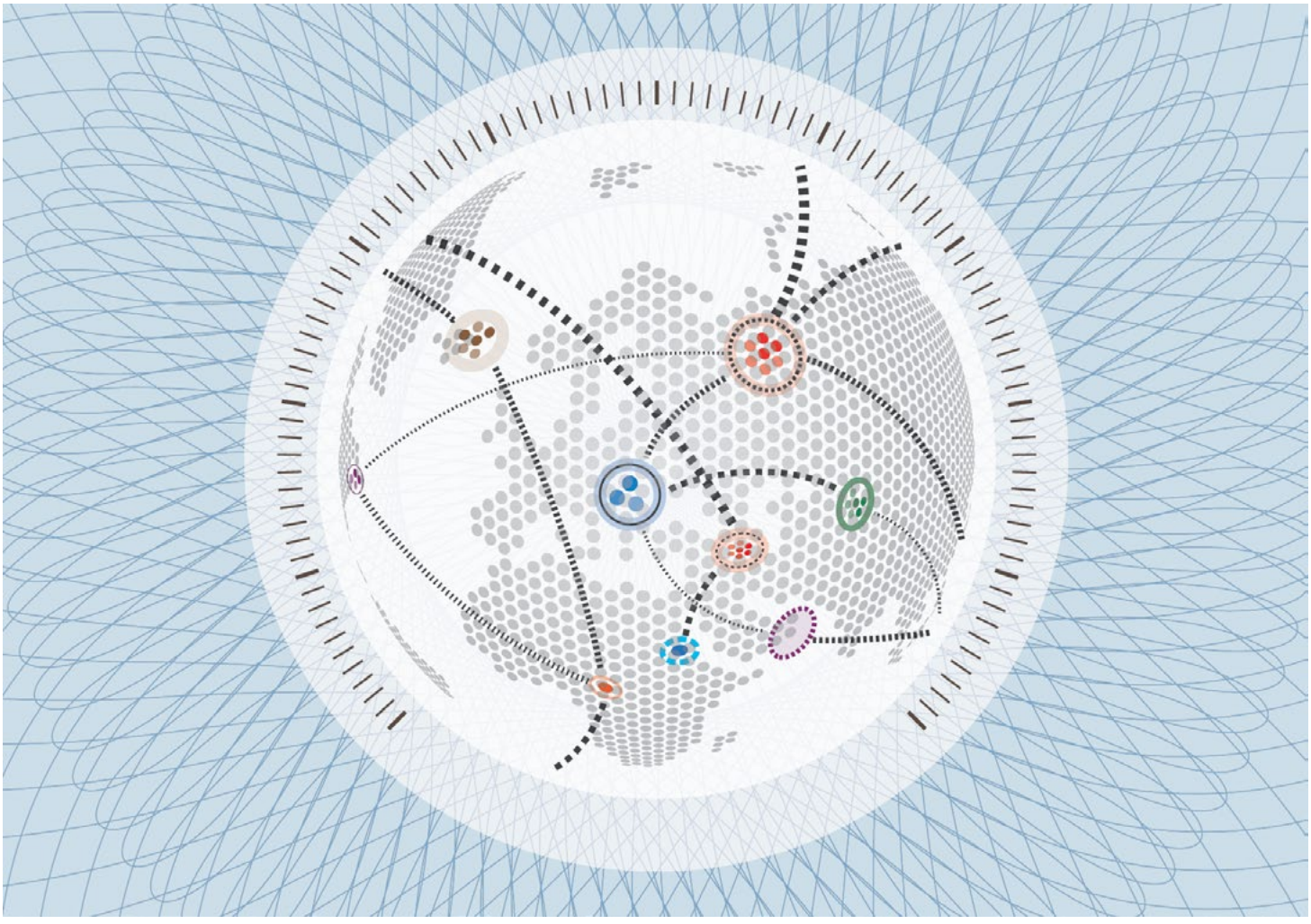
WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Insight Report

The Global Risks Report 2017

Subject topic: Emerging Technologies



Strategic Partner of the Report

ZURICH®

Part 3: Emerging Technologies

3.1: Understanding the Technology Risks Landscape

The emerging technologies of the Fourth Industrial Revolution (4IR) will inevitably transform the world in many ways – some that are desirable and others that are not. The extent to which the benefits are maximized and the risks mitigated will depend on the quality of governance – the rules, norms, standards, incentives, institutions, and other mechanisms that shape the development and deployment of each particular technology.

Too often the debate about emerging technologies takes place at the extremes of possible responses: among those who focus intently on the potential gains and others who dwell on the potential dangers. The real challenge lies in navigating between these two poles: building understanding and awareness of the trade-offs and tensions we face, and making informed decisions about how to proceed. This task is becoming more pressing as technological change deepens and accelerates, and as we become more aware of the lagged societal, political and even geopolitical impact of earlier waves of innovation.

Over the years *The Global Risks Report* has repeatedly highlighted technological risks. In the second edition of the *Report*, as far back as 2006, echoes of current concerns were noted in one of the technology scenarios we considered, in which the "elimination of privacy reduces social cohesion". This was classified as a worst-case scenario, with a likelihood of below 1%. In 2013, the *Report* discussed the risk of "the rapid spread of misinformation", observing that trust was being eroded and that incentives were insufficiently aligned to ensure the maintenance of robust systems of

