

[42] "Policy and strategy: High Level Group on Automotive Industry 'GEAR 2030,'" European Commission, February 5, 2019, https://ec.europa.eu/growth/sectors/automotive/policy-strategy_de.
[43] See, e.g., Urs Gasser, *et al,* "Multistakeholder as Governance Groups."

to a computer scientist than to a regulator. Fundamental assumptions about meaning must be scrutinized and challenged until all stakeholders can participate equally in discussing the challenges and developing solutions.

- **Take advantage of being a second (or third, or fourth, or…) mover.** AI is operating at different timescales around the world. This means that the pressing issues facing some regulators and policymakers today may seem like distant science fiction to others. This can be a distinct advantage for those who need not act immediately; as AI becomes more prevalent, such policymakers and regulators should actively learn from the experiences of both the public and private sector actors who have faced similar challenges previously. What worked in one country or region is unlikely to play out in the exact same way in another, but important lessons can still be learned through ongoing dialogue with businesses and policymakers who have grappled with similar challenges.
- **Keep an open door and an open mind.** When it comes to AI, no one understands all of the problems, let alone all of the solutions. Hearing from as many perspectives as possible will expose policymakers and regulators to issues that may not have been on their radar and creative solutions they may not have tried otherwise. And some of these solutions may not require law or regulation.

In putting these principles into action, there are a range of tools that policymakers and regulators can deploy to engage with diverse stakeholders in advancing AI governance. For example, policymakers and regulators can (1) incubate and engage in multistakeholder systems, (2) solicit public feedback on policies, and (3) utilize diverse forums as platforms for information exchange.

(1) Incubate and engage in multistakeholder systems

One way for policymakers and regulators to engage stakeholders is through either creating or engaging in multistakeholder systems. These can be either AI specific or based on a broader set of topics that encompass AI. In either case, these systems provide opportunities for policymakers, regulators, the private sector, and civil society to share and learn about emerging technologies, develop a common language, and build trust. AI-specific systems in particular can be effective in helping educate stakeholders, including policymakers and regulators about the current state of AI technologies, can help to articulate solutions to complex problems, and can help to develop synergies among different initiatives that aim at solving related issues.

(a) AI-specific Multistakeholder Systems. Participation in multistakeholder systems can be an effective way for policymakers and regulators to stay abreast of state-of-art AI issues.[44] Through sustained dialogue with the private sector, policymakers and regulators can learn about emerging AI applications. And through sustained dialogue with civil society and academics, policymakers and regulators can hear about societally beneficial AI applications, as well as emerging concerns.

[44] See National Science and Technology Council, Committee on Technology, "Preparing for the Future of Artificial Intelligence," Executive Office of the President, October 2016, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. The White House recommended that industry work with government to keep it updated on the general progress of AI.

For example, the Partnership on AI, originally formed by several technology companies, now includes academics and civil society organizations such as Human Rights Watch.[45] The Partnership serves as an open platform to educate its partners and advance the public understanding of AI, while supporting research, testing on AI tech, addressing privacy, transparency, security, and ethics concerns. Similarly, the World Economic Forum's Artificial Intelligence and Machine Learning Project, seeks to support governments, industry, and academia in co-developing policy frameworks to govern AI, that will help to test theories, extract lessons and scale their adoption globally.[46]

Multistakeholder systems can also help policymakers and regulators avoid duplicating existing efforts by coordinating across the myriad of AI-specific initiatives that have emerged over recent years. A recent example of this is the European AI-on-demand-platform[47] which aims to enhance the economic and social potential of AI by building synergies among the existing AI initiatives in the EU. The European AI-on-demand-platform aims at activating the AI community in Europe, serving as a hub of AI-related knowledge and tools, promote the integration of AI into applications, and facilitate access to data needed by AI algorithms.[48]

(b) Broad-Based Multistakeholder Systems. Because the impacts of AI are so widespread and diffuse, participation in more general multistakeholder systems can also help policymakers and regulators explore and respond to AI technologies' impacts.[49] For example, the Missing Maps project[50] of the Multi-stakeholder Forum on Science, Technology, and Innovation (STI) for the Sustainable Development Goals[51] at the United Nations, uses crowdsourcing to add missing areas

---

[45] See "Thematic Pillars," Partnership on AI, https://www.partnershiponai.org/thematic-pillars/. The Partnership was founded by major tech companies (Amazon, Apple, DeepMind, Google, IBM, Facebook, and Microsoft) and includes public and private partners, such as AI Now, ACLU, Amnesty International, Center for Democracy and Technology, Universities of Princeton University, University of Washington, Tufts University, UNICEF, Data and Society, Upturn, Open AI, and Humans Right Watch. The list of the initiatives the Partnership wants to address includes safety-critical AI; Fair, Transparent, and Accountable AI; and collaborations between people and AI systems.

[46] "Center for the Fourth Industrial Revolution: Projects," World Economic Forum, https://www.weforum.org/center-for-the-fourth-industrial-revolution/areas-of-focus.

[47] "Horizon 2020 - Work Programme 2018-2020 Information and Communication Technologies," European Commission, January 31, 2018, http://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-leit-ict_en.pdf,. 55-56.

[48] "The European Artificial Intelligence-on-demand-platform - Information day and brokerage event," European Commission, December 5, 2017, https://ec.europa.eu/digital-single-market/en/news/european-artificial-intelligence-demand-platform-information-day-and-brokerage-event.

[49] For more on the widespread impacts of AI technologies, see "Artificial Intelligence & Inclusion," ITS Rio, *et al*, https://aiandinclusion.org (last accessed May 2, 2018).

[50] "Missing Maps: crowdsourcing digital map creation," Global Innovation Exchange in collaboration with the United Nations, http://stisolutions4sdgs.globalinnovationexchange.org/innovations/missing-maps-crowdsourcing-digital-map-creation

[51] See "Multi-stakeholder Forum on Science, Technology and Innovation for the SDGs (STI Forum), 2017," STI Forum, https://sustainabledevelopment.un.org/TFM/STIForum2017. The STI Forum's objective is to facilitate discussions and multistakeholder partnerships to identify technology needs and to help transferring relevant technology for the Sustainable Development Goals. This platform is open to member states, UN organizations, civil society, academia, industry and private sector. The UN Inter-Agency Task Team on STI for SDGs (IATT) prepares

to maps so that governments and humanitarian organizations have accurate data during natural disasters or epidemics. Although not explicitly an AI issue, map data is critical to many AI technologies.[52]

Additionally, because many AI challenges will require solutions that span sectors and fields, broad-based multistakeholder systems can help ensure that the comprehensive solutions can be both developed and implemented with the support of all stakeholders. For example, the National Telecommunications and Information Administration (NTIA) has led several multistakeholder processes to help with a variety of complex challenges spanning cybersecurity,[53] unmanned aircraft systems,[54] Internet of Things,[55] and facial recognition technologies,[56] with the objective of developing best practices around those challenges.

(2) Soliciting Public Feedback on AI Policies

Another way that policymakers and regulators can engage diverse stakeholders on AI policies is through open calls for public feedback and dialogue about specific policy proposals. Such an approach can be particularly helpful in building trust between stakeholders and establishing legitimacy around proposed solutions. For example, when Brazil was considering the Marco Civil, an Internet civil rights frameworks, they used a variety of online tools between October 2009 to May 2010 to create a space for public consultation. Politicians, academics, artists, NGOs, private sector companies, individuals, and other stakeholders used online tools to blog, comment on, and debate the proposed legislative text.[57] Because of the openness of the process and the impact that stakeholder impacts had on the final text of the legislation "most stakeholders saw it as a uniquely legitimate piece of law."[58]

---

the work of the STI Forum by working with representatives from civil society, private sector and scientific community.

[52] Mimi Onouha, "Side-by-side images expose a glitch in Google's maps," Medum, June 6, 2017, https://qz.com/982709/google-maps-is-making-entire-communities-invisible-the-consequences-are-worrying.

[53] "Multistakeholder Process: Cybersecurity Vulnerabilities," National Telecommunications and Information Administration, December 15, 2016, https://www.ntia.doc.gov/other-publication/2016/multistakeholder-process-cybersecurity-vulnerabilities.

[54] "Multistakeholder Process: Unmanned Aircraft Systems," National Telecommunications and Information Administration, June 21, 2016, https://www.ntia.doc.gov/other-publication/2016/multistakeholder-process-unmanned-aircraft-systems.

[55] "Multistakeholder Process; Internet of Things (IoT) Security Upgradability and Patching," National Telecommunications and Information Administration, November 7, 2017, https://www.ntia.doc.gov/other-publication/2016/multistakeholder-process-iot-security.

[56] "Privacy Multistakeholder Process: Facial Recognition Technology," National Telecommunications and Information Administration, June 17, 2016, https://www.ntia.doc.gov/other-publication/2016/privacy-multistakeholder-process-facial-recognition-technology.

[57] See, e.g., Urs Gasser, *et al*, "Multistakeholder as Governance Groups."

[58] Carolina Rossini, Francisco Brito Cruz and Danilo Doneda, "The Strengths and Weaknesses of the Brazilian Internet Bill of Rights: Examining a Human Rights Framework for the Internet," Centre for International Governance Innovation and the Royal Institute of International Affairs, September 2015, https://www.cigionline.org/sites/default/files/no19_0.pdf, 7.

(3) Utilizing Diverse Forums for Information Exchange

Finally, there are numerous international forums that policymakers and regulators can utilize in order to learn from the experiences of peers who are struggling to address many of the same challenges. Policymakers and regulators can use these venues to learn from others' best practices around AI governance and identify areas where additional research or dialogue is needed. For example, the ITU's Global Symposium for Regulators (GSR) brings together the leaders of national telecommunication and information communications technologies regulatory authorities in order to share experiences and best practices regarding key regulatory challenges.[59] The World Summit on the Information Society is a multistakeholder platform for discussing ICT development.[60] And the Internet Governance Forum (IGF) is a UN-organized event that fosters dialogue around a variety of policy issues relating to Internet governance. These events are designed to foster dialogue, not implement solutions. While the IGF in particular has been criticized for being nothing more than a "talk shop," such information exchanges can be invaluable opportunities to learn from peers, particularly on emerging and complex technical issues.[61]

IV. Tools for Bridging the Digital Divide

Despite the opportunities AI technologies may offer, there is a real risk that—without thoughtful intervention—they may in fact exacerbate structural, economic, social, and political imbalances, and further reinforce entrenched inequalities. For regulators and policymakers around the world, uneven access to technology remains a major concern because of its potential impacts on social and economic inequality.[62] Although AI technologies can have global impacts, development has often been limited both geographically and sectorally, with a small number of companies driving forward these technologies with little input from different industries, disciplines, social classes, cultures, and countries. For that reason, there is a risk that increased reliance on AI may have unintended consequences that aggravate current disparities, particularly in countries that rely on industries at risk of being automated.[63] However, with

[59] "Global Symposium for Regulators," International Telecommunications Union, https://www.itu.int/en/ITU-D/Conferences/GSR/Pages/GSR.aspx.
[60] WSIS Forum 2018 (last visited June 15, 2018), https://www.itu.int/net4/wsis/forum/2018/.
[61] See, e.g., "Center for Democracy & Technology comments on IGF Istanbul and the future of the IGF," Center for Democracy & Technology, https://cdt.org/files/2014/10/CDT-comments-on-IGF-Istanbul-and-future-of-the-IGF.pdf.
[62] "ITU releases 2017 global information and communication technology facts and figures," ITU News, July 31, 2017, http://news.itu.int/itu-releases-2017-global-information-and-communication-technology-facts-and-figures/.
[63] Florence Jaumotte, Subir Lall, and Chris Papageorgiou, "Rising Income Inequality: Technology, or Trade and Financial Globalization?" International Monetary Fund, July 2008, https://www.imf.org/external/pubs/ft/wp/2008/wp08185.pdf, 16 ("[F]or economies that rely on low-skilled labour, automation could challenge their competitive advantage in the global labour market and exacerbate local unemployment challenges, impacting economic development.").

the help of strategic policies, AI technologies might be harnessed to overcome the persistent challenges posed by unequal access.

The use of AI to develop solutions to address inequalities is promising but also raises novel challenges for policymakers and regulators. For example, many companies are working to use AI and machine learning to provide alternative credit scoring to the unbanked in developing countries, using data gathered from cell phone usage to assess credit-worthiness.[64] While this has the potential to facilitate access to credit for many individuals, the quality of the input data depends on existing infrastructure, digital literacy, and cultural practices underlying cell phone usage. Innovative applications of AI technologies, such as alternative credit scoring, will inherit both the old challenges of the digital divide, such as disparities in digital infrastructure and technological literacy,[65] and face AI-specific challenges, such as a lack of representative data sets.[66]

Guiding Principles

Bridging the digital divide is not a new challenge for policymakers and regulators, but increased reliance on AI creates a greater sense of urgency. In prioritizing areas of focus, there are a few high-level principles that can guide policymakers and regulators:

- **Do not simply accept the status quo.** AI technologies may offer tremendous economic opportunities to entrenched, incumbent companies, just as they might reinforce existing biases and power relationships. But in this period of technological disruption, policymakers and regulators do not need to accept the status quo. Regulators and policymakers are in a privileged position to nudge how the power and economic gains of AI are shared.
- **Prioritize broad-based access to technology.** AI technologies are exacerbating existing needs for computers, broadband Internet, and data. In addressing these needs, policymakers and regulators should take an inclusive approach. For example, AI may have important implications for crop yields and resiliency, so policies that work to foster entrepreneurial ecosystems, such as funding for incubators, should consider regions beyond urban centers. Equity should be considered at every step: ICT regulators can work to close the digital divide by prioritizing equal access to a broad set of technologies with every policy decision.
- **Focus on entrepreneurship and innovation, not AI.** AI is simply one tool--one of many--that innovative companies will apply to build the next generation of new industries and jobs. Everything from intellectual property laws, the legality of non-compete agreements, educational

---

[64] Catherine Cheney, "How alternative credit scoring is transforming lending in the developing world," Devex, September 8, 2016, https://www.devex.com/news/how-alternative-credit-scoring-is-transforming-lending-in-the-developing-world-88487.

[65] Ben Shenglin, Felice Simonelli, Romain Bosc, Ruidong Zhang, and Wenwei Li, "Digital infrastructure: Overcoming Digital Divide in Emerging Economies," G20 Insights by Kiel Institute for the World Economy, January 15, 2018, http://www.g20-insights.org/policy_briefs/digital-infrastructure-overcoming-digital-divide-emerging-economies/. "[W]ithout proper education and skill training, the potential of digital technology cannot be fully tapped."

[66] See Emmanuel Letouzé, "Big Data for Development: Challenges & Opportunities," UN Global Pulse, Executive Office of the Secretary-General United Nations, May 2012, http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf, 42 (highlighting the challenges of using big data for development); Algorithmic Justice League, "Algorithmic Justice League," Algorithmic Justice League, https://www.ajlunited.org (highlighting bias in ML data sets).

opportunities, broadband access and more can affect innovation and entrepreneurship. The challenge for policymakers and regulators is that some of these factors may not be within their jurisdiction, and others may require significant financial investments, so cooperation and partnerships will be key.

In order to apply these principles into everyday practice, policymakers can work to close the digital divide by (1) supporting buildout of physical infrastructure, (2) supporting local ecosystems of entrepreneurship and start-ups, and (3) supporting capacity development at universities and through other training systems.

(1) Supporting Buildout of Physical Infrastructure

Disparities in the physical access to Internet infrastructure are a central source of the digital divide.[67] Because of the amount of data that many AI technologies require, the quality of digital infrastructure may impact the speed at which AI technologies can be deployed. For example, because a single autonomous vehicle will produce an estimated 4 terabytes of data every 90 minutes, AV systems rely on robust broadband Internet infrastructure.[68] Beyond broadband infrastructure, AI will also demand robust cloud capacity, electricity, and more. Creating this infrastructure will place significant demands on both private industry and the public sector. In working toward this infrastructure, policymakers and regulators will need to consider both funding and security:

V. **Infrastructure funding**. Direct investment can take several forms. One way is for the public sector to bear the brunt of the costs. Although it can be a risky and capital-intensive approach, because decisions are not driven by profit motives, policymakers and regulators can ensure that infrastructure is deployed in a manner that benefits all citizens, and can yield cheaper, higher-quality service than is available from the private sector. For example, when major telecommunications companies refused to provide broadband service to the town of Concord, Massachusetts, the town developed its own service at a cost of USD $3.9 million, which has begun to generate revenue for the town.[69] Such investments can be made easier when combined with other government infrastructure projects such as roads and electric improvement.[70] The other extreme is that the private sector can bear the brunt of costs, which

[67] See ITU, "Final Report: World Telecommunications Development Conference (WTDC-17)," 2018, https://www.itu.int/en/ITU-D/Conferences/WTDC/WTDC17/Documents/WTDC17_final_report_en.pdf; Calestous Juma, "How Can Africa Master the Digital Revolution?" Belfer Center for Science and International Affairs/World Economic Forum, April 1, 2016, https://www.belfercenter.org/publication/how-can-africa-master-digital-revolution. "Africa lags behind other regions in its use of core digital platforms such as the internet," partly due to the high prices for broadband resulting from undeveloped digital infrastructure." See also "2016 Broadband Progress Report," Federal Communications Commission, January 29, 2016, https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2016-broadband-progress-report. More than 34 million American still lack access to high-speed internet connectivity, 23 million of which are in rural areas.

[68] Kathy Winter, "For Self-Driving Cars, There's Big Meaning Behind One Big Number: 4 Terabytes," Intel Newsroom, April 14, 2017, https://newsroom.intel.com/editorials/self-driving-cars-big-meaning-behind-one-number-4-terabytes.

[69] David Talbot, Waide Warner, Susan Crawford, Jacob White, "Citizens Take Charge: Concord, Massachusetts, Builds a Fiber Network," Municipal Fiber Project, February 2017, https://dash.harvard.edu/bitstream/handle/1/30201055/2017-01_broadband.pdf?sequence=5.

[70] See UNESCAP, "A Study on Cost-Benefit Analysis of Fibre-Optic Co-Deployment with the Asian Highway Connectivity," Asia Pacific Information Superhighway (AP-IS) Working Paper Series, April 2018,

limits the ability of regulators and policymakers to exert influence over the way infrastructure is distributed. That said, policymakers and regulators do have some levers of influence, in particular when they can attach conditions through competitive bidding processes. For example, when Brazil auctioned off high-value LTE spectrum assignments in 2012, it tied ownership rights to rural coverage obligations.[71] Such conditions, however, must be carefully balanced so as not to deter the investments in the first place. For example, after a history of successful spectrum auctions, India had challenges selling off spectrum allocated for auction due to setting auction prices above well above market values in 2016.[72]

In between those extremes is the use of public-private partnerships (PPPs) to share both the burdens and benefits of digital infrastructure investments. Mats Granryd, Director General of the GSMA has stressed the important of PPPs for bridging the digital divide, noting that "it is more important than ever that governments and industry work together to ensure that all citizens benefit from this new era of hyper-connectivity."[73] Facilitating these partnerships sometimes requires legal and policy changes. For example, in 2005 Nigeria passed the Infrastructure Concession Regulatory Commission (ICRC) Act, and in 2009 approved the National Policy on PPPs, which together set clear guidelines for "project identification, evaluation, and selection," as well as "procurement, operation, maintenance, and performance monitoring."[74] While adopting of PPPs was slow in Nigeria,[75] this legislation has led to a rise in ICT PPPs in Nigeria over time.[76] Similarly, the World Economic Forum has launched the Internet for All initiative, aiming to bring Internet connectivity to 75 million people in Africa's Northern Corridor, and represents a collaboration between industry and the countries of Ethiopia, Kenya, Rwanda, South Sudan, and Uganda.[77]

---

http://www.unescap.org/sites/default/files/Cost-benefit%20analysis%20of%20fOC%20with%20Asian%20Highway.pdf, 10.

[71] "Delivering Digital Infrastructure Advancing the Internet Economy," World Economic Forum and Boston Consulting Group, April 2014, http://www3.weforum.org/docs/WEF_TC_DeliveringDigitalInfrastructure_InternetEconomy_Report_2014.pdf, 42.

[72] Salman SH, "2016 Spectrum Auction ends: Rs 65,789 Cr in bids, only 40% of spectrum sold," Medianama, October 7, 2016, https://www.medianama.com/2016/10/223-2016-spectrum-auctions-ends.

[73] "UN Broadband Commission Meets in Rwanda to Tackle Digital Divide," ITU News, May 8, 2018, http://news.itu.int/un-broadband-commission-brings-solutions-broadband-digital-connectivity-all/.

[74] Chidi Izuwah, "Nigeria blazes the trail for PPP disclosures with new web portal," World Bank Group Infrastructure and Public-Private Partnerships Blog, September 21, 2017, http://blogs.worldbank.org/ppps/nigeria-blazes-trail-ppp-disclosures-new-web-portal.

[75] Abdul Ganiyu Otairu, Abdullah A.Umar, Noor Amila Wan Abdullah Zawawi, Mahmoud Sodangi, and Dabo B.Hammad, "Slow Adoption of PPPs in Developing Countries: Survey of Nigerian Construction Professionals," Procedia Engineering, Volume 77, 2014, https://www.sciencedirect.com/science/article/pii/S1877705814009916, 194.

[76] "Sub-Saharan Africa Private Participation in Infrastructure Database Regional Snapshot," the World Bank, https://ppi.worldbank.org/snapshots/region/sub-saharan-africa.

[77] "Internet for All: A Key Initiative for Africa's Digital Transformation," press release, World Economic Forum, May 10, 2016, https://www.weforum.org/press/2016/05/internet-for-all-a-key-initiative-for-africa-s-digital-transformation/.

VI. **Cybersecurity**. Cybersecurity is a threat around the world,[78] but especially in developing countries experiencing rapid growth in digital infrastructure.[79] Because of the importance and sensitivity of data used in AI systems, these vulnerabilities will only be amplified with new AI technologies.[80] For that reason, policymakers should consider cybersecurity as a key component of all digital infrastructure projects. Fundamentally, this will require the establishment of legal frameworks protect privacy and allow for the redress of harm.[81] While regional entities, such as the African Union, have taken steps to establish such frameworks,[82] policymakers should also support legislation domestically.

(2) Supporting Local Ecosystems of Entrepreneurship and Startups

An important way for policymakers and regulators to address the digital divide is to promote entrepreneurship at home.[83] The entrepreneurial ecosystem concept is helpful for understanding conditions that are conducive to a vibrant market for innovation. The mass of perspectives and ideas that grow out of a successful entrepreneurial ecosystem will be crucial for the development of AI technologies in emerging markets. There are several approaches policymakers and regulators can try in order to entrepreneurial ecosystems that will enable AI development:

(a) **Government programs**. Government initiatives have shown promise in facilitating the growth of entrepreneurial ecosystems. For example, the Singapore government has a series of direct investment initiatives,[84] as well as the Startup SG program, which combines government investment with private equity. Similar efforts have occurred in the developing world. For example, the Botswana government established the Botswana Innovation Hub in 2012, which includes a campus to house companies that can benefit from seed funding provided by the Botswana Innovation Fund. The institutional structure provided by the Botswana Innovation Hub

[78] "Significant Cyber Incidents," Center for Strategic & International Studies, https://www.csis.org/programs/cybersecurity-and-governance/technology-policy-program/other-projects-cybersecurity.
[79] Mirko Hohmann, Alexander Pirang, and Thorsten Benner, "Advancing Cybersecurity Capacity Building: Implementing a Principle-Based Approach," Global Public Policy Institute, March 2017, http://www.gppi.net/fileadmin/user_upload/media/pub/2017/Hohmann__Pirang__Benner__2017__Advancing_Cybersecurity_Capacity_Building.pdf, 8-9.
[80] Barry Carin, "G20 safeguards vulnerabilities of digital economy, with financial sector focus," G20 Insights by Kiel Institute for the World Economy, November 20, 2017, http://www.g20-insights.org/policy_briefs/g20-safeguards-vulnerabilities-digital-economy-financial-sector-focus.
[81] Lilly Pijnenburg Muller, "Cyber Security Capacity Building in Developing Countries: Challenges and Opportunities," Norwegian Institute of International Affairs, 2015, https://brage.bibsys.no/xmlui/bitstream/id/331398/NUPI+Report+03-15-Muller.pdf, 9.
[82] "African Union Convention on Cyber Security and Personal Data Protection," African Union, June 27, 2014, https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection.
[83] "Digital Policy Playbook 2017: Approaches to National Digital Governance," World Economic Forum, September 2017, http://www3.weforum.org/docs/White_Paper_Digital_Policy_Playbook_Approaches_National_Digital_Governance_report_2017.pdf, 35.
[84] *See* "Grants," SME Portal, https://www.smeportal.sg/content/smeportal/en/moneymatters/grants.html.

has also attracted outside investment.[85]

(b) **Technology business incubators**. Business incubators are spaces that provide needed services to entrepreneurs, such as access to technology and marketing assistance, and are usually run for profit or as part of a university program. Incubators have served as important catalysts for entrepreneurial ecosystems, such as that in Bangalore, India.[86] One example of a successful incubator is SmartXChange in Durban, South Africa, having created over 3,000 jobs since its founding in 2004 and generating and average of six startups each year.[87]

(c) **Commercial transfer of research**. A symbiotic relationship between industry and academia is crucial for any entrepreneurial ecosystem. One way that policymakers can encourage such a link is by supporting legislation that facilitates the commercialization of research, such as laws that vest intellectual property rights with the institutions that fund research. For example, in 2009 Russia passed Federal Law 217, which allows universities to use IP generated with public funds to create private start-up companies, leading to the creation of 973 startups by June 2011.[88] The Bayh Dole Act (1980) in the United States has had similarly positive results, with more than 5,000 new companies founded since its passage based on university funded research.[89] IP frameworks that allow for commercialization of research both incentivize investment in future research and increase the number of startups working on cutting edge issues. Additionally, startups that emerge from university research create a pipeline for employment for university students and foster information exchange between educational institutions and the private sector. Together, these factors can contribute to a vibrant entrepreneurial ecosystem, which will be beneficial for a region attempting to break into the market for AI technologies.

(3) Supporting Capacity Development

Another critical opportunity for policymakers working to mitigate the digital divide is to encourage technological innovation through academic capacity development. Universities are particularly suited to serve as centers of capacity development by serving as the locus for investment in research, data collection, and training, given both their institutional function to further research in societal interest, as well as existing substantive capacity and expertise related to AI technologies.[90] That said, capacity development often occurs outside of the university setting, where training can be more targeted and efficient. For example, the Centers of Excellence in the ITU Academy, which provide important ICT training, are a mix of both universities and small training centers.[91] Similarly, the ITU's Digital Skills Toolkit highlights the role that

[85] Matshelane Mamabolo, "Botswana Innovation Hub to host US $3 million development programme," IT Web Africa, Aug. 21, 2017, http://www.itwebafrica.com/business-intelligence/508-botswana/239590-botswana-innovation-hub-to-host-us3-million-development-programme.
[86] Subrahmanya, "Comparing the Entrepreneurial Ecosystems," 55.
[87] Mail & Guardian
[88] Juan Julio Gutierrez and Paulo Correa, "Commercialization of Publicly Funded Research and Development (R&D) in Russia: Scaling up the Emergence of Spinoff Companies," Policy Research Working Paper, World Bank, Nov. 2012, https://openknowledge.worldbank.org/bitstream/handle/10986/19932/wps6263.pdf?sequence=1, 21.
[89] Catherine Kirby, "True Impact of Bayh-Dole Act," The McNair Center for Entrepreneurship & Innovation at Rice University's Baker Institute, Dec. 6, 2016, http://mcnair.bakerinstitute.org/blog/true-impact-bayh-dole-act/.
[90] "Capacity-building," Academic Impact, United Nations, https://academicimpact.un.org/content/capacity-building.
[91] ITU Academy (last visited June 15, 2018), https://academy.itu.int/index.php?lang=en.

that primary and secondary schools play in building digital skills and capacities.[92] Key considerations for capacity development include:

(a) **Capacity development should focus on both soft and hard skills**. In order to adapt to a labor market that integrates AI technologies, the next generation of workers will need to develop a broad range of new skills, as well as the ability to evolve skill sets throughout their careers at a rate previously unseen.[93] While technical or hard skills are crucial, equally important are business, entrepreneurial, and other soft skills (e.g., team working, curiosity, and communication).[94] AI technologies, instead of replacing occupations, will likely result in a greater emphasis on social skills within existing occupations that the technologies cannot replicate.[95] Accordingly, policymakers and regulators might consider soft skills capacity development through training programs. For example, the BEIGE Foundation in Ghana launched the BEIGE Talent initiative in 2015 to train recent university graduates in soft skills necessary to succeed in the workforce, from personal goal setting to customer care.[96] Moreover, both hard and soft skills development is particularly important for groups that have been traditionally underrepresented in the workforce, particularly women.[97] Policymakers and regulators should consider programs that target such groups to ensure that the benefits of an AI integrated economy are evenly distributed; this includes programs and policies to address the cases where education and training is insufficient to address labor displacement. In addition to skill building programs for populations, individual countries can invest in developing and retaining top expertise to promote R&D; for instance, the Canadian Institute for Advanced Research (CIFAR) recently announced a major investment in academic capacity for the country,[98] and Canada is developing programs such as the 150 Research Chairs Program to develop and attract leaders in the field.[99]

(b) **Ensure investments in educational capacity development are evenly distributed.** Policymakers and regulators should work to ensure that investments in capacity development are accessible to a broad range of students.[100] Concentrating resources in a small number of cities or universities, may limit the benefits of capacity development efforts to only those that have access to those institutions. Providing more resources to more institutions may require alternative

---

[92] "Digital Skills Toolkit," ITU, 2018, https://www.itu.int/en/ITU-D/Digital-Inclusion/Documents/ITU%20Digital%20Skills%20Toolkit.pdf.
[93] "Across nearly all industries, the impact of technological and other changes is shortening the shelf-life of employees' existing skill sets." "The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industry Revolution - Executive Summary," World Economic Forum, Jan. 2016, http://www3.weforum.org/docs/WEF_FOJ_Executive_Summary_Jobs.pdf, 3.
[94] "ITU-Academia Partnership Meeting: Developing skills for the digital era - Final Report," ITU Academy, Sept. 21, 2017, https://www.itu.int/en/ITU-D/Capacity-Building/Documents/ITU-Academia%20Partnership%20Meeting%202017/FinalReportITUAcademiaPartnershipMeeting201729Sept17.pdf, 4.
[95] "The Future of Jobs," World Economic Forum, 3.
[96] "The Soft Skills Imperative," The Adecco Group, Jan. 2017, https://www.adeccogroup.com/wp-content/themes/ado-group/downloads/the-adecco-group-white-paper-the-soft-skills-imperative.pdf, 10.
[97] Ibid., 5.
[98] Canadian Institute for Advanced Research, "Canada funds $125 million Pan-Canadian Artificial Intelligence Strategy," news release, Mar. 22, 2017, https://www.newswire.ca/news-releases/canada-funds-125-million-pan-canadian-artificial-intelligence-strategy-616876434.html.
[99] Ibid.
[100] ITU-Academia Partnership Meeting, ITU, 4.

funding approaches, such as those that Nigeria experimented with in 1998 and 2003, which allowed for non-governmental sources of funding.[101] Additionally, educational resources can be leveraged through coordinated and collaborative networks, such as the Global Network of Internet & Society Centers, which can act to fill capacity gaps, bridge disciplinary divides and facilitate meaningful interaction amongst diverse communities, and translate research into action by creating high quality channels of information for policy makers.[102]

(c) **Invest in Centers of Excellence (CoEs).** CoEs, such as those that are a part of the ITU Academy, are a structural alternative for facilitating capacity development within universities.[103] While CoEs can take many different forms, in the university context they are usually committed to research on a single topic and are often financially distinct entities within institutions. The specificity of subject matter enabled by the CoE design might be useful for capacity development focused on highly complex technologies such as AI. Additionally, the financial independence of CoEs may be particularly effective for capacity development in developing countries, where universities are often faced with budgetary constraints.[104] CoEs also provide an attractive format for investment from both the private sector[105] and international organizations.[106] One notable example includes the European Commission's recent announcement of a series of mechanisms to significantly increase capacity and investment in AI, including the formation of joint research excellence centers that strengthen coordinated research endeavors across countries.[107]

1. Addressing Information Asymmetries
2. Building Public-Private Partnerships
3. Bridging the Digital Divide
4. Sustaining a Competitive Environment

[101] Ibid., 23.
[102] "Global Network of Internet and Society Research Centers," http://networkofcenters.net/.
[103] Tomas Hellstrom, "Centres of Excellence as a Tool for Capacity Building - Draft Synthesis Report," Programme on Innovation, Higher Education and Research for Development (IHERD), OECD, https://www.oecd.org/sti/Draft_OECD%20synthesis%20report_final.pdf, 4,
[104] Ibid., 9.
[105] In March of 2018, the Indian Institute of Technology Kharagpur announced that it will set up an AI CoE with the help of an investment of more than $860,000 from Capillary Technologies Ltd. Subhankar Chowdhury, "IIT to have centre of excellence in AI," The Telegraph, Mar. 3, 2018, https://www.telegraphindia.com/calcutta/iit-to-have-centre-of-excellence-in-ai-212835.
[106] In 2014 the World Bank approved $150 million to finance 19 university-based CoEs in West and Central Africa to promote studies in STEM-related disciplines, agriculture, and health. The World Bank, "World Bank to Finance 19 Centers of Excellence to Help Transform Science, Technology, and Higher Education in Africa," press release, Apr. 15, 2014, http://www.worldbank.org/en/news/press-release/2014/04/15/world-bank-centers-excellence-science-technology-education-africa.
[107] "Communication Artificial Intelligence for Europe," European Commission, https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe, 8.

VII. Tools for Building and Sustaining a Competitive Environment

When France announced its AI strategy in March 2018, advisors to Emmanuel Macron noted that the country was lagging behind the US and China's AI prowess and would be unable to match those countries' investments in AI research, even while announcing a EUR 1.5 billion investment into AI research.[108] Similarly, the European Commission recently committed EUR 1.5 billion in research and innovation in AI technologies, the creation of AI research excellence centers, and more.[109] For those policymakers and regulators seeking to enhance their country's AI technologies, the task of becoming competitive can appear daunting, particularly with limited financial resources. And financial investments are not the only barriers to competition. Because AI is so data-intensive, network and lock-in effects can further complicate efforts to compete; companies like Google and Facebook train AI on their massive datasets and then use that AI to improve their products and attract more users and data.

In such an environment, policymakers and regulators face a difficult task of trying to build and sustain a level playing field for those developing and deploying AI-based tools. That said, creating a fertile, competitive environment for innovation has long been an important role for ICT regulators.[110] Moreover, although investments in AI research are important, they are not the only nor even the most important tool that policymakers and regulators have at their disposal for building and sustaining a competitive innovation landscape. The recent EU Communication on AI reflects this, highlighting financial investments, but also discussing education and training systems, as well as ethical and legal frameworks. It is this latter element where policymakers and regulators have the greatest opportunity to lay a foundation for supporting AI innovation. Of course, no set of approaches will instantly make a country competitive with China or the US in AI development, but these approaches may help support local AI innovation.

Well-crafted legal and technical frameworks can have an important influence on the competitiveness of the AI landscape. Because AI technologies are so dependent on data, that means that intellectual property laws can have a significant impact on how easy it is to develop new AI applications. The importance of data for AI means privacy and open data frameworks can mitigate lock-in effects. And similarly, technical and legal interoperability frameworks can impact the ease by which AI technologies can be used across national and regional boundaries.

Guiding Principles

- **Experiment with policy and support technical experimentation.** Just as AI is still being experimented with, so too should AI policy. It is tempting to think that the options are binary: either do not regulate or try to craft a complete regulation today. But instead, policymakers and regulators can create spaces that allow them to experiment in an iterative fashion with policies

[108] Nicholas Vinocur, "Macron's €1.5 billion plan to drag France into the age of artificial intelligence," Politico, April 14, 2018, https://www.politico.eu/article/macron-aims-to-drag-france-into-the-age-of-artificial-intelligence/.
[109] "Communication Artificial Intelligence for Europe," European Commission, https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.
[110] *See, e.g.,* "Competition and Price: Market Regulation for Information and Communications Technologies," ICT Regulation Toolkit, ITU, April 2012, http://www.ictregulationtoolkit.org/toolkit/2.

and regulatory approaches, that still allow for the development of new AI technologies, while still advancing core values of public safety, privacy, consumer protection, and due process.

- **Favor principles over rules.** With the pace of AI technological development, a rules-based approach to governance can become outdated before it even has a chance to take effect. Instead of rigid rules (e.g., vehicles must follow at a minimum distance of 200 feet), policymakers and regulators can adopt more flexible principles (e.g., vehicles must follow at a safe distance), which are more resilient to technical changes, without necessarily sacrificing policy objectives.[111]
- **Emphasize data sharing, collection, and measurement.** Data is important to AI competitiveness in several ways. First, data is an essential ingredient for machine learning, and fostering the collection and sharing of high-quality open datasets can boost the creation of AI technologies. Second, data is important to AI competitiveness because crafting effective governance approaches is difficult without a clear sense of the problem, and it is difficult to define the problem in the absence of baseline measurements and the ability to measure changes over time.

In putting these principles into action there are a range of tools that policymakers and regulators can promote a level playing field. In particular, by reducing legal frictions, reducing technical frictions, and promoting data sharing and metrics, policymakers and regulators can lay a foundation for a innovative, competitive AI industry.

(1) Reducing Legal Frictions for AI Innovation

One way for policymakers and regulators to level the playing field for local AI innovation is by carefully considering the entire legal and regulatory landscape and updating legal barriers that are based on outdated assumptions or norms and that are hindering AI development. Such obstacles might include intellectual property laws that limit smaller innovators from developing new AI technologies, and unclear or inflexible regulations that cannot adapt quickly enough to new, innovative AI applications. That does not mean that regulation of AI technologies is unnecessary, as the technology can create significant physical, economic, and social harms for which regulation may be necessary. What it does mean is that policymakers and regulators must try to optimize across several different values, including innovation, as they consider tradeoffs between potential AI governance interventions.

(a) Optimizing Intellectual Property Rules. When balancing across these myriad values, one dial that policymakers and regulators can adjust relates to the permissiveness (or lack thereof) of their intellectual property rules. Overly restrictive intellectual property rules can limit the opportunities for smaller AI startups, whereas overly permissive intellectual property rules can reduce incentives for innovation. Policymakers and regulators must work to deploy "IP in the right dosage" as the World Intellectual Property Organization describes the middle ground between anti-competitive, broadly extended IP and unprotected, underemployed IP.[112] As the

[111] For a good summary of the tradeoffs between principle and rule-based regulations, see Chris Brummer and Daniel Gorfine, "Fintech: Building a 21st-Century Regulator's Toolkit," Center for Financial Markets, Milken Institute, October 2014, http://assets1b.milkeninstitute.org/assets/Publication/Viewpoint/PDF/3.14-FinTech-Reg-Toolkit-NEW.pdf.

[112] "IP and Competition Policy," WIPO, http://www.wipo.int/ip-competition/en/ (last accessed May 2, 2018).

European Commission Communication on AI stated "Reflection will be needed on interactions between AI and intellectual property rights, from the perspective of both intellectual property offices and users, with a view to fostering innovation and legal certainty in a balanced way."[113]

One area of IP policy in which policymakers and regulators can add legal certainty is with regard to the IP rights of AI-generated outputs. Currently there is little guidance about whether to grant copyright protection to programmers, the public domain, or the AI itself.[114]

(b) Enhancing Regulatory and Technical Experimentation. Policymakers and regulators can create spaces that enable innovators to explore new applications without fear of legal punishment. Such spaces also allow policymakers and regulators to learn about emerging technologies, and develop new regulatory approaches. For example, since 2016 the Hong Kong Monetary Authority has used the Fintech Supervisory Sandbox in order to enable banks and technology firms to pilot nearly 30 fintech products without fully complying with applicable regulations.[115] Similarly, the UK's Financial Conduct Authority at the end of 2017 announced their third cohort of 18 business invited to test their products and services within their sandbox.

In other cases, simply providing greater clarity about regulatory standards can enable experimentation and innovation. A 2015 European Parliament report, for example, discussed the importance of developing standards regarding AI that would "provide predictable and sufficiently clear conditions under which enterprises could develop applications and plan their business models on a European scale."[116]

(2) Reducing Technical Frictions for AI Innovation

In addition to removing legal frictions, policymakers and regulators can also support AI innovation by reducing technical frictions. In particular, by facilitating data exchange and through establishing technical standards, policymakers and regulators can enable interoperability. Although there are notable exceptions, interoperability can in many conditions be a key driver of innovation in AI just as it is within the ICT context.[117]

---

[113] "Communication Artificial Intelligence for Europe," European Commission, https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.
[114] Andres Guadmuz, "Artificial intelligence and copyright," WIPO Magazine, October 2017, http://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html. For a primer on AI-generated art, see Jessica Fjeld and Mason Kortz, "A Legal Anatomy of AI-generated Art: Part I," JOLT Digest, November 21, 2017, http://jolt.law.harvard.edu/digest/a-legal-anatomy-of-ai-generated-art-part-i.
[115] "Fintech Supervisory Sandbox (FSS)", Hong Kong Monetary Authority, March 27, 2018, http://www.hkma.gov.hk/eng/key-functions/international-financial-centre/fintech-supervisory-sandbox.shtml.
[116] Motion for a European Parliament Resolution A8-0005/2017 by the Committee on Legal Affairs of the European Parliament, January 27, 2017, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//EN.
[117] Urs Gasser and John Palfrey, "Breaking Down Digital Barriers: When and How ICT Interoperability Drives Innovation," Berkman Publication Series, November 2007, https://cyber.harvard.edu/interop/pdfs/interop-breaking-barriers.pdf.

(a) Supporting Data Openness. The sharing of both technical data and training data can facilitate AI innovation by making it easier for startups and smaller competitors to enter the AI market. For example, Google has published the source code for their TensorFlow AI platform, which has been used by Airbnb, Uber, SAP, Snapchat, Qualcomm, and others in building their own AI systems.[118] Even more helpful can be the sharing of high-quality training data, which can be used in building AI systems. For example, the 2018 EU Communication on AI calls for the wider availability of privately held data. Of course, making private sector data sets public can create legal risks if the data contains personal or other sensitive data, necessitating legal protections in order to incentivize data sharing.[119] This is why the EU Communication on AI calls for a "new support centre for data sharing will provide public authorities and companies with legal and technical support when trying to access data from public sector bodies and companies."[120]

Another way to facilitate data sharing is through data harmonization efforts, increasing the ease by which data can be shared and used. One way policymakers and regulators can promote data standards is by adopting such standards internally. For example, the UK's Cabinet Office created open standard principles to develop common and secure IT infrastructure through agreed and open standards.[121]

(b) Supporting Technical Standards. The use of technical standards can also promote interoperability for developing a competitive AI landscape. Technical standards like single-sign on digital ID infrastructure, simplifies technical developments and encourages users to sign up for new new services.[122] Even something as simple as the EU mandate for a common standard for mobile phone chargers promotes greater competition in the mobile phone market by easing the costs of switching devices.[123] In the AI context, initiatives like the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems is working to create AI specific technical and ethical standards.[124] And the ITU-T Focus Group on Machine Learning for Future Networks including 5G is preparing technical reports and draft protocols that determine how AI can be used for things like network traffic management.[125]

(3) Developing metrics and tools for measurement of AI Impacts

---

[118] "TensorFlow," https://www.tensorflow.org (last accessed May 2, 2018).
[119] "Global Agenda Council on Cybersecurity," World Economic Forum, April 2016, http://www3.weforum.org/docs/GAC16_Cybersecurity_WhitePaper_.pdf, 18.
[120] "Communication Artificial Intelligence for Europe," European Commission, https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.
[121] "Policy Paper: Open Standards principles," UK Government, April 5, 2018, https://www.gov.uk/government/publications/open-standards-principles/open-standards-principles.
[122] Urs Gasser, "GSR discussion paper: Interoperability in the digital ecosystem," ITU, June 25, 2015, https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2015/Discussion_papers_and_Presentations/Discussionpaper_interoperability.pdf, 15.
[123] Ibid., 9.
[124] "Ethics in Action: The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems," IEEE, https://ethicsinaction.ieee.org (last accessed May 2, 2018).
[125] ITU-T, "Focus Group on Machine Learning for Future Networks including 5G," (last visited June 15, 2018), https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx.

Metrics and data are critical to the process of identifying the challenges of AI technologies, and developing the appropriate governance responses. The AI Index is one approach to addressing this challenge. The AI Index is a not-for-profit project that aims to collect data about the uses and progress of AI. In their 2017 Annual Report they observed that "without the relevant data for reasoning about the state of AI technology, we are essentially 'flying blind' in our conversations and decision-making related to AI."[126]

One key area in which metrics are important relate to the UN's Sustainable Development Goals, which represent 17 goals centered around ending poverty, protecting the environment, and ensuring widespread prosperity.[127] Increasingly, it is recognized that AI may have significant impacts on these SDGs, either as an enabler or a challenge.[128] As a result it is important to collect and make accessible data that enables policymakers and regulators to benchmark, observe, and respond to the impacts that AI is having in these important areas. For example, the UN has launched the Open SDG Data Hub, which currently includes data on 132 of the global indicators and over 460,000 records. Similarly, expert workshops focused issues of AI impact measurement in Asia has noted the importance of collecting, tracking, and making available AI-related data in areas such as employment, diversity and inclusion, and education.[129] Policymakers and regulators have an important role to play in helping to enable this data generation, collection, and sharing, starting with data available within government. This is critical for AI governance, because "the data that are easy to get may not be the most informative."[130] In particular, opening up additional data from global south and developing countries will provide important insights for AI governance that is currently lacking.

VIII. Conclusion

This report has identified several sets of tools that that policymakers and regulators can deploy to develop approaches to enhance AI's positive impacts and restrain its negative impacts. Some of the tools discussed above present inherent contradictions. Some involve greater regulation of AI technologies, others involve less. Some involve local action, others international. And some involve unilateral action, while other tools defer to collective, multistakeholder processes. These contradictions are not a flaw in the framework we have described; instead, they reflect the reality that governing often involves difficult tradeoffs.

The tools described above can help policymakers and regulators better understand the challenges that AI poses, and can help in developing innovative governance approaches to address those challenges. But there is no shortcut for resolving these tradeoffs, this framework is designed to help position

---

[126] "2017 Annual Report," Artificial Intelligence Index, November 2017, http://cdn.aiindex.org/2017-report.pdf, 59.
[127] "Sustainable Development Goals," UN, https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (last accessed May 2, 2018).
[128] XPRIZE, "How the AI XPRICE is helping achieve the SDGs," press release, ITU News, May 2, 2018, https://news.itu.int/how-the-ai-xprize-is-helping-achieve-the-sdgs/.
[129] "Global AI Dialogue Series: Observations from the China-US Workshop in Beijing (December 2, 2017)," Berkman Klein Center for Internet & Society at Harvard University, January 24, 2018, https://medium.com/berkman-klein-center/global-ai-dialogue-series-212279519169.
[130] "2017 Annual Report," Artificial Intelligence Index, November 2017, http://cdn.aiindex.org/2017-report.pdf, 51.

decisionmakers to be able to make those tradeoffs with greater confidence. No one has all of the answers when it comes to AI, but by experimenting with the tools identified here, decisionmakers will be better positioned to understand the problems, the potential approaches, and the necessary tradeoffs. Experimentation sometimes involves mistakes, but only through that process can policymakers and regulators ultimately help their constituents better adapt to AI's challenges and opportunities.

Appendix

Summary of Potential Approaches

---

1. **Addressing Information Asymmetries**
2. Building Public-Private Partnerships
3. Bridging the Digital Divide
4. Sustaining a Competitive Environment

Guiding Principles

- Create compelling opportunities for experts to join government.
- Reduce participatory friction for experts.
- Obtain hands-on experiences with AI technologies.

Potential Approaches

(1) Build internal capacity
  (a) Recruit individual expertise.
  (b) Build institutional expertise.
(2) Develop knowledge exchange interfaces with experts
  (a) Leverage academic expertise.
  (b) Activate stakeholder expertise.

---

1. Addressing Information Asymmetries
2. **Building Public-Private Partnerships**
3. Bridging the Digital Divide
4. Sustaining a Competitive Environment

Guiding Principles

- Develop a terminology, shared across all stakeholders.
- Take advantage of being a second (or third, or fourth, or…) mover.
- Keep an open door and an open mind.

Potential Approaches

(1) Incubate and engage in multistakeholder systems
  (a) Engage AI-specific multistakeholder systems.
  (b) Engage broad-based multistakeholder systems.
(2) Solicit public feedback on AI policies
(3) Utilize diverse forums for information exchange

---

1. Addressing Information Asymmetries
2. Building Public-Private Partnerships
3. **Bridging the Digital Divide**
4. Sustaining a Competitive Environment

Guiding Principles

- Do not simply accept the status quo.
- Prioritize broad-based access to technology.
- Focus on entrepreneurship and innovation, not AI.

Potential Approaches

(1) Support the buildout of physical infrastructure
  (a) Enable infrastructure funding.
  (b) Ensure privacy and cybersecurity.
(2) Support local ecosystems of entrepreneurship and startups
  (a) Develop government programs to support entrepreneurship.
  (b) Create technology business incubators.
  (c) Facilitate commercial transfer of research.
(3) Enhance capacity development
  (a) Focus on both soft and hard skills.
  (b) Ensure investments in educational capacity development are evenly distributed.
  (c) Invest in Centers of Excellence.

---

1. Addressing Information Asymmetries
2. Building Public-Private Partnerships
3. Bridging the Digital Divide
4. **Sustaining a Competitive Environment**

Guiding Principles

- Experiment with policy and support technical experimentation.
- Favor principles over rules.
- Emphasize data sharing, collection, and measurement.

Potential Approaches

(1) Reduce legal frictions for AI innovation
  (a) Optimize intellectual property rules.
  (b) Enhance regulatory and technical experimentation.
(2) Reduce technical frictions for AI innovation
  (a) Support data openness.
  (b) Support technical standards.
(3) Develop metrics and tools for measurement of AI Impacts

Artificial Intelligence (AI) for Development Series

Module on

AI, Ethics and Society

July 2018

Work in progress, for discussion purposes
Comments are welcome!
Please send your comments on this paper at: gsr@itu.int by 30 July 2018

This module was prepared by Michael Best, Director, United Nations University Institute on Computing and Society (UNU-CS), Professor, Sam Nunn School of International Affairs and the School of Interactive Computing, Georgia Institute of Technology, under the direction of the ITU/BDT Regulatory and Market Environment Division, in collaboration with the ITU/BDT Telecommunication Division and under close coordination with the Chief of the ITU/BDT Infrastructure, Enabling Environment, and E-Applications Department. We would like to thank the ITU General Secretariat and the ITU Standardization Bureau for their contributions.

Contents

1. Introduction

Artificial Intelligence is growing exponentially in its impact on human society. While the field of scientific inquiry and technical progress is roughly seventy-years-old (if we pin its origin to the 1950 work of Turing and the 1956 Dartmouth workshop), it is only now that we see AI impacting many of our lives on a daily basis. AI appears in the foreground as we interact with some fluidity, through voice recognition and natural language processing, with digital assistants like Siri and Alexa. And AI is present for many of us in the background, for instance as we use a credit card and our bank applies an AI based fraud detection algorithm while approving payment. It is not just the frequency with which we might interact with an AI today that makes it ubiquitous, it is also its broad range of applicability: from healthcare to education to agriculture to manufacturing to transportation to leisure and more.