Table 3.1.1: Twelve Key Emerging Technologies

Technology	Description
3D printing	Advances in additive manufacturing, using a widening range of materials and methods; innovations include 3D bioprinting of organic tissues.
Advanced materials and nanomaterials	Creation of new materials and nanostructures for the development of beneficial material properties, such as thermoelectric efficiency, shape retention and new functionality.
Artificial intelligence and robotics	Development of machines that can substitute for humans, increasingly in tasks associated with thinking, multitasking, and fine motor skills.
Biotechnologies	Innovations in genetic engineering, sequencing and therapeutics, as well as biological-computational interfaces and synthetic biology.
Energy capture, storage and transmission	Breakthroughs in battery and fuel cell efficiency; renewable energy through solar, wind, and tidal technologies; energy distribution through smart grid systems, wireless energy transfer and more.
Blockchain and distributed ledger	Distributed ledger technology based on cryptographic systems that manage, verify and publicly record transaction data; the basis of "cryptocurrencies" such as bitcoin.
Geoengineering	Technological intervention in planetary systems, typically to mitigate effects of climate change by removing carbon dioxide or managing solar radiation.
Ubiquitous linked sensors	Also known as the "Internet of Things". The use of networked sensors to remotely connect, track and manage products, systems, and grids.
Neurotechnologies	Innovations such as smart drugs, neuroimaging, and bioelectronic interfaces that allow for reading, communicating and influencing human brain activity.
New computing technologies	New architectures for computing hardware, such as quantum computing, biological computing or neural network processing, as well as innovative expansion of current computing technologies.
Space technologies	Developments allowing for greater access to and exploration of space, including microsatellites, advanced telescopes, reusable rockets and integrated rocket-jet engines.
Virtual and augmented realities	Next-step interfaces between humans and computers, involving immersive environments, holographic readouts and digitally produced overlays for mixed-reality experiences.

Source: The 12 emerging technologies listed here and included in the GRPS are drawn from World Economic Forum *Handbook on the Fourth Industrial Revolution* (forthcoming, 2017).

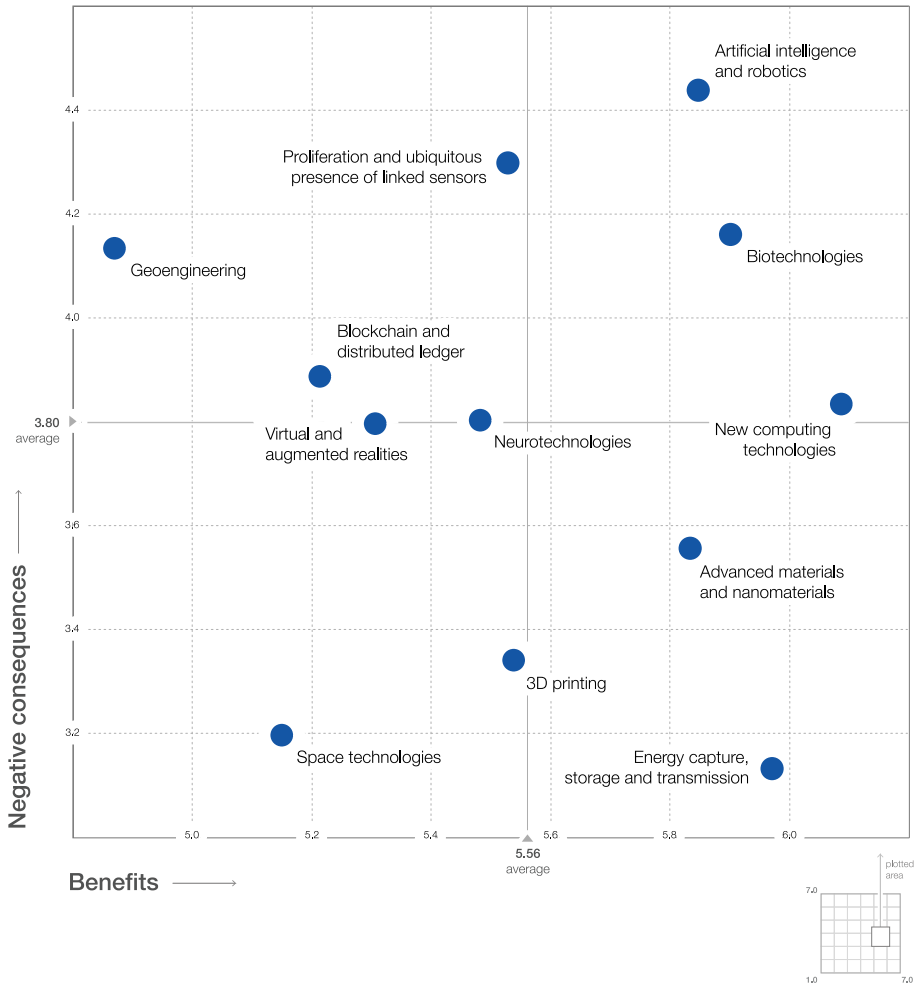
quality control or fact-checking. Four years later, this is a growing concern; in Chapter 2.1, the *Report* considers the potential impact of similar trends on the very fabric of democracy.

In 2015, emerging technology was one of the *Report*'s "risks in focus", highlighting, among other things, the ethical dilemmas that exist in areas such as artificial intelligence (AI) and biotechnology.

This year, the Global Risks Perception Survey (GRPS) included a special module on 12 emerging technologies (see Table 3.1.1). The results suggest that respondents are broadly optimistic about the balance of technological risks and benefits. Figure 3.1.1 shows that the average score is much higher for perceived benefits than it is for negative consequences. However, as Figure 3.1.2 makes clear, respondents still identify clear priorities for better governance of emerging technologies.

The remainder of this chapter highlights the particular challenges involved in creating governance regimes for fast-moving technologies, and then summarizes the key results of this year's GRPS special module on emerging technology. The chapter concludes with a discussion of the profound changes that new technologies will entail for businesses and of the cascading effects these changes may have on the global risk landscape.

Figure 3.1.1: Perceived Benefits and Negative Consequences of 12 Emerging Technologies



Source: World Economic Forum Global Risks Perception Survey 2016.

Note: See Appendix B for more details on the methodology.

Governance Dilemmas

How to govern emerging technologies is a complex question. Imposing overly strict restrictions on the development of a technology can delay or prevent potential benefits. But so can continued regulatory uncertainty: investors will be reluctant to back the development of technologies that they fear may later be banned or shunned if the absence of effective governance leads to irresponsible use and a loss of public confidence.

Ideally, governance regimes should be stable, predictable and transparent enough to build confidence among investors, companies and scientists, and should generate a sufficient level of trust and awareness among the general public to enable users to evaluate the significance of early reports of negative consequences. For example, autonomous vehicles will inevitably cause some accidents; whether this leads to calls for bans will depend on whether people trust the mechanisms that have been set up to govern their development.

But governance regimes also need to be agile and adaptive enough to remain relevant in the face of rapid changes in technologies and how they are used. Unexpected new capabilities can rapidly emerge where technologies intersect, or where one technology provides a platform to advance technologies in other areas.[1]

Currently, the governance of emerging technologies is patchy: some are regulated heavily, and others hardly at all because they do not fit under the remit of any existing regulatory body. Mechanisms often do not exist for those responsible for governance to interact with people at the cutting edge of research. Even where insights from the relevant fields can be combined, it can be hard to anticipate what second- or third-order effects might need to be safeguarded against: history shows that the eventual benefits and risks of a new technology can differ widely from expert opinion at the outset.[2]

To the extent that potential trade-offs of a new technology can be anticipated, there is scope for debate about how to approach them. There may be arguments for allowing a technology to advance even if it is expected to create some negative consequences at first, if there is also a reasonable expectation that other innovations will create new ways to mitigate those consequences.

Even if there is widespread desire to restrict the progress of a particular technology – such as lethal autonomous weapons systems – there may be practical difficulties in getting effective governance mechanisms in place before the genie is out of the bottle.

The growing popular awareness of the dilemmas associated with governing new technologies is revealed by media analysis: relevant mentions of such quandaries in major news sources doubled between 2013 and 2016. But which technologies should we be focusing on? In the latest GRPS, we asked respondents to assess 12 technologies on their potential benefits and adverse consequences, public understanding and need for better governance.

Technologies that Need Better Governance

Figure 3.1.1 plots respondents' perceptions of the potential benefits and negative consequences of the 12 technologies included in the GRPS. As noted above, the average score for benefits is much higher than it is for adverse consequences,[3] suggesting that respondents are optimistic about the net impact of emerging technologies as a whole.[4] Technologies considered to have above-average risks and below-average benefits, in the upper left quadrant of the figure, tended to be those where respondents felt least confident of their own assessments and also least confident of the public's understanding.

Three technologies occupy the upper-right quadrant of Figure 3.1.1, indicating an above-average score for both potential benefits and risks: *artificial intelligence (AI) and robotics, biotechnologies*, and *new computing technologies*. Analysis of media coverage resonates with respondents' high ranking for the risk associated with AI: from 2013 to 2016 there was a steady rise in reporting on whether we should fear AI technologies.[5] Respondents also cited artificial intelligence (AI) and robotics most frequently when asked how the 12 emerging technologies exacerbate the five categories of global risk covered by *The Global Risks Report*. As Figure 3.1.2 illustrates, this was seen as the most important driver of risks in the economic, geopolitical and technological categories.

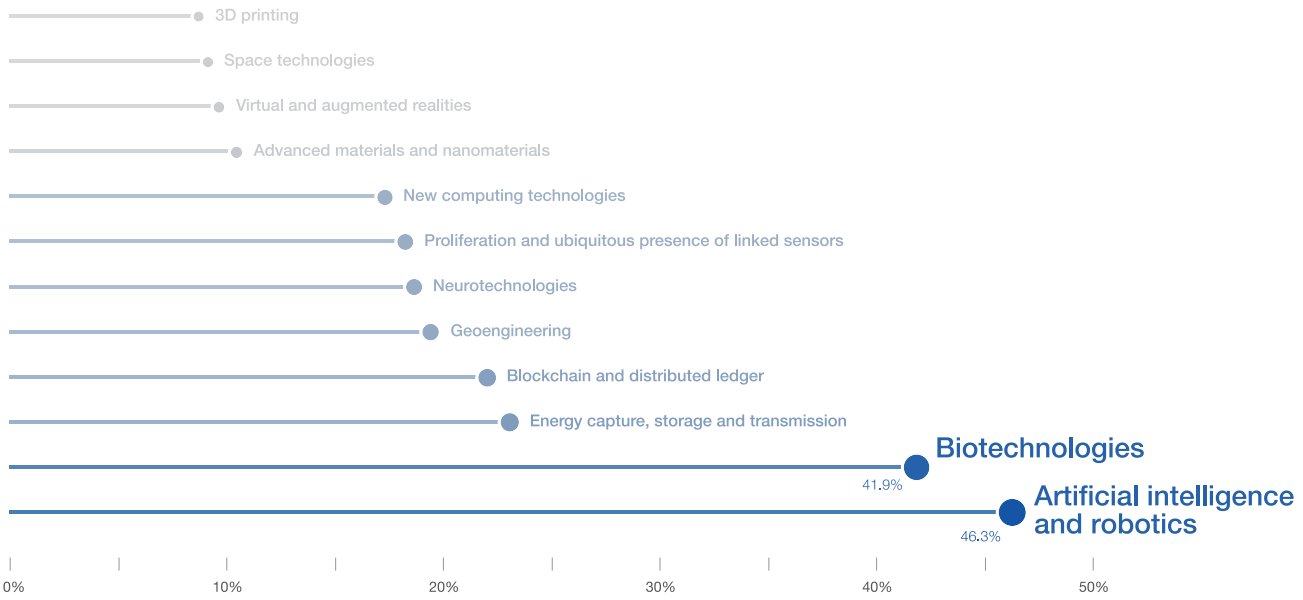
In Figure 3.1.3, two technologies stand out as requiring better governance in the view of GRPS respondents: both *artificial intelligence (AI) and robotics*

Figure 3.1.2: How Emerging Technologies Exacerbate Global Risks

Source: World Economic Forum Global Risks Perception Survey 2016.