Thus, for many of us, AI is newly ubiquitous. For all of us, however, AI has multiple valences; it can be an instrument for social and individual advancement and pleasure, or it can be an instrument for social and individual ills and discontent. Put simply, AI is replete with vast rewards and manifest risks. For example consider this utopian/dystopian dialectic: AI will either be the source of a broadly enjoyed ethical leisure society or the cause of massive unemployment. As Yudkowsky (2008) trenchantly put it, "after the invention of AI, humans will have nothing to do, and we'll starve or watch television."

These two factors – a growing global ubiquity and an emerging set of risks and rewards – is why AI presents such a wide array of increasingly sticky ethical and societal concerns. It is why, in particular, policymakers and political institutions must vigorously join the public debate over AI systems. Ultimately, policymakers need to be willing to speak, learn and act to enhance the rewards and mitigate the risks of increasingly ever-present artificial intelligences. These two factors are what motivated more than 8000 AI researchers and developers, including the likes of Elon Musk, Stephen Hawking, and Margaret Boden (https://futureoflife.org/ai-open-letter/) to argue that "[t]here is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase…. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls."

A sign of this development of AI's impact, and its clear benefits and potential pitfalls, is the growth of the ITU's AI for Good Global Summit (https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx), where industry, academia and government have been exploring the privacy, security and ethical questions that arise from AI.

These three factors – the growing scale and scope of AI, its bivalenced risks and rewards, and the central role for policymakers including ICT regulators – are the core foundations of the ethical and social issues overviewed in this module. *The goal of this module is to help ICT regulators and policymakers consider a few of the many core ethical and social issues that are emerging due to AI systems; these issues are developed here as a series of values and the ways that AI can positively or negatively impact these values. This module is not designed to offer policy prescriptions but instead will try to surface relevant values-based ethical and social issues. In doing so it raises some of the core ethical and social questions that policymakers and regulators*

must understand, track, and at times influence as AI continues its remarkable growth and development.

2. A Brief History of AI ethics and society

As mentioned above, the birth of modern AI is rooted in work from the 1950s, primarily undertaken in the USA and UK (see Box 2.1 for some emerging AI activities from other regions). The paper that perhaps best marks the creation of AI is Allan Turing's landmark *Computing Machinery and Intelligence* (Turing, 1950). In it he introduced what has come to be known as the Turing Test, an imitation game designed to demonstrate artificial intelligent behavior by a machine. Embedded within this parturition of AI, Turing already contemplates its ethical and social implications, arguing that an AI in principle *could* "[b]e kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong…" (and presumably could also be the opposite).

Attempts to encode ethical guidelines for AI even predates Turing's seminal work. Most famously we have Isaac Asimov's (1950) Three Laws, which seek to circumscribe the actions of a robot (and by extension any artificial intelligence) to ensure its social benefit and self-preservation. The Three Laws read:

1. "A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws."

In a sense, Asimov gives us the first policy designed to ensure modern artificial intelligences behave in a safe and ethical way. Today it is hard not to look back at this early AI policy with some envy: if only current realities would us to rely on such a simple, short and elegant policy.

While AI technologies have matured and developed at a remarkably rapid pace, our consideration of the ethical and social implications of AI systems have grown slowly from these storied beginnings. For instance, a decade ago Yudkowsky (2008) asked why there "aren't more AI researchers talking about safety?". In his influential essay, he argues for a "Friendly AI," an ethical and constrained artificial intelligence which benefits humanity and society. In the decade since the publication of his call for an ethical AI, many more researchers are indeed talking about the social and ethical implications of AI. Nevertheless, ethical and social thinking about AI has *not* kept pace with the rapid technological developments at hand. Moreover, the policy and regulatory community has often remained at a distance to this small but growing community of AI ethicists. We hope that this will change and that AI ethicists will be informed by and in turn will help to inform ICT policymakers. Indeed, as we will often repeat in this module: AI is moving at such a pace that it is critical for ICT policymakers and regulators to stay abreast of its growth along with any concomitant ethical and social dimensions to AI. To do otherwise could put our ICT systems (and more) at risk. Consider this module just one invitation to this conversation between ICT policymakers and regulators and AI ethicists.

BOX 2 .1: ARTIFICIAL INTELLIGENCE STARTUPS IN AFRICA AND LATIN AMERICA

Historically, artificial intelligence has been a project primarily within the USA and UK. Recently, China has made considerable strides in developing its AI capabilities (see Box 5.b). As a response to these realities, twenty-four EU ministers have signed a pledge to develop a "European approach" to artificial intelligence research, development, investment, and legal and ethical policies (Stupp, 2018). Stakeholders in Asia, Europe and North America are competing for AI dominance, but what about Latin America and Africa?

Countries in Africa are looking at artificial intelligence as a means to create new jobs and develop skills for the workforce (Novitske, 2018). Some argue that emerging states can leapfrog into the AI revolution (Samans & Zahidi, 2017). But challenges have been identified in the development of AI on the continent including: 1) weak communications and complimentary infrastructures, 2) limited government engagement, 3) a need for AI training and capacity building, 4) persistent "brain drain" among AI experts especially in Africa, and 5) limited gender diversity among the AI workforce (Brandusescu, Freuler, & Thakur, 2017).

Today, a combination of universities, domestic businesses and organizations are supporting AI in a number of African nations. Many USA headquartered tech companies are also investing and mentoring entrepreneurs through incubators, hubs and competitions like the Google Launchpad Accelerator, IBM AI Xprize and Microsoft's 4Afrika initiative. The extensive mobile infrastructure and embracing of mobile money creates ripe conditions for AI research and use. Companies in Africa are experimenting with AI to solve a variety of problems across sectors from finance, healthcare, manufacturing, agriculture and others. Here are some examples:

- In Nigeria, an AI start-up, Ubenwa, has developed a mobile app which uses a newborn's cry to detect childbirth asphyxia (http://ubenwa.com/). According to their website, Ubenwa's machine learning system takes an infant's cry as input, analyses the amplitude and frequency patterns of the cry and provides an instant assessment of whether the newborn has birth asphyxia or is at risk of it. Ubenwa has recently started conducting clinical validation exercises at hospitals in Nigeria and Canada.

- The Kenyan company, Farmdrive, is addressing the problem of financial exclusion faced by millions of family farmers across rural Africa (https://farmdrive.co.ke/). These smallholders are often excluded from the traditional banking system and face barriers to access capital. In 2014, Farmdrive was founded by two Kenyan women to leverage the power of disruptive tech to bridge the gap between small-scale farmers and financial institutions. According to their website, Farmdrive uses machine learning to construct alternative credit profiles for farmers, and decision-making tools for financial institutions that combine environmental data (weather and climate patterns, soil data), economic data (annual income, market data), agronomic and satellite data. By verifying and augmenting the self-reported financial history of the farmers with exogenous data, Farmdrive reduces risk for the banks. The company connects with its end user through a farmer-facing app that runs over SMS, allowing farmers access even should they lack data connectivity or data-enabled feature phones (LHoFT, 2017).
-

- According to Accenture, 78% of South African executives state that they need to boost their organization's competitiveness by investing in AI (Schoeman, Moore, Seedat, & Chen, 2017). One South African company, Data Prophet, strives to address this market need by providing artificial intelligence consulting services to South African, as well as international businesses in manufacturing, retail and financial sectors (https://dataprophet.com/). It's product suite, Omni, uses machine learning models to predict defects, faults or other quality control criteria in manufacturing. Convolutional neural networks and image recognition are also used to automate visual quality control inspection. According to their website, Omni further identifies and prescribes ideal process variables to optimize operating yield in manufacturing, thus improving the quality and efficiency of the business. Apart from Omni, Data Prophet also offers an AI chatbot called Mentat. The chatbot learns responses to customer service queries, escalating only unsolved queries to a human representative. The company claims to improve their clients call volume and costs by more than a fifth.

Similarly, across Latin America, many enterprises are using AI in novel ways to solve some of the toughest local and global challenges:

- *Operação Serenata De Amor* is an open-source and crowd-funded artificial intelligence project, pioneered by Brazilian data scientist Irio Musskopf (https://serenata.ai/). The project uses a combination of public data procured from government websites and agencies, and information from Google, Foursquare, Yelp and other websites, to "audit public accounts" and support citizen control of public finances and engagement with public officials. Brazilian government agency *Quota for Exercise of Parliamentary Activity* (CEAP) receives over 20,000 reimbursement claims every month from Brazilian Congress members (https://jarbas.serenata.ai/dashboard/chamber_of_deputies/reimbursement/). An AI system, named Rosie is used to analyze this enormous quantity of data and flag any irregularities or suspicious activities. The system then reports its findings to the lower house of Brazilian parliament and also flags it publicly on twitter, holding elected legislators accountable to Brazilian citizens (Matsuura, 2017). The *Serenata de Amor* team plan to gather and make public information such as the wealth of politicians, the donations received by campaigns, bills already proposed, and expenses for work and district projects, ahead of Brazil's national elections in 2018 (Monnerat, 2018). The organization further plans to develop a robot-journalist that will be able to write short articles about the bills that were flagged by Rosie.

- Based in Chile, Not Company is a food tech company that aspires to resolve world hunger, climate change and unsustainable food production through transforming the food we eat (http://www.thenotcompany.com/). The company has developed *Giuseppe*, an AI system that analyzes the molecular structure and properties of popular food ingredients and develops sustainable plant based substitute recipes for popular animal products like meat and milk (Al Jazeera, 2017). The company uses a machine learning algorithm to produce products that mimic taste, texture, nutrition and appearance of animal products, but require only a fraction of resources to produce and is more environmentally friendly (Baer, 2016).

Table 2.1

Country	Sector	Company Name	Project Description
Nigeria	Health	Ubenwa	Mobile app to detect childbirth asphyxia from a newborn's cry.
Kenya	Agriculture	Farmdrive	Profiling and decision support for credit to small-scale farmers.
South Africa	Consulting	Data Prophet	Manufacturing quality control and customer service chatbot.
Brazil	e-Government	*Operação Serenata De Amor*	Tools to enhance government accountability and transparency in their public expenditures and financial activities.
Chile	Food	Not Company	Application to develop plan based vegetarian substitutes for popular animal based recipes.

3. Is AI Different?

Above, we argue that AI is bivalenced – the technology presents both risks and rewards. But this surely is not unique to AI as, indeed, all technologies have both potential positive and negative impacts. Most insidiously, technologies also generally have unintended consequences. In this way, while designed for positive social impact they may instead unexpectedly result in negative outcomes. The ICT industry is replete with examples of this bivalenced tension. For example, we all know of the amazing social and economic benefits that have arisen from mobile telephony. But we also should reflect on the safety issues of texting or talking while driving; the criminal applications of anonymous "burner" phones, and so forth.

We have also argued that AI is expansive in its scale and scope. It can reach across most sectors and elements of society and appear in all number of possible systems. But is this unique to AI? Surely a similar argument could be made about, for instance, the Internet or mobile telephony. These technologies also have a broad range of applicability from healthcare to education to agriculture to manufacturing to transportation to leisure and more.

The mere fact that AI is bivalenced and broad is not, in and of itself, an argument that AI is socially or ethically a markedly different technology from those that have preceded it. The internet and mobile phones are bivalenced and broad as well. Nevertheless, there are some ways that AI *might,* in fact, be unique and different with potentially profound ethical and societal import.

First, we cannot lose sight that, definitionally, the goal of AI is the creation of intelligent machines. And intelligence qua intelligence is, in many ways, a different form of "technology." A set of AI thinkers have noted that the path towards intelligent machines may lead us to greater-than-human intelligence. The emergence of this sort of super artificial general intelligence could pose ethical and social challenges hitherto unknown. Indeed, some argue that the rise of a superintelligence (Bostrom, 2014) will present an explosive "existential AI threat" (Good, 1966),

or “singularity” (Vinge, 1993), beyond which it is impossible to predict the outcomes or even the continued existence of humanity. Undeniably, it is plausible reasoning that underpins this perceived threat. An AI might increase in its intelligence with vast speed due to, for instance, recursive self-improvement (Good, 1966): an AI is designed to learn efficiently, it applies this efficient learning to its own capacity to learn, which when enhanced becomes even better at learning efficiently, and so forth.

An AI explosion due to recursive self-improvement is a cautionary tale of technologies out of control. And some researchers have responded to this possible threat with a call for a moratorium on any relevant research pathways (Metzinger, Bentley, Häggström, & Brundage, 2018). Indeed, such a moratorium, and related approaches, are likely to be sensible policy responses.

Even if this feedback loop does not result in an AI explosion, the potential speed of AI change that it drives might be faster than we experience with other ICT areas. How can we be sure that ethical and social policies keep up with this pace of technological change? How can ICT policymakers position themselves to respond when and as required to AI development? A first step is for all ICT policymakers to commit to remaining knowledgeable and up-to-date with the cutting-edge state of AI ethics and society.

Beyond learning feedback loops, and the potentially unusual pace of change that results from this feedback, there are further ways that AI might be different from other current or emerging ICTs. For instance, AIs capacity to enhance or replace human agency with machine agency (and in doing so subsume moral agency directly), might be different. And any ability to outsource ethics itself to AI, with machine based ethical decision making and behavior, would be different. The ability for an AI to develop true autonomy could also be different. And so forth.

A superintelligent singularity or a moral or ethical AI are, indeed, ways that artificial intelligence might be different from other technologies today. But these are also directions that have more than a hint of a dystopian science fiction Hollywood film; they are not a likely reality. Therefore, beyond these few paragraphs, this module will not further consider these existential ethical and social implications for AI, instead focusing on more likely threats and promises. In doing so, we will be treating AI as more similar than dissimilar to other major recent technological innovations (e.g., the Internet and mobile phones).

What this overview should make clear is that we all are struggling to predict the future of AI and our gaze onto its potential impacts on humanity is hazy at best. The fact that AI is complex and its future is not entirely clear underpin the need for ICT regulators and policymakers to stay abreast of its development and conversant with any emerging social and ethical concerns. ICT policymakers will need to develop systems for multi-stakeholder consultation from which emerge dynamic, responsive, and as required, predictive and admonitory AI policies and public processes. It is only through an ongoing ethically-informed policy engagement that the best forms of AI are likely to flourish, and the most negative potential impacts of AI can be attenuated.

BOX 3.1 : THE AI & ETHICS ADMONITION:
AI is changing with enormous speed and is affecting many different elements of human society all at once. It is impossible to perfectly predict all of the many ways AI will impact the systems, infrastructures, and ethical and social areas of concern to the world's ICT policymakers and regulators. In order to respond quickly and effectively to ethical and social issues that arise out of new AIs, and in order to be proactive and prudent, ICT policymakers must remain up-to-date on AI social and ethical issues, engage in real-time and continuous multi-stakeholder and cross-institutional consultations and engagements on these issues, and maintain nimble and responsive policy mechanisms and procedures.

4. Framing AI, Ethics and Society

There are many potential ways to organize, or frame, the reciprocal role of AI on ethics and society. Different organizational structures would best reveal certain societal and ethical properties of AI and would most usefully structure a conversation on this topic.
For instance, one might taxonimize the various extant and nascent AI technologies, list for each technology the various societal and ethical considerations and contemplate the various policy positions relevant to this technology. For example, autonomous vehicles and drones could be considered as a stand-alone technology, as could natural language processing systems. While there is a natural ease to this particular framing device, there is a risk that it would become techno-centric allowing the systems to carry the conversation when it is the ethical and social issues that are most salient to policy decisions. Further, a technology focused approach might limit the policy responses, for instance requiring a specific regulation for each new technical innovation.

Another option is to organize the discussion around various economic sectors, for instance health, education, military, transportation, etc. But many social and ethical issues are cross-cutting; they impact multiple sectors at once. And, indeed, ICT policymakers are often themselves cross-cutting given the all-encompassing nature of communication and information infrastructures. Thus, sectors would seem to be an unparsimonious organizing principle.

Instead, in this module, we constitute a values-based organizing framework. One advantage of leading with values is that it naturally privileges what is most ethically and socially salient to a values-driven policymaker in order to arrive at a values-driven society. For instance, privacy is a widely regarded value, and a values framework would allow us to place privacy front-and-center in our considerations. Additionally, many values are rooted in certain universal principles including the Universal Declaration of Human Rights, adopted by the General Assembly of the United Nations and adhered to by member states. Admittedly, though, a values-based approach is not without its challenges as many values are not entirely universal in how they are defined or prioritized. For example, anti-discrimination of protected groups is a widely regarded value, though different cultures might differ in which groups they most assiduously strive to protect.

Nevertheless, we believe that a values framework offers the best chance at surfacing key ethical and social issues and so in this module we develop a bivalenced values framework for AI, ethics and society. Each *value* is presented and overviewed as it relates to artificial intelligence. For each value, we then offer examples of *rewards* – ways that AI might positively engage this value – along with *risks* – ways in which AI presents a challenge or threat to the value. We will also

explore *connections* – ways in which ICT policymakers and regulators have already considered how this value interfaces with types of information and communication systems. We also include some of the salient or most representative *questions* that each value exposes. As we repeatedly mention, AI is moving quickly and is affecting many parts of our lives. Some aspects to AI are new, and all aspects are dynamic. We are just now getting a handle on some of the most important ethical and social questions posited by these emerging AI technologies; it may be too early to have answers to many of these questions.

For this module we will examine just a few of the many critical values pertaining to artificial intelligence: 1) livelihood and work; 2) diversity, non-discrimination and freedoms from bias; 3) data privacy and minimization; 4) peace and security. We note that this list of values is quite similar to values that the IDRC have identified as potentially at risk, particularly in the Global South, due to the rise of artificial intelligence (M. L. Smith & Neupane, 2018).

4.1. Livelihood and Work

Value	Rewards	Risks	Connections
Livelihood and Work	Economic growth; new forms of work; expanded leisure	Enormous global labor disruptions; expanding unemployment and job competition	Similar claims were made about e-commerce, the internet generally, etc.

For most, work holds an intrinsic value. For all, a secure livelihood is paramount. Truly, livelihood and work encompass a set of human values which AI can and should positively support. But AI is also seen as a potential source of work and livelihood dislocations. This section will explore the positive and negative valences of AI as it relates to human livelihood and work.

Significant technological disruptions are commonplace in market economies going back to the Industrial Revolution and even before. And debates regarding the negative versus positive influence of technological change on work go back as far. The negative impact of technology on work, the displacement effect, occurs when human labor is replaced or undermined by technological systems. The positive effect of technology on work, the productivity effect, happens when demand for labor increases due to innovation and technological automation.

In general economists argue that while technological innovation often have a short-term displacement effect, in the long run, new technologies create an overall positive productivity effect on labor (Petropoulos, 2017). But, could AI be different compared to our experience with other technologies? Are we entering a new period of Artificial Unemployment (Boddington, 2017) different from what we have experienced from other technological changes? Or is AI creating new opportunities? Since the speed and scale of technological change and the scope of areas affected by AI is so vast, it may be harder to predict just how significant AI will be on labor, livelihood and work; AI may indeed be different than other technologies in this regard (Calo, 2017). As an earlier GSR discussion paper argued, “these applications are still moving from the lab to adoption, it is not feasible yet to quantify their impact at a macro-economic level,” (Katz, 2017). In this way, our admonition for ICT policymaker to be prepared, our call that they remain

up-to-date on the social and ethical issues associated with artificial intelligence, is particularly manifest when it comes to livelihood and labor issues.

4.1.1. Risks

Labor economists and related researchers have noted potentially massive workforce disruptions due to emerging AI technologies, including downward pressures on both the availability of jobs as well as on wages. According to one study, one-ninth of jobs in the USA could be affected due to self-driving transportation alone (Beede, Powers, & Ingram, 2017) and overall 47% of US workers have jobs at “high risk of potential automation” (Frey & Osborne, 2017). Are potentially half of all workers at risk of dislocation due to AI enabled automation technologies? Compared to earlier technology-driven disruptions, this would be an entirely new level of workplace dislocations.

Even if a substantial share of the US economy is impacted by AI, it is not clear if a similar scale of change will affect workplaces globally. Worldwide, McKinsey argues that “[w]hile about half of all work activities globally have the technical potential to be automated by adapting currently demonstrated technologies, the proportion of work actually displaced by 2030 will likely be lower, because of technical, economic, and social factors that affect adoption,” (Manyika et al., 2017). A recent World Bank World Development Report also warns of potentially enormous labor disruptions especially in low- and middle-income countries, but sites mitigating factors that should attenuate the impact (World Bank, 2016). The report states that “[t]wo-thirds of all jobs could be susceptible to automation in developing countries in coming decades, from a pure technological standpoint.” But it goes on to argue that “large-scale net job destruction due to automation should not be a concern for most developing countries in the short term,” due to the lower cost of labor (which creates less market pressure for labor substitution) and the slower pace of technological adoption in these economies. Nevertheless, the percentage of the economy susceptible to AI enabled automation in the Global South, even when taking wages and technology diffusion delays into account, is generally above 40% of the employment market (World Bank, 2016).

One sector of importance to many economies of the Global South, and at particular risk from automation, is the range of offshoring activities such as call centers and back-office business processing facilities (Katz, 2017). For example, as speech recognition and natural language processing systems continue to mature, their capacities may replace many corporate offshored call centers.

Studies have noted not just geographic variation on the potential impact of AI enabled automation on labor, but variation among the genders. The World Wide Web Foundation’s report on AI in low- and middle-income countries notes that automation based reduction in job opportunities will create even more pressures on women for employment as “men compete with women for fewer jobs,” (World Wide Web Foundation, 2017).

While AI may be a net benefit to elements of the economy, its negative impacts may be more widely felt. "[J]ob loss is more salient to people—especially those directly affected—than diffuse economic gains, and AI unfortunately is often framed as a threat to jobs rather than a boon to living standards," (Stone et al., 2016).

4.1.2. Rewards

It is fair to say that AI technologies will impact a large percentage of the global labor market; just when, how much, and for whom is a matter of some debate. But impacts will be felt by a lot of the economy and not in the far distant future. This impact on labor, however, need not be entirely negative; history tells us that new technologies serve to displace labor but usually (with time) in a way that is a net gain to overall employment. So short-term disruptions historically lead to productivity enhancing longer-term growth.

A common example, sited in many of the reports referenced above, is that of the bank ATM. When bank ATMs first came onto the scene, there was considerable concern as to the impact they would have on teller employment. The worry was that ATMs would automate away the bank teller job, displacing a non-trivial number of workers in many economies globally. Instead, while ATMs did disrupt the banking industry, overall they have had more positive than negative impact on employment. As an article in The Economist put it, "Replacing some bank tellers with ATMs, for example, made it cheaper to open new branches, creating many more new jobs in sales and customer service," (The Economist, 2016b).

Some authors have focused on the productivity enhancing possibilities that come when AI is partnered directly with humans around some particular work task. According to their report, "Gartner is confident about the positive effect of AI on jobs. The main contributor to the net job growth is AI augmentation — a combination of human and artificial intelligence, where both complement each other," (Gartner, Inc., 2017). Gartner is predicting that ultimately the job-creating benefits of AI will overwhelm any labor disruptions. They expect that "[i]n 2020, AI becomes a positive net job motivator, creating 2.3 million jobs while only eliminating 1.8 million jobs. In 2021, AI augmentation will generate $2.9 trillion in business value and recover 6.2 billion hours of worker productivity." Similarly, according to the panel report of the One Hundred Year Study on Artificial Intelligence, "AI will likely replace tasks rather than jobs in the near term and will also create new kinds of jobs. But the new jobs that will emerge are harder to imagine in advance than the existing jobs that will likely be lost…. It is not too soon for social debate on how the economic fruits of AI technologies should be shared," (Stone et al., 2016).

Another entirely radical viewpoint, articulated by some (perhaps on the fringe), is that AI's work supplanting capacities will be so unlimited as to render human labor itself superfluous. In this scenario, human activity will be taken over by leisure, art, interpersonal interactions, and rest.

Returning to the sober analysis of The Economist, they conclude by asking, "who is right: the pessimists (many of them techie types), who say this time is different and machines really will take all the jobs, or the optimists (mostly economists and historians), who insist that in the end technology always creates more jobs than it destroys? The truth probably lies somewhere in between" (The Economist, 2016a).

4.1.3. Connections

ICT policymakers and regulators will find some of these concerns (and promises) reminiscent of similar arguments lodged against e-commerce, the internet generally, and indeed ICTs more

generally. Market displacements, dislocations, and indeed growth are not new phenomena to ICT stakeholders. In order to manage the oncoming labor changes driven by AI, Manyika and Spence (2018) call us to attend to three priority areas:

1) skills and training,
2) labor market fluidity and dynamism (for instance ensuring that workers can easily move between jobs),
3) and income and transition support for those displaced.

Even though this list has been newly created to account for AI's emerging impacts, it should not be entirely unfamiliar to ICT policymakers who have looked to make similar accommodations for the labor impacts of previous information and communication technologies.

4.1.4. Key Questions to Consider for This Value

Are AI systems different enough from other technological change to upset the usual patterns where labor productivity effects ultimately outweigh labor displacement effects?
Will AI develop so quickly across so many economic sectors all at once that our economies will struggle to adapt and develop in time?
Can ICT infrastructure help to enhance specific labor productivity effects while mitigating displacement effects?
How will the ICT sector itself (e.g., customer support operations or network design) be assisted by AI engines?

4.2. Diversity, non-discrimination and freedoms from bias

Value	Rewards	Risks	Connections
Diversity, non-discrimination and freedoms from bias	Systems to support linguistic diversity, pre-literacy, physical disabilities, etc.	Learned bias (racial, gender, etc.); systems that unduly privilege majority populations	Universal service and common carriage; TTY and emergency response

In her seminal book, *Machines Who Think*, Pamela McCorduck (2004) maintains how, "I'd rather take my chances with an impartial computer," than cast her lot with a potentially biased human. We as a global community value diversity and eschew bias and discrimination against protected categories. Can we, however, claim that computers are indeed impartial? Are algorithms always free from discrimination and respectful of diversity? According to Kate Crawford, "[s]exism, racism and other forms of discrimination are being built into the machine-learning algorithms that underlie the technology behind many 'intelligent' systems that shape how we are categorized and advertised to," (Crawford, 2016).

One step in ensuring an AI is free of bias is for its designers and developers to be diverse and bias-free. Box 4.2 overviews the diversity challenges within the AI design and research communities. In addition, for an AI to be impartial also requires that the underlying data that

informs and trains the AI be non-discriminatory and inclusive. For example, machine learning error rates are usually inversely proportional to training data size. Therefore, a minority subpopulation within a group is at significant risk for increased error rates in an AI's decision making if its representation within the training data is small. If data is sampled evenly based on population sizes than small populations (minorities) will be weakly represented in the data and therefore subject to heightened error rates.

Put simply: the risks for bias in AI is probably greater due to the qualities of its datasets than for any "hand coded" biases of its algorithms. As the 100 year Study Panel put it, "though AI algorithms may be capable of making less biased decisions than a typical person, it remains a deep technical challenge to ensure that the data that inform AI-based decisions can be kept free from biases that could lead to discrimination based on race, sexual orientation, or other factors," (Stone et al., 2016).

BOX 4.2: WHERE ARE THE WOMEN? GENDER DISPARITIES IN AI RESEARCH AND DEVELOPMENT

The artificial intelligence community has a diversity problem. Microsoft researcher Margaret Mitchell has called AI a "sea of dudes" (Boddington, 2017). Kate Crawford, also a Microsoft researcher and NYU professor, asserts that AI has a "white guy problem" (Crawford, 2016). Crawford goes on to articulate why this matters: "Like all technologies before it, artificial intelligence will reflect the values of its creators. So inclusivity matters — from who designs it to who sits on the company boards and which ethical perspectives are included.
Otherwise, we risk constructing machine intelligence that mirrors a narrow and privileged vision of society, with its old, familiar biases and stereotypes."

The low level of female presence among AI researchers, developers and thought leaders might best epitomize this diversity challenge. Hannah Wallach, yet another Microsoft based AI researcher, has guessed that the entire field of machine learning is only 13.5% female (Weissman, 2016). To support those women who are already in the field, and increase the number of women who enter it, she co-founded the Women in Machine Learning (WiML) initiative (http://wimlworkshop.org). Since 2006, WiML has held regular events and now puts on an annual flagship workshop co-located with the NIPS conference.

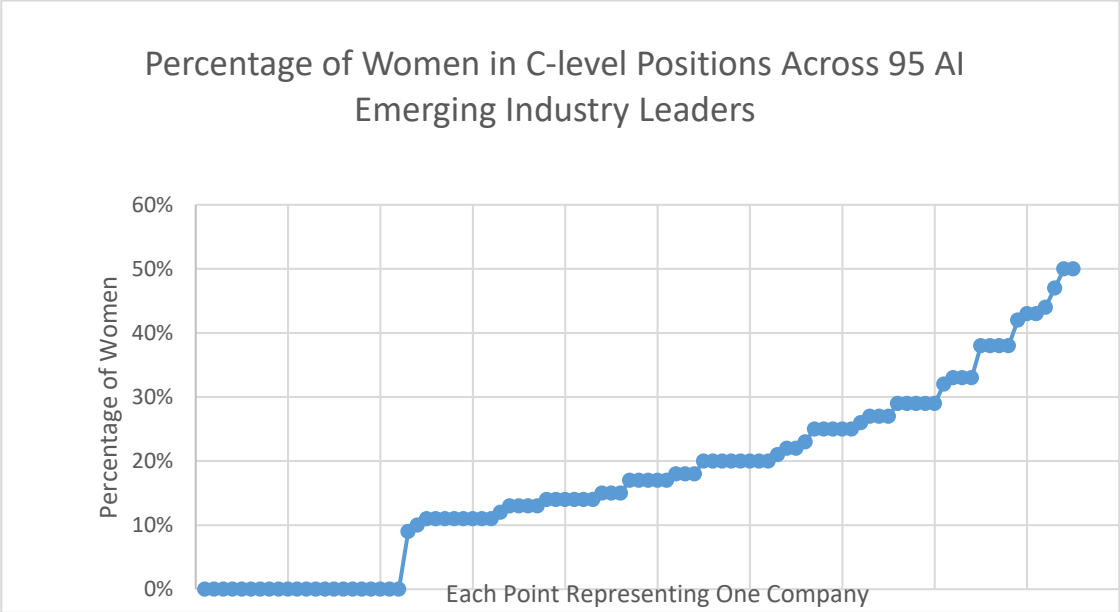
Wallach's estimate is depressingly low and underlines an enormous diversity challenge across the field. To better amass evidence as to this gender disparity, we have accumulated data on women participation in leadership among top AI companies, as well as scholarly presence among the top USA based university computer science faculty. Our new study finds that women represent a paltry 18% of C-level leaders among top AI startups across much of the globe and just 22% of faculty in top USA based university AI programs. While these percentages are slightly better than Wallach's overall industry estimate, we take no solace in them; clearly, females are overwhelmingly underrepresented among AI scholars and corporate leaders.

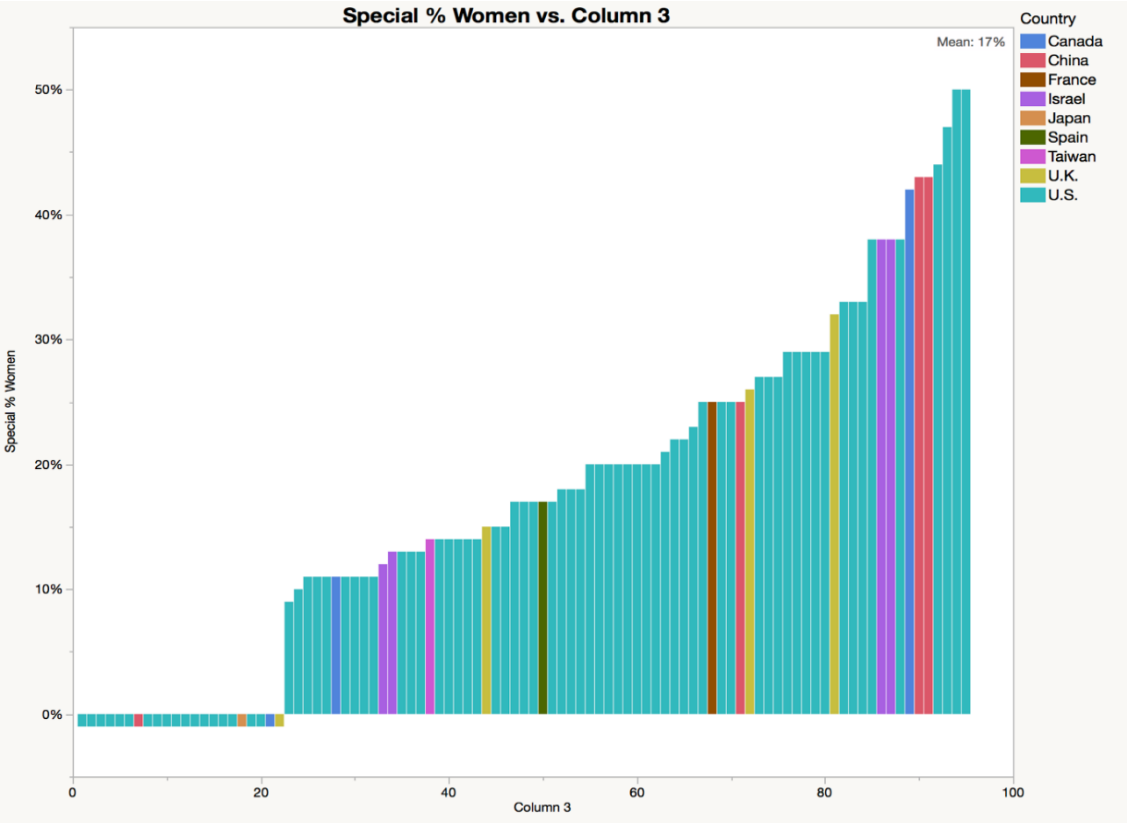
To calculate the percentage of women in executive management at leading AI startups we began with CB Insights' 2018 "AI 100", their ranking of the top 100 promising start-ups in Artificial Intelligence (https://www.cbinsights.com/research/artificial-intelligence-top-startups/). This ranking includes seed stage startups along with more advanced stage companies, and inclusion

criteria comprised factors such as investor profiles, technical innovation, business model, funding history and valuation. The executive board for each company, as listed on their website, was used to calculate the number of women in leadership positions. In case the website did not mention the executive management then searches across LinkedIn was used to establish those in leadership positions. Our calculations do not consider the business's Board of Directors, investors or advisors when gaging women in leadership positions.

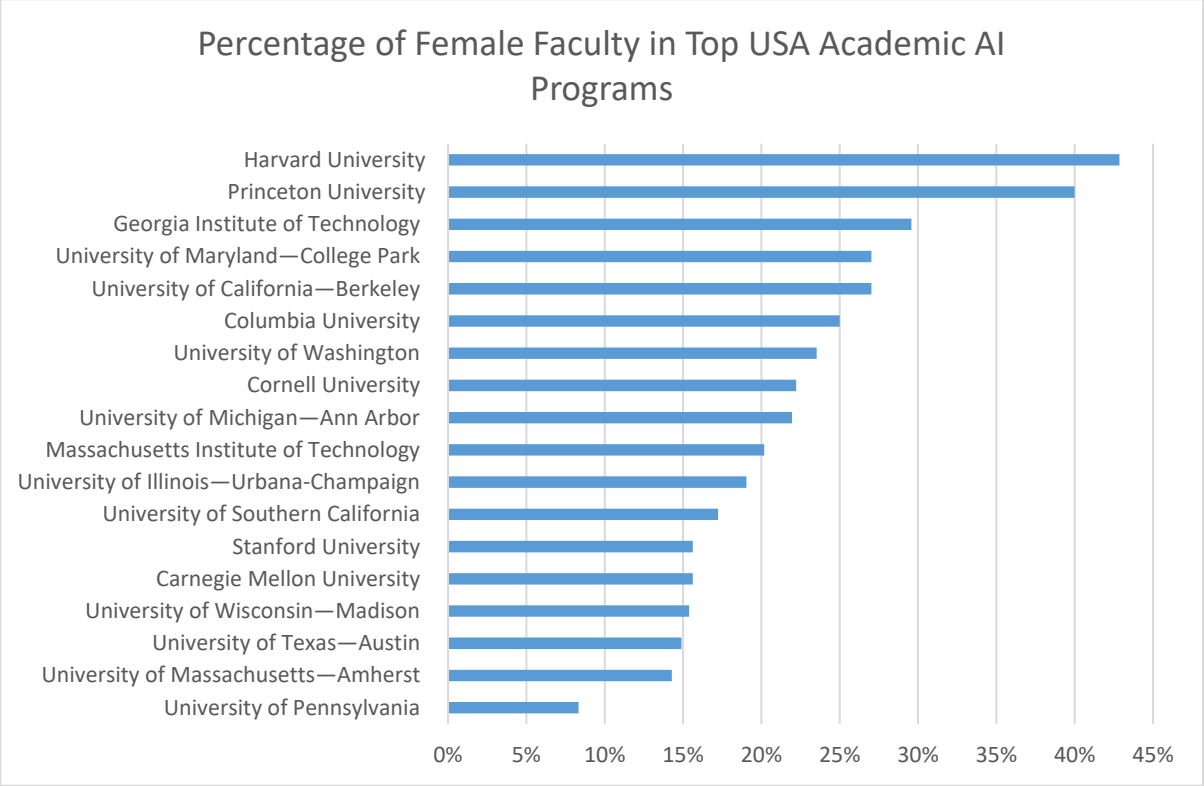
The AI 100 list includes companies from the USA, Canada, the UK, France, Spain, Japan, China, Taiwan, and Israel. We were able to establish the gender balance among executive management for all but five of these 100 companies. In only one instance was a C-level manager identified as non-binary and, for this calculation, they were not categorized.

Of the 95 companies we were able to establish data on, only two have an equal number of women to men in their C-level positions (e.g., gender parity) and none are majority female. Three in five have less than 20% women in their leadership team and one in five have no females at all. As stated above, females overall made up 18% of these AI leaders.





To compute the percentage of female professors at top US-based university AI programs we started with the US News & World Report 2018 ranking of best artificial intelligence graduate programs (https://www.usnews.com/best-graduate-schools/top-science-schools/artificial-intelligence-rankings). Using their list of the top 20, we calculated the number of faculty members (including adjuncts) from each university's website. For universities with AI Labs, we determined the faculty gender makeup directly from the lab's staff listing (e.g., Stanford University). In cases where the university did not have a separate AI Lab, the faculty's research interest, as stated on their website, was used as the inclusion criteria (e.g., Carnegie Mellon University). Some universities (e.g., Columbia University) subdivided research interest into AI, machine learning, robotics, etc. In these cases, the faculty for each related area was aggregated. We were able to obtain faculty gender information for all but two (UCLA and Cal Tech) of the top 20 programs. While the average, as noted above, was 22%, the percentage of female AI faculty ranged from a low of 8% (University of Pennsylvania) to a high of 43% (Harvard). No university had achieved gender parity among its AI faculty.



4.2.1. Rewards

First, let's consider a couple of the many ways that robust AI systems can support diversity and increase the potential for minority and exploited groups to thrive. For example, AI-based natural language translation and voice recognition systems can have a significant impact in countries with multiple languages, especially for those who communicate in a minority language which may reduce their political or economic engagement. This is the case in countries, such as Brazil or Mali, where the Portuguese or French language holds significant power. It is also true for countries with substantial linguistic diversity, such as Indonesia and Myanmar, where the majority language of Basa or Burmese holds significant power. Natural language translation systems, especially those from minority to majority languages or from local to European languages, has the promise to enhance diversity and support minority rights and inclusivity.

Of course for language translation to support linguistic diversity, natural language AI systems must become available for more languages. According to Google, their translate feature is currently available for over 100 languages (https://translate.google.com/intl/en/about/languages/index.html). If we consider Myanmar, a country of enormous linguistic diversity, their majority language of Burmese is available (though media reports have given the service mixed reviews (Weston, 2015)). However, beyond Burmese, none of the country's other major language groups are available. To belabor the obvious, natural language translation technologies can support linguistic inclusivity, but only if the language translation systems for a globally diverse set of languages are made available.

A related language technology, speech recognition, can also support diversity by providing text-to-speech readers for pre-literate populations and people with visual impairments. As the Web Foundation has put it, "These systems could also have an impact in places with high levels of illiteracy, allowing people to engage with the government or public service provision interfaces by spoken rather than by written means," (World Wide Web Foundation, 2017).

Beyond linguistic diversity, AI systems can help support other communities that are part of our diverse populations. For example, autonomous vehicles can help to enhance freedoms among people who have reduced mobility due to age or disability (Stone et al., 2016).

Policymakers and regulators have an interest in supporting diversity and ensuring inclusive access and benefit from ICTs across their populations. In this way, they should support ICT infrastructure and services that are inclusive and diversity-enhancing.

4.2.2. Risks

Bias and discrimination of AI systems against protected groups, including racial and gender groups, has received considerable recent attention. In an important piece of investigative journalism in the USA, ProPublica analyzed an AI risk assessment technology, called COMPAS, developed by the Northpointe software company (Angwin, Larson, Mattu, & Kirchner, 2016).

BOX 4.2.2: BIAS AND AI, THE CASE OF COMPAS

COMPAS is a system to assess the recidivism risk among criminal inmates in the USA being considered for parole. The COMPAS risk assessment system receives on input a set of features pertaining to an inmate housed within the US criminal justice system. This feature set includes personal details such as the inmate's education level, drug problems, age at first adjudication, number of prior arrests, etc. These inputs are then used to calculate a recidivism risk score, an attempt to predict the likelihood that the individual will reoffend in the future. An analysis from the independent news organization, ProPublica, showed that the likelihood of *false positives* – instances when the COMPAS system predicted future crime when in point of fact no such criminal act went on to occur – was higher for black individuals than for white individuals. As the ProPublica journalists described it, "The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants," (Angwin et al., 2016).

Note that this type of bias against black inmates was present even though race was *not* one of the features input into the system. In the USA, race is strongly correlated with other features that were input into the COMPAS technology; these surrogate features could be said to have statistically "stood in" for race in the system's analysis. Due to this surrogacy, simply excluding a protected feature from the AI system's inputs is not enough to prevent the statistical inference of this very feature. According to a Rand study, "The standard safeguard against algorithmic disparate impact effects is to hide sensitive data fields (such as gender and race) from learning algorithms. But the literature on modern reidentification techniques recognizes that learning algorithms can implicitly reconstruct sensitive fields and use these probabilistically inferred proxy variables for discriminatory classification," (Osoba, 2017).

The Northpointe corporation responded vigorously to ProPublica's reporting, highlighting independent studies arguing that COMPAS was fair in this particular way: the proportion of people classified as high-risk for recidivism who do reoffend is identical across racial groups (Corbett-Davies, Pierson, Feller, & Goel, 2016). In other words, Northpointe says their algorithm is not racially biased because their *true-positive* rates are identical across racial groups. ProPublica, in contrast, says that the COMPAS algorithm *is* racially biased because their false-positive rates are not identical across racial groups (and, specifically, blacks are more likely to be falsely classified as a future offender than whites).

According to Barocas and boyd (2017), "computer scientists and statisticians have debated the different qualities that an intuitive sense of fairness might imply: that a risk score is equally accurate in predicting the likelihood of recidivism for members of different racial groups; that members of different groups have the same chance of being wrongly predicted to recidivate; or that failure to predict recidivism happens at the same rate across groups." Scholars have noted that mathematically it is impossible to maintain all of these forms of fairness simultaneously; being fair in one way means necessarily abandoning another fairness criteria (Kleinberg, Mullainathan, & Raghavan, 2016).

The COMPAS example brings up many issues related to AI bias and fairness. How do we measure fairness and what are the trade-offs between various fairness values? What about when an AI is involved in life impacting critical-decision making areas, such as parole decisions within the criminal justice system? Does it need to be held to a higher standard? Who is accountable if the AI algorithm is free from bias but the system "learns" to be discriminatory due to its training data?

Many other examples of learned bias and discrimination in AI systems have appeared in the literature. As legal scholar Ryan Calo (2017) summarizes, "[t]he examples here include everything from a camera that cautions against taking a Taiwanese-American blogger's picture because the software believes she is blinking, to an image recognition system that characterizes an African American couple as gorillas, to a translation engine that associates the role of engineer with being male and the role of nurse with being female." These examples demonstrate the difficulties associated with bias and discrimination beyond critical-decision making machine learning systems.

4.2.3. Connections

Bias and non-discrimination are not entirely new to ICT policymakers and regulators. For instance, universal service provisions (USP) and public carrier principles are inclusive and non-discriminatory policies at their foundations. Some universal service obligations stipulate that all licensed operators must provide service of some minimal quality to all comers who meet some minimal requirements. Put simply, operators under such provisions are not allowed to discriminate against anyone making a legitimate request for their service.

Similarly, TTY and telecommunication relay services, obligated telecommunication facilities in many localities, support diverse and inclusive access to these benefits of ICT systems. These policies require that communication operators provide text-based services for users with hearing or voice disabilities.

In these ways, regulators have experience with ways to ensure that ICT providers are non-discriminatory in their practices and that they are inclusive in their services. The fact that demands on inclusivity and freedom from bias become heightened when dealing with life critical and high-risk systems also has some familiarity as it touches on some provisions of emergency response regulations (e.g. 999 or 911 services). Responding to the heightened need for care in critical decision making, Recital 71 (http://www.privacy-regulation.eu/en/r71.htm) of the European Union's General Data Protection Regulation (GDPR) states that people have the right to not be subject to decisions made by information processing systems, including AIs, that are critical in nature or have legal affect.

4.2.4. Key Questions to Consider for This Value

How can our policies best support AI systems that promote diversity and inclusivity?
How do we measure fairness and detect discriminatory action within an AI?
When an AI is involved in a life impacting critical-decision can we impose heightened requirements that it perform without discriminatory bias?
How can ICT operators use AI engines to help ensure non-discriminatory behavior and support diversity?

BOX 4.2.4 : CAN WE TRUST AI? - THE NEED FOR AI EXPLAINABILITY

"We are increasingly relying on machines that derive conclusions from models that they themselves have created, models that are often beyond human comprehension, models that 'think' about the world differently than we do," (Weinberger, 2017). With these words, technology author David Weinberger opens his essay on Alien Knowledge and the rise of AI systems whose decisions are beyond human explainability and understanding.

Consider a deep learning image analysis system designed to label pictures with terms that describe the predominant object on display (e.g., a "banana" or "car" or "tiger"). The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual competition around such a task, where competing systems machine label a publicly available image dataset (Russakovsky et al., 2015). In 2015, using a deep neural network learning algorithm, machine image labeling technology exceeded human performance in the ILSVRC competition. This technology, which performs beyond human capacities at this particular task, is an example of black-box AI; it is a form of Weinberger's Alien Knowledge. Its ability to classify images is based upon the computer system learning thousands of weights across a multi-layered neural network. The image classification that results from the application of these weights, upon the input of an image's many pixel values, admits to no human-understandable explanation. The system can correctly classify a tiger image as such, but it does so for no reason we can articulate; it is not because the object on display is "furry" and "with whiskers," for example. We cannot say exactly why the neural network weights correctly classify the image of the tiger. We can merely state that as an

outcome of processing a set of initial pre-labeled training images its many weights were set and hat it now, empirically, succeeds in labeling these cat images correctly.

Black-box (or Alien Knowledge) artificial intelligences have been the source of stunning recent successes and sit at the core of many of the most powerful AI systems in image analysis, speech recognition, natural language processing, game playing (e.g., Go, chess, or arcade games), etc. But they also are increasingly the source of consternation among some technology ethicists; in some instances, an impenetrable black-box might not be good enough and instead we may demand explainability of a system's decision. As Barocas and boyd (2017) ask us: "When is the ability to meaningfully interrogate a model sufficiently important to justify some cost in performance? What kinds of decisions—and real-world effects — drive data scientists to develop a model that they can explain, even if its decisions might be less accurate as a result?" In other writings, they answer their own questions by insisting that all critical life affecting applications demand decisions from algorithms that are fully explainable. In essence, they advocate a moratorium on black box decisions applied within these critical domain areas.

In a subsequent Wired article, Weinberger (2018) seems to consider Barocas and boyd's question on when an AI's outputs can and should be explainable, even at a cost in performance. He argues that, given the potential positive social benefits of powerful AI, explainability is an unnecessarily excessive goal if it comes at the cost of efficiency or effectiveness. "Demanding explicability sounds fine, but achieving it may require making artificial intelligence artificially stupid. And given the promise of the type of AI called machine learning, a dumbing-down of this technology could mean failing to diagnose diseases, overlooking significant causes of climate change, or making our educational system excessively one-size-fits-all. Fully tapping the power of machine learning may well mean relying on results that are impossible to explain to the human mind." Instead, for Weinberger it is enough for an AI to "meet its marks"; in other words, to empirically perform up to requirements and expectations (for accuracy, safety, etc.) independent of its knowability.

Other technologists have sought a middle ground: systems that offer the benefits of explainability while avoiding any algorithmic "dumbing-down" to achieve it. For instance, our colleagues at Harvard's Berkman Klein Center for Internet and Society (Doshi-Velez et al., 2017) call for what some others have referred to as algorithmic auditing. An audit is a method to systematically probe a black-box system with inputs designed to, in essence, reverse engineer an algorithmic explanation for its output choices. For instance, consider this example: "[S]uppose that the legal question is whether race played an inappropriate role in a loan decision. One might then probe the AI system with variations of the original inputs changing only the race. If the outcomes were different, then one might reasonably argue that race played a role in the decision," (Doshi-Velez et al., 2017).

TABLE 4.2.4: How to Trust AI: A Taxonomy of Know ability

Solution Meet its mark	Example Proponent	Reward	Risk
Ensure that AIs meet expectations on performance, safety, etc.	Weinberger, 2018	Does not risk pessimizing efficiency or effectiveness; embraces unknowable "Alien knowledge"	May mean some critical-decisions are made which cannot be explained; cannot ensure that decisions did not turn upon inappropriate bias
Auditability. Audit algorithms through varying input features	Doshi-Velez et al., 2017	Could get the best of both worlds: explainability with the power of deep learning black-boxes	Probably will not work in many cases; may be better at detecting specific cases of bias versus ruling out all potential bias
Explainability. Create new algorithms that are optimal and explainable	Angelino et al., 2017	Explainability without any of the performance detriments	Probably will not work in many cases; can cost in terms of efficiency and effectiveness
Red-line high-risk critical domain areas and demand explainability	Barocas & boyd, 2017	Ensures all critical-decisions arise from explainable systems	Excludes the potential benefits of AI from red-lined critical-decision domains

One concern with this input-varying approach to explainability is that it relies on an ability to vary eatures which may not be directly available given the very high-dimensioned inputs applied in deep learning applications. Indeed, the raw data input into these deep learning systems may be completely free of human-discernable features (such as race) and instead composed of, for instance, just a long series of pixel values. In these cases, the deep learning architecture relies on representation-learning methods that compose multiple levels of increasingly abstract data representations none of which, from input to output, are human discernable (LeCun, Bengio, & Hinton, 2015). Furthermore, consider the cases when the potential features of concern are not pre-known to those who wish to audit the algorithm, or when multiple features influence the

output through unexpected combinations such that single feature audits will never quiet reveal the full decision-making explanation.