Note: Respondents were asked to select the three emerging technologies that they believe will most significantly exacerbate global risks in each category.

Figure 3.1.3: Emerging Technologies Perceived as Needing Better Governance



Source: World Economic Forum Global Risks Perception Survey 2016.

Note: Respondents were asked to select the three emerging technologies that they believe most need better governance. The figure presents the percentage of respondents who selected each technology.

and *biotechnologies* were cited by more than 40% of respondents. These two technologies differ greatly in terms of the current state of their governance.

Biotechnologies, which involve the modification of living organisms for medicinal, agricultural or industrial uses, tend to be highly regulated.[6] Biotech became a global governance issue in 1992 with the Convention on Biological Diversity, now ratified by 196 countries.[7] AI and robotics, meanwhile, are only lightly governed in most parts of the world. As "general purpose technologies", in the words of economic historian Gavin Wright,[8] they have applications in many fields that already have their own governance regimes. For example, where machine learning is used in areas such as online translation, internet search and speech recognition, it comes under governance related to the use of data. Industrial robots are governed by International Organization for Standardization (ISO) standards,[9] while domestic robots are primarily governed by existing product certification regulations. There is increasing debate about the governance of AI given the risks involved, which are further discussed in Chapter 3.2.

The Disruptive Impact of Emerging Technologies

The potential of emerging technologies to disrupt established business models is large and growing. It is tempting to think of technological disruption as involving dramatic moments of transformation, but in many areas disruption due to emerging technologies is already quietly under way, the result of gradual evolution rather than radical change. Consider autonomous vehicles: we are not yet in a world of vehicles that require little or no human intervention, but the technologies that underpin autonomy are increasingly present in our "ordinary" cars.

As the technological changes entailed by the 4IR deepen, so will the strain on many business models. The automotive sector remains a good example. It has been clear for some time that car manufacturers need to plan ahead for a world in which many of the factors that determine current levels of car ownership may no longer be present. Increasing evidence of this planning is now starting to shape commercial decision-making. For example, in December 2016, Volkswagen launched a new "mobility services" venture, MOIA, in recognition of "an ever-stronger trend away from owning a vehicle towards shared mobility as well as mobility on demand".[10]

The deep interconnectedness of global risks means that technological transitions can exert a multiplier effect on the risk landscape. This does not apply only to newly emerging technologies: arguably much of the recent social and political volatility that is discussed in Parts 1 and 2 of this year's *Global Risks Report* reflects, in part at least, the lagged impact of earlier periods of technological change. One obvious channel through which technological change can lead to wider disruption is the labour market, with incomes pushed down and unemployment pushed up in affected sectors and geographical regions. This in turn can lead to disruptive social instability, in line with the GRPS finding this year that the most important interconnection of global risks is the pairing of unemployment and social instability.

Another prism through which to look at the interaction of risks and emerging technologies is that of liability – or, to put it another way, the question of who is left bearing which risks as a result of technological change. There are multiple potential sources of

disruption here. The insurance sector is an obvious example when talking about liability; just as car manufacturers must prepare for a future of driverless vehicles, so the reduction in accidents this future would entail means insurance companies must prepare for plummeting demand for car insurance.[11] But the idea of liability can also be understood more broadly, to include the kind of social structures and institutions discussed in Chapter 2.3 on social protection. Already there are signs of strain in these institutions, such as mounting uncertainty about the rights and responsibilities of workers and employers in the "gig economy". One of the challenges of responding to accelerating technological change in the 4IR will be ensuring that the evolution of our critical social infrastructure keeps pace.

Endnotes

[1] Alford, Keenihan, and McGrail 2012.

[2] Juma 2016.

[3] The overall average response for benefits to emerge from emerging technologies was 5.6, equating to a likelihood of above 55% and below 75%. This contrasts sharply with the average of 3.8 for negative consequences, equating to an assessed likelihood of between 25% and 45%.

[4] It is noteworthy that no single technology was, on average, assessed to present negative consequences at a higher likelihood than its benefits. The technology with the lowest net benefits in this regard was Geoengineering, with the fourth highest assessment of negative consequences overall and the lowest assessment of benefits. At the other end of the scale, the technology with the greatest assessed net benefit was Energy capture, storage and transmission.

[5] Quid analysis performed by the World Economic Forum on key search terms across major news sources, November 2016.

[6] In the United States, the White House Office of Science and Technology Policy issued its first federal framework for biotech regulation in 1986.

[7] United Nations 1992, Convention on Biological Diversity, Article 8.

[8] Wright 2000.

[9] See, for example, ISO 10218-1 (2011) and ISO 10218-2 (2011).

[10] Volkswagen 2016.

[11] KPMG 2015.

References

Alford, K., S. Keenihan, and S. McGrail. 2012. "The complex futures of emerging technologies: challenges and opportunities for science foresight and governance in Australia". *Journal of Futures Studies* 16 (4): 67–86.

Juma, C. 2016. *Innovation and Its Enemies: Why People Resist New Technologies*. New York: Oxford University Press.

Karembu, M., D. Otunge, and D. Wafula. 2010. *Developing a Biosafety Law: Lessons from the Kenyan Experience*. Nairobi: ISAAA AfriCenter.

KPMG. 2015. "Marketplace of change: Automobile insurance in the era of autonomous vehicles". White Paper, October 2015. https://home.kpmg.com/content/dam/kpmg/pdf/2016/05/marketplace-change.pdf

Nuffield Council on Bioethics. 2016. *Genome Editing: An Ethical Review*. London: Nuffield Council on Bioethics.