Alternatively, some technologists have tried to find AI solutions that perform as well as deep learning neural networks while also producing outputs that are transparent and easily understood by humans. Angelino and co-authors (2017) have developed a decision list predictive model and applied it to the same recidivism datasets used by the COMPAS tool introduced in the ProPublica analysis described in Box 4.2.2. They propose an algorithmic approach that can offer a rule-based human explanation of its decision processes along with a "certificate of optimality," proving that the algorithm is as accurate and efficient as alternative black-box solutions. While this type of approach may not succeed in all or even most AI application areas, it could be explicitly demanded within domains of high-risk and critical decision making; domains where, perhaps, no Aliens need apply.

4.3. Data Privacy and Minimization

Value	Rewards	Risks	Connections
Data privacy, protection and minimization	Privacy preserving decentralized systems; privacy expert systems; new privacy protecting policies	User profile data breaches; de-anonymization and privacy concerns that arise from basic digital records.	Data privacy concerns with mobile operator user data.

Today's newspapers are replete with stories of personal data loss at the hands of large online data brokers and social media platforms. These stories surface many issues of data privacy, protection and minimization (the principle that data acquired should be limited to just what is necessary for the particular purpose at hand and shall not grow beyond what is necessary). As discussed above and in previous modules, many of today's most significant AI systems are made possible through the acquisition and analysis of large corpora of (potentially personal) data. And the range and depth of personal data acquired by AI systems are on the rise. For example, voice-driven conversational assistants, such as Alexa (Echo), Siri, and Cortana, may be more likely to know detailed private information such as what you are eating. What is clear is that users, and policymakers, are increasingly sensitive to privacy issues that arise from artificial intelligences. Indeed in a recent survey, "[a]lmost three-quarters (71%) [of respondents] say they don't want companies to use AI that threatens to infringe on their privacy, even if it improves the customer experience," (genpact, 2017).

4.3.1. Risks

Many recent headline-grabbing incidents illustrate some of the challenges related to AI systems which acquire, store and analyze large amounts of personal user data. First are risks associated with the very business models underlying some online platforms, and how these business

practices incentivize the acquisition of highly personal user data and make possible user data releases, both accidental and purposeful. A second goes to the kind of personal information that can be inferred indirectly from data released by users.

The release of millions of social media users' profile information to a data science and machine learning corporation, Cambridge Analytica, has been widely reported across major news media. (The Guardian has been at the forefront of some of this reporting, see https://www.theguardian.com/news/series/cambridge-analytica-files.) This user profile data was used by clients of Cambridge Analytica, including high-profile political campaigns within the USA, to micro-target persuasive advertising to specific and precisely modeled user populations. In this specific incident, a reported breach of user data along with sophisticated machine learning approaches came together to produce powerful targeted communications that may have had considerable political impact. This raises ethical and societal concerns for AI. First are privacy concerns that arise from the user profile data release, generally without the knowledge of the users involved. Second is the role of machine learning in influencing a nation's political processes (a topic we will not further address in this module).

Jose Marichal (2012) argues how the acquisition and disclosure of private user data, made famous in the Cambridge Analytica story, is not a misfeature of social media platforms but instead is central to their business model. He simplifies these platforms to a system for connections and disclosures. *Connections* are mostly realized through social media's facility to network friends. These friends are self-selected, and research has shown mainly consist of intimate, strong-tie relations existing offline as well as online. Communication exists within this network of close connections, and it is these user communications that Marichal refers to as *disclosures*. The architecture of disclosure is the platform's purpose-built environment to systematically and, in some ways, insidiously encourage its users to not simply disclose but increasingly to disclose personal revelatory data. To Marichal, social media has become the "perfect machine to get you to reveal intimate (if sometimes banal) details about yourself to others," and in so doing, reveal these very details to the platform and ultimately their clients and advertisers.

Social media systems have perfected this architecture not with degraded voyeuristic interest; it is simply their business model. They capture and commodify a portfolio of these disclosures, often through profile modeling, and sell this on to their advertisers. They have no prurient interest in your personal data, but it is the acquisition and analysis of this unique and highly personal dataset that has allowed social media systems to become the world's largest micro-targeted advertising platforms.

The Cambridge Analytica incident perfect illustrates the clash between this business model, predicated on acquisition of increasingly personal user data, and its associated privacy risks. Increasingly, policymakers and technology leaders are arguing that this incident underlines the need for ethical AI privacy standards and regulations (Hern, 2018). We are now constantly reminded that we have "entrusted the most intimate parts of [our] digital life to a profit-maximizing surveillance machine," (Roose, 2018).

While the Cambridge Analytica story is predicated on the release of intimate *private information*, up to and including even private instant messages, it has also been shown that private information can be discerned even from basic *public information* such as "Likes" (Kosinski,

Stillwell, & Graepel, 2013). In the Cambridge Analytica case, intimate private user details were released to a third party (for the purpose of political persuasion). But researchers have now concluded that even basic seemingly non-private information, some of which may be by default on offer to the public at large, can be used to infer our most intimate of private details through machine learning techniques. Therefore, a data breach of private profile data may not even be required for an entity to obtain personally revelatory information about social media users.

In their study, Kosinski and co-authors (2013) determined that "Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender…." This information can be gleaned through machine learning techniques even though users never imagined they were releasing such information and certainly never consented to such a release. In the hands of an advertiser, one can imagine this information being used to target promotions that may be unwanted or even harmful when received by the platform user. And if used by other actors, information about sexual orientation or political views might pose an even more dire and direct threat to an individual.

4.3.2. Rewards

Sophisticated machine learning techniques along with social media platforms' architectures of disclosure working in concert have created a substantial challenge to our individual privacy. Happily, AI technologies are also attempting to enhance user privacy that has increasingly been put at risk. For instance, Thomson Reuters has teamed up with IBM's Watson division to develop the Data Privacy Advisor, an expert system able to provide specialized advice to privacy professionals on their obligations across multiple jurisdictions (B. Smith & Al-Kohafi, 2018). The system works through an IBM Watson conversational interface allowing users to pose questions through natural language queries.

A few additional privacy-reclaiming approaches have been suggested (and to some extent are underway) by policy and technology leaders. This includes a set of responses around the putative monopoly control that some of the largest social media platforms may enjoy. Breaking up these large companies, and the enhanced competition that this might introduce, may have privacy enhancing effect as customers select for platforms which meet their privacy needs. Second, technologists have increasingly been developing and piloting decentralized social media platforms. These initiatives include platforms such as Mastodon (https://joinmastodon.org/) which bills itself as "the world's largest free, open-source, decentralized microblogging network." While there are some doubts as to whether these new systems can take hold, grow their user base, avoid being gamed by bad actors, and generally succeed (Barabas, Narula, & Zuckerman, 2017), nonetheless, among some there remains hope in platform decentralization.

Finally, there has increasingly been a call for increased data privacy regulations, and the European GDPR offers a glimpse at this regulatory approach. The GDPR directs that when companies collect user data, they must:

- tell them what they are using it for,

- minimize the amount of data they collect and keep to that needed just for their expressly articulated purpose
- tell people just what data they have on them,
- allows users to correct or have removed any of their data held by the company,
- and explain the logic they have used for any decision made based on the user data (Meyer, 2018).

These provisions have privacy preservation and data minimization (and explainability) at their core. As the GDPR roles out it is likely to drive changes to machine learning systems and their use of large user profile datasets. It is worth considering if the GDPR feature set, for instance, could limit future Cambridge Analytica type events.

4.3.3. Connections

Privacy concerns, made manifest when AI is applied to social media profile data, has a lot in common with the privacy issues that impact telecommunications operators. While at first blush it may seem that mobile operators hold relatively basic digital records of their users, such as cell tower derived user position data, a relatively small amount of mobile location data can be used to uniquely identify individuals. In this way, even anonymized data can be relatively easily de-anonymized. Researchers have shown that "the uniqueness of human mobility traces is high, thereby emphasizing the importance of the idiosyncrasy of human movements for individual privacy. Indeed, this uniqueness means that little outside information is needed to re-identify the trace of a targeted individual even in a sparse, large-scale, and coarse mobility dataset," (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013). In turn, this locational data can be used to infer private details of the individual (Blumberg & Eckersley, 2009).

Thus the data privacy concerns that arise out of AI systems (including those based on social media profile data analysis) has significant similarities to the data privacy concerns already present with telecommunication user data (including mobile location data). Both data sets when subject to powerful analysis can turn even the public and seemingly most benign information into deeply personal details. This challenge will grow even for operator held user data sets as the capabilities of AI-driven analytics expands. AI engines, applied to an operator's user data, may result in intimate private user information almost at the scale held by social media platforms.

4.3.4. Key Questions to Consider for This Value

Can platform decentralization and private sector competition solve many of our data privacy woes?
How will policy responses, including the GDPR, support user privacy requirements?
Can privacy processes already in place for operator data assist us as AI grows in its analytic capabilities and data volume, sources, and services expand?
What are the new privacy risks and data protection imperatives for ICT operators when applying AI analytics to their large user datasets?

4.4. Peace and Physical Security

Value	Rewards	Risks	Connections
Peace and physical security	Systems for inclusivity and trust-building; peacekeeping situational awareness	Lethal Autonomous Weapons Systems	Special policy needs and realities in conflict stressed environments

Silicon Valley companies are increasingly experiencing internal debates as companies develop programs relevant to peace and warfare sectors. For instance at Google thousands of employees wrote a letter to their CEO calling for a company moratorium on "warfare technology" ," (Blumberg & Eckersley, 2009; Shane & Wakabayashi, 2018). These employees letter is indicative of a growing community of AI stakeholders calling on a moratorium on AI enabled warfare.

Alternatively, proponents have noted the promise AI holds for peace-preserving influences (for instance in fair resource distribution and climate mitigation) and conflict and crises response (Best, 2013). As one writer has put it, "AI holds much promise to enable the international community, governments and civil society to predict and prevent human insecurity. With increased connectivity, more sophisticated sensor data and better algorithms, AI applications may prove beneficial in securing basic needs and alleviating or stopping violent action," (Roff, 2017).

4.4.1. Risks

Even before Google employees sent a letter to their CEO, a large number of AI researchers (along with many thousand other endorsers) signed a letter calling for a moratorium on the development of AI-driven autonomous weaponry (https://futureoflife.org/open-letter-autonomous-weapons/). This letter concludes with the statement that, "[s]tarting a military AI arms race is a bad idea, and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control." As AI's relevance to warfare continues to grow, technology stakeholders are increasingly expressing concern especially around the development of Lethal Autonomous Weapons Systems (LAWS). LAWS are systems which, according to Regina Surber (2018), "once activated, would, with the help of sensors and computationally intense algorithms, identify, search, select, and attack targets without further human intervention."

Apprehension arises around the reduction of human control, which may make warfare seem lower risk or "easier," position AIs in situations that rely on human morality and judgment, and confuse issues of accountability. Some have called to moderate the autonomy of these AI systems, requiring instead "meaningful human control" or MHC. "At its most basic level, the requirement for MHC develops from two premises: 1. That a machine applying force and operating without any human control whatsoever is broadly considered unacceptable. 2. That a human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense," (Roff & Moyes, 2016). Among many scholars, a consensus may hold that people must always be meaningfully in the loop over kill decisions (Calo 2017).

The UN Convention on Certain Conventional Weapons (UN CCW), also known as the Inhumane Weapons Convention, has taken up the issue of AI-enabled autonomous weaponry. In 2014 UN

CCW convened its first informal Meeting of Experts; more recently they have stood up a Group of Governmental Experts (GGE) on LAWS. The GGE has continued to meet and explore the legality of LAWS under international law, methods for assigning responsibility and deciding questions of accountability with these autonomous systems, and consideration of the various international normative principles challenged by these technologies (Surber, 2017). This UN work is in its early stages, and much is needed to move the group towards an agreed to political declaration or international treaty.

However, others have argued that there is no practical method to restrict the development of autonomous AI enabled weaponry. Cummings (2017) argues that "[b]anning an autonomous technology for military use may not be practical given that derivative or superior technologies could well be available in the commercial sector." If the commercial sector is already undertaking an "arms race" to be the first to develop robust autonomous systems (e.g., driverless cars) then it might be practically impossible to restrict these systems from being applied in warfare settings.

4.4.2. Rewards

Can AI serve as a tool to reduce conflict and wage peace? Or will emerging AI's have unintended consequences that exacerbate war or, when placed in the wrong hands, actively erode peace?

The AI and Peace Consortium, currently incubated between Georgia Tech and Harvard's Berkman Klein Center, aims to explore the relationship of AI to peacemaking and peacekeeping through policy, social scientific and computational means. Collaborators will pursue novel research studies and interventions and convene stakeholders, scholars and decision-makers at workshops. Indeed the AI systems for inclusivity mentioned above are examples of the potential for AI to help with conflict mitigation, trust building, and post-conflict reconciliation (Best, Long, Etherton, & Smyth, 2011).

The UN Peacekeeping communities have also looked to AI to assist them in their mission to promote and establish peace in conflict stressed areas. Many have noted how peacekeeping suffers from insufficient access to and ineffective use of digital technologies (e.g., see Stauffacher, Weekes, Gasser, Maclay, & Best, 2011). Since the Brahimi Report (2000), which argues peacekeeping must be brought into the information age, operations have used ICTs but struggled to capture their full capabilities or keep pace with their rapid change. An ongoing program of relevant UN departments is exploring a broad and far-reaching platform for peacekeeping situational awareness. This platform relies on AI capabilities in data capture and validation; tracking, sensors and data integration; analysis; and visualization. Indeed, AIs could be responsive to three traditional security and peacekeeping challenges: "the inability to know about threats in advance; the inability to plan appropriate courses of action to meet these threats; and, the lack of capacity to empower stakeholders to effectively respond," (Roff, 2017). Indeed, Roff argues that artificial intelligences could automate most of the tasks associated with peacekeeping logistical support, supply chain management, forecasting and planning, and so forth.

4.4.3. Connections

In other venues we have examined the telecommunications policy process in conflict stressed environments (Best, 2011; Best & Thakur, 2009). We find that while there are many similarities in ICT policymaking between conflict stressed environments as compared with other locations, there are also differences. In particular, in conflict and immediate post-conflict states, policymaking has to contend with a weak and nascent institutional environment, intra-governmental competition, limited human and technical resources, the contested role of international actors such as the World Bank, and the dominance of elite groups in decision-making. While some of these factors are not unknown to many countries, especially in the Global South, they can be particularly germane in conflict-stressed environments.

The social and ethical concerns that arise out of AI systems applied in peace and security areas are likely to test many parts of a policymaking process. The distinctive risks associated, for instance, with LAWS means that policy and regulatory responses probably cannot just wait to "see what happens" but instead will need to be proactive and respond early. If a policy response is to be successful, traditional ICT infrastructures will almost certainly have to play a direct role.

4.4.4. Key Questions to Consider for This Value

Should AI, at least in the form of lethal autonomous weapons, be banned or tightly regulated?
How can we encourage the development of peace-enhancing AI systems?
Can traditional ICT infrastructures be brought to bear towards both of the above questions – somehow encouraging the best forms of peaceable AI while excluding the worse forms of AI supported warfare?
What are the special concerns for ICT policymakers operating within conflict stressed regions?

5. Conclusions

In this module we have proposed a bivalenced values framework for AI and have explored the risks and rewards associated with AI systems for four core example values: livelihood and work; diversity and non-discrimination; privacy and data minimization; and peace and security. Of course, this is just a starting set of values of importance to and impacted by AI. There are many other values salient to AI which should be considered when formulating policy and regulatory responses to emerging systems and that regulators and policymakers should keep in mind when examining the AI sector. The sections above offer just a handful of the many salient values that we hold, and which are impacted by and should drive ethical decisions around artificial intelligence technologies. Future work in AI, ethics and society might undertake similar analysis but for other values such as: 1) Economic freedoms and wealth; 2) democratic rights and civic engagement; 3) food security and healthy living; 4) leisure and entertainment 5) climate resilience; and 6) literacy and education.

In this module, we have also reviewed ways in which AI systems connect to systems and infrastructures well known to ICT policymakers and regulators. While there are ways that AI feels new – and without question it is particularly broad in scale and scope and is advancing with

unusual (nearly unprecedented) speed – it nonetheless shares plenty of features we encountered with the growth of mobile telephony or the internet. As we reflect on these connections between AI's social and ethical import and related values we encountered with other ICT infrastructures, we find a plethora of ways that AI impacts on areas already mandated to ICT policymakers:

- The ICT sector as a target or beneficiary of AI. For example, customer data retained by mobile and internet service providers can be subject to powerful de-anonymizing AI analysis increasing the import of data security and privacy among operators.
- The ICT sector as a tool for supporting the best forms of AI and responding to the worst. For example, operators may be best able to assist other stakeholders in identifying and responding to potentially harmful AIs released onto their networks.
- The ICT sector as a set of businesses directly employing AI, potentially in ways that have policy and regulatory relevance. For example, consider how much of operator customer support may move away from human agents (including offshored call offices) to AI chatbots.

While these are examples of ways AI is related to existing core ICT regulatory and policy areas, it is likely that in many locations ICT policy stakeholders will be asked to take on even more direct consideration of emerging AI issues. In order to be respond to existing mandated areas, and be ready for increasing and new considerations, ICT policymakers must remain informed, nimble, and conversant around the various social and ethical aspects of artificial intelligence. To do so, they must engage in real-time learning and consultation among multi-stakeholder cross-institutional coalitions.

This is already happening across a number of jurisdictions, among multi-lateral and professional societies, and within various companies. Box 5 overviews just some of the emerging policy reports and acts that are emerging.

With these connections in mind, and putting it simply, there are many many ways that artificial intelligence is already touching on areas of concern to ICT policymakers and regulators, and these associations are likely to grow, not diminish, with time. If useful, apply this module's bivalenced values framework to interrogate some of the many ways that AI is or may soon impact human ethics and society in areas relevant to ICT policymakers and regulators. It is these realities that drives us to our AI and Ethics Admonition. If they are not already, it is critically important that ICT policymakers and regulators engage with the many areas of ethical and social concern that AI touches.

BOX 5: A SAMPLE OF AI ETHICAL AND SOCIAL POLICY INITIATIVES

In the last few years a variety of initiatives have been launched to explore the ethical and social implications of AI, and to formulate policy responses to them. These initiatives can be loosely classified into three categories: governmental initiatives; initiatives started by tech companies; and those started by non-governmental organizations, academia and professional associations.

Example Governmental Initiatives

The US Government's *Preparing for the Future of AI Report* recommends re-evaluation of existing regulation with an eye towards adopting it to AI, as well as striking a balance between boosting

innovation, the costs and benefits to regulation and compliance, and the needs of public safety and fairness (National Science and Technology Council, 2016). The report also acknowledges that AI systems will need to transition cautiously from laboratories to real-life human environments, in order to avoid unsafe and unforeseen situations. Recently a FUTURE of Artificial Intelligence Bill was also introduced in US Congress that recommended forming a *Federal Advisory Committee On Development And Implementation Of Artificial Intelligence* (FUTURE of Artificial Intelligence Act, 2017). The bill sought to identify and eliminate possible bias in selection and processing of data used by AI algorithms, enhance the diversity in algorithm development, and identify applications of technology that could possibly have adverse consequences. It also explores how AI innovation will affect the privacy of individuals, create socio-economic changes, and how the government can best adopt AI to improve its own efficiency.

The European Union recently published its *Statement on AI, Robotics and 'Autonomous' Systems* that proposes nine fundamental principles for governing AI and further calls for creating an internationally recognized, common, ethical legal framework (European Group on Ethics in Science and New Technologies, 2018). It warns that the absence of a common AI regulation framework can result in "ethics shopping", and the relocation of AI development to regions with lower ethical standards. In April 2018 EU member states signed the *Declaration of Cooperation on Artificial Intelligence* and agreed on creating an ethical and legal framework based on fundamental rights and values enshrined in the EU charter including "privacy and protection of personal data" (European Council, 2018).

United Kingdom's House of Lords published a comprehensive report *AI in UK: Ready, Willing and Able?* (House of Lords' Select Committee on Artificial Intelligence, 2018), that discussed supporting and strengthening the AI industry through policy and education, as well as managing the loss of employment due to automation. It highlights the need for better legal and technical mechanisms allowing users to tailor control of their personal data while protecting privacy, instead of total data openness or total data privacy. The report endorses establishing data-trusts for the ethical sharing of data between organizations, and to counter data monopolization by a few technology companies globally. It also recomends developing a cross-sector ethical code of conduct, or "AI code" across private and public sectors, that includes creating an ethical board across companies pursuing AI development or use. The report argues that a blanket AI regulation will not be appropriate, and instead advocates for a more sector-wise regulation approach. Finally, the report also calls on research councils to request from university applicants demonstrations of implications of their AI research, and its potential misuse, along with measures taken to prevent misuse.

In 2017, China published its *New Generation AI Development Plan* (State Council, 2017), calling on financial and state resources to develop AI ensuring a "first mover advantage". The Plan also discusses challenges that AI will bring to Chinese society in employment, social stability, security risks and changing norms in international relations. It discusses developing a multi-level ethical framework for governing human-machine collaboration and deepening international cooperation to create artificial intelligence laws and regulations. The Development Plan recommends strengthening AI risk assessments with a long-term focus and establishing security monitoring & early warning mechanisms for AI.

Example Corporate Initiatives

Over the last years many Tech majors have developed principles and policies for AI research and development. For example, Microsoft states its four principles as: Fairness, Accountability, Transparency and Ethics (https://www.microsoft.com/en-us/ai/our-approach-to-ai). It speaks of honoring societal values and diversity of experience, though does elaborate on how it would implement these principles.

IBM has also articulated and AI position, declaring, for instance, that it will only develop artificial intelligence systems that augment human ability and not have any independent agency (https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/). It has also announced that IBM will be transparent about when and for what purposes it uses AI. IBM has proposed algorithmic responsibility and explanation based systems as a trust-building means to address bias in computer decision making (Banavar, 2016). They have also established an internal IBM Cognitive Ethics Board to advise and guide AI development and deployment.

In 2017 Google established the *DeepMind Ethics & Society* program, which has identified six key ethical challenges: Privacy, Transparency and Fairness; Economic Impact; Governance and Accountability; Managing AI Risk; AI Morality and Values; and Global Challenges (https://deepmind.com/applied/deepmind-ethics-society/research/). Their Responsible Development of AI document (https://www.blog.google/topics/ai/ai-principles/) recommends development of AI systems whose benefits will outweigh their risks. It states that Google will not pursue AI technology that supports surveillance, contravenes international law and human rights, or that can be used as a weapon. It also states that it will incorporate privacy safeguards and offer control over use of data and that its AI systems will affirm accountability by providing "appropriate opportunities for feedback, relevant explanations and appeal". Google has further stated that it will develop AI in accordance to the prevalent best practices and monitor all AI technologies post their deployment.

Technology companies have not only articulated individual AI principles, they have also been collaborating towards shared AI social and ethical policies. For example, the *Partnership on AI* is an initiative founded by Facebook, Amazon, Apple, Google, DeepMind, IBM and Microsoft (https://www.partnershiponai.org). The organization aims to advance public understanding and provide an inclusive platform for discussion and engagement with key stakeholders. It opposes "development and use of AI technology that would violate international conventions or human rights".

Other Example Initiatives

Among professional associations, the IEEE has convened a diverse set of experts to publish *Ethically Aligned Design* (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017). The report explores the application of ethical principles to AI, as well and includes legal frameworks for accountability. It also explores "embedding values into autonomous systems"; and concludes that such a system will need to be trained in values specific to the community it is deployed in, as well as in norms relevant to the operation it is designed for. The ACM US Public Policy Council (2017), meanwhile, has released a *Statement on Algorithmic Transparency and Accountability* that advises transparency of data used to train AI systems, as well as explainability of their decisions. The statement also suggests auditing systems

in case of harm, redressing groups adversely affected by algorithms, and holding the algorithm producing company accountable.

The *Future of Life Institute* has been at the forefront of exploring ethical challenges posed by AI. Asilomar Principles, published by the Institute, have been endorsed by thousands of AI researchers and reject the creation of an AI with undirected intelligence (https://futureoflife.org/ai-principles/). The principles advocates for people's right to access, manage and control their data and against any use of private data that would curtail real or perceived liberties. The principles also insist that if am AI system could cause harm then the cause must be identifiable. They address society at large calling for sharing of economic prosperity created through AI innovation, and strengthening social and civic processes using AI systems, rather than subverting them.

TABLE 5

			Type				Values Considered				
Sector	Country	Title	Policy	Act	Principles	Livelihood Jobs	Diversity Anti-Bias	Accountability	Privacy	Peace	Security
Government Initiatives	USA	Preparing for the Future of AI	X			X	X	X	X	X	X
	USA	FUTURE of AI Act		X		X	X	X	X	X	
	EU	Statement on AI, Robotics and Autonomous Systems	X			X	X	X	X	X	X
	EU	Declaration of Cooperation on Artificial Intelligence			X	X		X	X	X	
	UK	AI in UK	X			X	X	X	X	X	X

	China	New Generation AI Development Plan	X			X		X		X	X
	Microsoft	Our Approach to AI			X		X	X		X	
	IBM	Transparency and Trust in the Cognitive Era	X			X		X	X		
Corporate Initiatives	Google	Responsible AI Practices			X	X	X	X	X	X	X
	Facebook, Amazon, Apple, Google IBM, Microsoft	Partnership on AI			X	X	X	X	X	X	X
Other Examples	IEEE	Ethically Aligned Design Version 2	X			X	X	X	X	X	X
	ACM	Statement on Algorithmic Transparency and Accountability	X				X	X	X		
	Future of Life Institute	Asilomar AI Principles			X	X	X	X	X	X	X
	AINow	AINow Report 2017	X			X	X	X	X	X	

BOX 5. B : CHINA'S GREAT LEAP INTO AI

Realizing the potential of AI technology, the Chinese government has placed AI-related technologies as one of the strategic priorities for the next decade. President Xi Jinping, in his Report at the 19th Chinese Communist Party National Congress, declared that China is to become a "science and technology superpower"; four months before in July 2017, the Chinese State Council published the Next Generation Strategic Plan for AI technologies, in which it

specifically vocalized China's goal "to achieve global leadership in AI theories, technologies, and general application, as well as becoming a major AI innovation center worldwide" by 2030.[1] [2]

China is looking at AI as an enabler of the "Chinese Dream of the Great Rejuvenation of the Chinese People" and a crucial part to building "an innovative country.[3] Specifically, the State Council recognizes the profound impact AI technology can make in international geopolitics, economic prosperity, and societal development. For the Communist Party leadership, China has to undertake a national strategic initiative if it is to compete among the top international AI actors.[4]

Politically, the Chinese government considers leadership in AI technology a tool to "improve national competitiveness," especially as China "currently faces a complicated scene of national security and national competition."[5] In his 2017 and 2018 Annual Government Work Reports, Premier Li Keqiang enunciated China's plan to "secure core technology, develop top talent, and enforce high standards" in the near future.[6] Specific to the promotion of AI-related policymaking, the State Council plans to carry out research on the legal issues of AI, including its impacts on civil and criminal liability, privacy and intellectual property protection, safe use of information, and system accountability and transparency.[7]

At the societal level, the Chinese government believes the application of AI technology is a "new opportunity for societal construction."[8] The government considers AI technologies to be instrumental in the current "fully developing a moderately prosperous society."[9] The State Council plans to utilize AI technology to advance various societal issues ranging from education, medical care, environmental protection, urban management, legal counsel, and providing care

[1] 习近平在中国共产党第十 九次全国代表大会上的报告 (Xi Jinping's Report at the 19th Chinese Communist Party National Congress)
http://cpc.people.com.cn/n1/2017/1028/c64094-29613660-7.html

[2] 国务院印发《新一代人工智能发展规划》 (The State Council Issues "Next Generation Artificial Intelligence Development Strategic Plan")
http://www.gov.cn/xinwen/2017-07/20/content_5212064.htm

[3] 国务院关于印发新一代人工智能发展规划的通知 (Notice of The State Council's Issuance of Next Generation Artificial Intelligence Development Strategic Plan)
http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

[4] Ibid.

[5] Ibid.

[6] 2018 李克强总理政府工作报告 (Premier Li Keqiang's Annual Government Report in 2018)
http://www.xinhuanet.com/politics/2018lh/2018-03/22/c_1122575588.htm

[7] 国务院关于印发新一代人工智能发展规划的通知(Notice of The State Council's Issuance of Next Generation Artificial Intelligence Development Strategic Plan)
http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

[8] 国务院关于印发新一代人工智能发展规划的通知(Notice of The State Council's Issuance of Next Generation Artificial Intelligence Development Strategic Plan)
http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

[9] Ibid.

for the elderly.[10] The government hopes to leverage the perception, prediction, warning, and analyzing capabilities of AI systems based on big data to take an active role in policy-making and improve its ability of social management and stability maintenance.[11]

In the private sector, the three largest Chinese technology companies, Baidu, Alibaba, and Tencent are all making strong investments into AI research and development. Baidu, the Chinese search engine, is the best-equipped and most advanced in AI development. Like Google, Baidu has unique advantages in algorithm and data collection, and has a natural inclination to take a lead in AI technologies. Yanhong (Robin) Li, the Chairman and founder of Baidu, has declared AI to be its next "utmost priority" and expressed his optimism about AI as the "lever to new economic prospects" in a recent state media interview.[12] In February, the Chinese State Department requested Baidu to spearhead the National Engineering Lab on Deep Learning Technology and Application with collaborators including Tsinghua University, Beijing Aeronautics and Aerospace University, China Institute of Information, and others.[13] Baidu has currently developed two open-source platforms, DuerOS (Baidu's virtual assistant application) and Apollo (an open-source AI solutions platform) which is offering pilot tools for application development in finance, education, and medical services.[14]

Another major player in the private sector, Alibaba, is a relatively new actor in the development of AI. Unlike its American counterpart, Amazon, Alibaba—the biggest online retailer in China—established its own AI department just two years ago as an extension of its e-commerce platform. As of now, Alibaba's AI technologies are not yet at a level to compete with other global major players such as Microsoft or Google, and still primarily serve as a part of its powerful cloud computing and e-commerce network.[15]

Tencent, the developer of instant messaging services WeChat and QQ, with over 600 million daily active users, has also begun to take advantage of its data resources to become more vocal in the AI theater. Tencent's highest ranked leadership pays a significant amount of attention on its AI department. Tencent's CEO, Zhiping Liu, has repeatedly claimed that AI is a core technology in all of Tencent's products.[16] Tencent encourages every team on every project to expand its involvement in the AI sector, and to apply AI core technologies. At the same time, Tencent is building an experimental AI lab to research fundamental AI technologies. Currently, Tencent has

[10] Ibid.

[11] 国务院关于推进"互联网+"行动的指导意见(The Guiding Opinions Regarding the "Internet Plus" Initiative from the State Council)
http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm

[12] 李彦宏委员：用人工智能"撬开"关于未来的想象(Commissar Yanhong Li: Using AI to "Crack" Imaginations on the Future)
http://www.gov.cn/xinwen/2018-03/04/content_5270502.htm

[13] http://www.tsinghua.edu.cn/publish/thunews/9659/2017/20170303142537710950910/20170303142537710950910_.html

[14] 百度公开 AI 生态开放战略(Baidu Announces Its AI Ecosystem Strategies)
http://www.xinhuanet.com/tech/2017-07/06/c_1121271415.htm

[15] https://www.leiphone.com/news/201805/5sM1zwCCE1IBo5j7.html

[16] https://ai.tencent.com/ailab/腾讯总裁刘炽平：人工智能具有战略意义，加码投入不急于短期回报.html

invested a significant amount of capital in voice recognition, image recognition, computation visualization, voice processing, and deep learning.[17]

With national strategic leadership, a clear aim to become a global AI leader, and a number of highly invested major corporations, China is emerging as an AI behemoth. The many ways this will influence social and ethical issues of AI remain unclear.

[17] https://www.leiphone.com/news/201704/x1wlWNDfDZJqo3xz.html

References

ACM US Public Policy Council. (2017). *Statement on Algorithmic Transparency and Accountability*. Washington, DC: USACM. Retrieved from file:///Users/michaelbest/UNU/AI4Peace&Society/2017_usacm_statement_algorithms.pdf

Al Jazeera. (2017, June 3). Beyond meat: The end of food as we know it? Retrieved April 18, 2018, from https://www.aljazeera.com/programmes/talktojazeera/2016/02/meat-artificial-food-160205152233913.html

Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., & Rudin, C. (2017). Learning Certifiably Optimal Rule Lists. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 35–44). ACM.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. *ProPublica*, *23*.

Asimov, I. (1950). *I, Robot*. Greenwich, CT: Fawcett Publications.

Baer, D. (2016, April 26). This startup is using machine learning to create animal product substitutes. *Business Insider*. Retrieved from http://www.businessinsider.com/chilean-startup-uses-machine-learning-for-meat-subsitutes-2016-4

Banavar, G. (2016). Learning to Trust Artificial Intelligence Systems - Accountability, Compliance and Ethics in the Age of Smart Machines. *IBM*.

Barabas, C., Narula, N., & Zuckerman, E. (2017, September 8). Decentralized Social Networks Sound Great. Too Bad They'll Never Work. *WIRED*. Retrieved from https://www.wired.com/story/decentralized-social-networks-sound-great-too-bad-theyll-never-work/

Barocas, S., & Boyd, D. (2017). Engaging the ethics of data science in practice. *Communications of the ACM*, *60*(11), 23–25. https://doi.org/10.1145/3144172

Beede, D., Powers, R., & Ingram, C. (2017). *The Employment Impact of Autonomous Vehicles*. US Department of Commerce.

Best, M. L. (2011). Mobile Phones in Conflict-Stressed Environments: Macro, Meso and Microanalysis. In M. Poblet (Ed.), *Mobile Technologies for Conflict Management: Online Dispute Resolution, Governance, Participation*. London: Springer.

Best, M. L. (2013). Peacebuilding in a networked world. *Commun. ACM*, *56*(4), 30–32. https://doi.org/10.1145/2436256.2436265

Best, M. L., Long, W. J., Etherton, J., & Smyth, T. (2011). Rich Digital Media as a Tool in Post-Conflict Truth and Reconciliation. *Media, War & Conflict*, *4*(3), 231–249.

Best, M. L., & Thakur, D. (2009). The Telecommunications Policy Process in Post-conflict Developing Countries: The Case of Liberia. *INFO Journal*, *11*(2), 42–57.

Blumberg, A. J., & Eckersley, P. (2009). On locational privacy, and how to avoid losing it forever. *Electronic Frontier Foundation*, *10*(11).

Boddington, P. (2017). *Towards a Code of Ethics for Artificial Intelligence*. Springer.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press Oxford.

Brahimi, L. (2000). *Report of the Panel on United Nations Peace Operations*. Retrieved from http://www.un.org/en/ga/search/view_doc.asp?symbol=A/55/305

Brandusescu, A., Freuler, J. O., & Thakur, D. (2017). *Artificial Intelligence: Starting the Policy Dialogue in Africa*. Washington, DC: World Wide Web Foundation.

Calo, R. (2017). Artificial Intelligence Policy: A Primer and Roadmap. *UCDL Rev.*, *51*, 399.

Clark, J. (2016). Artificial intelligence has a 'sea of dudes' problem. *Bloomberg Technology*, *23*.

Corbett-Davies, S., Pierson, E., Feller, A., & Goel, S. (2016, October 17). A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/

Crawford, K. (2016). Artificial intelligence's white guy problem. *The New York Times*.

Cummings, M. L. (2017). *Artificial Intelligence and the Future of Warfare*. London: Chatham House.

DataProphet. (2018, April). Retrieved April 13, 2018, from https://dataprophet.com/

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*(1). Retrieved from http://www.nature.com/articles/srep01376

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., … Wood, A. (2017). Accountability of AI Under the Law: The Role of Explanation. *ArXiv Preprint ArXiv:1711.01134*.

European Council. (2018, April 10). Declaration of Cooperation on Artificial Intelligence.

European Group on Ethics in Science and New Technologies. (2018, March). Statement on Artificial Intelligence, Robotics and "Autonomous" Systems. European Group on Ethics in Science and New Technologies.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280.

Fundamentally Understanding The Usability and Realistic Evolution of Artificial Intelligence Act, H.R. 4625 § (2017).

Gartner, Inc. (2017). *Predicts 2018: AI and the Future of Work*. Stamford, CT: Gartner, Inc. Retrieved from https://www.commerce-associe.fr/wp-content/uploads/predicts_2018_ai_and_the_fut_342326.pdf

genpact. (2017). *The consumer: Sees AI benefits but still prefers the human touch*. Retrieved from http://www.genpact.com/downloadable-content/the-consumer-sees-ai-benefits-but-still-prefers-the-human-touch.pdf

Good, I. J. (1966). Speculations concerning the first ultraintelligent machine. In *Advances in computers* (Vol. 6, pp. 31–88). Elsevier.

Hern, A. (2018, April 15). Cambridge Analytica scandal "highlights need for AI regulation." *The Guardian*. Retrieved from http://www.theguardian.com/technology/2018/apr/16/cambridge-analytica-scandal-highlights-need-for-ai-regulation

House of Lords' Select Committee on Artificial Intelligence. (2018). *AI in the UK: Ready, Willing, and Able?*

Katz, R. L. (2017). *Social and Economic Impact of Digital Transformation on the Economy* (GSR-17 Discussion Paper). Geneva: ITU.

Kleinberg, J. M., Mullainathan, S., & Raghavan, M. (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores. *CoRR*, *abs/1609.05807*. Retrieved from http://arxiv.org/abs/1609.05807

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LHoFT, T. (2017, November 8). FinTech for All: Access to finance for Kenya's agriculture industry. Retrieved April 18, 2018, from https://medium.com/@The_LHoFT/fintech-for-all-access-to-finance-for-kenyas-agriculture-industry-f723ca420787

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., … Sanghvi, S. (2017). Jobs Lost, Jobs Gained: Workforce Transitions In A Time Of Automation. *McKinsey Global Institute*.

Manyika, J., & Spence, M. (2018, February 5). The False Choice Between Automation and Jobs. *Harvard Business Review*.

Marichal, J. (2012). *Facebook Democracy: The Architecture of Disclosure and the Threat to Public Life*. Farnham, UK: Ashgate Publishing Limited.

Matsuura, S. (2017, November 22). "Mesmo transparente, Brasil tem escândalos", diz criador de robô que analisa gastos públicos. Retrieved April 18, 2018, from https://oglobo.globo.com/sociedade/tecnologia/mesmo-transparente-brasil-tem-escandalos-diz-criador-de-robo-que-analisa-gastos-publicos-22097963

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. AK Peters Natick, MA.

Metzinger, T., Bentley, P. J., Häggström, O., & Brundage, M. (2018). *Should we fear the future of artificial intelligence?* Brussels: European Union.

Meyer, D. (2018, May 25). AI Has a Big Privacy Problem And Europe's New Data Protection Law Is About to Expose It. *Fortune*. Retrieved from http://fortune.com/2018/05/25/ai-machine-learning-privacy-gdpr/

Monnerat, A. (2018, January 12). Data scientists in Brazil working on the country's first robot-journalist to report on congressional bills. Retrieved April 18, 2018, from https://knightcenter.utexas.edu/blog/00-19182-data-scientists-brazil-working-country%E2%80%99s-first-robot-journalist-report-congressional-b

National Science and Technology Council. (2016). *Preparing for the Future of Artificial Intelligence*. Committee on Technology. Retrieved from http://itlaw.wikia.com/wiki/Preparing_for_the_Future_of_Artificial_Intelligence

Novitske, L. (2018). The AI Invasion is coming to Africa (and It's a Good Thing). *Stanford Social Innovation Review*.

Osoba, O. (2017). *An intelligence in our image: the risks of bias and errors in artificial intelligence*. Santa Monica, Calif: RAND Corporation.

Petropoulos, G. (2017, April 27). Do we understand the impact of artificial intelligence on employment? Retrieved June 4, 2018, from http://bruegel.org/2017/04/do-we-understand-the-impact-of-artificial-intelligence-on-employment/

Roff, H. M. (2017). *Advancing Human Security through Artificial Intelligence*. London: Chatham House.

Roff, H. M., & Moyes, R. (2016). Meaningful human control, artificial intelligence and autonomous weapons. In *Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons*.

Roose, K. (2018, March 28). Can Social Media Be Saved? *The New York Times*. Retrieved from https://www.nytimes.com/2018/03/28/technology/social-media-privacy.html

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252.

Samans, R., & Zahidi, S. (2017). *The Future of Jobs and Skills in Africa - Preparing the Region for the Fourth Industrial Revolution*. World Economic Forum.

Schoeman, W., Moore, R., Seedat, & Chen, J. (2017). *Artificial Intelligence - Is South Africa Ready?* Accenture.

Shane, S., & Wakabayashi, D. (2018, April 4). 'The Business of War': Google Employees Protest Work for the Pentagon. *The New York Times*. Retrieved from https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

Smith, B., & Al-Kohafi, K. (2018, January 29). How Thomson Reuters and IBM are bringing AI to data privacy professionals. Retrieved May 2, 2018, from https://www.ibm.com/blogs/watson/2018/01/thomson-reuters-ibm-bringing-ai-legal-professionals/

Smith, M. L., & Neupane, S. (2018). *Artificial intelligence and human development: Towards a research agenda*. Ottawa, Canada: IDRC.

State Council. (2017, July). New Generation of Artificial Intelligence Development Plan. State Council.

Stauffacher, D., Weekes, B., Gasser, U., Maclay, C., & Best, M. (Eds.). (2011). *Peacebuilding in the Information Age: Sifting hype from reality*. Geneva, Switzerland: ICT4Peace Foundation.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … Kraus, S. (2016). Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*.

Stupp, C. (2018, April 10). Twenty-four EU countries sign artificial intelligence pact in bid to compete with US & China. Retrieved April 18, 2018, from https://www.euractiv.com/section/digital/news/twenty-four-eu-countries-sign-artificial-intelligence-pact-in-bid-to-compete-with-us-china/

Surber, R. (2017). *Artificial Intelligence: Lethal Autonomous Weapons Systems and Peace Time Threats*. Zurich, Switzerland: ICT4Peace Foundation.

Surber, R. (2018). *Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats*. Zurich, Switzerland: ICT4Peace Foundation.

The Economist. (2016a, June 25). Automation and anxiety; The impact on jobs. *The Economist*, *419*(8995).

The Economist. (2016b, June 25). March of the machines; Artificial intelligence. *The Economist*, *419*(8995).

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically Aligned Design: A Vision for Prioritizing Humman Well-Being with Autonomous and Intelligent Systems* (No. Version 2). IEEE. Retrieved from http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433.

Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940022856.pdf

Weinberger, D. (2017, April 18). Our Machines Now Have Knowledge We'll Never Understand. *WIRED*.

Weinberger, D. (2018, January 28). Don't Make Artificial Intelligence Artificially Stupid in the Name of Transparency. *Wired*. Retrieved from https://www.wired.com/story/dont-make-ai-artificially-stupid-in-the-name-of-transparency/

Weissman, C. G. (2016, August 18). The Women Changing the Face Of AI. Retrieved April 10, 2018, from https://www.fastcompany.com/3062932/ai-is-a-male-dominated-field-but-an-important-group-of-women-is-changing-th

Weston, M. (2015, January 4). Digital Translation in an Analog Country. *Myanmar Business Today*, *12*(51). Retrieved from https://www.mmbiztoday.com/articles/digital-translation-analog-country

World Bank. (2016). *World Development Report 2016: Digital Dividends*. Washington, DC.

World Wide Web Foundation. (2017). *Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries*.

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global Catastrophic Risks*, *1*(303), 184.

Artificial Intelligence (AI) for Development Series

Report on AI and IoT in Security Aspects

July 2018

Work in progress, for discussion purposes
Comments are welcome!
Please send your comments on this paper at: gsr@itu.int by 30 July 2018

This Introductory module was prepared by Gyu Myoung Lee, under the direction of the ITU/BDT Regulatory and Market Environment Division and ICT Applications and Cybersecurity Division and under close coordination with the Chief of the ITU/BDT Infrastructure, Enabling Environment, and E-Applications Department. We would like to thank the ITU General Secretariat and the ITU Standardization Bureau for their contributions.

Contents

1. Introduction

ITU BDT has launched an Artificial Intelligence (AI) for Development Series to help Information Communications Technology (ICT) regulators (NRAs) prepare for AI, digital transformation and the digital world.

The Series includes an overall framework that will set the scene with modules on:

- 5G development examining the investment and infrastructure requirement (to support digital transformation, AI, Internet of Things (IoT), etc.);
- the social and economic impact of digital transformation;
- AI regulation for governance;
- AI for society and the security and data protection aspects linked to IoT and AI.

The series also includes a roadmap of actions.

The module on AI and IoT in Security Aspects examines the relationship between AI and IoT and analyse the security aspects linked to AI as the key component for the full realization of IoT. It also addresses the potential roles of competent national authorities, such as NRAs in ensuring that security and data protection aspects are taken into consideration or the roll-out of IoT services and applications, using AI features (e.g., machine learning).

In this regard, this report examines the relevance of AI in the current and future development of IoT and how security should be addressed, including data protection and privacy.

It covers the following objectives:

- Analyse the relation between AI and IoT and how AI is instrumental to unleash the full potential of IoT;
- Identify the current landscape and threats surrounding AI-enabled IoT. What are the main challenges and the most common attack vectors?;
- Provide use cases and good practices on securing IoT based applications, highlighting the ones that are AI enabled.
- Analyse how standardization processes can facilitate the development of securing IoT based applications and what are the standardization requirements to ensure that AI can support such deployment. That part will take into account the work of the ITU-T Study Group 20.
- Identify key security, privacy and trust challenges for IoT and AI as well as provide indications on what solutions (policy and/or technical) can be put in place to address those challenges.
- Analyse the role of national authorities on developing regulations that would promote safe use of IoT based applications, with specific focus on how data processes by AI (e.g., through machine learning) can be protected, and how privacy can be ensured.

2. Key technical trends for digital transformation in the future Introduction to AI

This section provides key technical trends for digital transformation from literature and highlight the important of AI and IoT technology. It also addresses potential risks and threats while the number of new technologies continues to grow.

Digital transformation is the change associated with the application of digital technology in all aspects of human society. The transformation stage means that digital usages inherently enable new types of innovation and creativity in a particular domain, rather than simply enhance and support traditional methods[1]. Key trends for digital transformation includes IoT, AI, 5G, Edge, Block-chain, and others.

Figure 1 shows data-driven IoT applications to leverage the massive amounts of data. Areas of IoT applications have been extended to various domains including consumers and industries. It generates a large volume of data continuously. Therefore, data-driven IoT applications is becoming significantly important in order to leverage the massive amounts of data from devices with emergence of big data and AI technology. Big data technology enables to capture, storage and analysis of data based on data collection through IoT. On top of Big data, AI can contribute to support intelligent applications without human intervention through data-based learning.

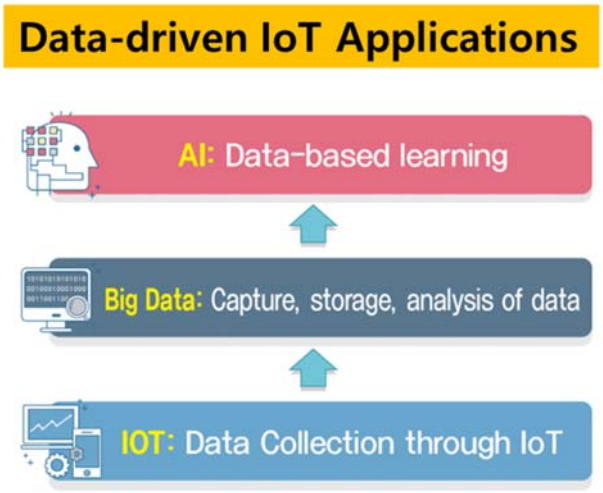


Figure 1. Data-driven IoT applications to leverage the massive amounts of data.

Big data can be thought of as an asset that is difficult to exploit. AI can be seen as a key to unlocking the value of big data; and machine learning is one of the technical mechanisms that underpins and facilitates AI. The combination of all three concepts can be called 'big data analytics'[2].

Recently IoT technology has been shifted to creating value through analytics and action from connection, sensing and communications for connecting devices. In this regard, data analytics and learning techniques are very essential to support IoT applications with optimization and autonomy from relatively simple sending and remote control (see Figure 2).

[1] http://www.emptrust.com/blog/impact-of-digital-in-new-hire-onboarding

[2] ICO, "Big data, artificial intelligence, machine learning and data protection," September 2017.

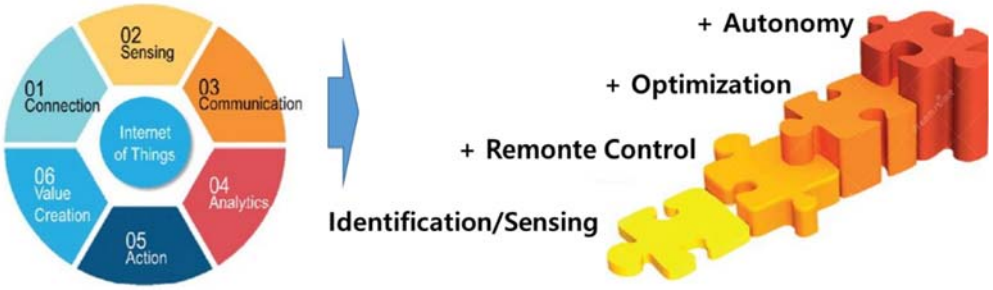


Figure 2. From connecting devices to creating value.

As one of key paradigm shifts, manufacturing processes are becoming increasingly digital, a new technical trend emerged recently; the digital twin as a sensor-enabled digital model of a physical object that simulates the object in a live setting. A digital twin can be defined, fundamentally, as an evolving digital profile of the historical and current behaviour of a physical object or process that helps optimize business performance. A digital twin is based on massive, cumulative, real-time, real-world data measurements across an array of dimensions[3]. These measurements can create an evolving profile of the object or process in the digital world that may provide important insights on system performance, leading to actions in the physical world such as changes in product design or manufacturing process.

The digital twin conceptual architecture in Figure 3 can rightly be thought of as an expansive or "under the hood" look at the enabling components that comprise the manufacturing process digital twin model, although the same basic principles may likely apply in any digital twin configuration. The conceptual architecture may be best understood as a sequence of six steps: create, communicate, aggregate, analyse, insight and act.

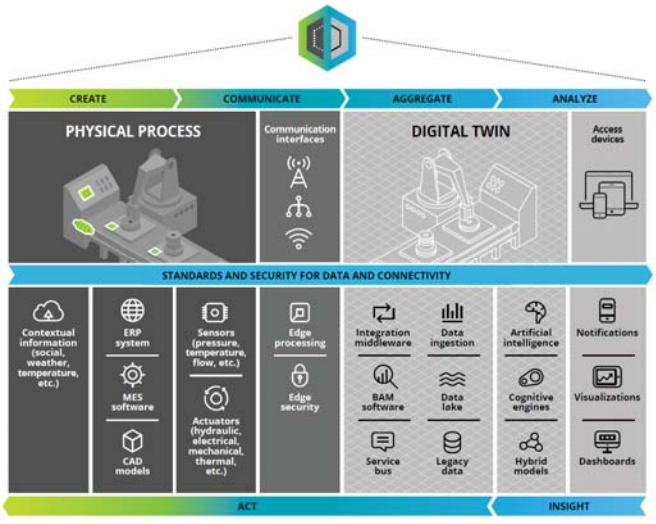


Figure 3. Digital twin conceptual architecture.

[3] "Industry 4.0 and the digital twin," Deloitte University Press, 2017.

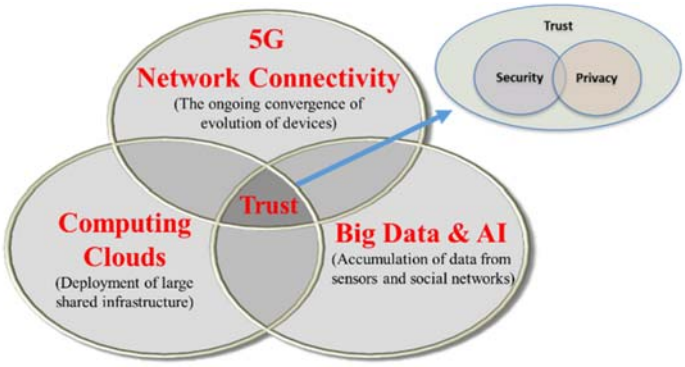


Figure 4. Security, privacy and trust issues in technology convergence.

For the future, there are many emerging technologies such as 5G, cloud computing, big data and AI. These technologies will be integrated to support more advanced features and provide significant benefits in terms of technological and societal viewpoints. In the technology convergence environment, security, privacy and trust will be common issues to be controlled and managed as shown in Figure 4.

Security concerns the confidentiality, availability and integrity of data or information. Security may also include authentication and non-repudiation. Systems need a variety of methods to prevent behaviours with malicious intents. Security mainly concerns technological aspects such as the confidentiality, availability and integrity. It also includes attack detection and recovery/resilience.

Privacy concerns the expression of or adherence to various legal and non-legal norms. In certain contexts this is often understood as compliance with data protection laws regarding the right to private life. Although it would be highly complex to map into personal data protection, the globally accepted privacy principles give a useful frame: consent, purpose restriction, legitimacy, transparency, data security and data subject participation[4]. Users need the protection of their personal information related to their behaviours and interactions with other people, services and devices. Privacy mainly concerns user aspects to support anonymity and restrictive handling of personal user data.

Trust is broader concept that can cover security and privacy. Generally, trust presents the confidence and the assurance that entities, users, systems, data and process behave as they are expected to. Therefore, trust can be considered as a way of achieving extra security and privacy objectives. Trust is an important feature in the decision-making process not only used by humans in daily life but also by applications and services in the ICT environment.

It's essential to prepare 4th industrial revolution with ICT. The combination of AI, data and networks is beginning to emulate human intelligence. As explained in the introduction, key technologies to support intelligence for IoT, big data and AI will be integrated with networks for exchanging information through connectivity between all actors. These solutions will be applied to various fields such as drones and robots, etc. Basically this concept has key characteristics such as real-time response, data in everything, supporting human decision-making and self-evolution. To achieve these

[4] Robinson, N., Valeri, L., Cave, J., Starkey, T., Graux, H., Creese, S. and Hopkins, P.P. (2010), *The Cloud: Understanding the security, privacy and trust challenges*, EU CORDIS, November.

goals, the following technology areas as challenges will be of significantly important as shown in Figure 5.

- Intelligence with hyper-connected distributed intelligence.
- Interoperability with open collaboration.
- Security and privacy with block-chain trust.

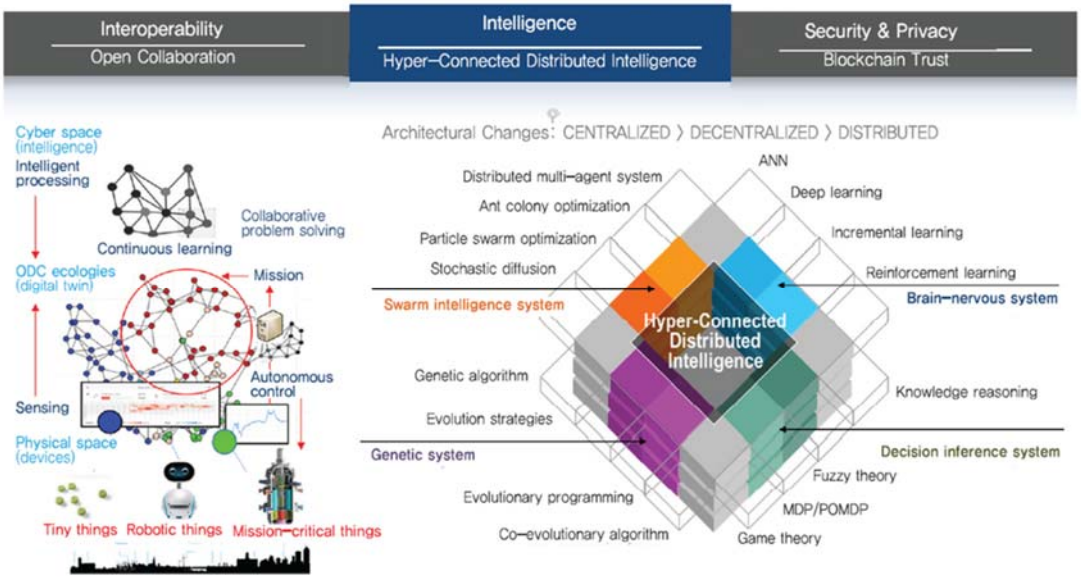


Figure 5. Challenges for 4th industrial revolution.

3. AI and Machine Learning

3.1 Introduction to AI and Machine Learning

In the summer of 1956, the dream of AI pioneers was to construct complex machines that possess the same characteristics of human intelligence. This is the concept that we think of as “General AI” - fabulous machines that have all the human senses and the ability to reason and think just like we do. General AI machines have remained in movies and science fiction novels. AI, as we currently know it, falls into the concept of “Narrow AI” which is defined as technologies that are able to perform specific tasks assigned by humans. Examples of narrow AI are: image classification, and face recognition on Facebook. These technologies exhibit some facets of human intelligence. Where does this intelligence come from? The answer is machine learning.

Machine learning is an approach to empower AI. At its most basic form, it is the practice of using algorithms to parse data, learn from it, and then make a decisions or predictions based on the collected data. It leverages algorithms to automatically model and find patterns in data, and these algorithms are heavily based on statistical and mathematical optimization. The optimization process often involves finding the smallest or largest value (minima or maxima) of a function. In a nutshell, machine learning is all about AI automatically learning a highly accurate predictive or classifier model, or finding unknown patterns in data, by leveraging optimization techniques. Machine learning algorithms are used primarily for the following types of output: Clustering (Unsupervised), Two-class and multi-class classification (Supervised), Regression: Univariate, Multivariate (Supervised), Anomaly detection (Unsupervised and Supervised), and Recommendation systems (aka recommendation engine).

Deep learning is a technique for implementing machine learning by using artificial neural networks. These artificial neural networks have discrete layers, connections, and directions of data propagation. Based on these general understanding of AI and machine learning, this report induces various leaning techniques.

Figure 6 illustrates basic concepts of AI and machine learning. Machine learning is the capability of a machine to learn without explicitly being programmed. On the other hand, AI is the capability of a machine to imitate intelligent human behaviour. For typical big data domain, it is concerned to ‘analyse’ on what happened and what will happen. With AI and networking, it is mainly concerning ‘act’ on which action should I take.

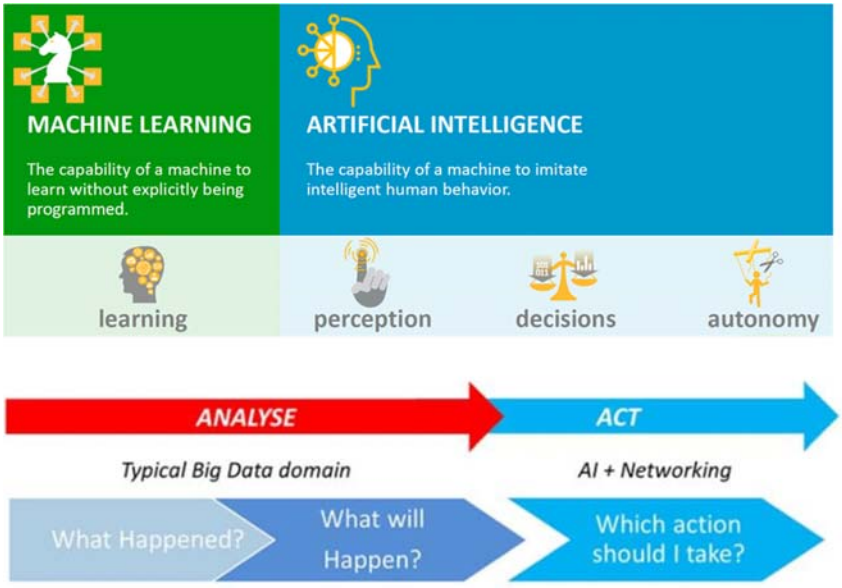


Figure 6. Artificial intelligence and machine learning.

3.2 Supervised learning

Supervised learning deals with datasets that are labelled[5]. That is, the dataset includes the features by which the event/object/thing is defined as well its desired output. In supervised learning, the data contains the response variable (label) being modelled, and with the goal being that you would like to predict the value or class of the unseen data.

Supervised learning is a machine learning technique for learning a function from training data. The output of the function can be a continuous value (called regression) or a class label of the input object (called classification). The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples (i.e., pairs of input and target output). The followings are the most common supervised algorithms: Decision Trees, Bagging, Boosting, Random Forest, K-NN, Linear Regression, Naive Bayes, Neural Networks, Logistic Regression, Relevance Vector Machine (RVM), and Support Vector Machine (SVM).

- Random Forest – An ensemble learning method for classification, regression which utilizes multitudes of decision trees and outputting of the mode of the classes (classification) and mean prediction (regression).
- Support Vector Machines – Another powerful set of Machine Learning algorithms used for classification of anomalies.
- K-nearest neighbour (K-NN) – Used heavily in pattern recognition and classification, they use the majority votes to align a data point to a specific class.

A good example of supervised leaning is classification for malicious email detection. The classification algorithm would be supplied a large dataset of emails as well as the labels ‘malicious’ or ‘safe’.

[5] Matt Lewis, “Rise of the machines: machine learning & its cyber security applications,” NCC Group Whitepaper, 2017.

3.3 Unsupervised learning

Unsupervised learning deals with unlabelled datasets. One of the most common application is to find groups/subsets within datasets, with applications being many and varied.

Unsupervised learning involves learning from data, but without the goal of prediction. This is because the data is not given with a target response variable (label), or someone chooses not to designate a response. The primary goal is to discover patterns, deep insights, understand variation, find unknown subgroups (amongst the variables or observations), from the data.

The two most commonly used techniques in unsupervised learning are principal component analysis (PCA) and clustering. PCA is the approach to learn what is called a latent variable model. Other notable latent variable modelling approaches include expectation-maximization algorithm (EM) and method of moments.

The followings are the most common unsupervised algorithms: BIRCH, Hierarchical, K-Means, DBSCAN, OPTICS, Mean-Shift, and Gaussian Mixed Model; they belong to Anomaly Detection Algorithms: Local outlier factor, and Isolation Forest; and they belong to Deep Learning Algorithms: Restricted Boltzmann machine, SOM, Autoencoder, and Generative Adversarial Networks.

3.4 Reinforcement Learning

Reinforcement learning is a reward-based learning system. Unlike most forms of machine learning, the learner of reinforcement learning is not told which actions to take. Instead, the learner must discover which actions yield the most reward by trying them. Actions may affect not only the immediate reward but also all subsequent rewards. Trial and error search, and delayed reward are the two most important distinguishing features of reinforcement learning. Reinforcement learning learns from immediate interaction with the environment, so it is different from supervised learning (learning from examples provided by a knowledgeable external supervisor).

3.5 Classification, clustering and regression

Machine learning problems can be categorised by their output or aim:

- Classification: inputs are mapped to user-specified outputs, such as emails to ‘malicious’ or ‘safe’.
- Clustering: inputs are grouped into clusters. The definitions of clusters are not known beforehand, unlike classification.
- Regression: a technique from the field of statistics used to estimate or predict outputs from a continuous – rather than discrete – set.
- Dimensionality reduction: the conversion of datasets with vast numbers of dimensions into datasets with fewer dimensions, resulting in more concisely-conveyed data for classification or regression tasks.

These are classification and regression algorithms: Decision Trees, Bagging, Boosting, Random Forest, K-NN, Linear Regression, Naive Bayes, Neural Networks, Logistic Regression, Relevance Vector Machine (RVM), and Support Vector Machine (SVM). These are cluster algorithms: BIRCH, Hierarchical, K-means, DBSCAN, OPTICS, Mean-shift, and Gaussian Mixed Model.

3.6 General application areas of AI

Classification learning systems are the area of the application which gives a new dimension for machine learning. AI systems have been widely applied in different domains: computer programming, game playing (AlphaGo), image recognition, speech recognition, medical diagnosis, agriculture, physics, email management, robotics, music, mathematics, natural language processing, etc.

The most recent AI trend is autonomous cars. An autonomous car (driverless car, self-driving car, robotic car) is a vehicle that is capable of sensing its environment and navigating without human input. Autonomous cars can detect surroundings using a variety of techniques such as radar, LIDAR, global positioning system (GPS), odometry, and computer vision. Google's self-driving car is an example of autonomous car project. For creating autonomous car, the system must be equipped with a strong AI[6].

IoT applications are generating data collected from various domains and industrial sectors. The data generated provides insights from the environments and applications that generated it. AI techniques provide the framework and tools to go beyond analytics of real time monitoring and automation use cases for IoT and move to IoT platforms that use concepts from AI and apply them to specific IoT use cases to provide smarter decision-making. AI-enabled IoT applications add a new layer of functionality and access, creating the next generation of smart homes/buildings, smart vehicles and smart manufacturing by providing intelligent automation, predictive analytics and proactive intervention[7].

In the IoT context, AI will support companies in finding the smart data and analyse the trends and patterns for better decision-making based on defined set of rules.

The AI techniques will enable cognitive systems to be integrated with IoT applications creating optimized solutions for each individual application. Cognitive IoT technologies will allow embedding intelligence into systems and processes, allowing businesses to increase efficiency, find new business opportunities, and to anticipate risks and threats so they can better address them. The IoT applications will gather and integrate data from many types of sensors and other sources, reason over data, and learn from the interactions, while creating communities of devices that share information. The information collected can be interpreted and managed by people, IoT applications or IoT platforms using cognitive systems in order to generate new and better services and use cases.

The data generated by edge devices combined with the unstructured data available from sources ranging from news Web sites and social networks can be combined using cognitive IoT capabilities at the edge at the cloud level.

The use on AI, swarm intelligence and cognitive technologies together with deep learning techniques for optimising the IoT services provides by IoT applications in smart environments and collaboration spaces will create solutions capable of transforming industries and professions.

[6] https://www.normshield.com/machine-learning-in-cyber-security-domain-1-fundamentals/

[7] Ovidiu Vermesan, et al., "IoT digital value chain connecting research, innovation and deployment," 2016.

4. IoT and Security – Security framework in Cyber-Physical Systems

This section discusses IoT and related security issues. For this, it presents a security framework in cyber-physical system (CPS) in order to understand the overall features and technical issues for securing IoT environment.

4.1 Cyber-Physical Systems for the IoT

The IoT can be characterised as a cyber physical system (CPS). A CPS conceptual model[8] is shown in Figure 7. This figure is presented here to highlight the potential interactions of devices and systems in a system of systems (SoS) (e.g., a CPS infrastructure). A CPS may be as simple as an individual device, or a CPS can consist of one or more cyber-physical devices that form a system or can be a SoS, consisting of multiple systems that consist of multiple devices. This pattern is recursive and depends on one's perspective (i.e., a device from one perspective may be a system from another perspective). Ultimately, a CPS must contain the decision flow together with at least one of the flows for information or action. The information flow represents digitally the measurement of the physical state of the physical world, while the action flow impacts the physical state of the physical world. This allows for collaborations from small and medium scale up to city/nation/world scale.

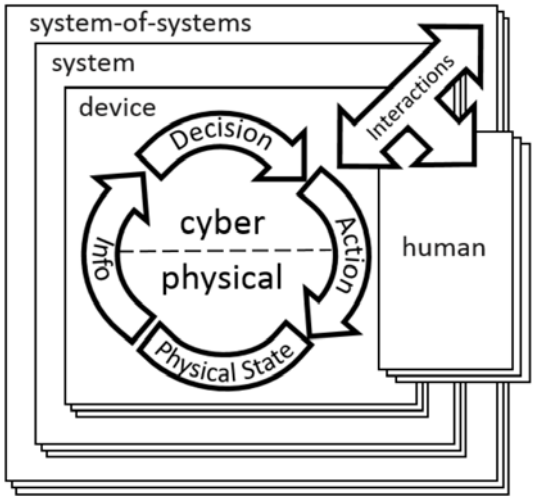


Figure 7. CPS conceptual model.

CPSs enable the physical world to merge with the virtual world by integrating computation and physical processes. A CPS facilitates tight integration between computation, communication, and control in its operation and interactions with the environment in which it's deployed. Now, in addition to embracing cyber and physical features, interest is growing in harnessing human and social factors in CPSs. For instance, one recently proposed cyber-physical-social system (CPSS)[9] embraces the cyberspace-enabled parallelism: a real system and its artificial counterparts run in parallel and interactively through cyberspace. Another proposal is a human-in the-loop CPS, which infers users' intent by measuring human cognitive activity through body and brain sensors. These studies take human and social features as important elements in CPSs mainly from the system automation and

[8] NIST, "Framework for cyber-physical systems" Release 1.0, May 2016.

[9] "A Data-Centric Framework for Cyber-Physical-Social Systems", IEEE IT Professional, Nov.-Dec. 2015.

control perspective. However, to grasp the fuller potential of a CPSS, data-centric realization is necessary.

Data-driven development will likely be a promising software paradigm in the coming decades. This will also lead to a revolution in the design and development of cyber-physical-social applications and services. With data-driven CPSS (D-CPSS), we can leverage the cross-space, multimodal data from heterogeneous data sources to better characterize the target (for instance, an event or object). The combined effects of tri-space data will also nurture numerous novel applications or services in urban environments. From a data-centric viewpoint, each CPSS follows a generic life cycle, consisting of data collection, processing, and usage. A four-layered architecture of a D-CPSS that follows this life cycle is shown in Figure 8. The resource management and cooperative sensing layers deal with data collection, whereas the data pre-processing and data analysis layers deal with data processing.

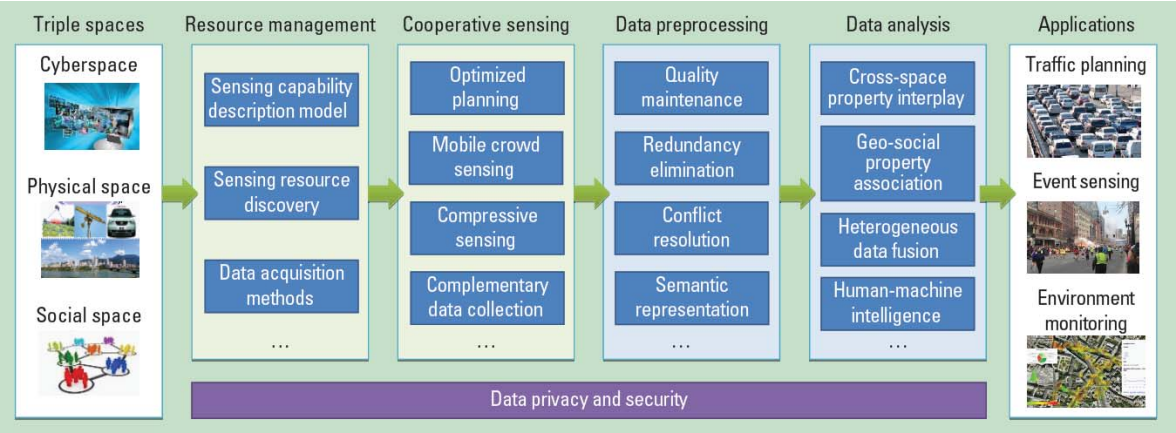


Figure 8. A layered architecture for data-driven cyber-physical-social systems (D-CPSS).

4.2 Potential risks in Cyber-Physical-Social Systems

There have been a larger number of emerging technologies been integrated in ICT infrastructures. These ICT infrastructure remain lacking in terms of having a proper platform and frameworks to ensure security and privacy, especially without the support of international standards. ICT systems, applications and services are significantly getting broader in scope, which require collaborations among various kinds of technologies and types of entities. ICT infrastructure including systems, applications and services is basically considered in CPS architecture in which the physical domain mainly consists of physical devices which interwork with each other through information and communication networks. The cyber domain is responsible for the delivery, storage and processing of data and information. While the social domain has become popular among people for sharing and showing their knowledge and become a new medium for connecting people in cyberspace.

While ICT infrastructure has grown in size and complexity, mechanisms to ensure security and privacy have not been able to keep in pace. Consequently, ICT infrastructures are at risk and vulnerabilities to a wide–range of threats at all component, device and system levels. Attackers will seek to exploit the device's physical, cyber, and social vulnerabilities by conducting various types of attacks targeted at the physical, cyber, social domains of CPS. There are many potential risks in ICT infrastructures as follows.

Risk at Physical Domain:

Smart devices and sensors have been more and more integrated into ICT infrastructures which are usually unattended by humans communicate with others via different media such as wired, wireless,

broadband. However, these physical components are usually resource constrained with limited computational capability and security mechanisms. Thus, they are extremely vulnerable to both external and internal attacks.

- Natural threats[10]: Earthquakes, hurricanes, floods, and fire could cause severe damage to the physical components and computer systems of ICT infrastructures. While few safeguard measures can be implemented against natural disasters, disaster recovery plans like backup and contingency plans are the best approaches to secure systems against natural threats.
- Physical Attacks: this kind of attack tampers with hardware components and device protocols such as insertion of valid authentication tokens into a manipulated device, inserting and booting with fraudulent or modified software, and environmental/side-channel attacks, both before and after of the device's deployment. The risk is also from access attacks that unauthorized persons gain access to networks or devices to which they have no right to access by means of eavesdropping[11], spoofing, packet sniffer and network ports scanning (reconnaissance attacks)[12].

Risk at Cyber Domain:

The cyber risks cover a large number of areas including cyber security, information security, data provenance, and privacy in which vulnerabilities, threats and cyber-attacks are analysed and managed. Cyber security and privacy mechanisms should protect the services, hardware resources, information and data, both in transition and storage, ensure both cyber networks and services are protected against unauthorized access from within the devices and externally.

- Cyber/Information Security Attacks[13]:
    - Attacks on the Core Network: Threats at mobile network operators such as impersonation of devices, traffic tunneling between impersonated devices, mis-configuration of the firewall in the modem, router, and gateways could be the target of several kinds of attacks such as denial of service (DoS). They may also include changing the device's authorized physical location in an unauthorized fashion or attacks on the radio access network, using a rogue device.
    - Configuration Attacks include fraudulent software update/configuration changes, mis-configuration by the owner, subscribers or users, mis-configuration or compromise of the access control lists.
    - Compromise of Credentials comprise of brute force attacks on tokens and (weak) authentication algorithms, physical intrusion, or side-channel attacks, and malicious cloning of authentication tokens.
    - User Data and Identity Privacy Attacks include eavesdropping for other users or devices data sent over the systems; masquerading as other user/subscribers device; users network ID or other confidential data revealed to unauthorized third parties.
    - Access Attacks: unauthorized persons gain access to networks or devices to which they have no right to access. There are two different types of access attack: the first is physical access, whereby the intruder can gain access to a physical device. The second is remote access, whereby the intruder gain access via IP-connected devices.

[10] H. G. Brauch, "Concepts of Security Threats, Challenges, Vulnerabilities and Risks," *Copying with Global Environmental Change, Disasters and Security*, vol. 5, pp. 61-106, 2011.

[11] G. H. I. Naumann, "Privacy features of european eid card specifications," *Network Security*, vol. 4, pp. 9-13, 2008.

[12] S. R. H. C. S. Ansari, "Packet sniffing: a brief introduction," IEEE Potentials, vol. 21, pp. 17-19, 2002.

[13] C.Wilson, "Botnets, cybercrime, and cyberterrorism:Vulnerabilities and policy issues for congress," CRS Report for Congress DTIC Document, Washington DC, 2008.

- Privacy Attacks[14]: Privacy protection in IoT has become increasingly challenging due to large volumes of information easily available through remote access mechanisms.
    - Data mining: enables attackers to discover information that is not anticipated in certain databases.
    - Cyber espionage: using cracking techniques and malicious software to spy or obtain secret information of individuals, organizations or the government.
    - Eavesdropping: listening to a conversation between two parties
    - Tracking: User's movements can be tracked by the devices unique identification number (UID). Tracking user's location would allow for attackers to pinpoint the user's location in situations in which they wish to remain anonymous.
    - Password-based attacks: attempts are made by intruders to duplicate a valid user password.
- Cyber-crimes: The Internet and smart objects are used to exploit users and data for materialistic gain, such as intellectual property theft, identity theft, brand theft, and fraud

Risk at Social Domain:

In the context of social domain, the entity incudes both human (users) and machine, and the social risks are from both human-to-human and human-to-machine interactions.

- Risk of lacking trust in interactions
    - Human-human interactions: If there is no trust among peoples, their interactions (e.g., exchanging data and information) would have no meaning due to lack of confidence with each other. If the people are not trustworthy, personal interactions do not invoke any response. The unclear decision making or unrealistic situation may be happening from low or broken trust in human relationships.
    - Human-machine interactions: When a human cannot trust a machine (e.g., delivering imprecise data from a machine to a human), meaningful human-machine interactions cannot be established and the potential benefits on system performance will be lost. The human-machine systems have always proved to be unpredictable and fallible, whereas the nature of the system is to function normally. It relies on technological dependency which accentuates risks.
- Risk of attacks in social world[15]: A malicious entity is dishonest and socially uncooperative in nature and can break the basic functionality of a ICT system. This malicious entity can perform the following attacks:
    - Self-promoting attacks: a malicious user can intentionally promote its importance (by providing good recommendations for itself) in order to be selected as the service provider.
    - Whitewashing attacks: a malicious entity can disappear and rejoin the application to wash away its bad reputation.
    - Discriminatory attacks: a malicious entity can discriminatively attack non-friends or nodes without strong social ties (without many common friends) because of human nature or propensity towards friends in social networks.

[14] R. H. Weber, "Internet of Things – New security and privacy challenges," *Computer Law & Security Review*, vol. 26, no. 1, pp. 23-30, 2010.

[15] F. B. J. G. Ing-Ray Chen, "Trust-based Service Management for Social Internet of Things Systems," *IEEE Transactions on Dependable and Secure Computing*, 2015.

- Bad-mouthing attacks: a malicious entity can ruin the reputation of another well-behaved entity by providing bad recommendations so as to decrease the chance of this good node being selected as a service provider. This is a form of collusion attacks in which collaboration among bad nodes is possible.
- Ballot-stuffing attacks: a malicious entity can boost the reputation of another bad node by providing good recommendations for it so as to increase the chance of this bad node being selected as a service provider. This is also a form of collusion attacks, i.e., it can collaborate with other bad entities to boost the reputation of each other.

Risks due to the integrity of Physical-Cyber-Social domains

- Mismatch in CPS environments: The CPS cannot be fully operable if there is a mismatch between the physical and cyber world. If the malfunction of a physical system does not notify at the responsible entities in the cyber world, there are some risks to prevent safety in a physical world. An intelligent human in a cyber world can avoid or reduce the risk of failures and minimize the unacceptable situation in a physical world. Time critical convergence applications such as smart grid and intelligent transportation systems require high trust between the cyber world and the physical world. Greater openness, in combination with hiding one's real identity in a physical world and making a false object in a cyber world, increases the risks that people are becoming victims of deception. They also include identity theft and exposure to inappropriate actions.
- Human errors: Without recognizing a set of rules and external conditions of a physical system, human actions may result in risks or failures. Human errors may be a primary cause or a contributing factor in risks and accidents. Intentional or unintentional human errors may cause serious problems in ICT infrastructures.
- Risk due to the complexity of ICT infrastructures:
  - A numerous number of ICT resources: Risks threaten us to cope with complexity of interactions and mechanisms of ICT infrastructures. The anonymous and/or malicious access of a large number of ICT resources causes irreparable damages and creates unpredictable dangers. It is essential to make ICT resources accessible to all the people with promises but with unknown dangers.
  - Complexity of network operation: There are a lot of algorithms for network resource optimization including efficient routing, congestion avoidance, and guaranteeing Quality of Service (QoS)/Quality of Experience (QoE). When unpredictable situations happen in a network, the out-of-service possibility increases. Natural disaster and distributed denial-of-service (DDoS) attacks are also part of the risks. While network control functions can arrange a by-pass or de-tour route to cope with overflowed traffic, the unexpected side effects like traffic fluctuation and domino effect may bring additional risks. To increase network survivability during network operation, networking protocols and Operations, Administrations, Maintenance, and Provisioning (OAM&P) functions should be re-designed to be trustworthy. Moreover, when a network infrastructure includes a cloud platform with a large volume of storage and processing capabilities, network instability is not only coming from traffic congestion. The operation of the cloud platform and high-level applications are additional harmful sources to increase network risks. The existing security functions including firewall and Deep Packet Inspection (DPI) may be replaced to provide a certain level of trust, through the implementation by a trust gateway system and trust-guaranteed network OAM functions.
  - Complexity of convergence services and applications: ICT based services and applications will continue to be heterogeneous, and this may lead to an increase in the number of convergence services that cover multiple service domains. Within the IoT and CPS environments, people, platforms and devices will be highly inter-connected by a dynamic

network while operating in heterogeneous environments. These kinds of highly connected environments increase the complexity of services and applications (which consume data and information from connected sensors, devices, etc.), and unknown potential risks may be incurred due to complex interactions. As ICT based applications and services will scale over multiple domains and involves multiple stakeholders, methods for assessing trust are needed to enable the users to have confidence to these services and applications.

- Risk in Data, information and knowledge process: Since future ICT infrastructures should provide data, information and knowledge process, the trust provisioning is essential. Data integrity refers to the maintenance and assurance accuracy and consistency in data. The failure of data aggregation is coming from any unintended changes to data as the results of storage, retrieval and processing operation for further information and knowledge. For example, if data stored in a cloud platform are shared by anonymous users, there may be a possibility for undesirable situations to happen. With a certain level of trust, data delivery and cognitive data, information, knowledge and wisdom (DIKW)[16] process may be effective and meaningful.

4.3 Security, privacy and trust in CPS

Traditional information technology (IT) cybersecurity provides information protection (integrity, confidentiality) and readiness for correct services (availability). CPS cybersecurity has the same goals as traditional IT cybersecurity though perhaps with different priorities but should also be focused on how to protect physical components from the results of cyber-attacks.

There are 10 steps to cybersecurity as follows[17].

- Network security: Protect your networks from attack. Defend the network perimeter, filter out unauthorised access and malicious content. Monitor and test security controls.
- User education and awareness: Produce user security policies covering acceptable and secure use of your systems. Include in staff training. Maintain awareness of cyber risks
- Malware prevention: Produce relevant policies and establish anti-malware defences across your organisation
- Removable media controls: Produce a policy to control all access to removable media. Limit media types and use. Scan all media for malware before importing onto the corporate system.
- Secure configuration: Apply security patches and ensure the secure configuration of all systems is maintained. Create a system inventory and define a baseline build for all devices.
- Managing user privileges: Establish effective management processes and limit the number of privileged accounts. Limit user privileges and monitor user activity. Control access to activity and audit logs.
- Incident management: Establish an incident response and disaster recovery capability. Test your incident management plans. Provide specialist training. Report criminal incidents to law enforcement.

[16] DIKW (Data, Information, Knowledge and Wisdom): This refers loosely to a class of models for representing purported structural and/or functional relationships between data, information, knowledge, and wisdom. "Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge". (Source: https://en.wikipedia.org/wiki/DIKW_Pyramid)

[17] National Cyber Security Centre, www.ncsc.gov.uk.

- Monitoring: Establish a monitoring strategy and produce supporting policies. Continuously monitor all systems and networks. Analyse logs for unusual activity that could indicate an attack.
- Home and mobile working: Develop a mobile working policy and train staff to adhere to it. Apply the secure baseline and build to all devices. Protect data both in transit and at rest.

Two challenges are typical for CPS cybersecurity[18]:

- Detection and prevention of deception attacks (e.g., attacks on sensors that can lead them to input malicious data to the cyber component and, as a result, to provide wrong, or even dangerous, output from the cyber component)
- Detection of compromised cyber components and prevention of incorrect cyber functioning (or failure to function)

These challenges are not unique to CPS; rather, their consequences are potentially more severe because they impact the physical and cyber world. More importantly, the means to prevent these problems include not only cybersecurity controls, but also safety and reliability controls that are not applicable to traditional IT systems. Thus, CPS cybersecurity requirements should be determined in conjunction with safety, reliability, and privacy requirements. In case of system failures or cyberattacks, CPS resilience should provide ways and means to continue not just IT services but also able to provide full CPS recovery. This can be done only through co-design of CPS cybersecurity, including privacy, with safety, reliability, and resilience. As a result, considerations of the traditional tenets of confidentiality, integrity, and availability are no longer the sole focus of cybersecurity for CPS. Neither is providing CPS cybersecurity simply a matter of prioritization and application of existing controls. Rather, it involves the tradeoff of risks. This process of risk management becomes even more critical when considering the potential impact of cybersecurity failures on the ability to deliver capability across the disciplines. In addition, to develop effective CPS cyber protection and mitigation actions, the nature, functions, and interactions of all three types of components of CPS – cyber, analogue, and physical – must be understood. CPS designers and integrators should take into consideration both the intended and unintended effects resulting from the combination of properties where the goals of each property may either contradict or compliment to their counterparts. Trade-off decisions should be considered in light of the system-of-systems objective, if known. This is much more challenging than it sounds.

Trustworthy systems

A main stakeholder goal for a system is that it be trustworthy in respect to the key system characteristics. The importance of each key system characteristic to a given deployment is unique to each system and achieving one can conflict with achieving another. Interactions between the key system characteristics must be understood based on drivers such as regulatory compliance, business process and industry norms, not in isolation[19].

[18] NIST, "Framework for cyber-physical systems" Release 1.0, May 2016.

[19] Industrial Internet consortium, "Industrial Internet of Things Volume G4: Security Framework," 2016

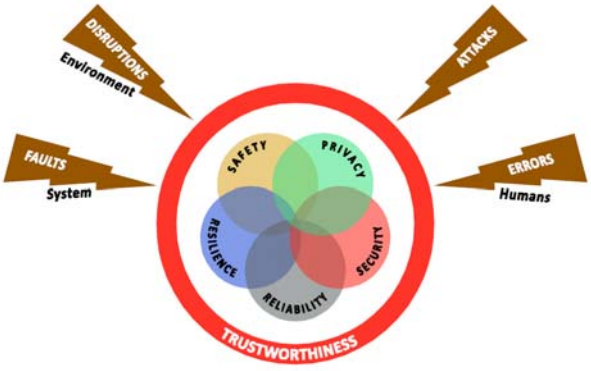


Figure 9. Trustworthiness of an Industrial IoT System.

Trustworthiness is the degree of confidence one has that the system performs as expected in respect to all the key system characteristics in the face of environmental disruptions, human errors, system faults and attacks.

ITU-T developed the first Recommendation (Y.3052) on trust in ICT. Recommendation ITU-T Y.3052[20] provides an overview of trust provisioning in information and communication technology (ICT) infrastructures and services. Recommendation ITU-T Y.3052 introduces necessity of trust to cope with potential risks due to lack of trust. The concept of trust provisioning is explained in the context of trusted ICT infrastructures and services. From the general concept of trust, the key characteristics of trust are described. In addition, a trust relationship model and trust evaluation based on the conceptual model of trust provisioning are introduced. Recommendation ITU-T Y.3052 then describes trust-provisioning processes in ICT infrastructures and services.

Trust provisioning is an integral function of physical, cyber and social trust that provides a valuable method of minimizing risks through identifying the trust characteristics of entities. Using trust provisioning, it is possible to develop trusted ICT infrastructures and services that cooperate with ICT applications in order to support these applications and services for better quality of services and experience by mitigating inherent and extraneous risks.

Figure 10 shows the concept of trusted ICT infrastructures and services. Three types of trust provisioning are classified into: physical trust for physical things (including sensors, actuators and devices); cyber trust for communication, computing and control; and social trust for stakeholders, which are mapped with trust in the physical, cyber and social worlds, respectively. In the trusted ICT world, trust entities may assume DIKW processes to minimize potential risks and to maximize the value of assets.

[20] ITU-T Y.3052, "Overview of trust provisioning in information and communication technology infrastructures and services," March 2017.

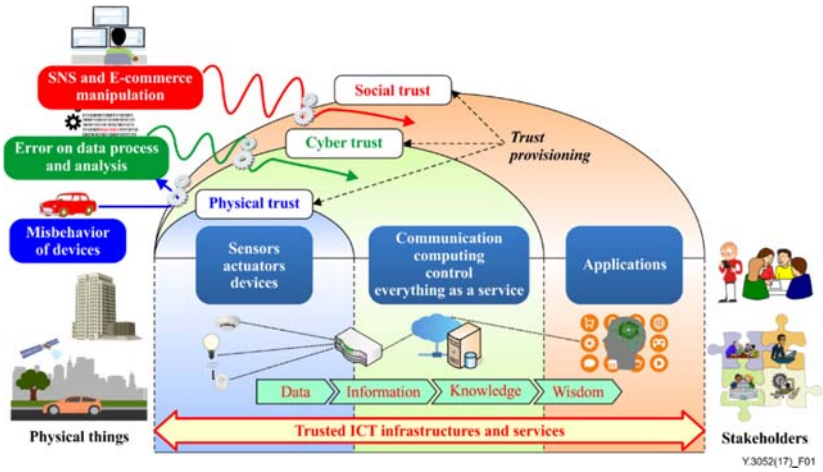


Figure 10. The concept of trusted ICT infrastructures and services.

From the perspective of trust provisioning, there are physical, cyber and social worlds. To build an ICT ecosystem, raw data from physical things in the physical world are produced by physical interfaces like sensors and actuators. In the cyber world, there are physical objects and logical objects. Physical objects are those mapping to hardware devices and equipment that have capabilities of data processing, data storage and communication, etc. Logical objects are algorithms, functions and software that work on computing, storage and networking components. In the social world, entities like humans, stakeholders and software agents, which are computer programs that act for a user, produce and consume various data and applications through user interfaces. Physical things, cyber objects and social entities interact to perform trusted ICT applications taking into consideration physical, cyber and social trust, respectively. Figure 11 shows the role of trust provisioning in the ICT world in realizing various trusted ICT applications.

Physical trust

Physical trust reflects various trust aspects of physical things, which can be measured by counting on theirs trustworthiness in terms of capability, integrity and cooperation. Its capability means the ability of the physical thing to perform its task with correct functionality. Its integrity means the state of the physical thing, being stable without trouble or breakdown. Its cooperation means that the physical thing works together with other physical things for their common purposes. Physical trust reflects trust propensity that is affected by risks related to the physical world.

Cyber trust

Cyber trust reflects various trust aspects of cyber objects, which can be measured by counting on their trustworthiness in terms of capability, integrity and cooperation. Its capability means that the ability of a cyber object is correct and certain to execute control, computing and communication. Its integrity means that data handled or provided by cyber objects are not accidentally or maliciously altered or destroyed during control, computing and communication. Its cooperation means how well the cyber object works together with other objects. Cyber trust reflects trust propensity that is affected by risks related to the cyber world.

Social trust

Social trust reflects various trust aspects of social entities. Social trust can be measured by considering its trustworthiness in terms of ability, honesty and benevolence. Its ability means human competence in the individual's activity. Its honesty implies that the social entity treats others honestly. Its benevolence means how nicely the social entity behaves to other social entities or how much the

social entity interacts with other entities for their kindness. Social trust reflects trust propensity that is affected by risks in the social world.

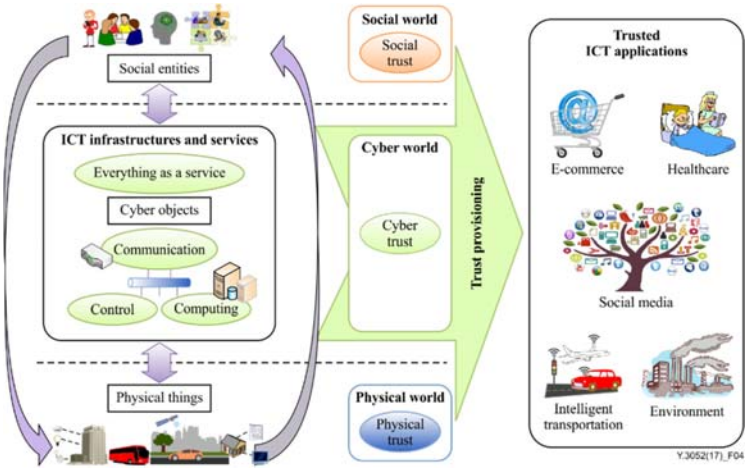


Figure 11. Trust provisioning in the ICT world for trusted ICT applications.

5. Security in IoT and AI

The IoT opens up devices for hostile parties ready to exploit every conceivable vulnerability. Thus, IoT is facing some of the greatest obstacles to its wide-spread adoption and deployment. For its global adoption, the IoT devices and networks must be safeguarded. This report lays out a framework for the adoption of AI and how it can be used to enforce the security of IoT devices and networks. One of the significant challenges of AI based security has been the inadequate resources for implementing AI techniques on resource starved IoT devices since the current approaches rely on the compute power of the Cloud to deploy AI algorithms. This approach is not feasible in practice since the target of attackers are devices and thus, securing these devices using AI techniques requires a completely different approach. This section presents some related work, discusses challenges and proposed areas of further research.

5.1 Introduction

IoT promises to integrate and connect every day object such as sensors, actuators and other physical objects to the Internet providing state-of-the-art intelligent services. The growth of IoT in the last few years has been very rapid with an estimated 50 billion devices to be connected by 2020[21][22]. IoT devices include for instance, internet connected cameras, smartphones, smartwatches and even bracelets that can share our physical activities with your friends and families. Along with the numerous benefits of utilizing IoT, it also comes with risks and security related concerns and issues. In this regards, security in various forms of attacks has been identified as one of the biggest weakness of IoT based platforms. This is due to the heterogeneity nature of these devices, communication protocols, data, as well as the humongous number of devices involved. Security issues such as jamming, spoofing, denial of services, eavesdropping, malwares in the form of viruses, Trojans, worms etc. are a great source of concern when it comes to designing and developing secured IoT systems. They present a variety of potential risks that they could be exploited to harm users or to even bring down an entire system via: (1) unauthorized access and misuse of personal information; (2) attacks facilitation on other systems; (3) risks to personal safety.

One of the major impacts of security issue in the IoT system is that it could undermine consumer confidence. Consumer confidence is necessary for the IoT technology to meet its full potential. The lack of trust and perceived risks of IoT devices may dissuade consumers to fully embrace IoT technology, creating a potential barrier in leveraging the benefits of IoT platforms in city development. These security issues are not new. They have been the major focus of infosec experts in the traditional computers and computer networks for decades[23]. However, in IoT they assume different dimensions since the conventional security mechanisms based on authentication, confidentiality, malware prevention, etc. cannot be directly deployed on IoT devices because of resource scarcity. IoT devices have prohibitively limited resources, battery lifetime, and even network bandwidth to run the traditional compute intensive security mitigation mechanisms. Thus, the lack of effective security measures enables malicious parties to access and misuse personal information, collected and transmitted through the IoT devices and network which is a challenge that needs to be urgently tackled.

A good example can be found within the smart home environment where IoT devices such as smart TV and smart phone are common. They enable consumers to browse the Internet, make purchase such as movies while disclosing sensitive information such as credit card details. This information could be stolen by identity thieves, who may effortlessly exploit the security weakness of IoT devices, in

[21] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT Security Techniques Based on Machine Learning," pp. 1–20, 2018.

[22] C. Tankard, "The security issues of the Internet of Things," *Comput. Fraud Secur.*, vol. 2015, no. 9, pp. 11–14, 2015.

[23] I. Andrea, C. Chrysostomou, and G. Hadjichristofi, "Internet of Things: Security vulnerabilities and challenges," in *Proceedings - IEEE Symposium on Computers and Communications*, 2016.

order to perpetrate fraud. Thus, in such smart home environment, the more the number of devices connected to the network, the more the vulnerabilities a malicious person could exploit to compromise personal information.

Another potential target is the network. Attack on any IoT device can facilitate attacks on the network to which it is connected and with potential to cause attack on several other connected devices. An attacked device can be used to launch denial of service attacks. In addition, considering the large number of IoT devices, the more devices the attackers can access, the more devastating the denial of service attack. Affected devices can also be used to send malicious messages via emails. There have been reports of people devices being hacked and their social media profiles are being accessed to post sensitive information or sometimes to defraud the user's social media friends.

Another important risk is that unauthorised access might be used to exploit the security vulnerability to create risks to physical safety. For example, smart cars can be hacked remotely, tampering its braking system causing physical road hazard.

Since the IoT integrates physical world, communication networks and applications, techniques for addressing security in this new computing infrastructure becomes a critical issue[24].

The goal of AI as a science is to make machines do things that would require intelligence if done by humans. The AI techniques based on machine learning for example, can recognize trends from past experiences and then able to make predictions. Therefore, security solutions based on AI techniques are expected to react more effectively to new threats than the traditional security approaches. Recently, AI based techniques such as machine learning and reinforcement learning, etc. are being suggested as effective solution to address security related issues in IoT platforms. AI based techniques have been explored broadly across the industry and applications as compute power and data availability increase.

In security, available of big data means AI techniques can be exploited to analyse and recognize patterns of security vulnerabilities to prevent such attacks. Thus, the ability of IoT based platform to learn from data to analyse, identify and mitigate security threats is an important feature that every IoT system should incorporate. These techniques are also more accurate in terms of assessment of potential malware threats from large quantity of data. In addition, AI is very suitable to detect and mitigate sophisticated attackers such as advanced persistent threats in which attackers can remain undetected for indefinite period. The rapid development in IoT and the so-called smart attacks have made it imperative to define IoT defence policy and determine various parameters in the security protocols for possible trade-off in the heterogeneous and dynamic networks.

However, one of the greatest challenges of IoT security is the question of how to classify the different types of data for detecting and scanning the IoT traffic that runs various protocols in order to identify patterns that represent security threats and then mitigate such cyber threats.

AI techniques that have been applied to address network security related issue can be broadly classified into three. Machine learning, deep learning and reinforcement learning.

5.2 IoT Security

There is no doubt that for us to push IoT for global penetration and adoption, a key challenge that must be tackled headlong is the issue of security. IoT platforms are expected to connect billions of devices, sensors, actuators and objects through the Internet allowing interactions between these objects, other entities and even humans. To allow this kind of interactions to be meaningful, IoT platforms must be provided with security guarantee to protect individual objects, information, data,

[24] M. Negnevitsky, "Artificial intelligence: a guide to intelligent systems. Pearson Education.," in *Artificial intelligence: a guide to intelligent systems*, 2005.

and services from security threats. Considering the ubiquitous nature of the IoT systems, protecting these systems against attacks is a complex process. This is because anyone can access these systems, anywhere and anytime. Moreover, an access to a single device by a malicious agent is enough to bring down the entire network of IoT systems. Furthermore, the heterogeneous nature of the billions of IoT devices exchanging data and information makes security issue a more difficult problem to address. The smart objects in IoT are connected to the global Internet with capability to communicate with several other objects with high probability of serious security breaches such as authenticity, and integrity problems[25].

With the wide range of facets that IoT system will impact, security issues if not appropriately tackled would have serious consequences, causing dame, disruption of operations and potential to cause loss of lives is quite high. Tankard[26] gave scenarios in which IoT security could come with serious consequences. In autonomous or even smart cars, these cars can be remotely hacked and this could cause massive loss of lives if security issues are not properly addressed.

In smart building, with smart heating, ventilation and air conditioning, lighting, door access control or even video surveillance, smart elevator, etc. all of which are interconnected with each other, any attack by malicious party could lead to loss of lives.

The security challenges in IoT could range from insufficient authentication, authorisation, insecure network services, lack of transport encryption, insecure cloud and edge interfaces, insecure mobile interface, poor security configurability problems, insecure software or firmware and even poor physical security. We should also note that most IoT devices have been developed without taking security into consideration partly because these devices have limited computational resources to execute security mechanisms. One of the key solution is to develop security solutions to the IoT by design. This allows security measure to be built into the IoT devices right from the start.

5.3 IoT Security Attacks

IoT systems including objects or things, networks, services and data are vulnerable to all kinds of attacks. IoT security can be defined as a set of technologies and processes designed to protect IoT devices, IoT networks, data and services from attacks, unauthorised access, change or destruction. In the traditional computing platform, cyber security minimally consists of anti-virus software, firewall and intrusion detection systems[27]. Considering security research in the academic, work on IoT security is still not well established as it is in the traditional computing environment. Most of the body of research consider the adoption of the traditional approaches to addressing the IoT security challenges. However, as said earlier, these approaches cannot be directly deployed on the IoT systems. First, in order to understand specific security issues in IoT, we analyse the peculiar features of this revolutionary computing paradigm that makes it unique and different from the existing computing platforms.

5.4 AI Techniques for IoT Security

The ability to automate the monitoring, management, and control of IoT security will be the driver for securing IoT devices and networks. AI techniques such as supervised and unsupervised machine

[25] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I.S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer, P. Doody, Internet of Things Strategic Research Roadmap, Cluster of European Research Projects on the Internet of Things, CERP-IoT, 2011.

[26] C. Tankard, "The security issues of the Internet of Things," *Comput. Fraud Secur.*, vol. 2015, no. 9, pp. 11–14, 2015.

[27] S. Sicari, A. Rizzardi, L. A. L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Networks*, 2015.

learning, reinforcement learning and even deep learning are being proposed to address security threats in IoT.

The following shows AI and machine learning as cyber tools.

- Behavioural Modelling: AI can be applied to real-time modelling of ALL Network Traffic, Log & Audit Files, Net Nodes, Servers and all “Smart IoT” Devices
- Zero-Day Attacks: AI modelling can mitigate risks of new malware that can no defined “signature”.
- Advanced Persistent Threats (APTs): Adaptive Learning Algorithms can detect the step-by-step penetration of APT malware (Phishing, Trojans, Adware, Botnets, etc.)

6. AI-based Privacy mechanism for Personal Data in the Internet of Things

6.1 Introduction

Privacy is a very broad and diverse notion for which literature offers many definitions and perspectives. With the increasing use and efficiency of electronic data processing, data privacy has become the predominant issue today, especially for the IoT. Data privacy is suitably defined as the appropriate use of data. When companies and merchants use data or information that is provided or entrusted to them, the data should be used according to the agreed purposes. The differences and relations between security and privacy is that security provides protection for all types of information, in any form, so that the information's confidentiality, integrity, and availability are maintained whereas privacy assures that personal information (and sometimes corporate confidential information as well) are collected, processed (used), protected and destroyed legally and fairly.

The evolving nature of the IoT regarding technologies and features and the emerging new ways of interaction with the IoT lead to specific privacy threats and challenges. Generally, data privacy in the IoT is the threefold guarantee to the subject for:

- Awareness of privacy risks imposed by smart things and services surrounding the data subject
- Individual control over the collection and processing of personal information by the surrounding smart things
- Awareness and control of subsequent use and dissemination of personal information by those entities to any entity outside the subject's personal control sphere.

Data privacy in the IoT captures in essence the idea of informational self-determination by enabling the subject (i) to assess its personal privacy risks, (ii) to take appropriate action to protect its privacy, and (iii) to be assured that it is enforced beyond its immediate control sphere. Data Privacy has been a hot research topic in different technology and application areas that are important enablers of the IoT vision.