Volkswagen. 2016. "MOIA: The Volkswagen Group's new mobility services company". Press release, 5 December 2016. https://www.volkswagen-media-services.com/documents/10541/4e91af8e-0b11-477c-a6fb-7ee089f1cc4d

Wright, G. 2000. "Review of Helpman (1998)". *Journal of Economic Literature* 38 (March 2000: 161–62; cited in Brynjolfsson, E. and A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York and London: W. W. Norton & Company.

Chapter 3.1 was contributed by Nicholas Davies, World Economic Forum, and Thomas Philbeck, World Economic Forum.

3.2: Assessing the Risk of Artificial Intelligence

Every step forward in artificial intelligence (AI) challenges assumptions about what machines can do. Myriad opportunities for economic benefit have created a stable flow of investment into AI research and development, but with the opportunities come risks to decision-making, security and governance. Increasingly intelligent systems supplanting both blue- and white-collar employees are exposing the fault lines in our economic and social systems and requiring policy-makers to look for measures that will build resilience to the impact of automation.

Leading entrepreneurs and scientists are also concerned about how to engineer intelligent systems as these systems begin implicitly taking on social obligations and responsibilities, and several of them penned an *Open Letter on Research Priorities for Robust and Beneficial Artificial Intelligence* in late 2015.[1] Whether or not we are comfortable with AI may already be moot: more pertinent questions might be whether we can and ought to build trust in systems that can make decisions beyond human oversight that may have irreversible consequences.

Growing Investment, Benefits and Potential Risk

By providing new information and improving decision-making through data-driven strategies, AI could potentially help to solve some of the complex global challenges of the 21st century, from climate change and resource utilization to the impact of population growth and healthcare issues. Start-ups specializing in AI applications received US$2.4 billion in venture capital funding globally in 2015 and more than US$1.5 billion in the first half of 2016.[2] Government programmes and existing technology companies add further billions (Figure 3.2.1). Leading players are not just hiring *from* universities, they are hiring the universities: Amazon, Google and Microsoft have moved to funding professorships and directly acquiring university researchers in the search for competitive advantage.[3]

Machine learning techniques are now revealing valuable patterns in large data sets and adding value to enterprises by tackling problems at a scale beyond human capability. For example, Stanford's computational pathologist (C-Path) has highlighted unnoticed indicators for breast cancer by analysing thousands of cellular features on hundreds of tumour images,[4] while DeepMind increased the power usage efficiency of Alphabet Inc.'s data centres by 15%.[5] AI applications can reduce costs and improve diagnostics with staggering speed and surprising creativity.

The generic term AI covers a wide range of capabilities and potential capabilities. Some serious thinkers fear that AI could one day pose an existential threat: a "superintelligence" might pursue goals that prove not to be aligned with the continued existence of humankind. Such fears relate to "strong" AI or "artificial general intelligence" (AGI), which would be the equivalent of human-level awareness, but which does not yet exist.[6] Current AI applications are forms of "weak" or "narrow" AI or "artificial specialized intelligence" (ASI); they are directed at solving specific problems or taking actions within a limited set of parameters, some of which may be unknown and must be discovered and learned.

Tasks such as trading stocks, writing sports summaries, flying military planes and keeping a car within its lane on the highway are now all within the domain of ASI. As ASI applications expand, so do the risks of these applications operating in unforeseeable ways or outside the control of humans.[7] The 2010 and 2015 stock market "flash crashes" illustrate how ASI applications can have unanticipated real-world impacts, while AlphaGo shows how ASI can surprise human experts with novel but effective tactics (Box 3.2.1). In combination with robotics, AI applications are already affecting employment and shaping risks related to social inequality.[8]

AI has great potential to augment human decision-making by countering cognitive biases and making rapid sense of extremely large data sets: at least one venture capital firm has already appointed an AI application to help determine its financial decisions.[9] Gradually removing human oversight can increase efficiency and is necessary for some applications, such as automated vehicles. However, there are dangers in coming to depend entirely on the decisions of AI systems when we do not fully understand how the systems are making those decisions.[10]

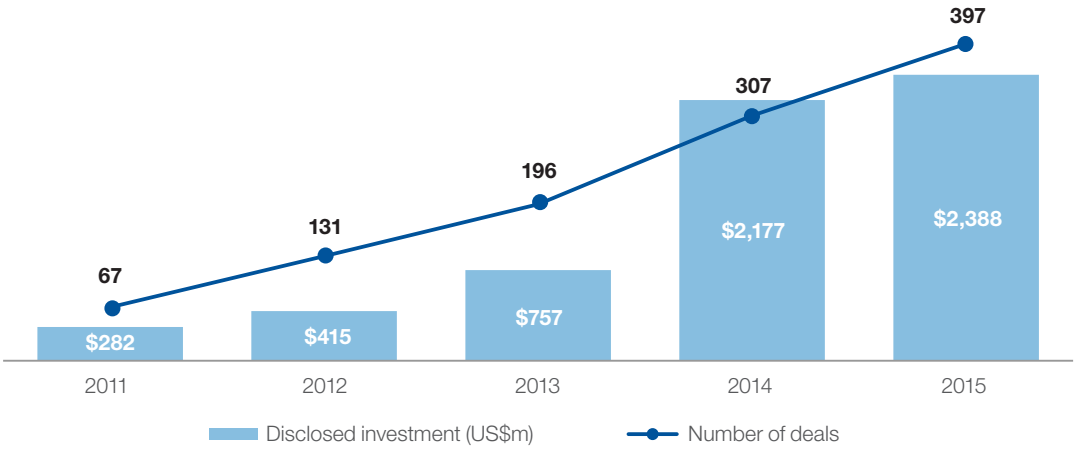
Risks to Decision-Making, Security and Safety

In any complex and chaotic system, including AI systems, potential dangers include mismanagement, design vulnerabilities, accidents and unforeseen occurrences.[11] These pose serious challenges to ensuring the security and safety of individuals, governments and enterprises. It may be tolerable for a bug to cause an AI mobile phone application to freeze or misunderstand a request, for example, but when an AI weapons system or autonomous navigation system encounters a mistake in a line of code, the results could be lethal.