Despite considerable contributions from research communities, arising privacy issues in the IoT have not been efficiently dealt. This is because the composition of a growing number of technologies and a range of changing features with an explosion in the number of smart things, interactions and inter-communications among users and things in the IoT. These new features of the IoT will aggravate privacy issues and introduce unforeseen threats that pose challenging technical problems.

In the complex IoT environment, privacy problems cannot be optimally solved due to their complexity. In these situations, AI has proven to be extremely useful and well-fitted to solve these problems. Artificial neural networks, evolutionary computation, clustering, fuzzy sets, multi-agent systems, data mining and pattern recognition are just a few examples of AI techniques that can be successfully used to solve some relevant privacy and security problems.

This report identifies the threats and challenges in data privacy in the IoT along with some state-of-the-art AI-based solutions as follows:

- briefly presents a Privacy reference model for the IoT
- introduces major privacy threats and challenges in the IoT
- describes some state-of-the-art approaches based on AI for dealing with such data privacy issues
- concludes the survey as well as propose prospective research directions for AI-based data privacy mechanisms in the IoT.

6.2 Privacy Reference Model for the IoT

6.2.1 Requirements related to IoT environment

Privacy includes the concealment of personal data as well as the ability to control what happens with this data[28]. The right to privacy can be considered as either a basic and inalienable human right, or as a personal right or possession. There are two main approaches for dealing with privacy challenges in the IoT:

- Privacy enhancing technologies (PET): PET refers to specific methods that act in accordance with the laws of data protection. PET is a system of ICT that measures the protection of informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal date[29]. The fulfilment of customer privacy requirements is quite difficult. A number of technologies have been developed in order to achieve privacy goals. PET can be any mechanisms that enhance the privacy[30].
- Legal course of action: Privacy legislation tries to draw boundaries to the evermore data-hungry business models of many Internet enterprises (e.g., data market places, advertising networks and e-commerce sites) and to define mandatory practices and processes for privacy protection.

The European Commission is aware of the security and privacy issues related to the RFID and the IoT. In particular, the Recommendation outlines measures to be taken for the deployment of RFID application to ensure that national legislation is complying with the EU Data Protection Directives 95/46, 99/5 and 2002/58 (No. 2). Member States should ensure that industry in collaboration with relevant civil society stakeholders develops a framework for privacy and data protection impact assessments (PIA; No. 4); this framework should be submitted to the Article 29 Data Protection Working Party within 12 months. The new General Data Protection Regulation (GDPR), adopted in 2016, replaces the EU Data Protection Directive (and the related national acts such as the UK DPA) as it came into force on May 25, 2018.

However, the level of privacy protection offered by legislation is insufficient, as day-to-day data spills and unpunished privacy breaches remain prevalence. The IoT will undoubtedly create new grey areas with ample of space to circumvent legislative boundaries.

6.2.2 Privacy Principles

The concepts of privacy and data protection must not be reduced to protection of data. In fact, the concepts have to be understood more broadly: they address the protection of human beings and their personal rights as well as democratic values of society. Keeping this in mind, privacy and data protection require safeguards concerning specific types of data since data processing may severely threaten informational privacy[31].

---

[28] Seda F. Gurses/Bettina Berendt/Thomas Santen, Multilateral Security Requirements Analysis for Preserving Privacy in Ubiquitous Environments, in: Bettina Berendt/Ernestina Menasalvas (eds), Workshop on Ubiquitous Knowledge Discovery for Users (UKDU '06), at 51–64; for privacy as freedom see Gus Hosein, Privacy as Freedom, in: Rikke Frank Jørgensen (ed.), Human Rights in the Global Information Society, Cambridge/Massachusetts 2006, at 121–147.

[29] https://en.wikipedia.org/wiki/Privacy-enhancing_technologies

[30] Fabian, supra note 6, 61 s; Benjamin Fabian/Oliver Gunther, Security Challenges of the EPCglobal Network, Communications of the ACM, Vol. 52, July 2009, 121–125, at 124 s.

[31] Danezis, George, et al. "Privacy and Data Protection by Design-from policy to engineering." arXiv preprint arXiv:1501.03726 (2015).

Several terms have been introduced to describe types of data that need to be protected. A term very prominently used by industry is "personally identifiable information (PII)", i.e., data that can be related to an individual. Similarly, the European data protection framework centres on "personal data". However, some authors argue that this falls short since also data that is not related to a single individual might still have an impact on the privacy of groups, e.g., an entire group might be discriminated with the help of certain information.

The ISO[32] and the OECD[33] have identified 11 privacy principles from privacy laws and regulations based on the international guidelines that have been defined to protect privacy. Wright and Raab extend that to 20 principles.

They argue that these principles should be considered as new products and services are developed[34]. Some of the principles are particularly applicable to IoT, such as "Right to confidentiality and secrecy of communications", "Consent and choice" and "People should not ... be denied goods or services or offered them on a less preferential basis". It seems that the IoT developers have not taken Wright and Raab's admonition to heart, hence the need for privacy-related IoT privacy-preserving solutions remains largely unfulfilled.

6.2.3 Privacy reference model

The privacy reference model is considered based on the IoT reference model proposed by ITU and IoT European Research Council (IERC) visions[35]. Here in the privacy model, there are 4 main types of entities namely: Smart Things, Subject, Infrastructure and Services with 5 different data flows including Interaction, Collection, Processing, Dissemination and Presentation.

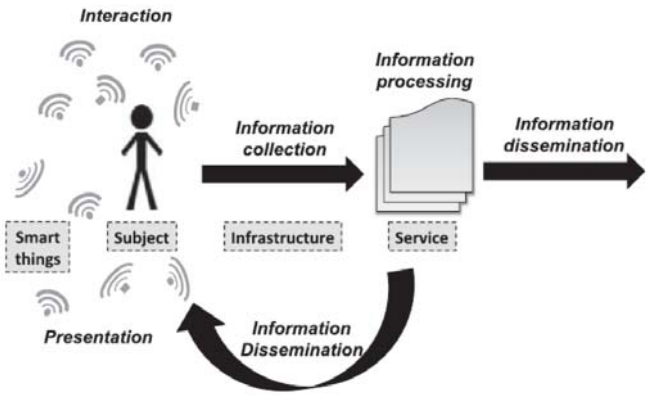


Figure 12. IoT reference model with relevant entities and data flows in a typical IoT application.

[32] International Organization for Standardization. Information technology security tech- niques privacy framework, iso/iec 29100, 2011.

[33] OECD. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. OECD Publishing, 2002

[34] D. Wright and C. Raab. Privacy principles, risks and harms. International Review of Law, Computers & Technology, 28(3):277–298, 2014.

[35] Internet of Things European Research Cluster (IERC). The Internet of Things 2012—New Horizons, 3rd edn.: Halifax, UK, 2012.

This abstract model is well aligned with other IoT models such as the models proposed by the IoT-i consortium[36], existing reference models[37] in Atzori et al. survey and IoT architectures. Considerable progress toward an explicit reference model has been made, for example, by EU FP-7 projects IoT-A[38] and CASAGRAS[39].

6.3 Privacy Threats and Challenges in the Internet of Things

The evolving nature of the IoT regarding technologies and features and the emerging new ways of interaction with the IoT lead to specific privacy threats and challenges.

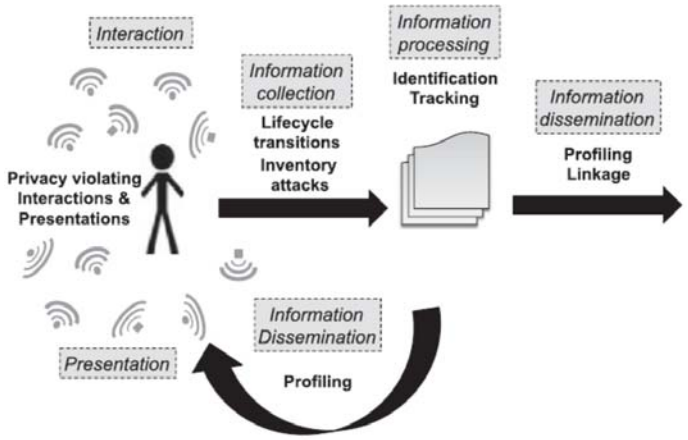


Figure 13. Threats in the reference model.

Figure 13 illustrates seven threat categories (phases) from the privacy reference model in IoT in Figure 12, namely: Identification, Localization and Tracking, Profiling, Privacy-violating Interaction and Presentation, Lifecycle transition, Inventory attacks, and Linkage[40].

6.3.1 Identification

Identification denotes the threat of associating a (persistent) identifier, for example, a name and address or a pseudonym of any kind, with an individual and data about him. The threat thus lies in associating an identity to a specific privacy-violating context, and it also enables and aggravates other threats, for example, profiling and tracking of individuals or combination of different data sources.

---

[36] Bauer M, Carrez F, Egan R, et al. IOT-I: Internet of Things Initiative: Public Deliverables – D1.2 First Reference Model White Paper, 2011

[37] Atzori L, Iera A, Morabito G. The Internet of Things: a survey. *Computer Networks* 2010; **54**(15): 2787–2805, doi:10.1016/j.comnet.2010.05.010

[38] IOT-A Consortium. Internet of Things architecture, 2011. Available at: http://bit.ly/124jw0M (Accessed 2012-10-12)

[39] Dunkels A, Vasseur J. IP for smart objects. Ipso alliance white paper, 2008

[40] Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. (2014). Privacy in the Internet of Things: threats and challenges. Security and Communication Networks, 7(12), 2728-2742.

The threat of identification is currently most dominant in the information processing phase at the backend services of our reference model, where huge amounts of information are concentrated in a central place outside of the subject's control. In the IoT, also the interaction and collection phase will become relevant because the impact of the evolving technologies and interconnection and interaction features aggravates the threat of identification. Identity protection and, complementary, protection against identification is a predominant topic in RFID privacy, but has also gained much attention in the areas of data anonymization[41], and privacy enhancing identity management. Those approaches (i.e., data anonymization, privacy enhancing identity management) are difficult to fit to the IoT: Most data anonymization techniques can be broken using auxiliary data, which are likely to become available at some point during the IoT evolution. Identity management solutions, besides relying heavily on expensive crypto-operations, are mostly designed for very confined environments, such as enterprise or home networks, and thus difficult to fit to the distributed, diverse, and heterogeneous environment of the IoT.

6.3.2 Localization and Tracking

Localization and tracking is the threat of determining and recording a person's location through time and space. Tracking requires identification of some kind to bind continuous localizations to one individual. Already today, tracking is possible through different means, for example, GPS, internet traffic, or cell phone location. Many concrete privacy violations have been identified related to this threat, for example, GPS stalking[42], disclosure of private information such as an illness[43], or generally the uneasy feeling of being watched[44]. However, localization and tracking of individuals is also an important functionality in many IoT systems.

The location privacy is the protection of location information of user's sensitive information such as residence location, behaviour, health status and other sensitive information. IoT devices have a built-in GPS system for positioning of location information. The user may issue a query to location based services (LBS) for location information. The query may be for a location of interest—for example, the nearest restaurant, hospital, park or other places. The query contains the identity and location of the user. The convenience of using LBS services creates issues of privacy risk. Based on the provided information, an adversary could easily link the identity and location of the user to get more private information. Security and privacy are a critical measure to consider for information gathering and broadcasting. This information and data must be secure from illegal and unauthorized access.

6.3.3 Profiling

Profiling denotes the threat of compiling information dossiers about individuals to infer interests by correlation with other profiles and data. Profiling methods are mostly used for personalization in e-commerce (e.g., in recommender systems, newsletters, and advertisements) and also for internal optimization based on customer demographics and interests. Existing approaches to preserve privacy

[41] Fung BCM, Wang K, Chen R, Yu PS. Privacy preserving data publishing: a survey of recent developments. *ACM Computing Surveys* 2010; 42 (4):14:1–14:53, doi:10.1145/1749603.1749605

[42] Voelcker J. Stalked by satellite—an alarming rise in GPS-enabled harassment. *IEEE Spectrum* 2006; **43**(7): 15–16, doi:10.1109/MSPEC.2006.1652998

[43] Chow CY, Mokbel MF. Privacy in location-based services: a system architecture perspective. *SIGSPATIAL Special* 2009; **1**(2): 23–27, doi:10.1145/1567253. 1567258.

[44] Toch E, Wang Y, Cranor L F. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Modeling and User-Adapted Interaction 2012; 22(1): 203–220, doi:10.1007/s11257-011-9110-z.

include client-side personalization, data perturbation, obfuscation and anonymization, distribution, and working on encrypted data[45][46].

In the IoT, identification and access control technologies provide essential infrastructure to link data between a user's devices with unique identities, and provide seamless and linked up services. At the same time, profiling methods based on linked records can reveal unexpected details about users' identity and private life, which can conflict with privacy rights and lead to economic, social, and other forms of discriminatory treatment. A balance must be struck between identification and access control required for the IoT to function and user rights to privacy and identity. Striking this balance is not an easy task because of weaknesses in cybersecurity and anonymisation techniques. The EU GDPR, set to come into force in May 2018, may provide essential guidance to achieve a fair balance between the interests of IoT providers and users. Through a review of academic and policy literature, this report maps the inherent tension between privacy and identifiability in the IoT. It focuses on four challenges: (1) profiling, inference, and discrimination; (2) control and context-sensitive sharing of identity; (3) consent and uncertainty; and (4) honesty, trust, and transparency. Sandra Wachter et. al examines the extent to which several standards defined in the GDPR will provide meaningful protection for privacy and control over identity for users of IoT[47].

6.3.4 Privacy-violating Interaction and Presentation

This threat refers to conveying private information through a public medium and in the process disclosing it to an unwanted audience. It can be loosely sketched as shoulder surfing but in real-world environments. Many IoT applications, for example, smart retail, transportation, and health care, envision and require heavy interaction with the user. In such systems, it is imaginable that information will be provided to users using smart things in their environment, for example, through advanced lighting installations, speakers, or video screens. Vice versa, users will control systems in new intuitive ways using the things surrounding them, for example, moving, touching, and speaking to smart things. However, many of those interaction and presentation mechanisms are inherently public; that is, people in the vicinity can observe them. This becomes a threat to privacy when private information is exchanged between the system and its user.

6.3.5 Lifecycle transition

Privacy is threatened when smart things disclose private information during changes of control spheres in their lifecycle. The problem has been observed directly with regard to compromising photos and videos that are often found on used cameras or smartphones—in some cases, disturbing data have even been found on ‘new’ devices. Because privacy violations from lifecycle transitions are mainly due to the collected and stored information, this threat relates to the information collection phase of the privacy reference model in IoT (Figure 12).

6.3.6 Inventory attacks

Inventory attacks refer to the unauthorized collection of information about the existence and characteristics of personal things. With the realization of the all-IP and end-to-end vision, smart things become query-able over the Internet. Whereas things can then be queried from anywhere by

[45] Spiekermann S, Cranor L. Engineering privacy. IEEE Transactions on Software Engineering 2009; 35(1): 67–82, doi:10.1109/TSE.2008.88

[46] Kobsa A. Privacy-enhanced web personalization. *The Adaptive Web*. Springer-Verlag: Berlin, Heidelberg, 2007; 628–670

[47] Wachter, Sandra. "Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR." (2017).

legitimate entities (e.g., the owner and authorized users of the system), non-legitimate parties can query and exploit this to compile an inventory list of things at a specific place, for example, of a household, office building, or factory. Even if smart things could distinguish legitimate from illegitimate queries, a fingerprint of their communication speeds, reaction times, and other unique characteristics could potentially be used to determine their type and model. With the predicted proliferation of wireless communication technology, fingerprinting attacks could also be mounted passively, for example, by an eavesdropper in the vicinity of the victim's house. Because inventory attacks are mainly enabled by the increasing communication capabilities of things, the threat arises in the information collection phase of the privacy reference model.

6.3.7 Linkage

This threat consists in linking different previously separated systems such that the combination of data sources reveals (truthful or erroneous) information that the subject did not disclose to the previously isolated sources and, most importantly, also did not want to reveal. Users fear poor judgment and loss of context when data that were gathered from different parties under different contexts and permissions are combined[48]. Privacy violations can also arise from bypassing privacy protection mechanisms, as the risks of unauthorized access and leaks of private information increases when systems collaborate to combine data sources. A third example of privacy violations through linkage of data sources and systems is the increased risk of re-identification of anonymized data. A common approach toward protecting privacy is working on anonymized data only, but the act of combining different sets of anonymous data can often enable re-identification through unforeseen effects[49][50].

6.4 AI-based Privacy Techniques and Mechanisms

Successful security success is about having the right combination of people, process, policy and technology. This can be achieved by developing a network management systems capable of intellectual reasoning, dynamic real time decision making, and self-adaptation and improvement based on experiences. The design of such efficient, dynamic and automated social network management framework requires support from the field of AI. Dealing with uncertainty and inconsistency has been a part of AI since its origins.

6.4.1 Traditional Privacy Preserving Approaches

In order to address the privacy concerns of end-users and privacy considerations of service providers, several approaches have been proposed by the research community:

a) Cryptographic techniques and information manipulation: Although researchers have spent many years proposing novel privacy-preserving schemes, cryptography is still the most dominant solution. However, due to limited storage and computation resources, cryptography often cannot offer adequate security protocols to safeguard end-users' data.

---

[48] Spiekermann S, Cranor L. Engineering privacy. *IEEE Transactions on Software Engineering* 2009; **35**(1): 67–82, doi:10.1109/TSE.2008.88

[49] Narayanan A, Shmatikov V. Myths and fallacies of "personally identifiable information". *Communications of the ACM* 2010; **53**: 24–26, doi:10.1145/ 1743546.1743558

[50] El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. *PLoS ONE* 2011; **6**(12) , doi:10.1371/journal.pone.0028071

b) Privacy awareness or context awareness: The solutions for the lack of privacy awareness have been primarily focused on relying individual applications to provide a basic privacy terms and conditions to the end-users. This practice is common among devices such as smart TVs, wearable fitness devices, and health monitor systems. For instance, in a recent research, a framework called SeCoMan was proposed to act as a trusted third party for the users as applications might not be reliable enough with the location information that they manage[51].

c) Access control: Access control is one of the viable solutions to be used in addition to encryption and raising privacy awareness. This gives users the power to manage their own data. An example of this approach is CapBAC, as proposed by Skarmeta, Hernandez, and Moreno[52]. It is essentially a distributed approach in which smart things themselves are able to make fine-grained authorization decisions.

d) Data minimization: The principle of "data minimization" means that the IoT service providers should limit the collection of personal information to what is directly relevant. They should also retain the data only for as long as it is necessary to fulfill the purpose of the services. In other words, they should collect only the personal data they really need and should keep it only for as long as they need it.

There are other proposed solutions that do not fall into the previous four categories, such as hitchhiking. This is a new approach to ensure the anonymity of users who provide their locations. Hitchhiking applications handle locations as the entity of interest. Because the knowledge of who is at a particular location is unnecessary, the fidelity tradeoff is removed[53]. Another example is the introspection technique that proactively protects users' personal information by examining the activities of the virtual machine (VM). It gathers and analyzes the CPU state of every VM, the memory contents, file I/O activity, network information that is delivered via hypervisor and detects malicious software on the VM. However, if IoT device loses integrity due to any malicious attack, it creates risks to the users' privacy[54].

6.4.2 Prospective AI-based Privacy Preserving

Capabilities of AI can be leveraged to deal with privacy challenges in the IoT.

AI-based Identification Management

- AI - is it the answer for identity management[55]?
- Could AI improve identity management and security[56]?

Identity and Access Management (IAM) is already a key weapon in the security arsenal of many organisations as a way to mitigate against data breaches and manage the additional risks that come with remote working and Bring Your Own Device (BYOD). And the take up of IAM solutions is set to gain even more momentum. IAM solutions enable a network or system to authenticate the identity of a user against a set of pre-prescribed credentials. Depending on the system being accessed, these

[51] A. Huertas Celdran, G. Clemente, J. Felix, M. Gil Perez, and G. Martinez Perez. Secoman: A semantic-aware policy framework for developing privacy-preserving and context-aware smart applications. IEEE Systems Journal, 99:1–14, 2013.

[52] A. F. Skarmeta, J. L. Hernandez-Ramos, and M. Moreno. A decentralized approach for security and privacy challenges in the internet ofthings. In Internet of Things (WF-IoT), 2014 IEEE World Forum on,pages 67–72. IEEE, 2014.

[53] K. P. Tang, P. Keyani, J. Fogarty, and J. I. Hong. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 93–102. ACM, 2006.

[54] C. Kang, F. Abbas, and H. Oh. Protection scheme for IoT devices using introspection. In Network of the Future (NOF), 2015 6th International Conference on the, pages 1–5. IEEE, 2015

[55] https://www.scmagazineuk.com/artificial-intelligence--is-it-the-answer-for-identity-management/article/531630/

[56] https://www.digitalcatapultcentre.org.uk/could-ai-improve-identity-management-and-security/

can range from a simple username and password to digital certificates, physical tokens, biometric passwords (such as fingerprints, iris scans, or facial recognition), or a combination of these features.

Traditionally, the strength of the authentication required depends on the sensitivity of the material being accessed, as well as the impact should these resources fall into unauthorised hands. Public information might require little or no authentication, while proprietary or classified data or accounts with administrative privileges will require stronger authentication, preferably using multiple factors.

While the above still holds true, recent thinking around best practice in IAM has moved on. The focus has shifted from authenticating identity to controlling access based on the principle of least privilege access. In practice, what that means is that every user – whether an individual, a device, a programme or a process – is given access only to the resources needed to fulfil their role.

Least privilege is an approach that acknowledges how serious the insider threat is to businesses and the fact that just because someone has established their identity as an employee with the right credentials should not mean unfettered access to company systems.

While this is sound in principle, least privilege and deciding who should have access to what and when can be difficult for organisations. One issue with least privilege in IAM is that users are usually given access privileges based on their role in an organisation, but employees rarely fit neatly into a single role. It is common that they may need special one-time access or each person fulfilling the same role may need slightly different types of access. Another challenge is that some organisations fail to extend the concept of least privilege access right across the organisation, monitoring those classified as privileged users, such as systems administrators.

But how might AI help? So often with data breaches it's not the management of the identity that causes the breach, but the transfer of credentials to some unknown party. While least privilege access control does afford some protection here, there are clearly insufficient. Identity management and access control have always been two sides of a coin, but in the future AI will be the glue to bind them together to much greater effect.

Moving on to biometric passwords, it's not difficult to conceive that AI could identify a user by using sight and sound. Rather than checking pre-defined credentials, a machine would be able to identify whether a person using visual and aural clues, granting access to this person accordingly.

AI also offers the potential for intelligent, real-time security by implementing fine-grained access control. Just because a user proved who they were at log on two minutes ago, should the system continue to believe they are who they say they are? Visual images and voice could obviously still play a part here, constantly monitoring users as they move around the network. However, in addition to behavioural factors and real-time, risk analysis can also come into play.

Working within a user's access permissions, AI systems could monitor in real-time whether a user is accessing or trying to access a part of the system they never normally would or suddenly downloading more documents than they generally would. The rhythm of a user's keyboard and mouse movements could be observed to identify irregular or unusual patterns. Taking this a step further it's not inconceivable that insights from an individual's online identity and activity – their social profile, groups they are part of, people they follow, websites they visit – could be used to determine a risk score. Drawing this data together, actions taken by the AI system could range from an alert being triggered, to specific areas of a corporate system being switched off for a user, to access being instantly revoked.

In the future, a truly intelligent system will know, understand, monitor and act drawing whatever clues it requires on a user. Identity and credentials will not be separate elements. An individual's identity will become their credentials. That should be the ultimate goal of any AI system.

AI and block-chain helping streamline data and identity management[57]

As we push further into the digital age, we unconsciously change how we perceive and expect services from government. This is particularly true of digital natives who, by birth right, have the "ask only once" expectation.

Digitization addresses two broad government objectives—to increase transparency and to improve speed of service for citizens. The two elements necessary to achieve this are data and identity. When combined effectively, they provide potent fuel to provide public services. The challenges of yesterday are the opportunities today and large data is easier to manage and share with citizens or agencies in order to effectively provide services. The key is to "un-duplicate" the work for citizens and public servants using disruptive technologies like AI, robotics, and block-chain.

With all the tools (networks) and fuel (data) available, government should aim for the "ask only once" point. Most data and identity is with the agencies or available to governments and same goes with citizens, this situation leaning towards transparency. To put it simply: citizens and governments know each other much better than they did a generation ago; we need ask fewer questions of one another and don't have to provide KYC at each stage of requesting a service.

The Aadhar card is one example of Indian government trying to seamlessly integrate services into citizens' everyday life. High school grades are now available in a government database and linked to each student's Aadhar number. Students should be allowed to send in applications without transcripts. The college can check everything with the Aadhar number.

Humans have historically spent a lot of time and resources producing things or providing services—from manpower to produce food to fossil fuels to ignite the industrial revolution. While we have succeeded in making the production of goods and food produce more efficient, we now have to better manage the huge amounts of data we generate and become leaner and more effective with data management. The future of data management means producing relevant data but not producing too much to manage. With tools like AI and block-chain we can avoid duplication of work and streamline identity management, while with biometrics we largely resolve issues relating to individual identity. The next step is to identify what data governments should manage in order to reach the "ask only once" point.

Although we seem to have everything to make "ask only once" a reality, we must still address how agencies secure this data. This is a challenge in democratic societies where privacy issues are respected; a recent Supreme Court judgment on privacy in India being a case in point. Citizens must have confidence in their agencies and this confidence must stem from legislation, much like it already does for the protection of physical property. The Estonian Tax and Customs Board's strong focus on transparency for example, has helped it become one of the most trusted organizations in the country.

The challenge is that unlike physical assets, our digital data is easy to access and potentially to tamper with. If government is to provide services seamlessly, it must first adequately secure our data with digital locks. This can be difficult insofar as much of it resides in the cloud and must be regulated by international treaties. Facebook and Google, for example, are often ordered to share information by national courts. Things become complicated when companies fall back on national data protection laws or rely on the unenforceability of national legislation.

Finally, most citizens today already have a digital footprint. It is in fact difficult to imagine that in a few more years anybody with citizenship will not have one. We are past the stage of asking who needs to be in the system. We are in the system.

A blockchain is a distributed immutable database consisting of a continuous growing list of blocks used to record transactions between peers in a network. The blockchain is then synchronized and

[57] https://www.capgemini.com/2017/09/ai-and-blockchain-helping-streamline-identity-management/

distributed across the network, playing as a distributed ledger. By nature, blockchain is inherently resistant to data modification because data in any given block cannot be altered retroactively as this would invalidate all hashes in the previous blocks in a blockchain; and break the consensus agreed among nodes in the network.

In Smart cities context, a GDPR-compliant data management solution for a Smart City platform leveraging Blockchain and Smart Contracts is envisioned. From citizens' point of view, personal data usage control should be strengthened by a trusted and transparent solution that enables citizens to verify if the data was accessed, processed, and transferred without violating consents, and to withdraw the consent arbitrarily whenever needed. From Smart City operators' perspective, the solution should provide a legal contract that the operators have received the consent from citizens for managing their personal data, and the agreement on data usage obligations from service providers.

In Smart Cities, operators and agencies are supposed to be trustworthy to manage all data from citizens and control the data usage from third-party service providers. Therefore, a permissioned blockchain is more suitable for a GDPR-compliant data management model rather than a public one, as many advantageous simplifications can be achieved such as it is not required to implement a cryptographic token for incentivizing mining and for determining financial stake. Transactions in a permissioned network can be considered immediately final, as the possibility of having to resolve a fork can be eliminated. In this regard, Smart City operators are in charge of restricting stakeholders to join the network, to read the ledgers, to propose transactions, and to engage in consensus process. In addition to high-speed transactions and finality, permissioned blockchain also benefits from improved privacy, control, and scalability. The adoption of the per-missioned blockchain has impact on privacy, anonymity, performance, and scalability.

AI-based Localization and Tracking

Some traditional AI-based approaches have been proposed for privacy preserving of location & tracking:

- Techniques based on anonymization - K-anonymity is one of the basic techniques for protection of privacy proposed for the first time by Sweeney[58]. The k-anonymity model addresses the re-identification problem during broadcasting sensitive information for the research objective. Gedik and Liu presented a new architecture for the protection of location privacy from several threats due to unrestrained practice of LBS[59]. This strategy contains a personalized k-anonymity prototype and a suite of algorithms based on anxiety to protect privacy. The distinctive feature of this design is the elastic personalization privacy to sustain k-anonymity for wide-ranging mobile clients. The prototype is designed to be on a trusted platform of an anonymization server. Wang et al. presented mobile user location privacy in active and varied scenarios, reinforcing it to articulate the location awareness and location privacy protection (L2P2) problem. The problem is additionally distributed into basic and enhanced problems, and a distinct algorithm offered for each problem[60]. AI can be used for improving the effectiveness of such techniques.
- Randomized noise-based techniques - Such techniques which are based on random noise added to the original location. This random noise changes or blurs the original location in such way that the adversary cannot acquire the actual location of the user. A location privacy preserving mechanism (LPPM) must contemplate three fundamental features: user privacy requirements, knowledge and abilities of adversary, and tolerated service quality. Shokri et al. introduced an

[58] Sweeny, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557–570.

[59] Gedik, B., & Liu, L. (2008). Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, *7*(1), 1–18.

[60] Wang, Y., Li, F., & Xu, B. (2012). L2P2: Location-aware location privacy protection for location-based services. In *INFOCOM, 2012 proceedings IEEE* (pp. 1996–2004).

optimum LPPM for LBS which gives users a service quality constraint against an adversary optimal inference algorithm[61]. The authors formalize mutual optimization with location privacy versus correctness of localization by using Stackelberg Bayesian games. It reports that an adversary could not observe that the location has been disturbed by the user.

AI-based Profiling

There are at least three possible ways of monitoring and profiling that offer grounds for discrimination in IoT systems: (a) data collection that leads to inferences about the person (e.g. Internet browsing behaviour); (b) profiling at large through linking IoT datasets (sometimes called 'sensor fusion'); and (c) profiling that occurs when data is shared with third parties that combine data with other datasets (e.g. employers, insurers).

Weaknesses of anonymisation to prevent profiling and resulting discrimination cause further problems. According to Gudgel, "there is special concern that if data is not anonymized then it could potentially be used to track specific individuals, linked to information in other databases, and possibly used to predict future behaviour."[62] Tracking data of the type that many IoT devices generate is notorious to open to re-identification and reverse engineering of identity. Following the assumption that data cannot be permanently anonymised without destroying its analytical value, non-technical methods may be necessary to prevent profiling and discrimination in the IoT. One potential solution is to treat all IoT generated data that refers to a user as personal data under data protection law, as it will always be possible in principle to link the data back to a person. This approach would ensure that the user would be able to exercise his/her rights granted under data protection law over all information that IoT devices create and manage. This concept would not prevent profiling as a result; but rather, extend the scope of existing user rights against privacy risks to cover all data, including inferences and profiles.

These problems will be exacerbated by the proliferation of machine learning in the IoT. Machine learning will lead to even less predictable inferences, while the complexity and opaqueness of machine learning algorithms can inadvertently hide discriminatory treatment from users. Systems operating as 'black boxes', for which the inputs, internal logic, and outputs may be unavailable or incomprehensible to individual users do not facilitate systematic observation, identification of harmful effects, or investigation of their causes. Machine learning can inadvertently and unknowingly reinforce existing biases and prejudices as a result[63].

European legislators have addressed the risks of profiling and discrimination, albeit often lacking detailed recommendations. European regulators have raised "major concerns in declarations on profiling," acknowledging that profiling offers grounds for discrimination, especially when datasets are combined. The European Commission has called for the creation of a set of guiding principles to govern IoT regulation, urging that "always being connected to the things around us has the potential to lead to more surveillance or more profiling by public authorities and private entities." Similarly, the European Data Protection Supervisor has voiced concerns that RFID tags used in IoT systems could lead to profiling by linking users to specific devices and usage records.

Similar concerns are reflected in the GDPR, especially in Article 21 (Right to object) and Article 22 (Automated individual decision-making, including profiling)[64].

AI-based Privacy-violating Interaction and Presentation

[61] Shokri, R., Theodorakopoulos, G., & Troncoso, C. Protecting location privacy: Optimal strategy against localization attacks. In *Proceedings of the 2012 ACM conference on computer and communications security* (pp. 617–627). ACM.

[62] John Gudgel, 'Objects of Concern? Risks, Rewards and Regulation in the Internet of Things"' (Social Science Research Network 2014) SSRN Scholarly Paper ID 2430780 12, July 2017.

[63] Brent Mittelstadt, 'Auditing for Transparency in Content Personalization Systems' (2016) 10 International Journal of Communication 12.

[64] Wachter, Sandra. "Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR." (2017).

Because such advanced IoT services are still in the future, privacy-violating interactions have not received much attention from research. Interaction mechanisms are, however, crucial to usable IoT systems, and privacy threats must consequently be addressed.

We identify two specific challenges that will have to be solved: First, we need means for automatic detection of privacy-sensitive content. It is easily imaginable that the provisioning of content and rendering it for the user are handled in two steps by two different systems: For example, company A generates recommendations for customers of a store, which are then delivered to the customer by company B's system either by special lighting and the use of speakers or through a push to his smartphone. How to choose between those two interactions mechanisms, one public one private? Should company A mark privacy-sensitive content or should company B detect it? How can company B (committed to privacy) protect itself from A's lax privacy attitude? Automatic detection of privacy-sensitive content can help to decide these questions. Second, with the previous point in mind, scoping will be necessary; that is, how can we scope public presentation medium to a specific subgroup of recipients or a specific physical area? This approach would prove useful to support users, which have no smartphone (or any other device providing a private channel for interactions and presentations). However, it will be difficult to accurately determine the captive audience of a particular presentation medium, separate the intended target group, and adjust the scope accordingly. For example, what if the target user is in the midst of a group of people? Applications for privacy-preserving pervasive interaction mechanisms are, for example, smart stores and malls, smart cities, and healthcare applications. Here, it would certainly be an achievement to provide similar levels of privacy as people would expect in the contexts of their everyday conversations, that is, interactions with their peers.

AI-based Linkage

The purpose of a privacy preserving technique for a privacy threat is to link data across organisations such that besides the linked records (the ones classified to refer to the same entities) no information about the sensitive source data can be learned by any party involved in the linking, or any external party. Challenges with this technique is that the linkage unit needs access to personal details (metadata might also reveal sensitive information). Also, collusion between parties, and internal and external attacks, make these data vulnerable.

Privacy-preserving record linkage (PPRL) is a solution that aims to overcome these drawbacks. No un-encoded data ever leave a data source; only details about matched records are revealed. However, provable security against different attacks PPRL is challenging (employs techniques from cryptography, databases, etc.)[65].

[65] Verykios VS, Karakasidis A and Mitrogiannis VK: Privacy preserving record linkage approaches. International Journal of Data Mining, Modelling and Management, 2009.

7. AI-based Trust mechanisms in the Internet of Things

7.1 Introduction

This section explains the latest advances and applications of trust solutions based on AI techniques for IoT through the various type of references including scientific publications, text books, and online articles. The scope of the survey includes, but is not limited to, titles in the areas of trust solutions based on AI techniques like supervised-unsupervised machine learning, reputation systems, reinforcement learning, deep learning, and multi agent systems.

First, in order to consider trust issues, it is required to understand basic concepts of trust in Section 4.3. Then, it is important to clarify that trust is neither a property of a trustor (e.g., trustor's preferences) nor a property of a trustee (e.g., trustee's trustworthiness and trustee's reputation). It is a relationship between the trustor and the trustee that is subjective and asymmetric which is derived from the triad of trustee's trustworthiness, trustor's propensity and environment's characteristics. Based on the clarification of the trust concept, a conceptual trust model in the IoT is proposed as illustrated in Figure 14. Then, a more specific trust definition in the IoT associated with the conceptual trust model is proposed as follows:

"Trust is the perception of a trustor on trustee's trustworthiness under a particular environment (within a period of time) so-called perceived trustworthiness."

According to the model illustrated in Figure 14, trust will be obtained by harmonizing the trustor's propensity and environment conditions into the trustee's trustworthiness. The harmonization is accomplished by aggregating both the observation of a trustor toward a trustee and the interactions between the two. It is worth to note that the environment conditions are reflected as risks taken during the observations and interactions. The trustor's propensity includes both requirements for the trust goal and the trustor's preferences about the trustee's trustworthiness whereas the environment conditions are the considerations for some factors such as vulnerabilities, threats and risks. The trust goal requirements with the environmental factors helps determining the set of TAs for deriving the perceived trustworthiness whereas the trustor's preferences is to help combining these TAs to obtain an overall trust value for making a decision. For example, trustor's preferences could be represented in forms of weights of TAs, indicate the levels of importance of the TAs when constructing trust. Trust as perceived trustworthiness is as an instance of trustee's trustworthiness respecting to a particular trustor and an environment, thus, even same a trustee and same an environment, different trustors might have different propensities of the trustee's trustworthiness. This illustrates the subjective characteristic of trust. Another important characteristic of trust is the context-dependence that can also be illustrated using this conceptual model as follows: with the same trustor and trustee, different environments might result in different TAs and different trustor's propensities.

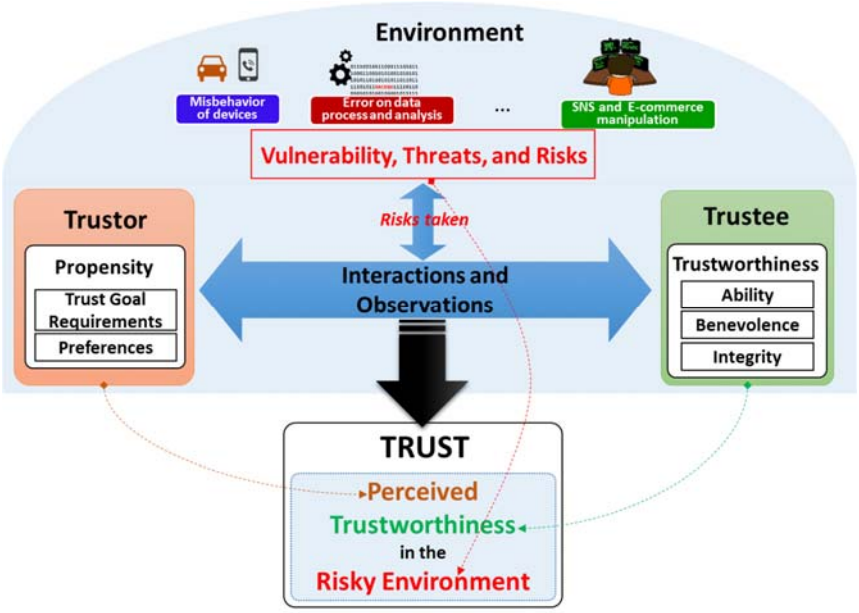


Figure 14. Conceptual Trust Model in the IoT environment.

Based on the conceptual model, the goal of any trust model is two-fold: (i) to specify and evaluate TAs of the trustworthiness of a trustee respecting to the trustor's propensity and the environment conditions; (ii) to combine the TAs to finalize the perceived trustworthiness as the trust value. From now on in this article, the term "trust" is referred to this conceptual model and it is interchangeably used with the term "perceived trustworthiness".

According to the proposed conceptual trust model, in order to quantify trust, it is necessary to investigate trustee's trustworthiness by specifying TAs associated with it. As mentioned above, trustworthiness is as a composite of a variety of TAs that illustrate different characteristics of the trustee. Despite a large number of TAs have been figured out in trust-related literature, TAs are mostly fallen into three categories as the three main dimensions of trustworthiness: Ability, Benevolence and Integrity. This classification is appropriate for consideration of trustworthiness in the IoT environment.

- Ability is a dimension of trustworthiness showing the capability of a trustee to accomplish a trust goal. An entity may be high benevolent and integrity for fulfilling a trust goal but the results may not be satisfactory if it is not capable. This term incorporates some other terms that have been used as TAs in many trust-related literature such as competence, expertness, and credibility.
- Benevolence is a dimension of trustworthiness showing to what extent a trustee is willing to do good things or not harm the trustor. Benevolence ensures that the trustee will have good intentions toward the trustor. This term incorporates some TAs such as credibility, relevance, and assurance as TAs.
- Integrity is a dimension of trustworthiness showing the trustee adheres to a set of principles that helps the trustor believe that the trustee is not harmful and not betray what it has committed to do. These principles can come from various sources such as fairness, or morality. This term incorporates some TAs such as honesty, completeness, and consistency.

7.2 The Challenge

Trust management technologies have been widely investigated in many fields including economics, sociology, and computer science [1-3].

For example, trust management systems established on policies are presented in [4], and [5], based on reputation systems are explained in [6], [7],and [8], authors in [9] and [10] have developed a trust model corresponding to their owner's social behaviour, [11], [12], [13] and [14] present a computational model for trust based on concepts like a community of interest, friendship, followers, similarity, information reliability, and social opinions.

However, the influence of a trust attribute on trust in above methods are determined by weighting factors. But the assessment of a proper weightage is a complex task since the trust is a varying quantity which depends on many factors, e.g. expectations of a trustor, time, context, etc. Thus, schemes that are more intelligent are required to find these weighting factors and a threshold that defines a trustworthy boundary. Further, existing technologies have investigated limited number of trust features, like packet forwarding ratio, QoS, privacy, reputation, feedbacks and some social attributes. Further, existing technologies heavily rely on readily available knowledge for trust assessment and dynamic nature of the trust is almost ignored in most systems.

7.3 AI for Trust Solutions

Complex characteristics in Trust make it challenging for traditional analysis, but ideal for the application of AI, machine learning techniques and big data analytics. The objective of AI is to investigate the very large volumes of data produced by various components in the IoT ecosystem, and transform the data into meaningful outputs such as trust based decision making, fault detection, service composition and generate ultimate wisdom.

In this regard, authors in [15] and [16] outline the requirements for robust probabilistic trust assessment using supervised learning and apply a selection of estimators to a real-world dataset, in order to show the effectiveness of supervised methods.

Another interesting work that applies machine learning techniques is found in [17]; in this work, they propose to use neural networks in order to provide a global reputation model using the distributed reputation evaluations.

The global reputation is determined by the neural network's output unit, a two class classification in this case. A trust model with a broader scope, not only considering reputation was introduced in [18]; in this work, the authors propose the use of a Bayesian-Network trust model to properly interact with trustworthy peers. Furthermore, authors in [4], [19] and [20] investigate more innovative models and solutions for privacy, security and data integrity based on statistical and deep learning concepts. Moreover, authors in [21] and [22] propose a regression based model which compares the variation of trustworthiness with respect to trust features in mobile ad-hoc networks (MANET) and WSN. Recently, authors in [23], [24] and [25] present several trust management frameworks based on reinforcement learning and multiclass classification techniques.

On the other hand, authors in [26, 27] discuss a generic machine learning framework called MetaTrust for identifying relevant features to determine trust. Specifically, a trustor uses its own previous transactions (with other agents) to build a knowledge base, and utilizes this to assess trustworthiness of a potential transaction based on associated features, which are capable of distinguishing successful transactions from unsuccessful ones.

From a different perspective, Tang et al. [28] addressed the issue of initial trust assessment using the homophily effect. Homophily suggests that similar users are more likely to establish trust relations.

For instance, people with similar tastes about fiction movies tend to trust each other. This work employs low-rank matrix factorization to study trust relations.

In [29], a trust model for multi-agent systems is developed to help the agent make optimal trust decisions over time in a dynamic environment. The target agent's behaviour is predicted according to the Hidden Markov Model (HMM) trust estimation module following the Q-learning greedy policy. ElSalamouny et al. [30] modelled the real dynamic behaviour of an agent by HMMs. They further justified the consistency of the model by measuring the difference between real and estimated predictive probability distributions using relative entropy. The works [31] and [32] demonstrate how HMM-based trust models are applied to distinct application scenarios: routing protocol design in mobile and ad-hoc networks (MANET) and Web service providers selection.

Many trust schemes for multi-agent e-marketplaces have been proposed to deal with the unfair rating problem. The Beta Reputation System (BRS) [33] calculates seller reputation using a probabilistic model based on the beta probability density function, which can be used to represent probability distributions of binary events. The personalized approach proposed by authors in [34] combines buyer's personal experience and the public knowledge held by the system. Several other approaches have also been proposed to deal with unfair ratings. Dellarocas [35] proposed a clustering-based algorithm to separate the advisor's ratings into two clusters (one cluster including lower ratings and another cluster including higher ratings). The ratings in the higher cluster are considered as unfairly high ratings and are discarded. However, this approach cannot effectively handle unfairly low ratings. The iCLUB approach [36] adopts a clustering technique (DBSCAN) to filter out dishonest advisors based on local and global information.

On the other hand, Wang et al. [37] describe a super-agent-based framework for Web service selection, where service clients with more capabilities act as super-agents. These super-agents maintain reputation information of the service providers and share such information with other service clients that have less capabilities than the super-agents. Also, super-agents maintain communities and build a community-based reputation for a service provider based on the opinions from all community members (service clients in a community) that have similar interests and judgement criteria as the super-agents or the other community members. A reward mechanism is also introduced to create incentives for super-agents to contribute their resources (to maintain reputation and form communities) and provide truthful reputation information.

While most of the works on trust evaluation in service oriented computing (SOC) have focused on accurately predicting trust scores, Conner et al. [38] present a trust model that allows each service client (with different trust requirements) to use different scoring functions over the same feedback data for customized evaluations. Rather than assuming a single global trust metric as with many existing reputation systems, they allow each service client to use its own trust metrics to meet its local trust requirements. They also propose a novel scheme to cache the calculated trust values based on recent client activity.

There is a different perspective to consider, that is, trust for AI. As a classical security application of machine learning, intrusion detection followed various learning-based approaches, in particular, anomaly detection [39, 40], rule inference [41-43] and supervised learning [44, 45]. Although most of the proposed methods performed well in controlled experiments, the practical intrusion detection systems, such as Snort [46] and Bro [47], are still rooted in the more conservative signature-based approach. Sommer and Paxson discussed several practical difficulties faced by learning-based intrusion detection systems [48]. Due to the extreme versatility of web applications, it is next to impossible to devise signatures for specific attack patterns.

7.4 References

[1] F. Huang, "Building social trust: A human-capital approach," Journal of Institutional and Theoretical Economics JITE, vol. 163, no. 4, pp. 552-573, 2007.

[2] G. Möllering, "The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension," Sociology, vol. 35, no. 2, pp. 403-420, 2001.

[3] S. P. Marsh, "Formalising trust as a computational concept," Ph.D. dissertation Ph.D. dissertation, Dept. Computing Science and Mathematics, University of Stirling, Stirling, Scotland, UK., 1994.

[4] F. Fei, S. Li, H. Dai, C. Hu, W. Dou, and Q. Ni, "A K-Anonymity Based Schema for Location Privacy Preservation," IEEE Transactions on Sustainable Computing, vol. PP, no. 99, pp. 1-1, 2017.

[5] T. Jim, "SD3: a trust management system with certified evaluation," presented at the Proceedings 2001 IEEE Symposium on Security and Privacy S&P, 2001, 2001.

[6] M. A. Azer, S. M. El-Kassas, A. W. F. Hassan, and M. S. El-Soudani, "A survey on trust and reputation schemes in ad hoc networks," presented at the Third International Conference on Availability, Reliability and Security ARES 08, 2008.

[7] U. Jayasinghe, N. B. Truong, G. M. Lee, and T.-W. Um, "RpR: A Trust Computation Model for Social Internet of Things," presented at the Smart World Congress , Intl IEEE Conferences on Ubiquitous Intelligence & Computing, 2016.

[8] Y. Zhang, H. Chen, and Z. Wu, "A Social Network-Based Trust Model for the Semantic Web," presented at the International Conference on Autonomic and Trusted Computing, Berlin, Heidelberg, 2006. Available: http://dx.doi.org/10.1007/11839569_18

[9] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl, and H.-W. Gellersen, "Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts," presented at the Proceedings of the 3rd international conference on Ubiquitous Computing, Atlanta, Georgia, USA, 2001.

[10] L. Atzori, A. Iera, and G. Morabito, "From "smart objects" to "social objects": The next evolutionary step of the internet of things," IEEE Communications Magazine, vol. 52, no. 1, pp. 97-105, 2014.

[11] Y. Hu, D. Wang, H. Zhong, and F. Wu, "SocialTrust: Enabling long-term social cooperation in peer-to-peer services," Springer Peer-to-Peer Networking and Applications, journal article vol. 7, no. 4, pp. 525-538, 2014.

[12] M. Nitti, R. Girau, L. Atzori, A. Lera, and G. Morabito, "A Subjective Model for Trustworthiness Evaluation in the Social Internet of Things," presented at the IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC, Australia, 2013.

[13] J. Zhan and X. Fang, "A novel trust computing system for social networks," presented at the IEEE Third International Conference onPrivacy, Security, Risk and Trust (PASSAT) and IEEE Third Inernational Conference on Social Computing (SocialCom), 2011.

[14] G. Yin, F. Jiang, S. Cheng, X. Li, and X. He, "Autrust: A practical trust measurement for adjacent users in social networks," presented at the Second International Conference on Cloud and Green Computing (CGC), 2012.

[15] S. Hauke, S. Biedermann, M. Mühlhäuser, and D. Heider, "On the Application of Supervised Machine Learning to Trustworthiness Assessment," in 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 2013, pp. 525-534.

[16] K. Zhao and L. Pan, "A Machine Learning Based Trust Evaluation Framework for Online Social Networks," in 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, 2014, pp. 69-74.

[17] S. Weihua and V. V. Phoha, "Neural network-based reputation model in a distributed system," in Proceedings. IEEE International Conference on e-Commerce Technology, 2004. CEC 2004., 2004, pp. 321-324.

[18] Y. Wang and J. Vassileva, "Bayesian network-based trust model," in Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), 2003, pp. 372-378.

[19] F. Jiang et al., "Deep Learning based Multi-channel intelligent attack detection for Data Security," IEEE Transactions on Sustainable Computing, vol. PP, no. 99, pp. 1-1, 2018.

[20] J. Shen, D. Liu, D. He, X. Huang, and Y. Xiang, "Algebraic Signatures-based Data Integrity Auditing for Efficient Data Dynamics in Cloud Computing," IEEE Transactions on Sustainable Computing, 2017.

[21] Y. Wang, Y.-C. Lu, I.-R. Chen, J.-H. Cho, A. Swami, and C.-T. Lu, "LogitTrust: A Logit Regression-based Trust Model for Mobile Ad Hoc Networks," presented at the Proceedings of the 6th ASE International Conference on Privacy, Security, Risk and Trust Cambridge, MA, 2014.

[22] Z. Li, X. Li, V. Narasimhan, A. Nayak, and I. Stojmenovic, "Autoregression Models for Trust Management in Wireless Ad Hoc Networks," presented at the IEEE Global Telecommunications Conference (GLOBECOM), Kathmandu, Nepal, 5-9 Dec. 2011, 2011.

[23] F. Boustanifar and Z. Movahedi, "A Trust-Based Offloading for Mobile M2M Communications," presented at the Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Toulouse, France, 2016.

[24] W. Li, W. Meng, L.-F. Kwok, and H. Horace, "Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model," Journal of Network and Computer Applications, vol. 77, pp. 135-145, 2017.

[25] A. Bolster and A. Marshall, "Analytical metric weight generation for multi-domain trust in autonomous underwater MANETs," presented at the IEEE Third Underwater Communications and Networking Conference (UComms), Lerici, Italy, 2016.

[26] L. Xin, G. Tredan, and A. Datta, "Metatrust: Discriminant analysis of local information for global trust assessment," in The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3, 2011, pp. 1071-1072: International Foundation for Autonomous Agents and Multiagent Systems.

[27] X. Liu, G. Tredan, and A. Datta, "A GENERIC TRUST FRAMEWORK FOR LARGE-SCALE OPEN SYSTEMS USING MACHINE LEARNING," Computational Intelligence, vol. 30, no. 4, pp. 700-721, 2014.

[28] J. Tang, H. Gao, X. Hu, and H. Liu, "Exploiting homophily effect for trust prediction," in Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 53-62: ACM.

[29] M. E. G. Moe, M. Tavakolifard, and S. J. Knapskog, "Learning trust in dynamic multiagent environments using HMMs," in Proceedings of the 13th Nordic Workshop on Secure IT Systems (NordSec 2008), 2008.

[30] E. ElSalamouny, V. Sassone, and M. Nielsen, "HMM-based trust model," in International Workshop on Formal Aspects in Security and Trust, 2009, pp. 21-35: Springer.

[31] Z. Malik, I. Akbar, and A. Bouguettaya, "Web services reputation assessment using a hidden markov model," in Service-Oriented Computing: Springer, 2009, pp. 576-591.

[32] M. E. Moe, B. E. Helvik, and S. J. Knapskog, "TSR: Trust-based secure MANET routing using HMMs," in Proceedings of the 4th ACM symposium on QoS and security for wireless and mobile networks, 2008, pp. 83-90: ACM.

[33] A. Josang and R. Ismail, "The beta reputation system," in Proceedings of the 15th bled electronic commerce conference, 2002, vol. 5, pp. 2502-2511.

[34] J. Zhang and R. Cohen, "Evaluating the trustworthiness of advice about seller agents in e-marketplaces: A personalized approach," Electronic Commerce Research and Applications, vol. 7, no. 3, pp. 330-340, 2008.

[35] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in Proceedings of the 2nd ACM conference on Electronic commerce, 2000, pp. 150-157: ACM.

[36] S. Liu, J. Zhang, C. Miao, Y.-L. Theng, and A. C. Kot, "iCLUB: An integrated clustering-based approach to improve the robustness of reputation systems," in The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3, 2011, pp. 1151-1152: International Foundation for Autonomous Agents and Multiagent Systems.

[37] Y. Wang, J. Zhang, and J. Vassileva, "A SUPER-AGENT-BASED FRAMEWORK FOR REPUTATION MANAGEMENT AND COMMUNITY FORMATION IN DECENTRALIZED SYSTEMS," Computational Intelligence, vol. 30, no. 4, pp. 722-751, 2014.

[38] W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou, and K. Nahrstedt, "A trust management framework for service-oriented environments," in Proceedings of the 18th international conference on World wide web, 2009, pp. 891-900: ACM.

[39] P. Laskov, C. Schäfer, I. Kotenko, and K.-R. Müller, "Intrusion detection in unlabeled data with quarter-sphere support vector machines," Praxis der Informationsverarbeitung und Kommunikation, vol. 27, no. 4, pp. 228-236, 2004.

[40] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection," in Proceedings of the 2003 SIAM International Conference on Data Mining, 2003, pp. 25-36: SIAM.

[41] W. Lee and S. J. Stolfo, "Data Mining Approaches for Intrusion Detection," in USENIX Security Symposium, 1998, pp. 79-93: San Antonio, TX.

[42] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," ACM transactions on Information and system security (TiSSEC), vol. 3, no. 4, pp. 227-261, 2000.

[43] M. V. Mahoney and P. K. Chan, "Learning nonstationary models of normal network traffic for detecting novel attacks," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 376-385: ACM.

[44] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on, 2002, vol. 2, pp. 1702-1707: IEEE.

[45] S. Mukkamala, A. H. Sung, and A. Abraham, "Intrusion detection using ensemble of soft computing paradigms," in Intelligent systems design and applications: Springer, 2003, pp. 239-248.

[46] M. Roesch, "Snort: Lightweight intrusion detection for networks," in Lisa, 1999, vol. 99, no. 1, pp. 229-238.

[47] V. Paxson, "Bro: a system for detecting network intruders in real-time," Computer networks, vol. 31, no. 23-24, pp. 2435-2463, 1999.

[48] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in Security and Privacy (SP), 2010 IEEE Symposium on, 2010, pp. 305-316: IEEE.

8. Case Study – Securing IoT based Applications

This section identifies various use cases for securing IoT based applications and investigate related case study or good practices from industries and academia for AI enabled applications.

The secure cyber-physical transportation system (CPTS) is an important case on the security of cyber physical systems, which is for secure IoT environment, and it is important to avoid traffic congestion and traffic accidents. Figure 15 show an example of

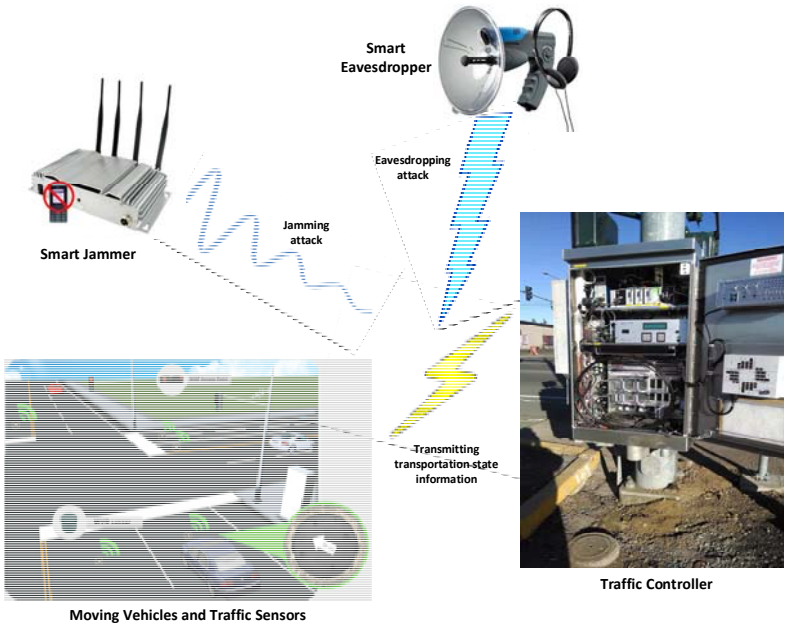


Figure 15. An example of cyber-physical transportation system.

CPTS combine the capabilities of sensing, control, communication and computing together, and in such systems, the sensors send the sensing information to traffic controllers through the open and wireless communication media. The open property of wireless communications makes it easy to be under the threat of security attacks, e.g., eavesdropping and jamming, as shown in Figure 15, and these security attacks will cause serious traffic accidents to result in deaths and economic loss. More importantly, the eavesdroppers and jammers are smart. They can adjust the power strategy to maximize the negative effects on the wireless communications, according to the feature of information transmission power. It is a challenge to actively extract and learn the features of malicious attacks, e.g., the above-mentioned feature: power strategy, which is dynamically changed with the change of the information transmission power. The wireless communications between sensors and controllers are open-mode. The open property makes the wireless communications easy to be under the threat of eavesdropping and jamming attacks. Moreover, the eavesdropper and jammer are smart in adjusting their power strategy to maximize the negative effects on the wireless communications.

Business continuity management (BCM) ensures continuation of an organization's business processes by utilizing data collected from the organization's IT and operation technology (OT) systems. The objective of BCM is to provide an advanced risk assessment processed from collected IT and OT data and to implement necessary measures to mitigate the impact to the organization's business processes, see Figure 16.

The BCM IoT platform will gather incident information from various security systems (i.e. IT systems) as well as planned and actual production data from production control systems (i.e. OT systems) utilizing sensor fusion technology. The BCM IoT platform will import threat intelligence information from other organizations to acquire insights into the situations of other interdependent systems and cohesive knowledge of the current and future attacks. The platform will analyse the incident information and perform risk analysis of the incident. It will also create security measures such as risk mitigation plans that will minimize the effect to the production activities. The BCM IoT platform will implement security measures, such as isolation of the affected subsystems or interruption of production lines. The IoT platform will analyse the production data to create an optimal production plan in response to affected capabilities of each production site.

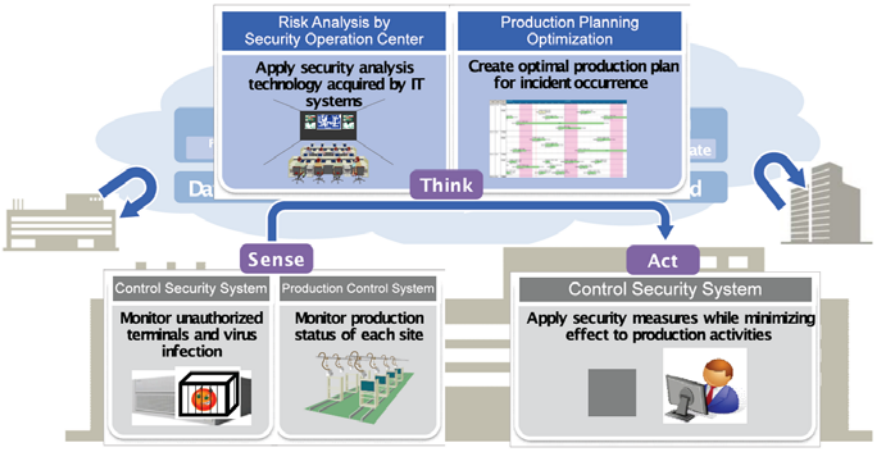


Figure 16. Industrial domain use case[66].

Smart City solutions bring together a number of heterogeneous IoT platforms that collectively contain a wide variety of sensors and data sources. These include temperature, humidity, noise, gas, and motion sensors, cameras, mobile devices, network sniffers, smart meters, water meters and a plethora of others devices that collectively monitor the dynamics of a city and optimize city operations while also enhancing citizen services. The Smart City IoT platform will transform the multi-modal sensing information from these various IoT platforms into cross-domain and real-time information mashups, using semantic interoperability. Advanced data mining and machine learning techniques will easily access a variety of different platforms and their operating environments to provide applications for residents and multiple agencies and enable intelligent actions. The Smart City platform of platforms will include real-time applications to enhance public safety, improve city mobility, optimize utility usage and enhance the plethora of citizen services that involve physical objects. Smart City platforms and the platform of platforms will rely heavily on smart and secure sensing to optimize services. They will use a combination of data from the public, private and personal sectors (anonymized as necessary) to seamlessly gain a more holistic view of the Smart City environment. They will use cross-domain communication techniques to bring together the disparate IoT systems deployed by individual agencies in a geopolitical entity and enable cross-domain cooperation and optimization, see Figure 17.

The Smart City platforms and platform of platforms will rely heavily on semantic disambiguation and contextualization of information to support advanced data processing and next-generation analytics that are developed for, and focused exclusively on, optimization of citizen services. These platforms

[66] IEC, "IoT 2020: smart and secure IoT platform," White Paper 2016.

will leverage advanced connectivity such as 5G and in-memory databases to move and process the vast amounts of data generated by the plethora of devices within their geopolitical boundary. The platforms will support edge-aware stream processing to handle processing topologies on parallel-networked systems. The Smart City platform of platforms and individual platforms will use next-generation technology to create smarter and more secure citizen quality of life environments.

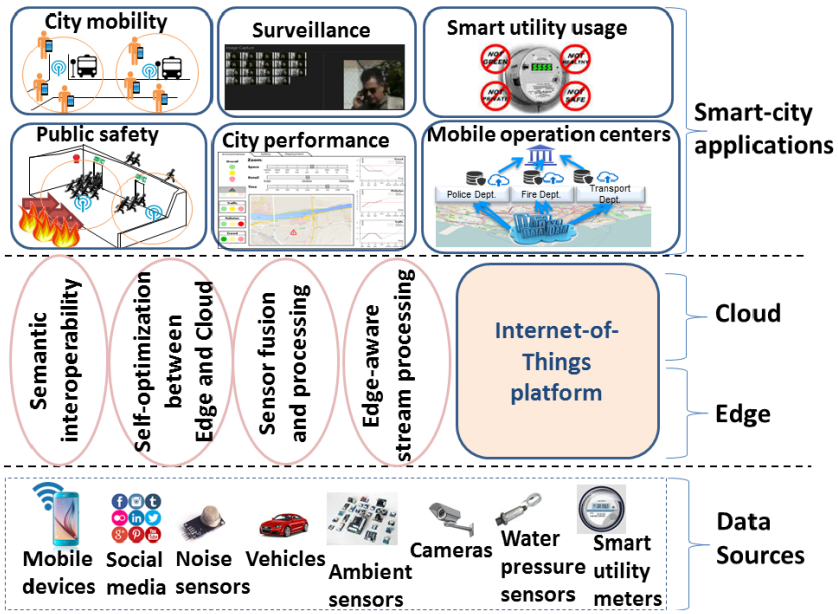


Figure 17. Smart city optimization[67].

[67] IEC, "IoT 2020: smart and secure IoT platform," White Paper 2016.

9. Challenges for Global Standardization

This section identifies the necessity of global standardization considering big technical trends towards the use of AI in various IoT applications while ensuring security and privacy. Then, it clearly identifies important standardization items and challenges to stimulate related activities in related standardization bodies.

For future standardization, it's necessary to investigate potential topics. Autonomous IoT with big data and AI and cyber security including trust with bloc-chain are very important topics. For IoT standardization, global interoperability for cross-border applications will be always important in support of various standardization bodies and fora in different sectors for vertical application domains and telecommunications. Furthermore, open source projects and alliances will be of interest. The most important technological trends is technology convergence among 5G for network connectivity, computing resources and platforms for Big data and AI while supporting security, privacy and trust including governance. It will ultimately target trustworthy autonomous IoT applications for supporting increasing intelligence like human brain.

As driving force for changes, there are three keyword: Data, Network and AI – i.e., DNA as shown in Figure 18. The important thing is to link between data and AI considering data from IoT and diffusion of AI. There are very important roles to support data-driven networking and services. Ultimately, standardization of trustworthy data-driven ICT combining data and AI. In fact, there are common features to support these technical requirements. However, rather than unlimited number of solutions, it is required to standard critical technologies globally with a fermentation and assembling approach.

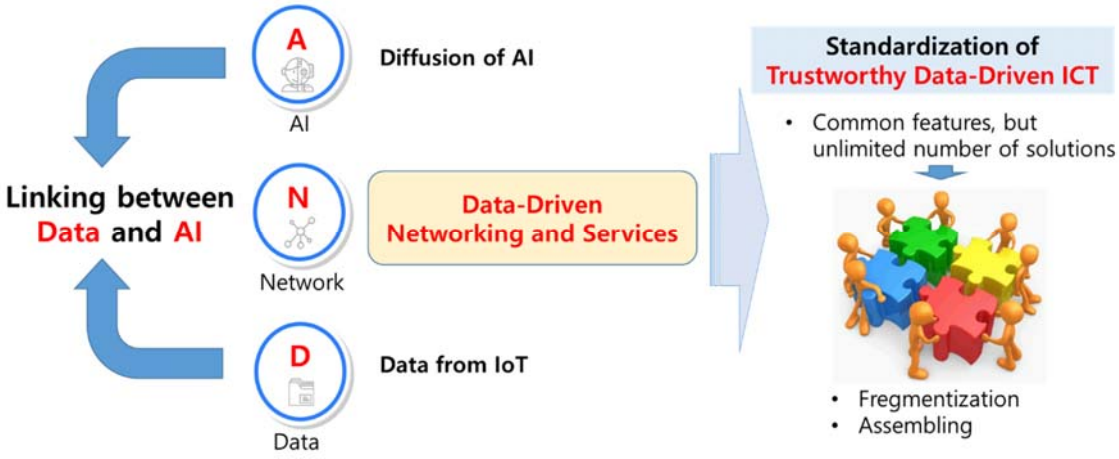


Figure 18. Future standardization towards trustworthy data-driven ICT.

Considering AI and IoT with security, privacy and trust, there are many potential items for future standardization as follows.

- DNA core technologies
  - A new networking paradigm – Data-driven networking
  - DNA platform – Technology convergence (IoT+Big Data+Cloud+AI)
  - Data-Information-Knowledge-Wisdom (DIKW) process
- Use of AI in ICT infrastructures and services (trustworthy autonomous ICT)

- Automotive control and management in networking and services
- Operational efficiency in Things + Processing + Communications + Storage
- Data-driven applications with AI (linking between data and AI)

- Security, privacy and trust including regulatory issues
  - Trust in DNA, particularly human-technology interface including social aspects

For trust technology, ITU-T Q16/13 made a preliminary version of roadmap for future trust standardization as shown in Figure 19. Based on the Data-Information-Knowledge concept, at the 1st stage, basic issues and key features on trust have been focused. Then, core technical solutions for trust provisioning from ICT infrastructures and services perspectives at the 2nd stage should be standardized. Finally, technology deployment as well as new services and business aspects on trust-based networks and eco-platforms are necessary. For this, there will be many related issues with other groups in ITU-T.

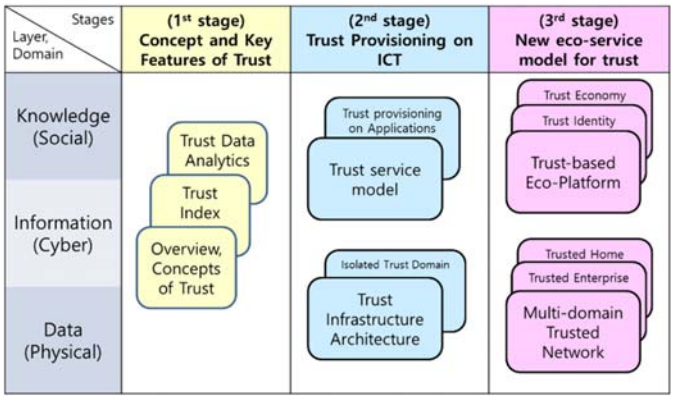


Figure 19. Roadmap for future trust standardization.

The social and economic value of data is mainly reaped during two moments: first when data is transformed into knowledge (gaining insights) and then when it is used for decision making (taking action). The knowledge is accumulated by individuals or systems through data analytics over time. So far data processing, management and interpretation for awareness and understanding have been considered as fundamental processes for obtaining the knowledge. As shown in Figure 20, trust is positioned throughout the whole process including belief between knowledge (i.e., awareness and understanding) and action.

- Trusted data collection and aggregation - Since an ICT convergence ecosystem is widely deployed and the number of data sources and types are dramatically increased, the trustworthiness of data itself comes to the fore. Data collection and aggregation should be trustworthy. However, as the vulnerability of the data sources, it is important to detect wrong data caused by various reasons. Since collection and aggregation of false data will cause damage and waste of system resources, trust metrics and models, which are able to use as criteria for checking trustworthiness, should be well managed to achieve trusted data collection and aggregation.

- Trusted data process and analysis

  Data processing means data manipulation such as filtering, fusion or mining. When the huge amounts of data are collected to a system, these data should be processed and analyzed in trustworthy ways. Data process and analysis mainly occurs in cyber domain (for example, utilizing cloud computing for Big-data analysis), however, it also can be done in physical domain as well as

social domain. Each domain has their own intelligence to process incoming data to create new useful information. This information is usually propagated to different entities and domains, so there are some ways to check whether given data process and analysis mechanism is trustworthy or not. Measurable trust value should be defined to analyze trust of entities, and it is also important to find appropriate trust evaluation mechanisms for analyzing trust values for a specific domain. Trust bootstrapping is also an important issue. When a new entity is joined to an existing system, the system is able to score trust values for a newcomer.

- Trustworthy decision-making, action and data dissemination - ach entity makes their decisions, takes necessary actions and disseminates data based on data process and analysis. Also, each domain has different criteria for making a decision and taking an action, so different trust mechanism should be considered on each step. The result of decision-making or action can be disseminated among physical, cyber and social domains. In the dissemination stage, the feature of data should be considered; for example, some data can be related with user privacy, or they can have urgent information. These kinds of data characteristics should be considered for trustworthy data dissemination.

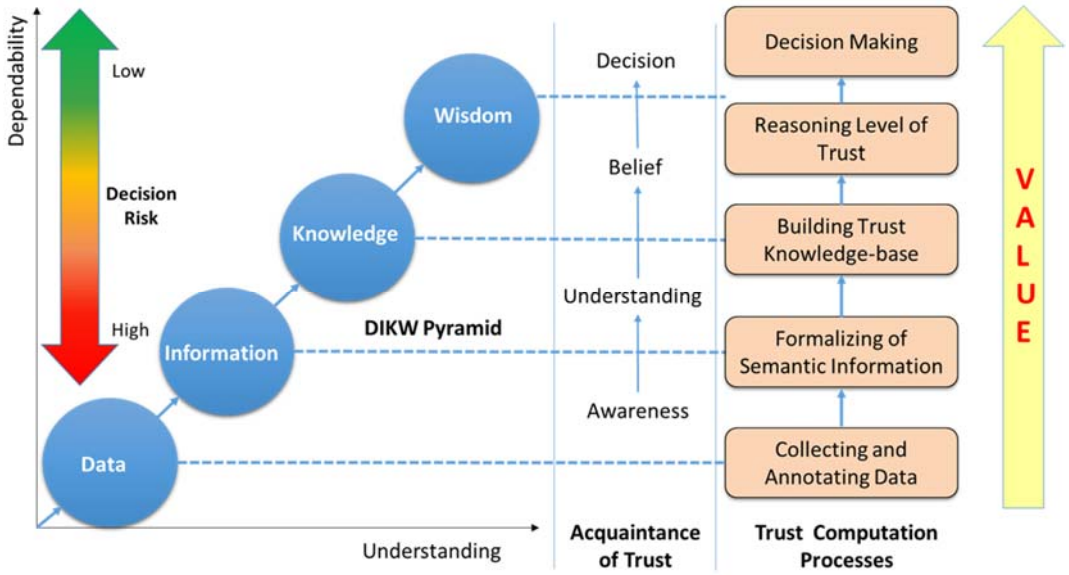


Figure 20. Trusted computing from raw data to actionable knowledge.

Recently decentralization is one of the emerging technologies that implements proof of trust using block-chains but applying it to overall holistic network security is yet to be studied thoroughly. Assuming a trustworthy model is in place, several functional benefits become apparent. For example, domain-based trust can be represented as a function of time and credibility score from number of other domains. This is computed by verification of correctness and that data remained unimpaired while transiting a domain. New routing paradigms (and corresponding business incentives) could emerge as now domains may choose to transit through systems that are better in trust-scores. A side effect of such approach will be finding a balance of cost and security within the network[68]. As one of

[68] Huawei, "Internet 2030, Towards a new Internet for the Year 2020 and beyond," 2018

tools to support trust, block-chain should be considered as an important standardization item taking into account the 4th industrial revolution.

10. Considerations for promoting safe use of IoT based application with AI

This section presents required actions to promote safe use of IoT based application with AI. It presents considerations for data protection, ensuring privacy, ethical, policy and regulatory issues.

10.1 Data protection, privacy and ethical considerations

The key characteristics of big data analytics still represent a step change in the processing of personal data. The analysis of big data using techniques made possible by AI creates implications for data protection, and it can be more challenging to apply the data protection principles when using personal data in a big data context. These implications arise not only from the volume of the data but from the ways in which it is generated, the propensity to find new uses for it, the complexity of the processing and the possibility of unexpected consequences for individuals.

Article 8 of the EU's Charter of Fundamental Rights enshrines data protection as a fundamental right, which reinforces the necessity to ensure the implementation of high standards for data protection, privacy and information security.

ITU-T Focus Group on Data Processing and Management (FG-DPM) is developing deliverables on data security, privacy and trust issues including governance. The deliverable on framework of security and privacy presents key considerations of data protection[69]. Data protection within the context of smart cities can be seen from different points of view. First of all, one has to bear in mind that when we look at smart cities they are subject to different legislations in many parts of the world and therefore their regulation is not uniform.

If we take into account the European situation, starting from an user-centred point of view, and taking into account the GDPR, we can derive important principles of significant importance for smart cities[70]. The GDPR enshrines a set of fundamental principles and norms that are always to be taken into account in the context of smart cities. Among them: lawfulness; fairness; transparency; purpose limitation; Data minimisation; accuracy; storage limitation; integrity and accountability. The GDPR also provides detailed norms for the collection of consent. The GDPR is more prescriptive when it comes to the conditions for consent, however the new rules transpose into law what was already required by certain supervisory authorities. According to article 4(11) of the GDPR consent means any "freely given, specific, informed, and unambiguous indication of the data subject's wishes by which he or she, by statement or by clear affirmative action, signifies agreement to the processing of personal data relating to him or her". The GDPR also details the requirement for the processing of personal data of underage persons and processing of special categories of data. The GDPR sets out obligations provides for obligations towards the facilitation of the exercise of the data subject's right to information such as access to personal data, rectification and erasure, right to data portability. The legal provisions also enable the data subject to restrict processing of his data under certain circumstances and detail processes for objection and seeks to protect the individual vis-à-vis automated decision making mechanisms.

The GDPR will require i) the use of Privacy by design (PbD), privacy by default and the use of the Privacy Impact Assessment (PIA) in the design and management of ICT solutions using personal data, and ii) the appointment of data protection officers (DPO).

Many researchers have argued that, by way of an "intelligence explosion" sometime in the 21st century, a self-improving AI could become so vastly more powerful than humans that we would not be able to stop it from achieving its goals. General super-intelligence would be capable of independent initiative and of making its own plans, and may therefore be more appropriately

[69] ITU-T FG-DPM, "Draft Technical Report – Framework of security and privacy in DPM," DPM-O-067.

[70] For a general scenario of legal framework see [Create-IoT-17].

thought of as an autonomous agent. Since artificial intellects need not share our human motivational tendencies, it would be up to the designers of the super-intelligence to specify its original motivations. In theory, a super-intelligent AI would be able to bring about almost any possible outcome and to thwart any attempt to prevent the implementation of its top goal, many uncontrolled unintended consequences could arise. It could kill off all other agents, persuade them to change their behaviour, or block their attempts at interference[71].

10.2 Technical considerations and challenges

There are several tools and approaches, including anonymization, PIAs and privacy by design, that can help organisations to ensure their processing complies with data protection legislation and minimises the impact on privacy. It needs the trend towards organisations developing their own ethical principles and building relationships of trust with the public, because putting this into practice will assist compliance with data protection requirements. Recent moves towards setting up 'councils of ethics', within organisations and nationally, are a positive development that should also support this.

10.2.1 Privacy by Design and Privacy by Default

Usually it is not the technology that increases the risks for privacy, data protection and security, but the way it is developed and applied. The negative consequences of this practice for privacy, data protection and security will significantly increase, if applied to IoT systems. Therefore mechanisms are needed to ensure that no unwanted processing of personal data takes place and that individuals are informed of the processing, its purposes, the identity of the processor and of how to exercise their rights. At the same time processors need to comply with the data protection principles as data minimisation, purpose limitation etc., which might be especially challenging in an IoT environment, where automatic communication without human intervention between objects and between objects and persons is at the core of the system.

Definition - Privacy-by-Default

The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons (art. 25.2 in GDPR).

Definition - Privacy-by-Design

Institutionalisation of the concepts of privacy in organisations and integration of these concepts in the design, and life cycle of systems.

The GDPR defines it at a methodology according to which: "Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymization, which are designed to implement data-protection principles, such as data minimisation in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and to protect the rights of data subjects (art. 25.1 in GDPR).

[71] https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence

On a technical level an appropriate process to protect personal data should be defined. For example, privacy policies (at the user level) that can be pushed/built inside the objects (and translated according to the technical specificities of each object) and finally be enforced by the object with appropriate mechanisms to ensure data protection. The technical challenge is however, to enable objects with limited processing power and / or memory to receive and respect such policies.

The possibilities for individuals to exercise their data subject rights need to be enhanced. It needs to be ensured that clear, easily understandable information on the data processing of IoT systems, their objects, functions and purposes, is provided to individuals. Mechanisms need to be found to make individuals aware of the processing and to provide information on the processor, the purpose of the processing and possibilities to exercise data subject rights, as most IoT applications are expected to operate in the background, invisible to and unrecognised by the individual.

Information on how to build privacy-friendly applications needs to be provided to IT engineers, system designers and standardisation bodies, to ensure that the concepts of Privacy by design and Privacy by default settings get implemented in practice. Data protection officers might have an important role here, provided that defined minimum requirements ensure that they themselves are sufficiently trained.

Challenges addressed with this option include i) design considerations for IoT technologies, ii) the risks are context-aware and situational, iii) traceability / profiling / unlawful processing, iv) exercising data protection rights for individuals and compliance with DP legislation for organisations, and v) loss of user control / difficulty in making decisions.

10.2.2 Human/technology interface considering ethics

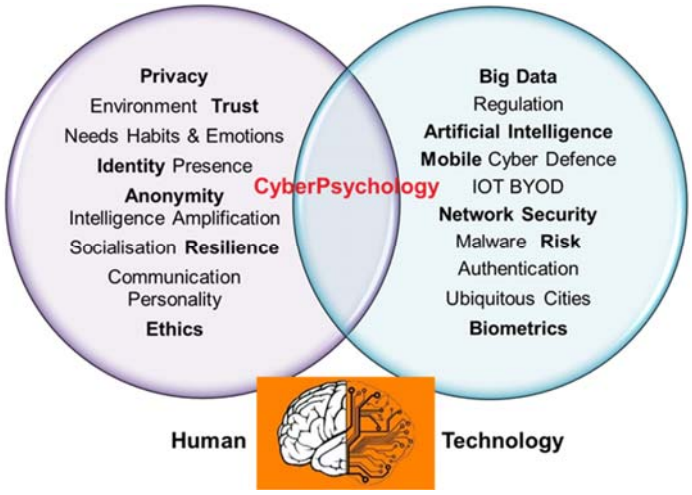


Figure 21. Deliberating insight at the human/technology interface[72].

Beyond conventional security solutions, it's very important to consider transparency, data protection, privacy preserving, policy and regulatory issues including ethics.

The biggest threat to network security is company employees and firms need to adapt management structures to deal with leaderless environments says a leader in sphere of what is now known as cyber-

[72] Source: Mary Aiken, "The CyberPsychology of Cyber Security" RCSI

psychology (see Figure 21), a world with no authority for authority, leaving it very difficult to establish robust controls. Worse still, this was a world where people assume that they do not need leaders and everything should be done by consent. And one in which technology was evolving much fast than the required corporate psychology required to cope best with this change, meaning that there was tension between traditional and leaderless corporations.

An ethics point of view, the sheer complexity of human value systems makes it very difficult to make AI's motivations human-friendly. Unless moral philosophy provides us with a flawless ethical theory, an AI's utility function could allow for many potentially harmful scenarios that conform to a given ethical framework but not "common sense". It needs an AI design that avoids several types of unintended AI behaviour including self-delusion, unintended instrumental actions, and corruption of the reward generator.

10.2.3 Policy and regulatory considerations

Main goals of any regulatory approach should therefore be i) to ensure full compliance of IoT technology and applications with the Charter of Fundamental Rights and ii) to minimise potential barriers for the adoption of IoT technology in order to benefit of it's full economic potential.

Lack of trust makes consumers hesitate to buy online and adopt new services, including public e-government services. If not addressed, this lack of confidence will continue to slow down the development of innovative uses of new technologies, to act as an obstacle to economic growth and to block the public sector from reaping the potential benefits of digitisation of its services, e.g. in more efficient and less resource intensive provisions of services. This is why data protection plays a central role in the Digital Agenda for Europe, and more generally in the Europe 2020 Strategy.

Standards could provide presumption of conformity with the legal requirements and could be used for certification; in addition, they could provide definitions for clear information to individuals on how to exercise their data protection rights, mindful that clear information is a prerequisite for informed consent.

On the other hand standards are voluntary and non-binding. Therefore the tool might possibly be too “weak” for the intended outcome. Other regulatory measures which are more binding might also be needed. A recommendation could be envisaged and comitology procedures could be launched to build a European privacy risk management tool (methodology + best practices).

According to the GDPR, which is probably the most stringent legislation at the global level, many cities will need to comply with legal accountability obligations to European Data protection law. As data controllers, cities will be required to implement appropriate technical and organizational measures to ensure and be able to demonstrate that data processing is performed in accordance with the GDPR, and review and update those measures when necessary. In each case cities will be called to evaluate which measures will be appropriate. This will depend on the nature, scope, context and purpose of the processing and also the risks for rights and freedoms of individuals. Regulators around the word are also embracing the concept of accountability as a key principle in data process management (privacy regulators in Canada, Hong Kong, Australia have issued “Accountability Guides” or “Privacy Governance Frameworks” in order to assist the private sector).

Basic frameworks to deal with IoT solutions, also in the context of smart cities, have been recently elaborated. Hyper-connectivity in fact entails a great deal of risks and calls for the development of a proper legal framework. The regulatory ecosystem of course will vary according to the different kind of activities in place. For what may concern data protection and privacy the Article 29 Working party has specifically raised the issue of privacy and security issues raised by the IoT.

10.2.4 Risk management

Data processing and management in Smart Cities and IoT generate uncertainty and may constitute a threat for the stability of the system, the efficiency of the operations, or the wealth, the health or the dignity of the persons. Thus, it is the responsibility of the ecosystem to measure the risks linked to data, and to manage it.

Different methodologies/frameworks already exist to manage the risks. We can cite for example Coso Enterprise Risk Management[73] (Committee of Sponsoring Organizations of the Treadway Commission). We can also cite the Operational Risk Framework as sponsored by the Basel Committee on Banking Supervision[74]. Even if Coso emphasizes on Governance and Communication, while Basel standard emphasizes on calculation accuracy and the losses modelization, both methodologies have the same canvas to manage the risks linked to data processing and management in Smart Cities and IoT[75].

[73] https://www.coso.org/Documents/2017-COSO-ERM-Integrating-with-Strategy-and-Performance-Executive-Summary.pdf

[74] https://www.bis.org/publ/bcbs195.pdf

[75] ITU-T FG-DPM, "Draft technical specification on risk management in DPM for IoT and smart cities," FG-DPM-O-64.

11.Conclusion

This report has examined the relevance of AI in the current and future development of IoT and how security should be addressed, including data protection and privacy as follows.

- It has provided key technical trends for digital transformation from literature and highlight the important of AI and IoT technology. It has also addressed potential risks and threats while the number of new technologies continues to grow.
- It has provided basic concepts and background of AI and Machine Learning including general application areas of AI.
- It has discussed IoT and security issues, including a security framework in CPS in order to understand the overall features and technical issues for secure IoT environment.
- It has surveyed security in IoT and AI, AI-based privacy mechanism for personal data, and AI-based trust mechanisms.
- It has identified the necessity of global standardization considering big technical trends towards the use of AI in various IoT applications while ensuring security and privacy. Then, it has also identified important standardization items and challenges to stimulate related activities in related standardization bodies.
- It has presented required actions to promote safe use of IoT based application with AI. It presents considerations for data protection, ensuring privacy, ethical, policy and regulatory issues.

In the area of security, privacy and trust research in IoT, AI plays an important role. Recently, the increasing amount of data as well as the rich metadata brought by large-scale Web applications (e.g., social media, e-commerce, recommender systems) has led to a new trend of applying formerly unutilized machine learning methodologies, such as deep learning, to more precisely model trust. In this report, we have discussed potential benefit of integrating machine learning and other AI technologies with trust and security concepts and its crucial role in data-driven applications, service composition, social networking, recommendation systems and security aspects. Further, it is observed that there is an emerging trend in scientific community in developing many research activities at the junction of trust and AI techniques. To understand the prevailing issues related to trust, and provide more smart solutions, it is essential to bring researchers in these two communities closer to each other.

In the context of the IoT, standards should be elaborated specifically on the aspects of good application design, user application interfaces and on tools for individuals to play their part in security, data protection and privacy for the IoT for people. In order to make Privacy by design and Privacy by default a reality, the consideration of data protection requirements should become a mandatory design goal in standardisation, as standards can serve as a multiplier for privacy friendly application design.

Setting the Scene for 5G: Opportunities & Challenges

Setting the Scene for 5G: Opportunities & Challenges

2018

Acknowledgments

This report was prepared under the supervision of Mr Kemal Husenovic Chief of ITU/BDT Infrastructure, Enabling Environment, and E-Applications Department with the contribution of Mr Iqbal Bedi, Intelligens Consulting, under the direction of Ms Sofie Maddens, Head of the ITU/BDT Regulatory and Market Environment Division, in collaboration with the ITU/BDT Telecommunication Technologies and Network Division.

Significant contribution was also provided by the ITU Telecommunication Standardization Bureau (TSB) and the ITU Radiocommunication Bureau (BR). The ITU Telecommunication Development Bureau (BDT) team included István Bozsóki, Desiré Karyabwite and Nancy Sundberg. The BR team included Mario Maniewicz, Philippe Aubineau, Sergio Buonomo, Joaquin Restrepo, Diana Tomimura and Nikolai Vassiliev. The TSB team included Bilel Jamoussi. Martin Adolph, Denis Andreev, Cristina Bueti, Tatiana Kurakova and Hiroshi Ota.

The views expressed in this report are those of the author and do not necessarily reflect the opinions of ITU or its Membership.

ISBN

978-92-61-27581-5 (Paper version)

978-92-61-27591-4 (Electronic version)

978-92-61-27601-0 (EPUB version)

978-92-61-27611-9 (Mobi version)

Please consider the environment before printing this report.

Foreword

I am pleased to present this report on Setting the Scene for 5G: Opportunities & Challenges, prepared in collaboration with the ITU's Standardization and Radiocommunication Bureaux. The report sets out the landscape for ICT policy-makers, national regulatory authorities (NRAs) and operators as 5G technologies move ever closer. 5G has the potential to be transformative for citizens, businesses, governments and economies. Investment is key but there are many factors to take into account before investments can be committed.

This report helps navigate through the issues related to 5G and provides a measured, practical approach to policy-makers looking to make important investment decisions in the months and years ahead. The sixteen key issues presented are essential reading and constitute a powerful embarkation point as we set out to grapple with 5G opportunities and challenges.

This report further helps demystify the hype surrounding 5G, recognizes the great potential of 5G and makes a host of recommendations designed to help policy-makers, regulators and operators work together effectively, both to meet the challenges and to benefit from the many opportunities that this new technology represents. It also recommends that policy-makers improve availability and quality of 4G networks until the case for 5G networks becomes clearer and more compelling.

I thank my colleagues Mr Chaesub Lee, Director of ITU Telecommunication Standardization Bureau, and Mr François Rancy, Director of ITU Radiocommunication Bureau, for their invaluable contribution in making this report a useful ITU tool to provide guidance as policy-makers and NRAs seek to harness the benefits of the digital economy.

Brahima Sanou
Director, Telecommunication Development Bureau, ITU

Table of Contents

List of Tables, Figures and Boxes

Figures

Boxes

Abbreviations and acronyms

Various abbreviations and acronyms are used through the document; they are provided here for simplicity.

Abbreviation/acronym	Description
2G, 3G, 4G, 5G*	Refers to different generations of mobile standards
5GIA	5G Infrastructure Association
AI	Artificial Intelligence
AV	Autonomous Vehicle
BEREC	Body of European Regulators for Electronic Communications
BNG	Broadband Network Gateway
CAGR	Compound Annual Growth Rate
CAV	Connected Autonomous Vehicle
CCTV	Closed-Circuit TV
CPRI	Common Public Radio Interface
C-RAN	Cloud/centralized Radio Access Network
DAN	Data Aware Networking
EC	European Commission
EMBB	Enhanced Mobile Broadband
EMC	Electromagnetic Compatibility
EMF	Electromagnetic Field
EU	European Union
FCC	Federal Communications Commission
FG ML5G	Focus Group on Machine Learning for Future Networks, including 5G
FMC	Fixed Mobile Convergence
FTTH	Fibre to the Home
FTTP	Fibre to the Premise
FUTEBOL	Federated Union of Telecommunications Research Facilities for an EU-Brazil Open Laboratory
GSMA	The GSM Association
HAPS	High Altitude Platform Systems
ICNIRP	International Commission on Non-Ionizing Radiation Protection

* https://www.itu.int/en/publications/Documents/tsb/2017-IMT2020-deliverables/mobile/index.html#p=1

Abbreviation/acronym	Description
ICT	Information and Communications Technology
IMT-2020	International Mobile Telecommunication 2020 standards
IoT	Internet of Things
ITU	International Telecommunication Union
ITU-BDT	ITU Telecommunication Development Bureau
ITU-D	ITU Development Sector
ITU-R	ITU Radiocommunication Sector
ITU-T	ITU Standardization Sector
LTE-A Pro	Long-Term Evolution Advanced Pro
MCS	Mission Critical Services
MER	Main Equipment Room
MIMO	Multiple Input, Multiple Output
MIoT	Massive Internet of Things
mmWave	Millimeter Wave
MMTC	Massive Machine-type Communications
NFV	Network Function Virtualization
NISA	National Information Society Agency (Korea Rep. of)
NRA	National Regulatory Authority
OTN	Optical Transport Network
PMP	Point-to-Multipoint
PPP	Public–Private Partnership
Q.NBN	Qatar National Broadband Network
RAN	Radio Access Network
RF EMF	Radio Frequency Electromagnetic Field
ROF	Radio-over-Fibre
RRU	Remote Radio Unit
SDN	Software-Defined Networking
SMP	Significant Market Power
TCO	Total Cost of Ownership
TIM	Telecom Italia Mobile

* https://www.itu.int/en/publications/Documents/tsb/2017-IMT2020-deliverables/mobile/index.html#p=1

Abbreviation/acronym	Description
URLLC	Ultra-Reliable and Low-Latency Communication
WHO	World Health Organization
WLAN	Wireless Local Arean Network
WRC-19	World Radiocommunication Conference 2019
WTDC	World Telecommunication Development Conference

* https://www.itu.int/en/publications/Documents/tsb/2017-IMT2020-deliverables/mobile/index.html#p=1

Unless otherwise stated, policy-makers refers to NRAs, local (municipal or state) or national (federal) government agencies.

Executive Summary

Expectations of 5G are high, with many assuming it will deliver a transformative promised land – an improved end-user experience, new applications, new business models and new services riding swiftly on the back of gigabit speeds, improved network performance and reliability. 5G networks and services, standing as they do on the shoulders of successful 2G, 3G and 4G mobile networks, are forecast by independent economic studies to deliver very significant economic gains.

Caution: high levels of investment needed

However, despite the potential benefits, there is concern that 5G is premature and notes of caution are being sounded. Operators are sceptical about the commercial case given the high-levels of investment needed to deploy 5G networks.[1] The report estimates the cost to deploy a small cell-ready 5G network – assuming fibre backhaul is commercially feasible – can range from USD 6.8 million for a small city to USD 55.5 million for a large, dense city.

Danger of increasing digital divide

A viable case for investment in 5G can be made for densely populated urban areas – always the most commercially attractive regions for operators. More challenging will be a commercial argument for investing in 5G networks outside such areas, especially in the early years of 5G deployment. As a result, rural and suburban areas are less likely to enjoy 5G investment, and this will potentially widen the digital divide.

Balanced view is needed

As long as the investment case for 5G remains uncertain, industry and policy-makers should remain cautious and should consider enhancing the availability and quality of existing 4G networks in the run up to 5G. The need for 5G is not immediate. Policy-makers and operators should only consider deploying 5G networks where there is demand or a robust commercial case in favour of doing so.

Policy-makers' actions will make a difference

Where demand exists alongside high 5G deployment costs, policy-makers can use a range of legal and regulatory actions to facilitate 5G network deployment. These include:

- Supporting the use of affordable wireless coverage (e.g. through sub-1 GHz bands) to reduce the digital divide;
- Commercial incentives such as grants, or PPPs to stimulate investment in 5G networks.

5G: 16 key issues for policy-makers to consider

This report highlights 16 key issues – and responses – for policy-makers to consider as they formulate strategies to stimulate investment in 5G networks. Together they represent powerful means of calibrating an overall approach across major aspects of migration and, where appropriate, embarking on a judiciously facilitated, accelerated transition to 5G.

[1] https://www.techradar.com/news/eu-backed-groups-warns-about-5g-claims

Key issues for consideration

No.	Summary	For consideration…
1)	Investment case	Policy-makers may consider undertaking their own independent economic assessment of the commercial viability of deploying 5G networks
2)	4G network strategy	Until the case for 5G networks can be clearly made, policy makers may consider enhancing the availability of and boosting the quality of 4G networks
3)	Harmonize spectrum	NRAs may consider allocating/assigning globally harmonized 5G spectrum bands
4)	Spectrum roadmap	NRAs may consider adopting a spectrum roadmap and a predictable renewal process
5)	Spectrum sharing	NRAs may consider allowing sharing to maximize efficient use of available spectrum, particularly to benefit rural areas
6)	Spectrum pricing	NRAs may consider selecting spectrum award procedures that favour investment
7)	700Mhz spectrum	Policy-makers may consider supporting the use of affordable wireless coverage (e.g. through the 700 MHz band) to reduce the risk of digital divide
8)	Fibre investment incentives	Policy-makers, where the market has failed, may consider stimulating fibre investment and passive assets through PPPs, investment funds and the offering of grant funding, etc.
9)	Fibre tax	Policy-makers may consider removing any tax burdens associated with deploying fibre networks to reduce the associated costs
10)	Copper migration to fibre	Policy-makers may consider adopting policies/financial incentives to encourage migration from copper to fibre and stimulate deployment of fibre
11)	Wireless backhaul	Operators may consider a portfolio of wireless technologies for 5G backhaul in addition to fibre, including point-to-multipoint (PMP), microwave and millimeter wave (mmWave) radio relays, high altitude platform systems (HAPS) and satellites
12)	Access/sharing of passive infrastructure	Policy makers may consider allowing access to government-owned infrastructure such as utility poles, traffic lights and lampposts to give wireless operators the appropriate rights to deploy electronic small cell apparatus to street furniture NRAs may consider continuing to elaborate existing duct access regimes to encompass 5G networks allowing affordable fibre deployments
13)	Access costs	Policy-makers/NRAs may consider ensuring reasonable fees are charged to operators to deploy small-cell radio equipment onto street furniture
14)	Asset database	Policy-makers may considerholding a central database identifying key contacts, showing assets such as utility ducts, fibre networks, CCTV posts, lampposts, etc. This will help operators cost and plan their infrastructure deployment more accurately
15)	Wayleave (rights of way) agreements	Policy-makers may agree upon standardized wayleave agreements to reduce cost and time to deploy fibre and wireless networks
16)	5G test beds	Policy-makers may consider encouraging 5G pilots and test beds to test 5G technologies, and use cases, and to stimulate market engagement

1 Introduction

The ITU has highlighted 5G networks and artificial intelligence (AI) as fields of innovation necessary for enabling smarter societies. 5G is the next generation of mobile standards and promises to deliver improved end-user experience by offering new applications and services through gigabit speeds, and significantly improved performance and reliability. 5G networks are expected to be enhanced with AI to make sense of data, manage and orchestrate network resources and to provide intelligence to connected and autonomous systems.

To this end, the ITU is developing "IMT for 2020 and beyond", setting the stage for 5G research activities emerging around the world. The ITU has also established the Focus Group on Machine Learning for Future Networks, including 5G (FG ML5G).[1] This Focus Group is studying the use cases, services, requirements, interfaces, protocols, algorithms, ML-aware network architecture and data formats.

This report has been prepared as part of the overall framework of AI reports to help governments, information and communications technology (ICT) regulators or national regulatory authorities (NRAs) prepare for AI and 5G digital transformation.

This report reviews expectations of 5G and examines the infrastructure and investment requirements on the private and public sectors as they prepare for 5G. It is designed to support emerging use cases and services, and to help all sectors meet the expected performance (gigabit data rates), low latency and high reliability requirements of these services, ensuring that end users reap in full the economic benefit that 5G is expected to offer.

In addition, the report looks at the transition strategies used by wireless operators to upgrade 4G networks to 5G – particularly in urban areas where 5G rollouts are likely to be prioritized – and the various political, strategic and tactical challenges that can hold back deployment of 5G networks. While significant steps are being taken towards 5G in developed economies, consideration is also given to the challenges that will be faced by wireless operators in less developed economies.

Also included in this report is a high-level cost model to estimate the potential capital investment required by a wireless operator to upgrade to a 5G network and the potential models that can be used by NRAs to incentivize investment in 5G. Finally, based on interviews with operators and supplemented by secondary research, the report draws on real examples of the role policy-makers can play as facilitator, enabler and coordinator in preparing for 5G development, to speed up deployment and reduce the cost of deployment.

The remainder of this document is structured as follows:

- Section 2 examines 5G, its evolution and what it can deliver over and above existing wireless technologies, including economic and wider societal benefits.
- Section 3 explains 5G spectrum requirements and the technologies to support 5G networks and how operators are expected to evolve to 5G networks.
- Section 4 describes the key challenges of rolling out 5G networks from an infrastructure and spectrum policy perspective.
- Section 5 provides examples of how policy-makers are starting to work through the issues associated with deploying 5G networks.
- Section 6 explores the investment requirements of developing 5G networks and potential approaches to incentivizing investment in them.

[1] https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx

- Section 7 recommends actions for policy-makers in NRAs and governments, helping them simplify and reduce costs as they move towards implementation.

2 5G overview

This section introduces the role of the ITU in developing 5G standards as well as the potential benefits that 5G can generate. While the ecosystem is not fully developed, 5G may not yet be an appropriate consideration across all regions. In addition, there is some concern that the initial deployment of 5G in dense urban areas may increase the digital divide.

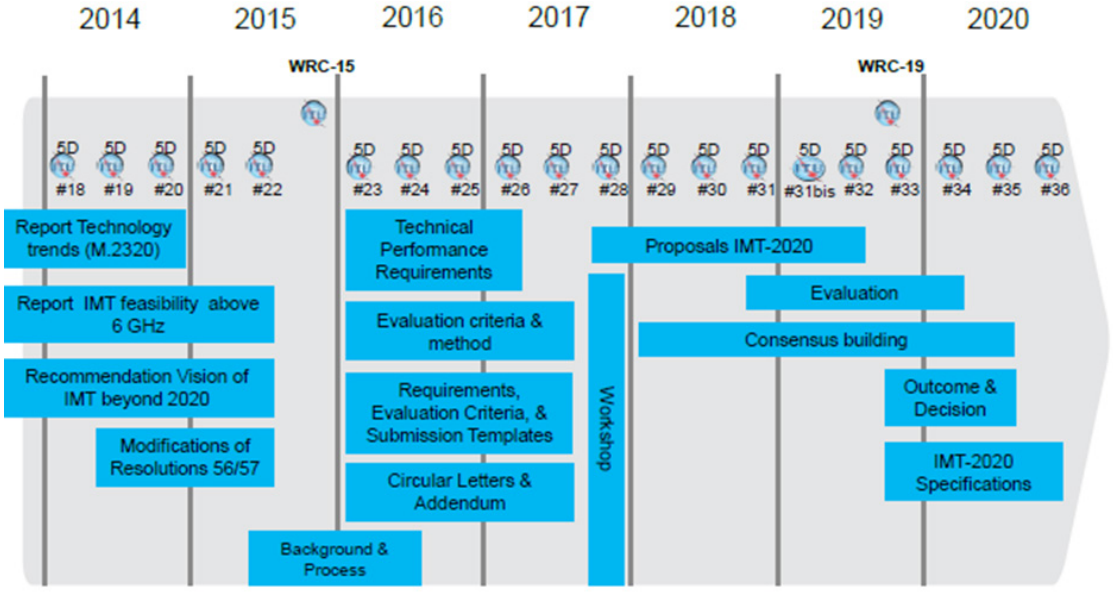
2.1 The role of the ITU

5G is the next generation of mobile standards being defined by the ITU. IMT-2020 (5G) is a name for the systems, components, and related elements that support enhanced capabilities beyond those offered by IMT-2000 (3G) and IMT-Advanced (4G) systems.