Machine-learning algorithms can also develop their own biases, depending on the data they analyse. For example, an experimental Twitter account run by an AI application ended up being taken down for making socially unacceptable remarks;[12] search engine algorithms have also come under fire for undesirable race-related results.[13] Decision-making that is either fully or partially dependent on AI systems will need to consider management protocols to avoid or remedy such outcomes.

AI systems in the Cloud are of particular concern because of issues of control and governance. Some experts

Figure 3.2.1: Global Financing for AI Start-Ups, 2011–2015



Source: CB Insights 2016.

Box 3.2.1: Artificial Intelligence and the Future of Warfare - by Jean-Marc Rickli, Geneva Centre for Security Policy

One sector that saw the huge disruptive potential of AI from an early stage is the military. The weaponization of AI will represent a paradigm shift in the way wars are fought, with profound consequences for international security and stability. Serious investment in autonomous weapon systems (AWS) began a few years ago; in July 2016 the Pentagon's Defense Science Board published its first study on autonomy, but there is no consensus yet on how to regulate the development of these weapons.

The international community started to debate the emerging technology of lethal autonomous weapons systems (LAWS) in the framework of the United Nations Convention on Conventional Weapon (CCW) in 2014. Yet, so far, states have not agreed on how to proceed. Those calling for a ban on AWS fear that human beings will be removed from the loop, leaving decisions on the use lethal force to machines, with ramifications we do not yet understand.

There are lessons here from non-military applications of AI. Consider the example of AlphaGo, the AI Go-player created by Google's DeepMind division, which in March last year beat the world's second-best human player. Some of AlphaGo's moves puzzled observers, because they did not fit usual human patterns of play. DeepMind CEO Demis Hassabis explained the reason for this difference as follows: "unlike humans, the AlphaGo program aims to maximize the probability of winning rather than optimizing margins". If this binary logic – in which the only thing that matters is winning while the margin of victory is irrelevant – were built into an autonomous weapons system, it would lead to the violation of the principle of proportionality, because the algorithm would see no difference between victories that required it to kill one adversary or 1,000.

Autonomous weapons systems will also have an impact on strategic stability. Since 1945, the global strategic balance has prioritized defensive systems – a priority that has been conducive to stability because it has deterred attacks. However, the strategy of choice for AWS will be based on swarming, in which an adversary's defence system is overwhelmed with a concentrated barrage of coordinated simultaneous attacks. This risks upsetting the global equilibrium by neutralizing the defence systems on which it is founded. This would lead to a very unstable international configuration, encouraging escalation and arms races and the replacement of deterrence by pre-emption.

We may already have passed the tipping point for prohibiting the development of these weapons. An arms race in autonomous weapons systems is very likely in the near future. The international community should tackle this issue with the utmost urgency and seriousness because, once the first fully autonomous weapons are deployed, it will be too late to go back.

propose that robust AI systems should run in a "sandbox" – an experimental space disconnected from external systems – but some cognitive services already depend on their connection to the internet. The AI legal assistant ROSS, for example, must have access to electronically available databases. IBM's Watson accesses electronic journals, delivers its services, and even teaches a university course via the internet.[14] The data extraction program TextRunner is successful precisely because it is left to explore the web and draw its own conclusions unsupervised.[15]

On the other hand, AI can help solve cybersecurity challenges. Currently AI applications are used to spot cyberattacks and potential fraud in internet transactions. Whether AI applications are better at learning to attack or defend will determine whether online systems become more secure or more prone to successful cyberattacks.[16] AI systems are already analysing vast amounts of data from phone applications and wearables; as sensors find their way into our appliances and clothing, maintaining security over our data and our accounts will become an even more crucial priority. In the physical world, AI systems are also being used in surveillance and monitoring – analysing video and sound to spot crime, help with anti-terrorism and report unusual activity.[17] How much they will come to reduce overall privacy is a real concern.

Can AI Be Governed – Now or in the Future?

So far, AI development has occurred in the absence of almost any regulatory environment.[18] As AI systems inhabit more technologies in daily life, calls for regulatory guidelines will increase. But can AI systems be sufficiently governed? Such governance would require multiple layers that include ethical standards, normative expectations of AI applications, implementation scenarios, and assessments of responsibility and accountability for actions taken by or on behalf of an autonomous AI system. AI research and development presents issues that complicate standard approaches to governance, and can take place outside of traditional institutional frameworks, with both people and machines and in various locations. The developments in AI may not be well understood by policy-makers who do not have specialized knowledge of the field; and they may involve technologies that are not an issue on their own but that collectively present emergent properties that require attention.[19] It would be difficult to regulate such things before they happen, and any unforeseeable consequences or control issues may be beyond governance once they occur (Box 3.2.2).

One option could be to regulate the technologies through which the systems work. For example, in response to the development of automated transportation that will require AI systems, the U.S. Department of Transportation has issued a 116 page policy guide.[20] Although the policy guide does not address AI applications directly, it does put in place guidance frameworks for the developers of automated vehicles in terms of safety, control and testing.

Box 3.2.2: Aligning the Values of Humans and AI Machines - by Stuart Russell, University of California, Berkeley

Few in the field believe that there are intrinsic limits to machine intelligence, and even fewer argue for self-imposed limits. Thus it is prudent to anticipate the possibility that machines will exceed human capabilities, as Alan Turing posited in 1951: "If a machine can think, it might think more intelligently than we do. … [T]his new danger … is certainly something which can give us anxiety."