International Mobile Telecommunication 2020 standards (IMT-2020):

- Set the stage for 5G research activities that are emerging around the world
- Define the framework and overall objectives of the 5G standardization process
- Set out the roadmap to guide this process to its conclusion by 2020 (see Figure 1).

Figure 1: Detailed timeline and process for ITU-R IMT-2020



Annex A provides an overview of the work the ITU has undertaken on 5G.

2.2 What is 5G?

At the highest level, 5G is an opportunity for policy-makers to empower citizens and businesses. 5G will play a key role in supporting governments and policy-makers in transforming their cities into smart cities, allowing citizens and communities to realize and participate in the socio-economic benefits delivered by an advanced, data-intensive, digital economy.

5G promises to deliver improved end-user experience by offering new applications and services through gigabit speeds, and significantly improved performance and reliability. 5G will build on the successes of 2G, 3G and 4G mobile networks, which have transformed societies, supporting new services and new business models. 5G provides an opportunity for wireless operators to move beyond providing connectivity services, to developing rich solutions and services for consumers and

industry across a range of sectors – and at affordable cost. 5G is an opportunity to implement wired and wireless converged networks, and offers in particular opportunities in integrating network management systems.

Commercial 5G networks are expected to start deployment after 2020, as shown in Figure 2, as 5G standards are finalized.[1] By 2025, the GSM Association (GSMA) expects 5G connections to reach 1.1 billion, some 12 per cent of total mobile connections. It also forecasts overall operator revenues to grow at a CAGR of 2.5 per cent, to reach USD 1.3 trillion by 2025.[2]

5G is also expected to increase data rates dramatically and reduce latency compared to 3G and 4G. 5G is expected to significantly reduce latency to below 1ms, suited to mission-critical services where data are time-sensitive. Its high-speed capability means 5G networks can provide a range of high-speed broadband services and offer an alternative to last-mile access such as FTTH or copper connections.

Box 1: Role of IMT 2020 (5G) and beyond

The framework of the future development of IMT for 2020 and beyond is described in detail in ITU-Recommendation M.2083-0. This states that IMT systems should continue to contribute to the following:

- Wireless infrastructure to connect the world: Broadband connectivity will acquire the same level of importance as access to electricity. IMT will continue to play an important role in this context as it will act as a key pillar in enabling mobile service delivery and information exchange. In the future, private and professional users will be provided with a wide variety of applications and services, ranging from infotainment services to new industrial and professional applications.
- New ICT market: The development of future IMT systems is expected to promote the emergence of an integrated ICT industry which in turn drives economies around the globe. Some possible areas include: the accumulation, aggregation and analysis of big data; the delivery of customized networking services for enterprise and social network groups on wireless networks.
- Bridging the digital divide: IMT will continue to help closing the gaps caused by an increasing digital divide. Affordable, sustainable and easy-to-deploy mobile and wireless communication systems can support this objective while effectively saving energy and maximizing efficiency.
- New ways of communication: IMT will enable sharing of any type of content anytime, anywhere and through any device. Users will generate more content and share this content without being limited by time and location.
- New forms of education: IMT can change methods of education by providing easy access to digital textbooks or cloud-based storage of knowledge on the Internet, boosting applications such as e-learning, e-health, and e-commerce.
- Promote energy efficiency: IMT enables energy efficiency across a range of sectors of the economy by supporting machine-to-machine communication and solutions such as smart grid, teleconferencing, smart logistics and transportation.

1 In 2012, the ITU's Radiocommunications sector (ITU-R) embarked on a programme to develop international mobile telecommunication (IMT) standards for 5G by 2020.

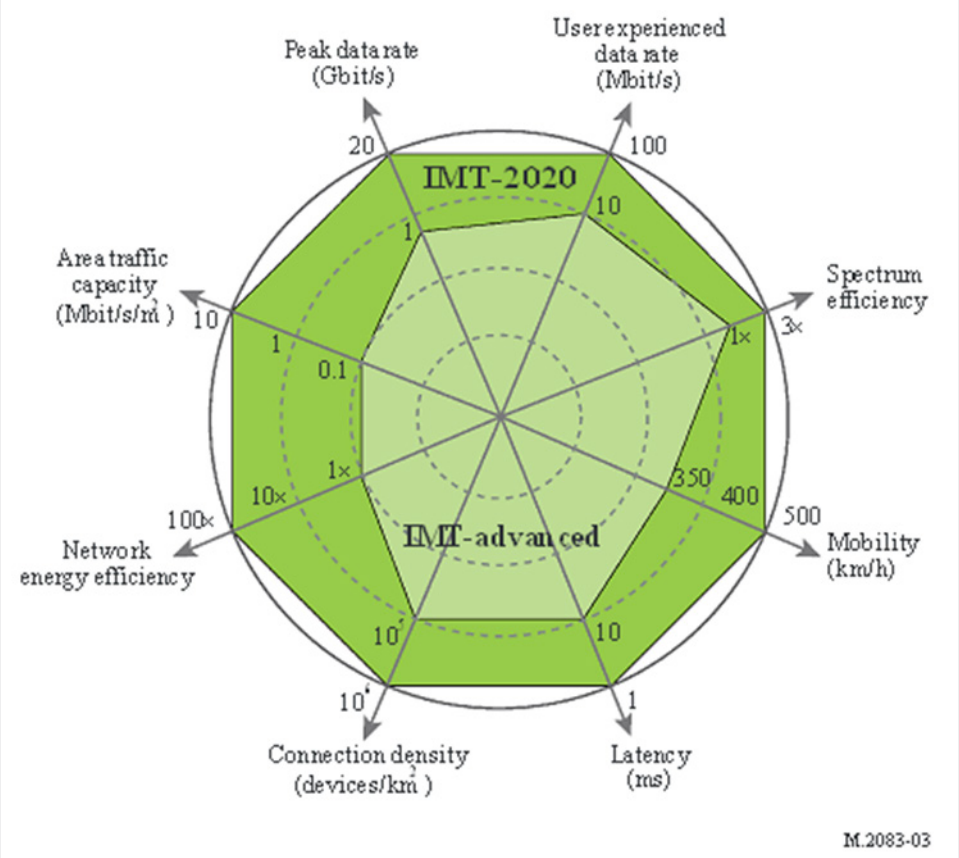
2 "The 5G era: Age of boundless connectivity and intelligence automation", GSMA Intelligence, 2017: https://www.gsmaintelligence.com/research/2017/02/the-5g-era-age-of-boundless-connectivity-and-intelligent-automation/614/

Box 1: Role of IMT 2020 (5G) and beyond (continued)

- Social changes: Broadband networks make it easier to quickly shape and share public opinions for a political or social issue through social network service. Opinions formed by a huge number of connected people due to their ability to exchange information anytime anywhere will become a key driver of social change.
- New art and culture: IMT will support artists and performers in creating works of art or in participating in group performances or activities, such as a virtual chorus, flash mob, co-authoring and song writing. Also, people connected to a virtual world are able to form new types of communities and establish their own cultures.

The targets set for IMT-2020 are described below.

Enhancement of key capabilities from IMT-Advanced to IMT-2020



The peak data rate of IMT-2020 for enhanced mobile broadband is expected to reach 10 Gbit/s. However under certain conditions and scenarios IMT-2020 would support up to 20 Gbit/s peak data rate, as shown in Figure 3. IMT-2020 would support different user-experienced data rates covering a variety of environments for enhanced Mobile Broadband. For wide area coverage cases, e.g. in urban and sub-urban areas, a user-experienced data rate of 100 Mbit/s is expected to be enabled. In hotspot cases, the user-experienced data rate is expected to reach higher values (e.g. 1 Gbit/s indoor).

Box 1: Role of IMT 2020 (5G) and beyond (continued)

Spectrum efficiency is expected to be three times higher compared to IMT-Advanced for enhanced mobile broadband. The achievable increase in efficiency from IMT-Advanced will vary and could be higher in some scenarios (for example five times subject to further research). IMT-2020 is expected to support 10 Mbit/s/m2 area traffic capacity, for example in hot spots.

The energy consumption for the radio access network of IMT-2020 should not be higher than IMT networks deployed today, even as it delivers enhanced capabilities. The network energy efficiency should therefore be improved by a factor at least as great as the envisaged traffic capacity increase of IMT-2020 relative to IMT-Advanced for enhanced mobile broadband.

IMT-2020 would be able to provide 1 ms over-the-air latency, capable of supporting services with very low latency requirements. IMT-2020 is also expected to enable high mobility up to 500 km/h with acceptable QoS. This is envisioned in particular for high-speed trains.

Finally, IMT-2020 is expected to support a connection density of up to 106/km2, for example in massive machine-type communication scenarios.

Source: ITU-R Recommendation M.2083-0

Figure 2: Evolution of mobile networks

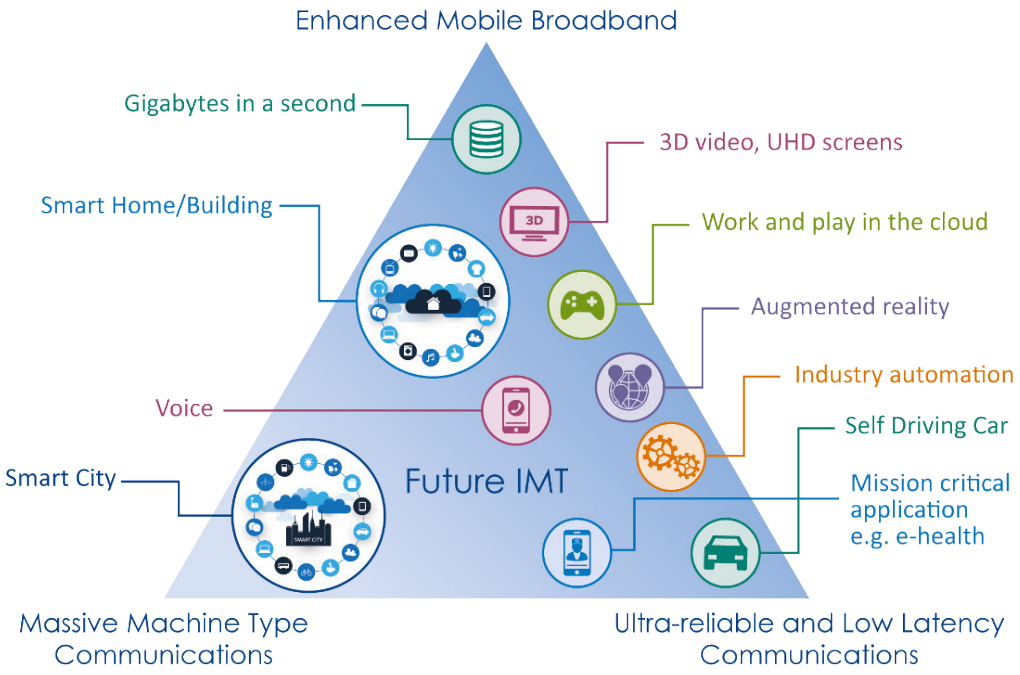
	1G	2G	3G	4G	5G
Approximate deployment date	1980s	1990s	2000s	2010s	2020s
Theoretical download speed	2kbit/s	384kbit/s	56Mbit/s	1Gbit/s	10Gbit/s
Latency	N/A	629 ms	212 ms	60-98 ms	< 1 ms

2.3 5G use cases

The high speeds and low latency promised by 5G will propel societies into a new age of smart cities and the Internet of Things (IoT). Industry stakeholders have identified several potential use cases for 5G networks, and the ITU-R has defined three important categories of these (see Figure 3):

1. **Enhanced mobile broadband (eMBB)** – enhanced indoor and outdoor broadband, enterprise collaboration, augmented and virtual reality.
2. **Massive machine-type communications (mMTC**) – IoT, asset tracking, smart agriculture, smart cities, energy monitoring, smart home, remote monitoring.
3. **Ultra-reliable and low-latency communications (URLLC)** – autonomous vehicles, smart grids, remote patient monitoring and telehealth, industrial automation.

Figure 3: 5G usage scenarios



eMBB is expected to be the primary use case for 5G in its early deployments, according to wireless operators. eMBB will bring high-speed mobile broadband to crowded areas, enable consumers to enjoy high-speed streaming for in-home, screen and mobile devices on demand, and will allow enterprise collaboration services to evolve. Some operators are also considering eMBB as the last-mile solution in those areas lacking copper or fibre connections to homes.

5G is also expected to drive the evolution of smart cities and IoT through the deployment of a considerable number of low-power sensor networks in cities and rural areas. The security and robustness built into 5G will make it suitable for public safety as well as for use in mission-critical services, such as smart grids, police and security services, energy and water utilities, and healthcare. Its low latency performance characteristics make it suitable for remote surgery, factory automation and the control of real-time processes.

5G's low latency and safety characteristics will play well in the evolution of intelligent transport systems, enabling smart vehicles to communicate with each other, and creating opportunities for connected, autonomous cars and trucks. For example, an autonomous vehicle (AV) operated via a cloud-based, autonomous driving system must be able to stop, accelerate or turn when told to do so. Any network latency or loss in signal coverage preventing the message from being delivered could result in catastrophic consequences. However, wireless operators believe that AVs have a significant way to go before they come into service, despite ongoing pilots and trials.

Box 2: 5G and fixed mobile convergence (FMC)

FMC is a networking solution in any given configuration, providing services and applications to the end user regardless of fixed or mobile access technologies and independent of the user's location. The concept of FMC has been implemented since 2005. With the move towards 5G, the FMC solution acquires additional flavour.

Recommendation ITU-T Y.3101, the IMT-2020 network envisages an access network-agnostic architecture, the core of which will be a common unified core network for new radio access technologies for IMT-2020, as well as existing fixed and wireless networks (e.g. wireless local area network (WLAN)). The access technology-agnostic unified core network is expected to be accompanied by common control mechanisms, decoupled from access technologies.

Emerging information and communications technologies (for example virtualization, cloud, software-defined networking (SDN), network function virtualization (NFV)) are transforming telecommunication operators' fixed and mobile networks to achieve high resource utilization and network flexibility, which in turn contribute to network functions' convergence in an IMT-2020 network.

To this end, ITU-T SG13 approved the Recommendation ITU-T Y.3130 (01/2018) that specifies service-related requirements such as unified user identity, unified charging, service continuity and guaranteed quality of service support – as well as network capability requirements such as control plane convergence, user data management, capability exposure and cloud-based infrastructure, to support fixed mobile convergence in IMT-2020 networks.

Currently, ITU-T SG13 continues to investigate different facets of the FMC approach. This includes FMC service scheduling – a network capability to collect information from application layer, network layer and user layer to generate service scheduling policies (i.e. traffic scheduling, access selection, etc.) in the FMC network which supports multiple RAT accesses.

In the context of IMT-2020, FMC represents capabilities that provide services and applications to end users regardless of fixed or mobile access technologies being used and independently of the users' location.

2.4 Socio economic implications of 5G

There are few third party studies that have looked at the economic impact of investments in 5G. Nevertheless, it is possible to draw upon some third-party forecasts to estimate the impact that 5G could have on economic output.

The ITU suggests that policy-makers undertake an independent economic benefits assessment since third party estimates are not endorsed by the ITU.

One report estimates that 5G will underwrite USD 12.3 trillion of global economic output by 2035, with the greatest growth in sales activity coming from manufacturing because of an anticipated increase in spending on 5G equipment. This is followed by sales growth in the ICT sector driven by higher expenditure on communications services. Investment in the value chain is expected to generate a further USD 3.5 trillion in output and provide support for 22 million jobs by 2035.[3]

[3] "The 5G Economy", IHS economics and IHS technology, January 2017: https://cdn.ihs.com/www/pdf/IHS-Technology-5G-Economic-Impact-Study.pdf

The European Commission (EC) estimates the total cost of 5G deployment across the 28 Member States will be EUR 56 billion, resulting in benefits of EUR 113.1 billion per annum arising from the introduction of 5G capabilities, and creating 2.3 million jobs. It is also estimated that benefits are largely driven by productivity in the automotive sector and in the workplace generally. Most of the benefits are expected in urban areas while only 8 per cent of benefits (EUR 10 billion per annum) will be realized in rural areas.[4]

Other reports have also indicated significant economic benefits and productivity enhancements resulting from investment in 5G networks.[5] Such estimates set out to provide a quantification of the benefits of 5G while assuming ideal investment conditions. The true economic benefit for each country will vary depending upon market structure and the availability of digital and supporting economic infrastructure.

Key finding: Policy-makers may consider undertaking their own economic assessment of the commercial viability and economic impact of 5G networks.

Despite the potential economic benefits, the industry remains sceptical about the commercial case for investment in 5G. Given the significance of required investment, scepticism remains among some European operators over 5G hype and question whether they can make money from it. These concerns are supported by the 5G Infrastructure Association (5GIA), an EU-backed body, and by senior telecoms executives cautioning against premature 5G launch announcements.[6]

Many 5G announcements – some are highlighted in this report – are simply regional 5G pilots and trials rather than full-scale commercial deployments. There is some way to go before the investment case for operators can be made robustly and before any large scale commercial deployment can commence.

Key finding: Until the investment case for 5G is demonstrable, industry and policy-makers may consider approaching 5G investment with caution and should continue to enhance the availability and quality of existing 4G networks.

2.5 Digital divide

The industry takes the view that initial deployment of 5G networks will be in dense urban areas and will offer services such as enhanced mobile broadband (eMBB) – it will be commercially challenging to deploy 5G networks in rural areas where demand tends to be lower – consequently, rural areas may be left behind, thereby increasing the digital divide.

However, the use of sub-1 GHz frequency spectrum if available, can counteract this in rural areas. This part of the spectrum allows mobile operators to cover wide areas at lower cost than with higher frequency spectrum.

While data speeds and network capacity in this part of the spectrum are not as high as in higher frequency bands, the sub-1 GHz spectrum will enhance the coverage of rural networks.

Key finding: Local authorities and regulators should recognize the risk of increasing the digital divide and support commercial and legislative incentives to stimulate investment in affordable wireless coverage through sub-1 GHz spectrum, where possible.

---

4 "Identification and quantification of key socio-economic data to support strategic planning for the introduction of 5G in Europe", European Commission, 2016: https://connectcentre.ie/wp-content/uploads/2016/10/EC-Study_5G-in-Europe.pdf

5 "5G mobile – enabling businesses and economic growth", Deloitte, 2017; "Tech-onomy: Measuring the impact of 5G on the nation's economic growth", O2 Telefonica (UK), 2017

6 https://www.techradar.com/news/eu-backed-groups-warns-about-5g-claims

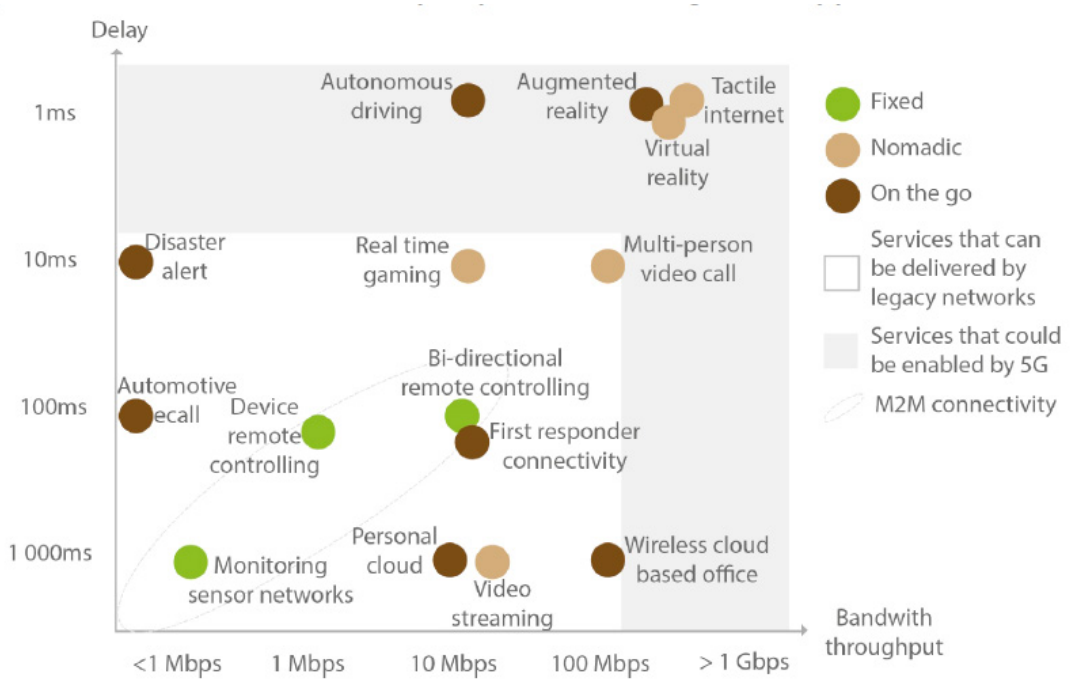
3 5G technology and spectrum requirements

Radio spectrum, backhaul, softwarization of core networks and radio access networks will be vital in early deployment of 5G networks particularly where enhanced mobile broadband is concerned.

3.1 Radio access networks

Most outdoor 4G mobile network deployments are currently based on macro-cells.[1] However, macro-cells that cover large geographical areas will struggle to deliver the dense coverage, low latency and high bandwidth required by some 5G applications (as shown in Figure 4).

Figure 4: Bandwidth and latency requirements for 5G applications



Source: GSMA Intelligence, 2015.

To deliver the dense coverage and high capacity network required by 5G, wireless operators are now investing in the densification of their 4G radio access network (RAN) – particularly in densely populated urban areas – by deploying small cells. Small cells, while serving a much smaller geographical area than a macro cell, increase network coverage, capacity and quality of service. See Figure 5.

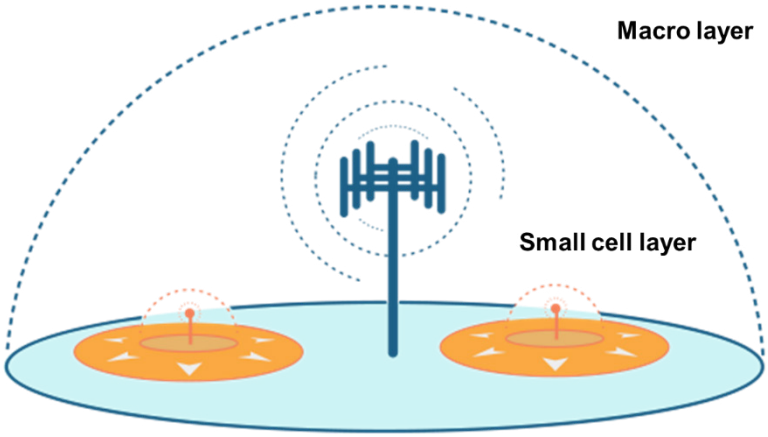
The deployment of small cells is one way of boosting the capacity and quality of existing 4G networks while laying the foundation for commercial 5G networks and early eMBB services. Small cells are already being used by some wireless operators to boost the capacity and coverage of their existing 4G networks particularly in a dense urban setting, see Box 3 as an example.

Small cells boost network capacity without the need for additional spectrum, making them attractive to operators with a low spectrum holding or where spectrum is scarce. Furthermore, the industry view is that the deployment of small cells in dense urban to boost existing 4G network quality is likely to support the anticipated high capacity requirements of 5G networks and early eMBB services.[2]

[1] https://www.mobileworldlive.com/blog/blog-global-base-station-count-7m-or-4-times-higher/

[2] TechUK: https://goo.gl/Q58ZA8- FCC: https://www.fcc.gov/5G- ITU: https://www.itu.int/ITU-T/workprog/wp_item.aspx?isn=14456

Figure 5: Macro versus small cell networks



Box 3: Aberdeen

In September 2017, independent tower specialist Wireless Infrastructure Group, in collaboration with Telefónica, launched Europe's first small cell network supporting cloud RAN (C-RAN) for faster and higher capacity mobile services in the city centre of Aberdeen.

Source : http://www.wirelessinfrastructure.co.uk/city-of-aberdeen-paves-the-way-for-5g/

Due to the dense coverage that small cells need to provide, small cell antennae need to be installed onto street furniture – bus shelters, lampposts, traffic lights, etc. These are often accompanied by a street cabinet to accommodate the operator radio equipment, power and site connectivity. Figure 6 shows an example of an antenna system mounted onto a lamppost as well as its supporting street cabinet.

Figure 6: An example of a small cell antenna system and a street cabinet

Massive MIMO (multiple input, multiple output) scales up to hundreds or even thousands of antennae, increasing data rates and supporting beamforming, essential for efficient power transmission. Massive MIMO increases spectral efficiency and in conjunction with dense small cell deployment, will help operators to meet the challenging capacity requirement of 5G.[3]

[3] IEEE: https://ieeexplore.ieee.org/document/7881053/

3.2 Core networks

End-to-end flexibility will be one of the defining features of 5G networks.[4] This flexibility will result in large part from the introduction of network softwarization where the core network hardware and the software functions are separated. Network softwarization – through network functional virtualization (NFV), software defined networking (SDN), network slicing and Cloud-RAN (C-RAN) – aims to increase both the pace of innovation and the pace at which mobile networks can be transformed.

- **NFV** – replaces network functions on dedicated appliances – such as routers, load balancers, and firewalls, with virtualized instances running on commercial off-the-shelf hardware, reducing the cost of network changes and upgrades.
- **SDN** – allows the dynamic reconfiguration of network elements in real-time, enabling 5G networks to be controlled by software rather than hardware, improving network resilience, performance and quality of service.
- **Network slicing** – permits a physical network to be separated into multiple virtual networks (logical segments) that can support different RANs or several types of services for certain customer segments, greatly reducing network construction costs by using communication channels more efficiently.
- **C-RAN** – is presented as a key disruptive technology, vital to the realization of 5G networks. It is a cloud-based radio network architecture that uses virtualization techniques combined with centralized processing units, replacing the distributed signal processing units at mobile base stations and reducing the cost of deploying dense mobile networks based on small cells.

For the last few years, Telefónica has been focusing its efforts on virtualizing its core network based on SDN/NFV in preparation for 5G in Argentina, Mexico and Peru, see Box 4.

Box 4: Telefonica investing in SDN and NFV

Operators such as Telefónica are already investing in SDN and NFV as part of their gradual transition to 5G – which is likely to reduce core network costs in the long term. Telefónica has an ambitious plan to virtualize its network end-to-end, across access, aggregation and backbone domains under its UNICA programme.

Source: https://www.telefonica.com/documents/737979/140082548/Telefonica_Virtualisation_gCTO_FINAL.PDF/426a4b9d-6357-741f-9678-0f16dccf0e16?version=1.0

Other technology enhancements being considered include signal coding techniques which provide improved spectral efficiency and the high-speed performance required by 5G. In addition, edge computing is increasingly important for real-time and very latency-sensitive applications. Edge computing brings data closer to end-user devices, providing computing power with very low latency for demanding applications. This speeds up the delivery of actionable data, cuts down on transport costs and optimizes traffic routes.

3.3 Backhaul

Backhaul networks connect the radio network (RAN) to the core network. The ultra-high capacity, fast speeds and low latency requirements of 5G require a backhaul network capable of meeting these high demands. Fibre is often considered the most suitable type of backhaul by mobile operators due to its longevity, high capacity, high reliability and ability to support very high capacity traffic.

[4] ITU: http://news.itu.int/5g-update-new-itu-standards-network-softwarization-fixed-mobile-convergence/

However, fibre network coverage is not ubiquitous in all cities where 5G is expected to launch initially – and even less so in suburban and rural areas. Building new fibre networks in these areas can often be prohibitive in terms of cost for operators. In this case, a portfolio of wireless backhaul technologies should be considered in addition to fibre, including point-to-multipoint (PMP) microwave and millimeter wave (mmWave). PMP is capable of downstream throughput of 1Gbit/s and latency of less than 1ms per hop over a 2-4 km distance. mmWave has significantly lower latency and is capable of higher throughput speeds.

While most focus is being given to terrestrial technology, there is also a role for high altitude platform systems (HAPS) and satellite technology in 5G. HAPS and satellite systems (including non-geostationary constellations) can deliver very high data rates (> 100 Mbit/s – 1 Gbit/s) to complement fixed or terrestrial wireless backhaul networks outside major urban / suburban areas and can deliver video transmission to fixed locations. HAPS and satellites may be integrated with other networks rather than function as a standalone network to provide 5G, thereby augmenting the 5G service capability and addressing some of the major challenges regarding the support of multimedia traffic growth, ubiquitous coverage, machine-to-machine communications and critical telecom missions.[5]

Key finding: A portfolio of wireless technologies may be considered in addition to fibre, including point-to-multipoint (PMP) microwave, millimeter wave (mmWave), HAPS and satellites.

In summary, a realistic 5G backhaul strategy is likely to consist of a portfolio of technologies. Each approach should be considered on its own merits in light of the performance needs, available infrastructure and the likely return on investment.

3.4 Fronthaul

Conventionally in a 4G wireless network, the fronthaul link exists between radio frequency (RF) function and the remaining layer 1, 2 and 3 (L1/L2/L3) functions. Recommendation ITU-T Y.3100 defines fronthaul as “a network path between centralized radio controllers and remote radio units (RRU) of a base station function”. This architecture allows for the centralization of all high layer processing functions at the expense of the most stringent fronthaul latency and bandwidth requirements. The increase in data rates in 5G makes it impractical to continue with the conventional Common Public Radio Interface (CPRI) fronthaul implementation. Allocating more processing function to RRU would relax the latency and bandwidth requirements – but fewer processing functions can then be centralized. It is thus critical that the new functional-split architecture take into account technical and cost-effective tradeoffs between throughput, latency, and functional centralization.[6]

The following documents specify or describe technologies that can be used for fronthaul:

- Supplement 55 to G-series Recommendations “Radio-over-fibre (RoF) technologies and their applications”
- Supplement 56 to G-series Recommendations “OTN transport of CPRI signals” describes alternatives for mapping and multiplexing CPRI client signals into the OTN
- Recommendation ITU-T G.987 series: 10-Gigabit-capable passive optical networks (XG-PON)
- Recommendation ITU-T G.9807 series: 10-Gigabit-capable symmetric passive optical network (XGS-PON)
- Recommendation ITU-T G.989 series: 40-Gigabit-capable passive optical networks 2 (NG-PON2)
- Draft Recommendation ITU-T G.RoF “Radio over Fibre systems” (under development)

[5] EMEA Satellite Operators Association: https://gscoalition.org/cms-data/position-papers/5G%20White%20Paper.pdf

[6] G-series Technical Report on “Transport network support of IMT-2020/5G” (GSTR-TN5G): http://www.itu.int/pub/publications.aspx?lang=en&parent=T-TUT-HOME-2018

- Draft Supplement to G-series Recommendations (G.sup.5GP) "5G wireless fronthaul requirements in a PON context" (under development)
- Recommendation ITU-T G.709(.x) series: Optical transport network (OTN) beyond 100 Gbit/s
- Draft Recommendation ITU-T G.ctn5g: Characteristics of transport networks to support IMT-2020/5G (under development)
- Draft Supplement to G-series Recommendations G.Sup.5gotn: Application of OTN to 5G transport (under development)
- Recommendation ITU-T G.695: Optical interfaces for coarse wavelength division multiplexing applications
- Recommendation ITU-T G.698.4: Multichannel bi-directional DWDM applications with port agnostic single-channel optical interfaces
- Recommendation ITU-T G.959.1: Optical transport networks physical layer interfaces

3.5 Spectrum for 5G

More spectrum bandwidth will be required to deploy 5G networks (than 4G) to the high capacity requirements, increasing the need for spectrum. In consequence, the industry is making concerted efforts to harmonize 5G spectrum. ITU-R is coordinating the international harmonization of additional spectrum for 5G mobile systems development (Box 5). ITU's Standardization Sector (ITU-T) is playing a key role in producing the standards for the technologies and architectures of the wireline elements of 5G systems.

Box 5: ITU-R technical feasibility of IMT in the frequencies above 24 and up to 86 GHz

The ITU-R investigates the technical feasibility of future 5G spectrum in the frequencies above 24 and up to 86 GHz based on recently conducted (and still ongoing) studies carried out by many sector members. Solutions based on MIMO and beamforming are becoming increasingly feasible with higher frequencies. Bands below and above 6 GHz could be used in a complementary manner for the year 2020 and beyond. The ITU is expected to decide on the additional spectrum for IMT in the frequency range between 24 GHz and 86 GHz at the World Radiocommunication Conference in 2019 (WRC-19).

New spectrum bands under study for WRC-19:

Existing mobile allocation	No global mobile allocation
24.25 – 27.5 GHz	31.8 – 33.4 GHz
37 – 40.5 GHz	40.5 – 42.5 GHz
42.5 – 43.5 GHz	
45.5 – 47 GHz	47 – 47.2 GHz
47.2 – 50.2 GHz	
50.4 GHz – 52.6 GHz	
66 – 76 GHz	
81 – 86 GHz	

5G use cases could potentially be met by a variety of spectrum frequencies. For example, low-latency and short-range applications (suited to dense urban areas) are likely to be suitable for mmWave frequency (above 24 GHz). Long-range, low-bandwidth applications (more suited to rural areas) are likely to be suitable for sub-1 GHz frequencies. While the lower frequencies have better propagation characteristics for better coverage, the higher frequencies support higher bandwidths due to the large spectrum availability at mmWave bands. Huawei, for example has proposed a multi-layer spectrum approach, which summarizes this approach best (see Box 6).

The challenge for NRAs will be to select globally harmonized spectrum bands for 5G. The best way to achieve this goal will be to take into account the WRC-19 relevant decisions for higher bands, as well as WRC-07 and WRC-15 decisions for lower bands.

While the EC has earmarked the 700 MHz spectrum as essential to achieve wide-area and indoor coverage for 5G services,[7] it could be used differently in parts of Africa to enhance 4G coverage. It is expected that by 2020, only 35 per cent of the Sub-Saharan population will be covered by 4G networks, with many rural areas enjoying little or no 4G mobile coverage. This compares to a global average of 78 per cent.[8] For this reason, policy-makers in Sub-Saharan Africa might well consider using 700MHz spectrum as the ideal way forward to increasing rural 4G coverage rather than using this for 5G.

Key finding: Policy-makers may consider making available low frequency spectrum (e.g. in the 700MHz) to ensure mobile broadband can be provided to rural areas.

Box 6: An operator's perspective – Huawei's multi-layer spectrum approach

- **Coverage layer** – exploits spectrum below 2 GHz (e.g. 700 MHz) providing wide-area and deep indoor coverage.
- **Coverage and capacity layer** – relies on spectrum in the 2 – 6 GHz range to deliver the best compromise between capacity and coverage.
- **Super data layer** – relies on spectrum above 6 GHz and mmWave to address specific use cases requiring extremely high data rates.

Source: http://www.huawei.com/en/about-huawei/public-policy/5g-spectrum

The GSMA expects the 3.3–3.8 GHz spectrum to form the basis of many initial 5G services, particularly to offer enhanced mobile broadband. This is because the 3.4-3.6 GHz range is almost globally harmonized – and therefore well positioned to drive the economies of scale needed for low-cost devices.

Key finding: Policy-makers may consider grouping together to reach agreement on harmonized spectrum bands for 5G. NRAs would then benefit from an exchange of best practices in regard to shaping markets through the granting of spectrum licences.

[7] EC: https://ec.europa.eu/commission/commissioners/2014-2019/ansip/blog/700-mhz-must-digital-single-market_en
[8] GSMA: https://www.gsma.com/mobileeconomy/sub-saharan-africa-2017/

4 Key challenges in rolling out 5G

This section reviews the key challenges faced by telecom operators rolling out 5G networks. Particular focus is given to how appropriate regulation and government policy might help wireless operators to deploy small cells, fibre backhaul as well as the use of spectrum.

4.1 Small cell deployment challenges

In some countries, regulation and local authority policy have slowed the development of small cells through excessive administrative and financial obligations on operators, thus blocking investment. Constraints to deploying small cells include prolonged permitting processes, lengthy procurement exercises, excessive fees and outdated regulations that prevent access. These issues are described in Box 7 and in more detail below:

- **Local permitting and planning processes**: the time taken by local authorities to approve planning applications for small cell implementations can take 18 to 24 months (Box 7), resulting in delays.
- **Lengthy engagement and procurement exercises**: local authorities have used lengthy procurement processes lasting 6 to 18 months to award wireless providers with exclusive rights to deploy small cell equipment onto street furniture, costing time and expense.
- **High fees and charges to access street furniture**: local authorities currently charge high fees to use street furniture. According to the American Consumer Institute, one city set a USD 30 000 application fee to attach small cell equipment onto a utility pole; another locality imposed a USD 45 000 fee.
- **Human exposure to radiofrequency electromagnetic fields (EMF):** exposure limits differ across countries, and in some cases are unduly restrictive. ITU recommend that if radio frequency electromagnetic field (RF EMF) limits do not exist, or if they do not cover the frequencies of interest, then ICNIRP (International Commission on Non-Ionizing Radiation Protection) limits should be used. Where new antennae are added, all regular steps should be taken during the deployment phase to respond to any public concern. One factor contributing to public concern is the visibility of antennae, particularly on rooftops. Here, multi-band antennae can be used to reduce visual impact by maintaining the same number of antennae on rooftops. Without any spectrum or technology re-farming strategy, the 5G network will increase the localized exposure resulting from wireless technologies, at least during the transition period. It is thus important to include national authorities at an early stage in establishing how 5G can be deployed and activated – and how compliance with national limits can best be assessed and met. This has already proved difficult in countries where exposure limits are more restrictive than those recommended by the World Health Organization (WHO), based on the ICNIRP RF-EMF exposure guidelines.[1]
- **Access and code powers**: wireless operators[2] may not have the right to install small cell or radio apparatus onto street furniture such as lampposts. In the UK, for example, the code has been updated to overcome these limitations, but is non-binding, meaning its impact might be debatable.

Many of these local rules and regulations are prohibiting the rapid and cost-effective roll-out of small cells in city centres where 5G is initially expected to be most in demand. Policy-makers that offer streamlined and flexible regulatory processes stand to benefit the most from the innovation and economic growth that 5G will bring.

---

1 From Supplement ITU-T K.Suppl.9 "5G technology and human exposure to RF EMF": https://www.itu.int/rec/T-REC-K.Sup9/en

2 Code powers are statutory entitlements for specified telecoms operators to have the right to install, maintain, adjust, repair or alter their apparatus on public and private land. In order to have the benefit of these rights, a telecoms provider must apply to and be recognized by OFCOM, the UK telecoms regulator.

Box 7: Industry viewpoint on barriers to deploying small cells

Telecom providers such as Crown Castle, AT&T, Sprint, T-Mobile and Verizon have all described experiencing significant regulatory barriers from local authorities – these include excessive fees, prohibitions on small cell placement, unreasonable aesthetic restrictions and prolonged permitting processes. According to Crown Castle, its small cell deployments usually take 18 to 24 months to complete, from start to finish, largely due to the need to obtain local permits for the installation of the devices.

Source: https://goo.gl/6UaKJ4

Small cells have also yet to be deployed on a significant scale in Asia, although wireless operators in Japan and Korea (Rep. of) have densified their networks using macro cell C-RAN. C-RAN deployments are possible in Japan and Korea (Rep. of) because of the widespread availability of fibre backhaul, which may not be the case in other markets.

4.2 Fibre backhaul

Deploying fibre backhaul networks for small cells – to support high data rates and low latency – will be one of the largest challenges faced by operators due to the poor availability of fibre networks in many cities.

The UK, for example, has one of the lowest fibre penetration rates in Europe at 2 per cent penetration. This compares to a European average of around 9 per cent.[3] To incentivize investment in fibre networks, the UK Government has introduced a five-year relief from business rates on new fibre networks infrastructure.[4]

Where it is not cost effective to deploy fibre backhaul, operators should consider wireless backhaul technologies. A portfolio of wireless technologies including PMP, mmWave and satellite should be considered in addition to fibre where this is the case.

Key finding: Policy-makers may consider removing tax burdens to reduce investment cost associated with fibre in order to facilitate the deployment of 5G networks.

Some of the other challenges faced by operators are described in Box 8.

Key finding: Local authorities may consider agreeing upon standardized wayleave agreements to reduce the cost and time of deploying fibre networks.

4.3 Spectrum

The allocation and identification of globally harmonized spectrum across a range of frequencies requires coordination among the global community, regional telecommunication organizations and NRAs. This represents one of the largest challenges for NRAs in the successful deployment of 5G networks. Harmonized allocation has many advantages since it minimizes radio interference along borders, facilitates international roaming and reduces the cost of equipment. This overall coordination is the main objective of the ITU-R in the process of World Radiocommunication Conferences (WRCs).

[3] https://www.ispreview.co.uk/index.php/2017/02/uk-shunned-2017-ftth-ultrafast-broadband-country-ranking.html

[4] The UK's valuation office agency has recently undertaken a revaluation of business rates which will increase the tax likely to be paid by fibre operators. High business rates could adversely affect the commercial model to deploy fibre connectivity to support small cell deployment.

Box 8: Barriers to deploying fibre networks

- **Refused planning permission:** lack of early engagement between operators and local authorities can result in planning permission being refused. Local authority policy on the siting and aesthetics of street cabinets can also increase costs and delays while alternative solutions are sought.
- **Complex wayleave processes**: wayleave agreements allow operators to install telecoms infrastructure on public or private land. Landowners using a procurement process to grant wayleaves add risk, time and expense to the process. In addition, using bespoke wayleaves are expensive. Local authorities using wayleaves to generate revenues create an additional barrier to investment.

Source: Intelligens Consulting, 2018

For WRC-19, this process is currently at the stage of building consensus on the allocation and identification for IMT of large contiguous blocks of world-harmonized radio spectrum above 24 GHz, where large bandwidths are available. WRC-19 decisions on this topic will be based on ITU-R studies on extensive sharing and compatibility between the mobile service and the incumbent services in these and in adjacent frequency bands.

A number of NRAs in developed countries are considering the 700 MHz, 3.4 GHz and 24 GHz bands for initial deployment of 5G to satisfy coverage and capacity requirements.

Consideration should also be given to the sharing of spectrum to make more efficient use of what is available. Traditionally, NRAs have allocated spectrum to mobile operators on an exclusive basis. However, due to growing need, sharing can provide a means to improve the efficient use of existing spectrum.

Key finding: Policy-makers should consider using globally harmonized spectrum to maximize the efficient use of available spectrum.

Further consideration also needs to be given to the licensing and usage models for 5G spectrum, particularly above 24 GHz. Traditionally, mobile spectrum divided into small bandwidths (e.g. 5 MHz, 10 MHz, 20 MHz) has been scarce, and can therefore attract a high price at auction. Spectrum above 24 GHz is more readily available, so scarcity is less of an issue. This will influence commercial models and spectrum auctions. NRAs should consider what licensing models they should use (see also Section 5.7). National examples of approaches to spectrum sharing have been published by the ITU: for instance in the ITU Report on WTDC-14, Resolution 9.

Key finding: Policy-makers may consider a spectrum roadmap that supports exclusive, shared and unlicensed models with a predictable renewal process. Policy-makers should avoid artificially high 5G spectrum prices and instead opt for procedures that favour investment when awarding spectrum.

4.4 Other factors

- **Device availability** – the availability of devices compatible with 5G standards and spectrum will be vital in creating end-user demand for 5G services in the initial launch. Manufacturers are currently developing technology that embeds 5G, 4G, 3G and 2G onto a single chip and is expected to become available from 2019, and after 2020 for globally harmonized standards.
- **Coordination of industry verticals** – the telecoms industry is a well-tuned and formal ecosystem comprising device and chip manufacturers, equipment vendors and retail and wholesale

operators. Collaboration within this ecosystem is therefore relatively straightforward when developing new standards and services.

- **Net neutrality** – BEREC, the European telecoms regulator, has published final guidelines on how to strengthen net neutrality by requiring Internet service providers to treat all web traffic equally, without favouring some services over others. However, 17 mobile operators including Deutsche Telekom, Nokia, Orange, Vodafone and BT lobbied heavily for BEREC to adopt a more relaxed interpretation of the rules, saying these "create significant uncertainties around 5G return on investment". Furthermore they stated that they would *not* introduce high-speed 5G networks unless BEREC took a softer approach to net neutrality.[5]

[5] "5G manifesto for timely deployment of 5G in Europe", various industry players, July 2016: http://telecoms.com/wp-content/blogs.dir/1/files/2016/07/5GManifestofortimelydeploymentof5GinEurope.pdf

5 What does 'good' look like?

This section reviews what can be learned from the way in which wireless providers, NRAs and governments across the world are working through the deployment issues associated with deploying 5G networks.

5.1 Streamlining small cell deployments

Bills have been proposed in Illinois, Washington State, Florida and California to streamline the deployment of small cell equipment on street furniture. These bills restrict local government fees – and some go further to ensure no exclusive arrangements are made with wireless providers.

Key finding: Federal and state governments should work with local municipal authorities to ensure reasonable fees are charged to deploy small cell radio equipment on street furniture.

Box 9: Streamlining the deployment of small cells

In September 2017, a California bill was passed streamlining small cell deployment by permitting its use and making such deployment no longer subject to a local discretionary permit or with specified criteria. The new legislation standardizes small cell deployments across the state. In addition, the bill:

- Grants providers non-discriminatory access to public property
- Allows local governments to charge permit fees that are fair, reasonable, non-discriminatory and cost-based
- Limits the costs charged by local governments of attaching equipment to USD 250
- Stops local governments putting an unreasonable limit on the duration of the permit on the telecom facility

A similar approach has been proposed in a bill in Florida, requiring an authority to process applications for siting small cell equipment on utility poles on a non-discriminatory basis and approving applications within set time-scales. The bill also proposes that authorities may not enter into any exclusive arrangements entitling providers to attach equipment to authority utility poles. Furthermore, the bill states that authorities may not charge more than USD 15 per year, per utility pole.

In Washington State, a bill proposes to authorize the installation of small cell facilities on publicly owned assets and limits charges to USD 500 per annum. In Illinois, a bill proposes that local government may not prohibit, regulate or charge operators to deploy small cell wireless equipment.

Sources: California SB-649, 2017; Florida SB-596, 2017; Washington SB-5711, 2017; Illinois SB-1451, 2017

Key finding: Local authorities may consider improving access to government-owned street furniture and streamlining engagement processes as alternatives to lengthy procurement exercises.

5.2 Policy intervention - fibre and spectrum

Leading economies like the UK have low fibre penetration, according to the FTTH Council, because of underinvestment in pure fibre networks. Box 10 describes the actions being taken by UK policy-makers to improve fibre penetration in the run up to 5G.

Box 10: UK fibre investments

In 2016, the UK Government announced a GBP 740 million challenge fund to invest in local full fibre networks to support the development of 5G. The fund is now being distributed through a competitive process to local authorities across the UK.

Sources: Federal Ministry of Transport and Digital Infrastructure, Germany, 2017; "a 5G Strategy for Germany", Federal Government of Germany, 2017; Department of Culture Media and Sport, Government of United Kingdom, 2016

The Australian Government has identified a clear 5G policy agenda to speed up the deployment of digital infrastructure and availability of 5G spectrum (see Box 11).

Box 11: 5G working group, Australia

The Australian Government is developing a 5G Directions Paper which outlines a 5G policy approach for Australia including the establishment of a 5G working group to facilitate ongoing dialogue with industry. The paper highlights actions which make spectrum available in a timely manner and which streamlines arrangements to allow wireless providers to deploy digital infrastructure more quickly and at lower cost.

Source: "5G-Enabling the Future Economy", Department of Communications and the Arts, Australia, 2017

Key Finding: Where market failure has occurred, governments may consider stimulating investment in fibre networks and passive assets through setting up PPPs, investment funds and offering grant funds, etc.

5.3 Infrastructure sharing

Where fibre is the preferred method of backhaul, it may not be commercially attractive. Modest levels of duct sharing, and re-use can generate significant savings in the development of fibre networks. Regulatory policies that promote infrastructure sharing and re-use can help significantly lower 5G deployment costs – although they can be complex to implement (see Box 12).

A study undertaken by Vodafone suggests that the duct access regime is commonly used by NRAs in France, Spain and Portugal, ensuring minimum bureaucracy and maximum transparency to all parties. In contrast, where SMP infrastructure access has been mandated, as in the UK and Germany, many of these detailed provisions are lacking.[1]

Key finding: Policy-makers may consider continuing the duct access regime to encompass 5G networks, thereby helping reduce the cost of investing in 5G fibre backhaul networks.

Commercially led network-sharing agreements are preferred by most NRAs and seem to have gained significant market traction. These can speed up the deployment and reduce costs for 5G networks where network sharing ranges across mobile infrastructure as well as fibre (see Box 13).

[1] "Best practice for passive infrastructure access", WIK-Consult, 2017: https://www.vodafone.com/content/dam/vodafone-images/public-policy/reports/pdf/best-practice-passive-infrastructure-access-050517.pdf

Box 12: Mandated network sharing

- In November 2017, the Netherlands passed a bill designed to accelerate broadband roll-outs. It mandated all owners/administrators of networks and related infrastructure to comply with reasonable requests for shared access and/or coordinated network deployment, and to share information about their infrastructure.
- Indonesia's Ministry of Communications and Information Technology is working toward new rules to encourage the development of passive infrastructure sharing such as ducts, poles, towers, cabinets, etc.
- UK telecoms regulator Ofcom is running a market consultation to mandate the incumbent operator and significant market player BT to offer duct fibre access to rival operators. Previous attempts to mandate dark fibre access failed.
- In Italy, ultra-fast broadband legislation has enabled TIM and UTILITALIA (the federation of electricity, gas, water and environment companies) to sign a memorandum of understanding to facilitate the use of pre-existing infrastructures of more than 500 local utility operators to deploy fibre networks.

Sources: https://goo.gl/kqYCRM (Netherlands), https://goo.gl/vWq7aD (Indonesia), https://goo.gl/vdFxz9 (Ofcom, UK), https://goo.gl/m24g32 (Italy)

Box 13: Commercially driven network sharing

- In Spain, telecoms operator MASMOVIL has passed the ten million household threshold using a fibre network that it shares with Orange Espana through a network-sharing pact.
- In Portugal, Vodafone and operator NOS have signed an agreement to deploy and share a fibre network that will be marketed to around 2.6 million homes and businesses. The two companies provide access to each other's networks on agreed commercial terms.
- New Zealand's wholesale network operator, Chorus, is calling on the government to begin formulating plans for a single 5G mobile network – one which can be shared by all service providers, a more sustainable approach than having a separate 5G network for each of the country's three mobile operators.
- Vodafone Cameroon has recently signed a 'strategic national network sharing agreement' with CamTel, allowing Vodafone to use CamTel's existing network infrastructure in Douala and Yaounde and to expand its coverage to new locations across the country.
- Telenor Denmark and Telia Denmark have signed a services contract with Nokia to manage their shared mobile networks run by one infrastructure company (TT-Netvaerket).
- Econet Wireless (Zimbabwe), has stated it is open to infrastructure sharing, under an equitable 'one-for-one' infrastructure.

Sources: https://goo.gl/u2fojb (Spain), https://goo.gl/bT9hZ4 (Portugal), https://goo.gl/vh4LGP (New Zealand), https://goo.gl/AAbapS (Cameroon), https://goo.gl/JmuSnJ (Denmark), https://goo.gl/iSb4sq (Zimbabwe)

The use of independent wholesale infrastructure providers for the provision of small cell networks has increased over the last few years, reducing deployment costs, promoting retail competition and increasing service coverage. Wireless provider Crown Castle (US) for example, increased its small cell

revenues by over 40 per cent between 2015 and 2016[2] as mobile operators move to densify their networks in preparation for 5G roll-outs.

5.4 Transition to fibre

At present, lower wholesale copper access prices are competitive when set against the price of fibre services, adversely affecting the take-up of fibre. There is no consensus on the most appropriate approach to pricing during the transition from copper to fibre. NRAs should consider allowing incumbents *to withdraw copper-based access products* as soon as they offer fibre-based access services, to prevent the undermining of the business case for more expensive fibre services (see Box 14).

Box 14:Transition to fibre

- The Government of Australia has imposed a deadline of 2020 by which all premises are to be migrated from copper to fibre. In 2014, Telstra (Australia) began to switch off services being delivered over its copper networks. The government-funded NBNCo initiative, which has driven wholesale fibre connectivity across Australia, will switch off copper networks in areas where NBNCo already provides fibre services.
- Verizon (US) has requested regulatory permission to migrate its copper network in selected markets from 2018. Verizon delivers services via its fibre infrastructure and wishes to cease maintaining the copper facilities in Virginia, New York, New Jersey, Pennsylvania, Rhode Island, Massachusetts, Maryland and Delaware.
- ComReg, the Irish telecoms regulator, has launched a consultation on the potential of its incumbent operator, Eir, to transition from copper in some parts of the country, particularly in areas of extensive fibre coverage.
- Singtel (Singapore) announced plans to discontinue its copper-based ADSL network in April 2018 as it accelerates fibre-based service adoption for its business and residential customers in the city.
- Chorus (New Zealand) is set to get regulatory relief from its copper network under plans to deregulate the copper network where it competes with fibre access networks from 2020.

Sources: https://goo.gl/2YVKsd (Australia), https://goo.gl/VCyfap (US), https://goo.gl/X3EeKa (Ireland), https://goo.gl/mRKu1C (Singapore), https://goo.gl/n6kqVb (New Zealand)

Key finding: NRAs may consider policies and financial incentives to encourage migration from copper to fibre and to stimulate the deployment and take-up of fibre services.

5.5 Addressing local planning challenges

Operators have often cited that it would be helpful to have a central database showing all available infrastructure and utility assets, such as existing local authority or utility ducts, fibre networks, CCTV posts, lampposts, etc. Such a database should also identify key contacts and the process for securing access to the assets. Such databases already exist in Portugal and Spain and may exist in other countries.

Key finding: Local authorities may consider holding a central database identifying key contacts, showing assets such as utility ducts, fibre networks, CCTV posts, lampposts, etc. to help operators cost and plan their infrastructure deployment more accurately.

[2] https://www.telegeography.com/products/commsupdate/articles/2017/02/07/tower-talk-a-guide-to-the-latest-major-cell-site-developments/index.html

Standardized wayleave agreements used among local authorities can significantly reduce the cost and time to implement fibre networks such as that developed by the City of London Corporation (see Box 15).

Box 15: City of London standardized wayleave agreements

In 2015, the City of London Corporation recognized that a key reason for the lack of fibre investment was the complex wayleave process. The corporation developed a standardized wayleave toolkit to facilitate the delivery of fibre infrastructure effectively and efficiently. The toolkit is now available to all local authorities in London.

Source: http://news.cityoflondon.gov.uk/standardised-toolkit-helps-london-businesses-get-faster-access-to-broadband/

Local authorities should also standardize the processes giving operators the appropriate permission to undertake relevant street works when laying fibre networks or deploying small cell equipment onto street furniture (Box 16). It is also best practice to undertake consultations with the market to understand potential issues and solutions arising in deployment.[3]

Key finding: Local authorities may consider undertaking market consultations or soft market testing to identify best practices for deploying 5G networks prior to committing to formal procurement processes.

Box 16: Efficient planning processes

In 2015, the City of Centennial (Colorado, US) permitting office was authorized to require the co-location of underground facilities upon the filing of a major right of way permit request by telecoms operators. The right of way policy has allowed the city to coordinate investments, saving time and costs.

In Kentucky (US), a guide was issued on fibre planning to communities and utilities. The guide included advice on streamlining survey requirements, permit applications and developing pole attachment agreements.

Sources: City of Centennial, Colorado, 2015; https://goo.gl/FswzSv (Kentucky)

5.6 Spectrum harmonization

The focus for early 5G applications has been on the bands above 24 GHz and below 6 GHz (see Box 17). NRAs should coordinate their proposals on the millimetre bands to maximize the opportunity for global spectrum harmonization.

In order to prepare European positions for WRC-19, EU ministers for example, agreed in December 2017 a roadmap for the roll out of 5G technology across Europe. The roadmap will provide consensus over the harmonization of 5G spectrum bands and how they will be allocated to operators across Europe.

[3] http://webarchive.nationalarchives.gov.uk/20100402171309/ http://www.ogc.gov.uk/documents/Early_Market_Engagement_Guidance.pdf

Box 17: 5G spectrum proposals by some NRAs

- **Ofcom, UK:** working closely with European NRAs, Ofcom has proposed the use of spectrum in the 700 MHz, 3.4 GHz and 24 GHz bands for 5G use. Ofcom has also proposed to change the authorization regime in the 64–66 GHz band to licence-exempt and expand the use cases for the 57–66 GHz band.
- **The FCC, US**: has identified almost 11 GHz of spectrum for flexible use wireless broadband – 3.85 GHz of licensed spectrum in the 28 GHz, 37 GHz and 39 GHz bands.
- **MIIT, China**: plans to allocate 5G mmWave spectrum in the 24–27 GHz and 37–42 GHz bands, in addition to the 3.3–36 GHz and 4.8–5 GHz bands for 5G.
- **KCC, Korea (Rep. of)**: will start to auction off 5G spectrum in the 3.5 GHz and 28 GHz bands in
- **ACMA, Australia:** announced plans for a multi-band spectrum auction to be launched before the end of 2017, comprising lots from the 1800 MHz, 2 GHz, 2.3 GHz and 3.4 GHz bands.

Note that WRC-15 excluded the 28GHz band from the scope of study toward the international harmonization of millimetric bands for IMT by WRC-19.

Sources: https://goo.gl/kpPnTy (UK), https://goo.gl/Mc5wZx (US), https://goo.gl/bdusHx (China), https://goo.gl/pGz5jG (South Korea), https://goo.gl/1aK5LY (Australia)

5.7 Spectrum licensing

The design of selection procedures and conditions attached to 5G licences can significantly impact the structure of mobile markets – by enhancing competition or limiting it.

Traditionally, NRAs have granted spectrum licenses to mobile operators giving exclusive rights to offer voice or data services. In some cases, the licence may come with population and time-based coverage obligations. The licensed spectrum allows mobile operators to plan and invest in mobile infrastructure with certainty, and should include conditions that ensure that the allocated spectrum is used effectively, particularly in rural areas.

Licensed, shared-access spectrum can improve spectrum utilization in rural areas. For example, the granting of spectrum use to some secondary users in such areas will not interfere with the primary licence-holder's radio signals.

Current examples of shared spectrum include aeronautical telemetry, broadcast and wireless cameras. This shared licence model may well provide the 5G ecosystem with sufficient flexibility to make good use of spectrum now underutilized by other services to provide additional capacity at lower cost.

Following a study, ITU-R has approved regulatory tools to support enhanced, shared use of the spectrum[4] – as well as spectrum management principles, challenges and issues related to dynamic access to frequency bands via radio systems employing cognitive capabilities.[5]

Spectrum auctions have traditionally awarded exclusive spectrum rights to wireless operators paying the highest fees. Policy-makers view auctions favourably as a means of generating significant incomes. However, auctions can be counterproductive in that they reduce funds available for infrastructure,

[4] See Report ITU-R SM.2404: https://www.itu.int/pub/R-REP-SM.2404
[5] See Report ITU-R SM.2405: https://www.itu.int/pub/R-REP-SM.2405

diluting economic impact.[6] As 5G investment becomes more critical to the digital economy, it will be important for NRAs to select spectrum award procedures favouring investment in infrastructure and maximizing economic impact.

Unlicensed spectrum enables NRAs to allow access to spectrum – but this arrangement leads to uncertainty as regards tenure of investments, because of obligations to operate on a basis of non-interference and non-protection. In addition, controlling interference can be difficult, if not impossible to manage. For this reason, unlicensed spectrum is more appropriate in high-frequency bands – such as the mmWave band with poorer propagation characteristics – and with low-power equipment to meet stringent limits of primary services, and for more localized usages. In view of these factors, the use of unlicensed spectrum may be considered by NRAs for instance in small cell deployments.

The GSMA consider that licensed spectrum is essential in guaranteeing high-quality 5G services, while unlicensed spectrum can play a complementary role in enhancing user experience.[7]

Key finding: Policy-makers may consider the use of licensed, unlicensed and shared spectrum to create a balanced spectrum ecosystem – one that encourages investment, makes efficient use of spectrum and promotes competition.

5.8 5G pilots

Policy-makers in governments and NRAs are encouraging early technology pilots to promote early investment in 5G networks and infrastructure, and to aid their understanding of 5G technologies (see Box 18).

Box 18: Government-led 5G initiatives

- The Government of Korea (Rep. of), via the NISA, established 5G pilot networks at the 2018 Winter Olympics, providing futuristic experiences such as augmented reality-based navigation.
- A GBP 17.6 million government grant has been awarded to a consortium led by the University of Warwick to develop a UK central test bed for connected autonomous vehicles (CAVs). Small cells will be deployed along a route through Coventry and Birmingham where the CAVs will be tested.
- The FCC (US) has encouraged applications from the research community for experimental licences for radio frequencies not granted or assigned, to promote innovation and research through experiments in defined geographic areas.
- The EC Horizon 2020 work programme (2018-2020) is promoting innovation in 5G involving the EU, China, Taiwan, China and the US. Activities include end-to-end testing of cross-border connected and automated mobility, and 5G trials across multiple vertical industries.
- The Federated Union of Telecommunications Research Facilities for an EU-Brazil Open Laboratory (FUTEBOL), is creating research that promotes experimental telecommunication resources in Brazil and Europe. FUTEBOL will also demonstrate use cases based on IoT, heterogeneous networks and C-RAN.

[6] Additional information on the economic aspects of spectrum management can be found in Report ITU-R SM.2012: https://www.itu.int/pub/R-REP-SM.2012

[7] "5G Spectrum Public Policy Position", GSMA, 2016: https://www.gsma.com/iot/iot-knowledgebase/gsma-public-policy-position-5g-spectrum/

Box 18: Government-led 5G initiatives (continued)

- The Russian Ministry of Communications concluded an agreement with Rostelecom and Tattelecom to create an experimental 5G zone in the hi-tech city of Innopolis.

Sources: https://goo.gl/JWFBCY (Korea Rep. of), https://goo.gl/FnLZCd (UK), https://goo.gl/wNVZqs (US), https://goo.gl/iXkYQ o (Europe), https://goo.gl/VNeDwn (EU-Brazil), https://goo.gl/4DySs2 (Russia)

In addition, the telecoms sector, comprising operators, vendors and research institutes, has been participating in 5G test beds independently of NRA or government intervention (see Box 19).

Box 19: Commercially led 5G test beds

- Telstra (Australia) is working with Ericsson on key 5G technologies including massive MIMO, beamforming, beam tracking and waveforms. Telstra and Ericsson achieved download speeds of between 18 Gbit/s and 22 Gbit/s during the first live trial of 5G in Australia. Optus also completed a 5G trial with Huawei, reaching the fastest speeds in Australia so far of 35 Gbit/s.
- Italian mobile operator Wind Tre, Open Fibre (Italy's wholesale fibre operator) and Chinese vendor ZTE have announced a partnership to build what they say will be Europe's first 5G pre-commercial network in the 3.6– 3.8 GHz band. They will also collaborate with local universities, research centres and enterprises to test and verify 5G technical performance, network architecture, 4G/5G network integration and future 5G use cases – including augmented reality or virtual reality, smart city, public safety and 5G healthcare. The pilot project will run until December 2021.
- A 5G pilot network was deployed in and around the Kazan Arena stadium (Russia) for the World Cup 2018 football tournament in a project led by MegaFon. Rostelecom is also partnering with Nokia on a 5G pilot wireless network located at a Moscow business park to test various 5G usage scenarios.
- Verizon (US) announced it is planning 5G tests in several US cities. The roll-outs will be based on wireless backhaul rather than fibre. AT&T also indicated that it will launch 5G fixed-wireless customer trials based on its recent trials in Austin where it achieved 1 Gbit/s speeds and sub-10 milliseconds latency. The tests will be conducted using equipment from Ericsson, Samsung, Nokia and Intel.
- Comsol plans to launch South Africa's first 5G wireless network. Comsol's trial will test the performance of 5G in real-world conditions using small cells in addition to macro solutions. It is likely that Comsol will offer fixed-wireless service to compete with fibre-to-the-home (FTTH) services.
- Huawei and NTT DOCOMO achieved a 4.52 Gbit/s downlink speed over 1.2km. Huawei supplied one of its 5G base stations, which supports massive MIMO and beamforming technologies in addition to its 5G core network.

Sources: https://goo.gl/cWTC31 (Australia), https://goo.gl/tYspR9 (Italy), https://goo.gl/EQftwd (Russia), https://goo.gl/yxaoyy (US), https://goo.gl/VeuiaW (South Africa), https://goo.gl/Teq6e2 (Japan)

Key finding: Policy-makers may consider encouraging 5G pilots and test beds to test 5G technologies and use cases and to stimulate market engagement.

6 Example of costs and investment implications

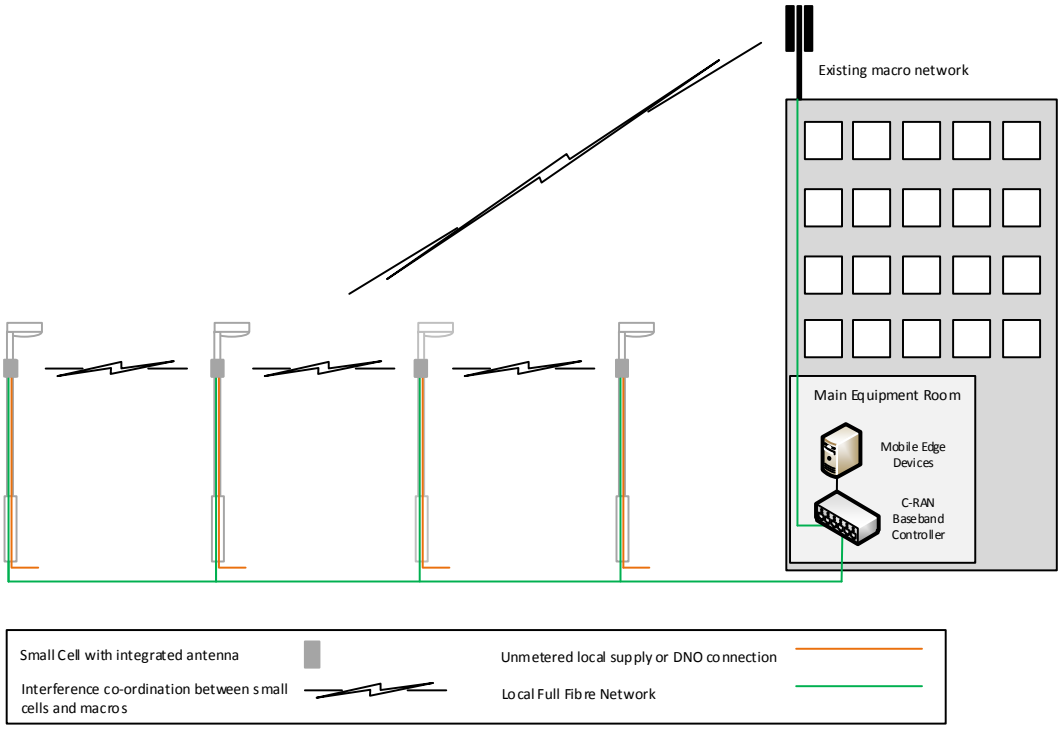
The deployment of small cells in dense urban areas is likely to represent a key investment for mobile operators in the run up to 5G. This section develops an example of a high-level cost model to estimate the potential investment required by a wireless operator to deploy a 5G-ready small cell network.

6.1 Overview

In the run up to 5G, operators are likely to focus on enhancing existing 4G coverage in urban areas through the deployment of small cells. These will increase the amount of network capacity available, improve street level coverage and enhance overall network quality, as would be required by 5G networks. Most of these deployments will occur in densely populated urban centres or cities.

For the purpose of this exercise we have assumed a small cell network is deployed by an independent wireless operator on a wholesale basis to mobile operators. This approach reduces the total cost of ownership (TCO) to mobile operators and increases the attractiveness of small cells to mobile operators. A typical small cell solution that is currently in deployment across parts of Europe and the US is depicted in Figure 7. Although this approach assumes a fibre backhaul strategy, wireless backhaul can be considered in cases where it is not commercially feasible to deploy a fibre backhaul network.

Figure 7: Typical neutral host wholesale small cell solution



The solution is comprised of the following elements:

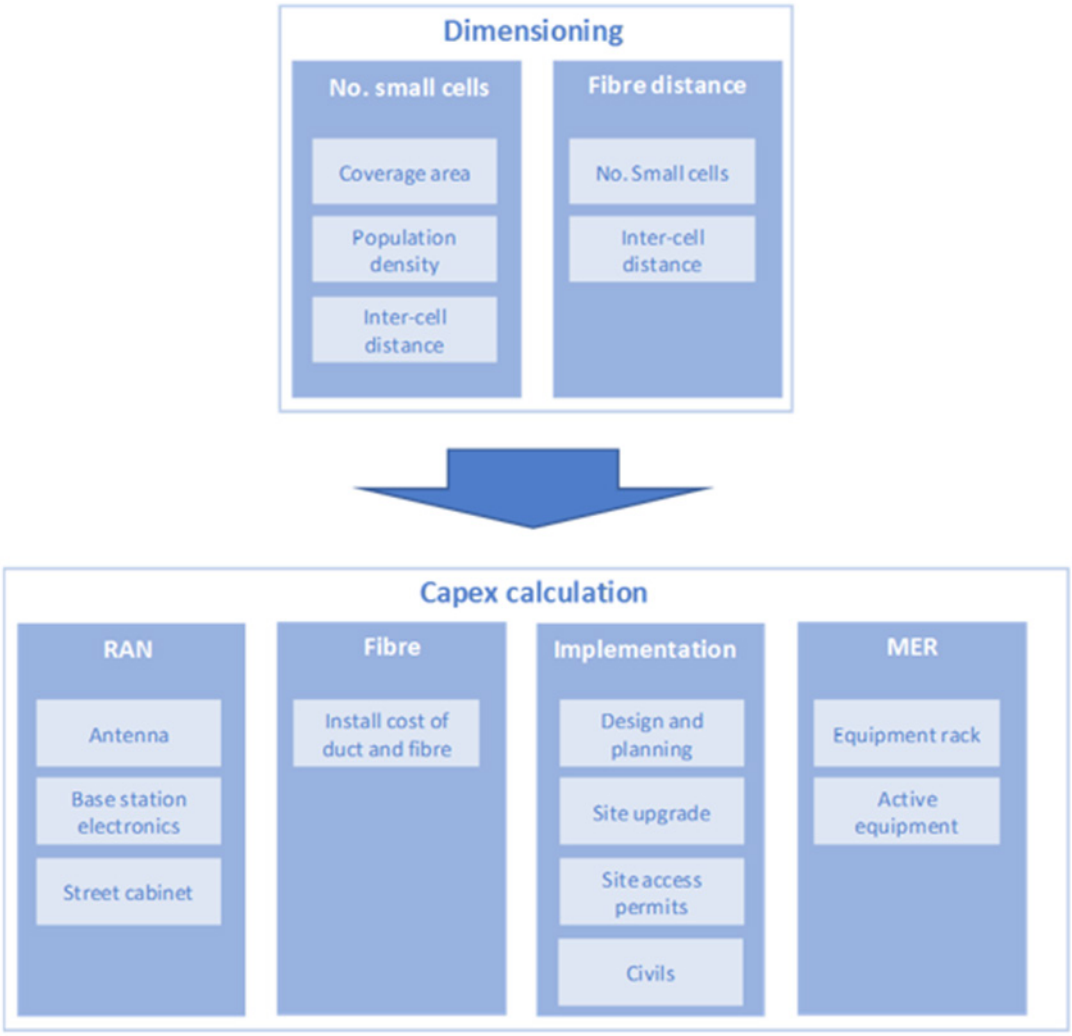
- **Antennae** – discreet high-performance antenna system which shapes the mobile operator's signal to maximize service performance for end users.
- **Street lights** – deployment of antennae on existing street lights to minimize aesthetic disruption.

- **Street cabinets** – shared accommodation hosting mobile operator radio equipment, battery backup and control equipment.
- **Fibre network** – high speed fibre that connects the radio network with the core network. Note that in some cases it may be more cost effective to use wireless backhaul.
- **Main Equipment Rooms (MER)** – A series of localized, shared main equipment room and a central point of interconnect to mobile operator backhaul networks.

6.2 Methodology

The focus of the model is to understand the initial capital expenditure investment of deploying a small cell network; the model only takes capital expenditure into consideration and excludes operating costs such as electricity, rent and maintenance. Being a wholesale model, it does not include mobile operator radio equipment costs as these will be provided by each operator. Due to the uncertainty surrounding the cost of 5G spectrum and investment in NFV/SDN, these costs are also excluded – as are site acquisition costs which can vary significantly from one city to another. Figure 8 shows that there are two steps to developing the cost model: dimensioning and CAPEX calculation.

Figure 8: Typical neutral host wholesale small cell solution



Network dimensioning estimates the number of small cells and amount of fibre needed and are calculated based on the required coverage area, population density and intercell cell-site distances. The outputs of the dimensioning step are then used to determine the total investment CAPEX required

to implement the small cell solution for the RAN, fibre, the main equipment room, implementation and design.

The model assumes the following cost elements:

- **RAN**, which includes the cost of antenna, street cabinet and base station electronics such as battery backup and network maintenance modules.
- **Implementation costs**, which include design and planning costs, site upgrade costs, permit costs and civils costs to lay street cabinets.
- **Fibre network**, which includes the provision of 144 fibre and new ducts along the route of the activated street assets.
- **Main equipment room (MER),** comprising a single rack and termination equipment to provide an interconnection between the mobile operators and the dark fibre network in a co-location site.

Note that the actual costs may vary according to each country as labour costs, exchange rates, equipment costs and taxes will vary in each country. The cost model assumes a Western country with a highly competitive market comprised of four mobile operators, advanced levels of 4G coverage and low urban fibre density.

6.3 Scenarios

The above methodology is used in two scenarios to provide an estimate of the cost of deploying a fibre connected small cell solution in the central business district – scenario 1 is a large, dense city and scenario 2 is a small, less dense city. In both, it is assumed that the city benefits from advanced levels of macro 4G coverage and the network demand characteristics are such that the investment case for 5G based small cells connected by fibre is commercially attractive.

Scenario 1 – large densely populated city

In this scenario the following assumptions are made:

- Proposed urban coverage area: 15 sq km
- Population density of coverage area: 12 000 people per sq km
- Inter-site small cell distance: 150 m.

Scenario 2 – small medium density city

- Proposed urban coverage area: 3 sq km
- Population density of coverage area: 3 298 people per sq km
- Inter-site small cell distance: 200 m.

A larger, denser city puts a higher strain on the mobile network, and therefore requires small cells to be sited closer together. For this reason, the distance between small cell sites is lower in scenario 1 compared to scenario 2.

6.4 Results

Figure 9 and Figure 10 show that the CAPEX required to deploy a fibre-connected small cell network can range from USD 6.8 million for a small city to USD 55.5 million for a large, dense city. The cost of deploying a small cell network in a dense city is greater per square kilometer because of the greater density of small cells deployed, owing to the shorter distance between small cell sites.

Figure 9: CAPEX for scenario 1 – large dense city

Item	Value
Total CAPEX (USD millions)	**55.5**
Number of small cell sites	1 027
Cost per square km (USD millions)	3.7
CAPEX per site (USD thousands)	54.1

Figure 10: CAPEX for scenario 2 – small less dense city

Item	Value
Total CAPEX (USD millions)	**6.8**
Number of small cell sites	116
Cost per square km (USD millions)	2.3
CapEx per site (USD thousands)	58.6

The total CAPEX incurred by each operator will vary according to population, population density, current 4G coverage and the proposed coverage area. In addition, the cost of fibre deployment will be lower in cities where there is a high availability of, and easy access to, dense fibre networks or ducts. Where wireless backhaul is more cost-effective than fibre, the backhaul costs will be significantly reduced. In cities where the existing macro network density is high (e.g. in Madrid where site access is less restrictive than in other cities), there will be less need for small cells. Similarly, mobile operators with large spectrum assignments need not densify their networks as much with small cells.

Figure 11 provides a breakdown of the cost components for scenario 1 and scenario 2 and shows that implementation costs are the most significant cost element. In regions where labour costs are low, deployment costs will be less than those estimated in this report.

Figure 11: Contribution to CAPEX

Small cell distance	Scenario 1	Scenario 2
RAN equipment (antenna, street cabinet, base station electronics, battery backup and network maintenance modules)	25%	24%
Implementation costs (design and planning costs, site upgrade costs, permit costs and civils costs to lay street cabinets)	50%	46%
Fibre (provision of 144 fibre along the route of activated street assets)	25%	30%
MER (single rack and termination equipment)	<0.1%	<0.1%

6.5 Independent cost estimates

The above costs – in particular CAPEX per site – are in line with industry estimates. AT&T estimate that the deployment costs can range from USD 20 000 to USD 50 000 per site assuming fibre backhaul for sites, something AT&T has in abundance.[1,2] According to Nokia, site CAPEX is estimated to be between USD 40 000 to USD 50 000 for a site that requires trenching and power.

[1] https://www.rcrwireless.com/20170814/carriers/att-small-cell-cactus-antenna-concealment-tag17
[2] http://www.telecompetitor.com/cfo-extensive-fiber-assets-firstnet-give-att-an-advantage-on-5g-backhaul/

Work undertaken by independent analysts estimates a total cost of ownership of GBP 71 billion to build a ubiquitous 5G network in the UK delivering 50 Mbit/s, built in 2020 and operated until 2030. This reduces to GBP 38 billion when infrastructure sharing is encouraged.[3]

Other reports estimate the cost of deploying 5G across the US as being in the order of USD 300 billion. In Europe investment costs are expected to range between EUR 300 billion to EUR 600 billion according to one mobile operator.[4]

Although these reports do not state the frequency spectrum used to derive the analysis, it is assumed that much of the cost results from network densification (through small cell deployment) – necessary for the smaller cell sizes required because of the use of higher mMWave frequency spectrum used by 5G, e.g. above 24GHz (mentioned in Section 3.5).

6.6 Investment models

Given the considerable CAPEX investment required in deploying 5G, operators face major challenges in making the investment case for 5G. Policy-makers will need to consider alternative investment models (for example PPPs, loans, challenge funds and investment vehicles) to ensure high upfront CAPEX costs are not a barrier for wireless providers.

Some examples of government interventions have already been described in Section 5, which include a range of PPP programmes. These programmes can either be: i) publicly led, where the government builds and owns fibre networks, as in Qatar; or ii) privately led, where the government partly funds the development of fibre networks in partnership with the market, as in Germany.

Other approaches include offering grants to local authorities, as in the UK, to construct and upgrade passive assets (such as ducts, fibre networks, data centres, street furniture, etc.). Governments can also offer low-cost loans to operators in return for a guaranteed investment from the operators, as in Malaysia.

Where operators prefer to access capital from private markets, governments can set up investment funds in collaboration with established private sector fund managers to provide operators with equity. The equity would then support operator network expansion programmes.

Many other PPP models for incentivizing investment in telecom networks do exist and have been written about extensively.[5]

Not all 5G deployments require government intervention. Some small cell and pre-5G deployments to date have been privately financed, as demonstrated in previous sections.

[3] https://www.itrc.org.uk/wp-content/PDFs/Exploring-costs-of-5G.pdf
[4] http://www.lightreading.com/mobile/5g/how-much-will-5g-cost-no-one-has-a-clue/a/d-id/733753
[5] "Investment strategies for broadband deployment and access to the digital economy", ITU, 2016

7 Conclusion

Until the investment case for 5G is compelling, the industry and policy-makers should approach investment with caution while enhancing the availability and quality of existing 4G networks.

5G is expected to play a key role in digital economies, improving economic growth, enhancing citizens' life experiences and creating new business opportunities.

Despite such benefits, care must be taken in establishing the commercial case and whether 5G is a real priority for the economy. A 5G investment decision must be backed by a sound investment case.

Operators are sceptical about the return on investment because of high investment levels. They are currently investing in 5G test beds and pilot networks in large dense cities with advanced 4G deployments and with supporting infrastructure more suited to network economics.

This 'city led' strategy is likely to have an adverse impact on the digital divide since the case for 5G in rural areas is less convincing. Local authorities and regulators should recognize this risk and should counter it. This can be done by supporting commercial and legislative incentives to stimulate investment for the provision of fibre networks and affordable wireless coverage through the use of sub-1 GHz bands.

An overhaul of the regulatory, government and local authority approach to digital policy is needed to boost the roll-out of 5G networks. Importantly, this includes ensuring affordable access to public assets thereby strengthening the commercial case to invest in small cell infrastructure and 5G spectrum.

Annex A

In preparation for WRC-19, ITU-R is undertaking sharing and compatibility studies in the frequency bands agreed at WRC-15, and which could potentially be identified for implementation of IMT-2020 (5G).

ITU-R Study Group 5

ITU-R Study Group 5 (Terrestrial systems) is responsible for the overall radio system aspects of IMT systems and for studies related to the land mobile service, including wireless access in the fixed service.

Recommendations and reports developed by ITU-R include:

- ITU-R M.1457 "Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-2000 (IMT-2000)". Specifications for IMT-2000.
- ITU-R M.2012 "Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications Advanced (IMT-Advanced)". Specifications for IMT-Advanced.
- ITU-R M.2083 "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond", includes a broad variety of capabilities associated with envisaged usage scenarios. Furthermore, it addresses the objectives of the future development of IMT-2020, which includes further enhancement of existing IMT and the development of IMT-2020.
- ITU-R M.2370 "IMT Traffic estimates for the years 2020 to 2030". As traffic demand for mobile broadband communications represented by IMT is increasing, the transport network in the mobile infrastructure is becoming an important application that requires special consideration.
- ITU-R M.2375 "Architecture and topology of IMT networks", offers an overview of the architecture and topology of IMT networks and a perspective on the dimensioning of the respective transport requirements in these topologies – assisting relevant studies on the transport network in the mobile infrastructure.
- ITU-R M.2376 "Technical feasibility of IMT in bands above 6 GHz", expects that usage of higher frequencies will be one of the key enabling components of future IMT.
- ITU-R M.2410 "Minimum requirements related to technical performance for IMT-2020 radio interface(s)", describes the key requirements related to the minimum technical performance of IMT-2020 candidate radio interface technologies.
- ITU-R M.2411 "Requirements, evaluation criteria and submission templates for the development of IMT-2020", describes the requirements and the submission process of the technologies.
- ITU-R M.2412 "Guidelines for evaluation of radio interface technologies for IMT-2020", provides guidelines for the evaluation of the radio interface.

Additional documentation can be found at: *https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/Pages/default.aspx*.

The standardization activities within ITU also cover the needs for backhauling in support of 5G development – including studies of several radiocommunication solutions, such as satellite communications, the use of high speed radio relays and high-altitude platform stations (HAPS).

ITU-T Study Group 13

ITU-T Study Group 13 (Future networks) is ITU's lead group for 5G wireline studies and continues to support the shift to software-driven network management and orchestration. The group is progressing draft 5G standards, addressing subjects including network architectures, network capability exposure, network slicing, network orchestration, network management-control, and frameworks to ensure high quality of service.

5G wireline standards developed by ITU-T Study Group 13 and approved in 2017-2018 include:

- Recommendation ITU-TY.3071 "Data Aware Networking (Information Centric Networking) – Requirements and Capabilities" will support ultra-low latency 5G communications by enabling proactive in-network data caching and limiting redundant traffic in core networks.
- Recommendation ITU-T Y.3100 "Terms and definitions for IMT-2020 network" provides a foundational set of terminology to be applied universally across 5G-related standardization work.
- Recommendation ITU-T Y.3101 "Requirements of the IMT-2020 network" provides general principles of the IMT-2020 network, then specifies requirements for overall non-radio aspects of the IMT-2020 network from both the service and network operation points of view.
- Recommendation ITU-T Y.3102 "Framework of IMT-2020 network" specifies the framework for overall non-radio aspects of the IMT-2020 network: the key features of the IMT-2020 network and architectural design considerations.
- Recommendation ITU-T Y.3111 "IMT-2020 network management and orchestration framework" establishes a framework and related principles for the design of 5G networks.
- Recommendation ITU-T Y.3112 "Framework for the support of Multiple Network Slicing" describes the concept of network slicing and the high-level requirements and high-level architecture for multiple network slicing in IMT-2020 network, illustrated by the use cases.
- Recommendation ITU-T Y.3110 "IMT-2020 network management and orchestration requirements" describes the capabilities required to support emerging 5G services and applications.
- Recommendation ITU-T Y.3150 "High level technical characteristics of network softwarization for IMT-2020". Taking from global recognition of the usefulness of network slicing technology, as the most typical substantiation of the network softwarization approach, this Recommendation describes how network softwarization and network slicing contribute to IMT-2020 systems, explores network slicing from two viewpoints: vertical and horizontal aspects, details network slicing for mobile fronthaul/backhaul, introduces the advanced data-plane programmability, and capability exposure.
- Recommendation ITU-T Y.3130 "Requirements of IMT-2020 fixed mobile convergence" specifies service related requirements such as unified user identity, unified charging, service continuity and guaranteed quality of service support, and network capability requirements such as control plane convergence, user data management, capability exposure and cloud-based infrastructure, to support fixed mobile convergence in IMT-2020 networks.
- ITU-T Supplement 35 to Y.3033-series "Data-aware networking- scenarios and use cases" lists a set of service scenarios and use cases supported by data-aware networking (DAN) including: 1) content dissemination; 2) sensor networking; 3) vehicular networking; 4) automated driving; 5) networking in a disaster area; 6) advanced metering infrastructure in a smart grid; 7) proactive video caching; 8) in-network data processing; 9) multihoming; and 10) traffic engineering. It provides informative illustrations and descriptions of how DAN can be designed, deployed and operated to support DAN services. In addition, the benefits of data aware networks to the scenarios and use cases, as well as several migration paths from current networks to data-aware networks, are elaborated.
- Supplement 44 to ITU-T Y.3100 series "Standardization and open source activities related to network softwarization of IMT-2020" summarizes open-source and standardization initiatives relevant to ITU's development of standards for network softwarization.
- Supplement 47 to the ITU-T Y.3070-series Recommendations "Information-Centric Networking - Overview, Standardization Gaps and Proof-of-Concept" provides the overview of information-centric networking and describes the fifteen standardization gaps and five proof-of-concept based on the ICN related contents investigated by ITU-T Focus Group on IMT-2020 (FG IMT-2020) during 2015-2016.

ITU-T Study Group 15

In addition, ITU standardization work on the wireline elements of 5G systems continues to accelerate. ITU-T Study Group 15 (SG15- Transport, access and home) develops standards for providing transport support for 5G systems.

SG15 work related to 5G includes:

- G-series Technical Report (GSTR-TN5G) "Transport network support of IMT-2020/5G".
- Supplement 55 to G-series Recommendations "Radio-over-fibre (RoF) technologies and their applications" provides general information on radio over fibre technologies and their applications in optical access networks. This technology is used in the radio shadow.
- Supplement 56 to G-series Recommendations "OTN transport of CPRI signals" describes alternatives for mapping and multiplexing CPRI client signals into the OTN. This Supplement relates to Recommendations ITU-T G.872, ITU-T G.709/Y.1331, ITU-T G.798 and ITU-T G.959.1.
- Recommendation ITU-T G.987 series: 10-Gigabit-capable passive optical networks (XG-PON)
- Recommendation ITU-T G.9807 series: 10-Gigabit-capable symmetric passive optical network (XGS-PON)
- Recommendation ITU-T G.989 series: 40-Gigabit-capable passive optical networks 2 (NG-PON2)
- Recommendation ITU-T G.RoF "Radio over Fibre systems" (under development)
- New Supplement to G-series Recommendations (G.sup.5GP) "5G wireless fronthaul requirements in a PON context" (under development)
- Recommendation ITU-T G.9700 series: Fast access to subscriber terminals (G.fast)
- Recommendation ITU-T G.709 series: Optical Transport Network (OTN)
- Draft Recommendation ITU-T G.ctn5g: Characteristics of transport networks to support IMT-2020/5G (under development)
- Draft Supplement to G-series Recommendations G.Sup.5gotn: Application of OTN to 5G transport (under development)
- Recommendation ITU-T G.695: Optical interfaces for coarse wavelength division multiplexing applications
- Recommendation ITU-T G.698.4: Multichannel bi-directional DWDM applications with port agnostic single-channel optical interfaces
- Recommendation ITU-T G.959.1: Optical transport networks physical layer interfaces

In addition, SG15 develops standards on network synchronization for supporting 5G networks (Recommendation ITU-T G.8200 series).

ITU-T Study Group 12

Related work underway in ITU-T Study Group 12 (performance, quality of service, quality of experience), includes:

- Draft Recommendation ITU-T G.IMT2020: QoS framework for IMT-2020. Review of SG12 QoS frameworks in the context of IMT-2020.
- Draft Recommendation ITU-T Y.cvms: Considerations for realizing virtual measurement systems. As network service providers seek to take advantage of the scale, flexible deployment, and cost reductions first realized in cloud computing, they have begun to define new architectures for their infrastructure in order to realize network function virtualization (NFV). At the same time, measurement functions will be implemented for deployment as virtual functions. This

document makes recommendations in key areas such as on-demand deployment and accuracy considerations. Development of virtualized measurement systems in areas highly relevant to SG12 work are in the early stages, so this Recommendation is timely.

- Draft Recommendation ITU-T G.QoE-5G: Quality of experience (QoE) factors for new services in 5G networks.

In addition, SG12 is developing Recommendations concerning the quality of experience of augmented reality (AR) and virtual reality (VR), which are among the most talked about 5G use cases.

ITU-T Study Group 11

ITU-T Study Group 11 (Protocols and test specifications) is studying the 5G control plane, relevant protocols and related testing methodologies.

- Supplement 67 to Q-series Recommendations "Framework of signalling for software-defined networking" enables the development of a signalling protocol(s) capable of supporting traffic flows in SDN environment.
- Recommendations ITU-T Q.3710-Q.3899 series on Signalling requirements and protocols for SDN.
- Recommendation ITU-T Q.3315 "Signalling requirements for flexible network service combination on broadband network gateway". As the key position to offer broadband network services, the broadband network gateway (BNG) should be able to support flexible service combination, new services introduction and provisioning. Q.3315 describes the signalling requirements, based on the service platform BNG architecture, needed to achieve outstanding benefits like easy deployment of network services, fine grained network services, etc.

ITU-T Study Group 5

ITU-T Study Group 5 (Environment, climate change and circular economy) has assigned priority to its emerging study of the environmental requirements of 5G systems. ITU-T SG5 is developing a series of international standards (ITU-T Recommendations), Supplements and Technical Reports that will study the environmental aspects related to: electromagnetic compatibility (EMC), electromagnetic fields (EMF); energy feeding and efficiency, and resistibility. The ITU-T Recommendations and Supplements developed by ITU-T SG5 include:

- Supplement ITU-T K.Suppl.8 "Resistibility analysis of 5G systems" analyses 5G system resistibility requirements for lightning and power fault events.
- Supplement ITU-T K.Suppl.9 "5G technology and human exposure to RF EMF" contains an analysis of the impact of the implementation of 5G mobile systems with respect to the exposure level of electromagnetic fields (EMF) around radiocommunication infrastructure.Supplement ITU-T K.Suppl.10 "Supplement ITU-T K.Suppl.10 "Analysis of electromagnetic compatibility aspects and definition of requirements for 5G mobile systems" provides guidance on the EMC compliance assessment considerations for 5G systems. It focuses on possible emission and immunity requirements for 5G systems.
- Supplement ITU-T K.Suppl.14 "The impact of RF-EMF exposure limits stricter than the ICNIRP or IEEE guidelines on 4G and 5G mobile network deployment" provides an overview of some of the challenges faced by countries, regions and cities which are about to deploy 4G or 5G infrastructures. It also provides information on a simulation on the impact of RF-EMF limits that was carried out in Poland as an example of a wider phenomenon, which is applicable to several other countries, which have set limits that are stricter than those contained in the ICNIRP or IEEE guidelines.
- Recommendation ITU-T L.1220 "Innovative energy storage technology for stationary use- Part 1: Overview of energy storage" introduces an open series of documents for different families of

technologies (battery systems, super-capacitor systems, etc.) that will be enriched progressively as new technologies emerge that may have a possible significant impact in the field of energy storage.

- ITU-T L.Suppl.36 to ITU-T L.1310 "Study on methods and metrics to evaluate energy efficiency for future 5G systems" analyses the energy efficiency issues for future 5G systems.

Earlier the ITU-T Focus Group IMT-2020 produced a set of technical reports elaborating the different facets of the 5G wireline aspects "ITU-T Focus Group IMT-2020 deliverables flipbook, 2017":

https://itu.int/en/publications/Documents/tsb/2017-IMT2020-deliverables/mobile/index.htm

The preparatory work of ITU-T for the introduction on IMT-2020 is depicted in the "5G Basics flipbook, 2017":

https://itu.int/en/publications/Documents/tsb/2017-IMT2020-deliverables/mobile/index.htm

See ITU-T webpages at: https://itu.int/en/ITU-T/

International
Telecommunication
Union
Telecommunication
Development Bureau
Place des Nations
CH-1211 Geneva 20
Switzerland

ISBN: 978-92-61-27591-4

9 789261 275914

Published in Switzerland
Geneva, 2018

ICT Regulatory Tracker 2018

#ITUdata

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Afghanistan	15	20	19	19.33	73.33
Albania	18	16	25	24	83.00
Algeria	18	16	16	11.5	61.50
Andorra	6	8	8	0	22.00
Angola	14	20	20	10.67	64.67
Antigua and Barbuda	8	11.5	8	13.33	40.83
Argentina	17	20	21	28	86.00
Armenia	19	19.5	20	27	85.50
Australia	19	21.5	26	28	94.50
Austria	18	16.5	28	27	89.50
Azerbaijan	8	13.5	24	25	70.50
Bahamas	19	18.5	26	25.33	88.83
Bahrain	17	18	26	26.33	87.33
Bangladesh	17	20	15	22.67	74.67
Barbados	17	12.5	18	21	68.50
Belarus	6	11.5	11	16	44.50
Belgium	18	19	30	27	94.00
Belize	17	18.5	20	7.33	62.83

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Benin	16	16	21	12	65.00
Bhutan	15	20	16	18.33	69.33
Bolivia (Plurinational State of)	9	9	8	8.5	34.50
Bosnia and Herzegovina	19	21	27	26	93.00
Botswana	18	22	19	26	85.00
Brazil	16	18.5	26	28	88.50
Brunei Darussalam	15	17	17	12.33	61.33
Bulgaria	19	16.5	28	28	91.50
Burkina Faso	19	19	20	26	84.00
Burundi	11	18	12	23	64.00
Cabo Verde	17	20	23	21.33	81.33
Cambodia	13	17	14	21.33	65.33
Cameroon	17	18	16	13	64.00
Canada	19	16.5	30	20	85.50
Central African Rep.	14	18	9	17	58.00
Chad	15	16	13	14.33	58.33
Chile	14	20	18	27	79.00
China	7	11	16	15	49.00
Colombia	15	15	22	27	79.00
Comoros	17	19	24	22.33	82.33

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Congo (Rep. of the)	17	17	22	19.67	75.67
Costa Rica	19	16	26	24	85.00
Croatia	19	19	28	28	94.00
Cuba	2	12	14	5	33.00
Cyprus	18	16	28	23.67	85.67
Czech Republic	17	17	30	25	89.00
Côte d'Ivoire	17	15.5	14	15.33	61.83
Dem. Rep. of the Congo	14	20	20	25.33	79.33
Denmark	18	18	28	23.67	87.67
Djibouti	0	2.5	2	0	4.50
Dominica	11	15.5	20	26	72.50
Dominican Rep.	19	19.5	28	28	94.50
Ecuador	20	18.5	21	26	85.50
Egypt	15	20.5	21	24.33	80.83
El Salvador	19	14.5	14	26	73.50
Equatorial Guinea	13	15	13	9.33	50.33
Eritrea	8	11	4	2	25.00
Estonia	14	20	26	27	87.00
Eswatini	19	19	14	7.33	59.33
Ethiopia	7	12	8	2	29.00
Fiji	13	14	19	17	63.00
Finland	18	17	30	27	92.00

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
France	18	20	30	26	94.00
Gabon	15	17	16	15	63.00
Gambia	20	19	16	18.67	73.67
Georgia	18	16.5	30	28	92.50
Germany	16	20.5	30	27	93.50
Ghana	18	21	22	27	88.00
Greece	20	17	28	26.33	91.33
Grenada	14	17	20	23	74.00
Guatemala	12	12.5	10	18.67	53.17
Guinea	16	18	22	12.33	68.33
Guinea-Bissau	10	10	8	18	46.00
Guyana	18	18	15	11	62.00
Haiti	14	19.5	10	15	58.50
Honduras	17	19	26	20	82.00
Hong Kong, China	18	18.5	20	27.33	83.83
Hungary	19	22	28	28	97.00
Iceland	18	18	22	28	86.00
India	18	14.5	20	23	75.50
Indonesia	16	13.5	18	25	72.50
Iran (Islamic Republic of)	19	19	28	16	82.00
Iraq	17	21.5	16	3.33	57.83
Ireland	20	19	30	28	97.00

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Israel	8	11.5	28	24	71.50
Italy	18	22	30	27.33	97.33
Jamaica	19	12.5	19	28	78.50
Japan	8	11.5	26	27	72.50
Jordan	19	20	24	21.5	84.50
Kazakhstan	6	10	14	24	54.00
Kenya	18	21.5	21	27	87.50
Kiribati	13	18.5	4	12	47.50
Korea (Rep. of)	18	22	20	21.67	81.67
Kuwait	20	19	12	12	63.00
Kyrgyzstan	16	16.5	16	26	74.50
Lao P.D.R.	0	12	17	7.67	36.67
Latvia	18	16.5	30	26	90.50
Lebanon	8	18	5	0.67	31.67
Lesotho	16	17.5	16	18.33	67.83
Liberia	17	20	22	12.33	71.33
Libya	2	2.5	0	0	4.50
Liechtenstein	14	14	24	26.33	78.33
Lithuania	19	21	28	27	95.00
Luxembourg	18	17	22	26	83.00
Madagascar	17	17.5	18	17	69.50
Malawi	18	22	20	27	87.00

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Malaysia	18	22	24	23	87.00
Maldives	13	20	12	8.33	53.33
Mali	18	18	18	26.33	80.33
Malta	19	20	28	28	95.00
Marshall Islands	2	6.5	4	3	15.50
Mauritania	17	19	18	17	71.00
Mauritius	18	20.5	15	27.33	80.83
Mexico	19	17	26	28	90.00
Micronesia	0	4	4	0	8.00
Moldova	19	17.5	26	26	88.50
Monaco	0	15	8	12	35.00
Mongolia	18	19	18	14.67	69.67
Montenegro	19	19	28	28	94.00
Morocco	18	19.5	24	27	88.50
Mozambique	16	10.5	16	15.17	57.67
Myanmar	6	17	17	23.67	63.67
Namibia	19	17	22	12.67	70.67
Nauru	10	11.5	6	23	50.50
Nepal (Republic of)	18	17	11	22	68.00
Netherlands	19	18	28	28	93.00
New Zealand	17	13.5	22	28	80.50
Nicaragua	18	18	12	26	74.00

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Niger	15	20	20	19	74.00
Nigeria	17	20	20	21.33	78.33
North Macedonia	18	20	30	19	87.00
Norway	20	18.5	30	27	95.50
Oman	17	19	28	26.33	90.33
Pakistan	20	19	22	27	88.00
Palestine	4	11.5	13	13.67	42.17
Panama	19	21	20	26	86.00
Papua New Guinea	16	19.5	12	11	58.50
Paraguay	18	15.5	12	16.33	61.83
Peru	18	13	28	28	87.00
Philippines	16	12	17	22	67.00
Poland	16	17.5	28	27	88.50
Portugal	19	18	30	27	94.00
Qatar	14	18	21	16.67	69.67
Romania	18	19	28	27	92.00
Russian Federation	4	11	13	14	42.00
Rwanda	20	20	18	24.33	82.33
Saint Kitts and Nevis	5	15	6	20	46.00
Saint Lucia	16	18	24	27	85.00
Saint Vincent and the Grenadines	17	18	18	27	80.00

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Samoa	14	17	22	13.33	66.33
San Marino	0	4	2	16	22.00
Sao Tome and Principe	16	17	21	21	75.00
Saudi Arabia	19	22	29	22	92.00
Senegal	19	19	24	18	80.00
Serbia	20	19.5	26	27	92.50
Seychelles	6	12	16	28	62.00
Sierra Leone	16	19	14	7	56.00
Singapore	17	21.5	26	27	91.50
Slovakia	15	18.5	28	26.67	88.17
Slovenia	20	18.5	28	27	93.50
Solomon Islands	9	14	8	3.67	34.67
Somalia	14	19	10	24	67.00
South Africa	17	17	24	13.33	71.33
South Sudan	12	17	12	13.67	54.67
Spain	16	14	28	28	86.00
Sri Lanka	18	20	15	9.33	62.33
Sudan	15	20	18	18.67	71.67
Suriname	15	17	18	9.67	59.67
Sweden	19	20	24	26	89.00
Switzerland	18	18.5	30	27	93.50

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Syrian Arab Republic	19	15	15	6.33	55.33
Tajikistan	2	6	2	4	14.00
Tanzania	20	21	19	25	85.00
Thailand	20	19.5	22	19.83	81.33
Timor-Leste	13	21	3	5	42.00
Togo	15	22	20	12	69.00
Tonga	1	11	15	22.67	49.67
Trinidad and Tobago	18	19	22	26.33	85.33
Tunisia	19	16	25	14.67	74.67
Turkey	19	19.5	30	26	94.50
Turkmenistan	0	6	0	1.67	7.67
Tuvalu	0	4.5	0	5	9.50
Uganda	17	20	22	27	86.00
Ukraine	17	17.5	23	24	81.50
United Arab Emirates	19	21	27	16	83.00
United Kingdom	20	20	28	27	95.00
United States	19	17.5	28	24	88.50
Uruguay	17	17	20	13	67.00
Uzbekistan	7	6.5	2	6.33	21.83
Vanuatu	17	14.5	14	25.67	71.17
Venezuela	20	21.5	16	25	82.50

Cluster	C1: Regulatory Authority	C2. Regulatory Mandate	C3. Regulatory Regime	C4. Competition Framework	Overall Score
Max Score:	**20**	**22**	**30**	**28**	**100**
Country					
Viet Nam	10	19	24	13	66.00
Yemen	0	3	4	4	11.00
Zambia	19	18	15	19.67	71.67
Zimbabwe	20	19	18	17	74.00