So far, the most general approach to creating generally intelligent machines is to provide them with our desired objectives and with algorithms for finding ways to achieve those objectives. Unfortunately, we may not specify our objectives in such a complete and well-calibrated fashion that a machine cannot find an undesirable way to achieve them. This is known as the "value alignment" problem, or the "King Midas" problem. Turing suggested "turning off the power at strategic moments" as a possible solution to discovering that a machine is misaligned with our true objectives, but a superintelligent machine is likely to have taken steps to prevent interruptions to its power supply.

How can we define problems in such a way that any solution the machine finds will be provably beneficial? One idea is to give a machine the objective of maximizing the true human objective, but without initially specifying that true objective: the machine has to gradually resolve its uncertainty by observing human actions, which reveal information about the true objective. This uncertainty should avoid the single-minded and potentially catastrophic pursuit of a partial or erroneous objective. It might even persuade a machine to leave open the possibility of allowing itself to be switched off.

There are complications: humans are irrational, inconsistent, weak-willed, computationally limited and heterogeneous, all of which conspire to make learning about human values from human behaviour a difficult (and perhaps not totally desirable) enterprise. However, these ideas provide a glimmer of hope that an engineering discipline can be developed around provably beneficial systems, allowing a safe way forward for AI. Near-term developments such as intelligent personal assistants and domestic robots will provide opportunities to develop incentives for AI systems to learn value alignment: assistants that book employees into US$20,000-a-night suites and robots that cook the cat for the family dinner are unlikely to prove popular.

Scholars, philosophers, futurists and tech enthusiasts vary in their predictions for the advent of artificial general intelligence (AGI), with timelines ranging from the 2030s to never. However, given the possibility of an AGI working out how to improve itself into a superintelligence, it may be prudent – or even morally obligatory – to consider potentially feasible scenarios, and how serious or even existential threats may be avoided.

The creation of AGI may depend on converging technologies and hybrid platforms. Much of human intelligence is developed by the use of a body and the occupation of physical space, and robotics provides such embodiment for experimental and exploratory AI applications. Proof-of-concept for muscle and brain–computer interfaces has already been established: Massachusetts Institute of Technology (MIT) scientists have shown that memories can be encoded in silicon,[21] and Japanese researchers have used electroencephalogram (EEG) patterns to predict the next syllable someone will say with up to 90% accuracy, which may lead to the ability to control machines simply by thinking.[22]

Superintelligence could potentially also be achieved by augmenting human intelligence through smart systems, biotech, and robotics rather than by being embodied in a computational or robotic form.[23] Potential barriers to integrating humans with intelligence-augmenting technology include people's cognitive load, physical acceptance and concepts of personal identity.[24] Should these challenges be overcome, keeping watch over the state of converging technologies will become an ever more important task as AI capabilities grow and fuse with other technologies and organisms.

Advances in computing technologies such as quantum computing, parallel systems, and neurosynaptic computing research may create new opportunities for AI applications or unleash new unforeseen behaviours in computing systems.[25] New computing technologies are already having an impact: for instance, IBM's TrueNorth chip – with a design inspired by the human brain and built for "exascale" computing – already has contracts from Lawrence Livermore National Laboratory in California to work on nuclear weapons security.[26] While adding great benefit to scenario modelling today, the possibility of a superintelligence could turn this into a risk.

Conclusion

Both existing ASI systems and the plausibility of AGI demand mature consideration. Major firms such as Microsoft, Google, IBM, Facebook and Amazon have formed the Partnership on Artificial Intelligence to Benefit People and Society to focus on ethical issues and helping the public better understand AI.[27] AI will become ever more integrated into daily life as businesses employ it in applications to provide interactive digital interfaces and services, increase efficiencies and lower costs.[28] Superintelligent systems remain, for now, only a theoretical threat, but artificial intelligence is here to stay and it makes sense to see whether it can help us to create a better future. To ensure that AI stays within the boundaries that we set for it, we must continue to grapple with building trust in systems that will transform our social, political and business environments, make decisions for us, and become an indispensable faculty for interpreting the world around us.

Chapter 3.2 was contributed by Nicholas Davies, World Economic Forum, and Thomas Philbeck, World Economic Forum.

Endnotes

[1] Russell, Dewey, and Tegmark 2015.
[2] CB Insights 2016.
[3] Mizroch 2015.
[4] Martin 2012.
[5] Clark 2016.
[6] Bostrom 2014.
[7] Scherer 2016.
[8] Frey and Osborne 2015.
[9] Sherpany 2016.
[10] Bostrom 2014; Armstrong 2014.
[11] Wallach 2015.
[12] Hunt 2016.
[13] Chiel 2016.
[14] Maderer 2016.
[15] Talbot 2009.
[16] Russell, Dewey, and Tegmark 2015, p. 111
[17] Bloomberg 2016.
[18] US regulatory policy is aimed at end products such as automated vehicles rather than the underlying technical system or its development.
[19] Scherer 2016, p. 359.
[20] U.S. Department of Transportation 2016.
[21] Cohen 2013.
[22] Kelly 2016.
[23] Bostrom 2014, Chapter 3.
[24] Conversation with Aldo Faisal, Senior Lecturer in Neurotechnology, Imperial College London, 29 September 2016.
[25] Yirka 2016.
[26] Lawrence Livermore National Laboratory 2016.
[27] Hern 2016.
[28] Kime 2016.

References

Armstrong, S. 2014. *Smarter than Us: The Rise of Machine Intelligence*. Berkeley, CA: Machine Intelligence Research Institute.

Bloomberg. 2016. "Boston Marathon Security: Can A.I. Predict Crimes?" *Bloomberg News*, Video, 21 April 2016. http://www.bloomberg.com/news/videos/b/d260fb95-751b-43d5-ab8d-26ca87fa8b83

Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

CB Insights. 2016. "Artificial intelligence explodes: New deal activity record for AI startups". Blog, 20 June 2016. https://www.cbinsights.com/blog/artificial-intelligence-funding-trends/

Chiel, E. 2016. "'Black teenagers' vs. 'white teenagers': Why Google's algorithm displays racist results". *Fusion*, 10 June 2016. http://fusion.net/story/312527/google-image-search-algorithm-three-black-teenagers-vs-three-white-teenagers/

Clark, J. 2016. "Google cuts its giant electricity bill with deepmind-powered AI". *Bloomberg Technology*, 19 July 2016. http://www.bloomberg.com/news/articles/2016-07-19/google-cuts-its-giant-electricity-bill-with-deepmind-powered-ai

Cohen, J. 2013. "Memory implants: A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories." *MIT Technology Review.* https://www.technologyreview.com/s/513681/memory-implants/

Frey, C. B. and M. A. Osborne. 2015. "Technology at work: The future of innovation and employment". *Citi GPS: Global Perspectives & Solutions*, February 2015. http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work.pdf

Hern, A. 2016. 'Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft. *The Guardian Online*, 28 September 2016. https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms

Hunt, E. 2016. "Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter". *The Guardian*, 24 March 2016. https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter

Kelly, A. 2016. 'Will Artificial Intelligence read your mind? Scientific research analyzes brainwaves to predict words before you speak". *iDigital Times*, 9 January 2016. http://www.idigitaltimes.com/will-artificial-intelligence-read-your-mind-scientific-research-analyzes-brainwaves-502730

Kime, B. "3 Chatbots to deploy in your busines". *VentureBeat*, 1 October 2016. http://venturebeat.com/2016/10/01/3-chatbots-to-deploy-in-your-business/

Lawrence Livermore National Laboratory. 2016. "Lawrence Livermore and IBM collaborate to build new brain-inspired supercomputer", Press release, 29 March 2016. https://www.llnl.gov/news/lawrence-livermore-and-ibm-collaborate-build-new-brain-inspired-supercomputer

Maderer, J. 2016. "Artificial Intelligence course creates AI teaching assistant". *Georgia Tech News Center*, 9 May 2016. http://www.news.gatech.edu/2016/05/09/artificial-intelligence-course-creates-ai-teaching-assistant

Martin, M. 2012. "C-Path: Updating the art of pathology". *Journal of the National Cancer Institute* 104 (16): 1202–04. http://jnci.oxfordjournals.org/content/104/16/1202.full

Mizroch, A. 2015. "Artificial-intelligence experts are in high demand". *Wall Street Journal Online*, 1 May 2015. http://www.wsj.com/articles/artificial-intelligence-experts-are-in-high-demand-1430472782

Russell, S., D. Dewey, and M. Tegmark. 2015. "Research priorities for a robust and beneficial artificial intelligence". *AI Magazine* Winter 2015: 105–14.

Scherer, M. U. 2016. "Regulating Artificial Intelligence systems: Risks, challenges, competencies, and strategies". *Harvard Journal of Law & Technology* 29 (2): 354–98.

Sherpany. 2016. "Artificial Intelligence: Bringing machines into the boardroom", 21 April 2016. https://www.sherpany.com/en/blog/2016/04/21/artificial-intelligence-bringing-machines-boardroom/

Talbot, D. 2009. "Extracting meaning from millions of pages." *MIT Technology Review*, 10 June 2009. https://www.technologyreview.com/s/413767/extracting-meaning-from-millions-of-pages/

Turing, A. M. 1951. "Can digital machines think?" Lecture broadcast on BBC Third Programme; typescript at turingarchive.org

U.S. Department of Transportation. 2016. *Federal Automated Vehicles Policy – September 2016*. Washington, DC: U.S. Department of Transportation. https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016

Wallach, W. 2015. *A Dangerous Master*. New York: Basic Books.

Yirka, B. 2016. "Researchers create organic nanowire synaptic transistors that emulate the working principles of biological synapses." *TechXplore*, 20 June 2016. https://techxplore.com/news/2016-06-nanowire-synaptic-transistors-emulate-principles.html

3.3: Physical Infrastructure Networks and the Fourth Industrial Revolution

Since the appearance of railways and canals, industrial revolutions have been characterized by the transformation of physical infrastructure networks as much as by production methods. Now the Fourth Industrial Revolution (4IR) is shaking up the interdependent set of critical physical infrastructure networks on which we all depend, including transport (road, rail, waterways, airports); energy (electricity, heat, fuel supply: gas, liquid and solid); digital communications (fixed, mobile); water (supply, waste water treatment, flood protection); and solid waste (collection, treatment, disposal). This process brings huge opportunities for innovation, but also complex risks.

The Economic Characteristics of Infrastructure Networks

The value of a physical infrastructure network increases with its scope. In communications (transport, digital), the more people a network connects, the more useful it becomes. In resource networks (energy, water), connecting more people can help build resilience and leverage economies of scale. Costs are high relative to returns in the early stages of building a network, and also later when connecting geographically remote areas with low population density: extending coverage to such areas usually requires government intervention, although 4IR technologies may shake up that economic logic by drastically cutting the costs of connectivity.

Because physical infrastructure networks are often natural monopolies as a result of barriers to entry, the public sector typically either provides those barriers or regulates them on behalf of their users. Regulators have to tread the delicate line between setting affordable tariffs and ensuring that capital can be found to invest in maintaining and renewing networks. The pendulum has swung between private and public capital funding of infrastructure: for example, private financiers backed the creation of railway networks in Europe and North America in the 19th century, some losing their shirts. But much of today's ageing physical infrastructure in advanced economies was built with public funding during the 20th century. Britain led the way in utility privatization in the 1980s and 1990s, and it has generally improved asset management and reduced costs for customers. On the other hand, private finance has typically shied away from large and risky new assets, such as nuclear reactors. Uncertainties related to the 4IR play a part in that reluctance.

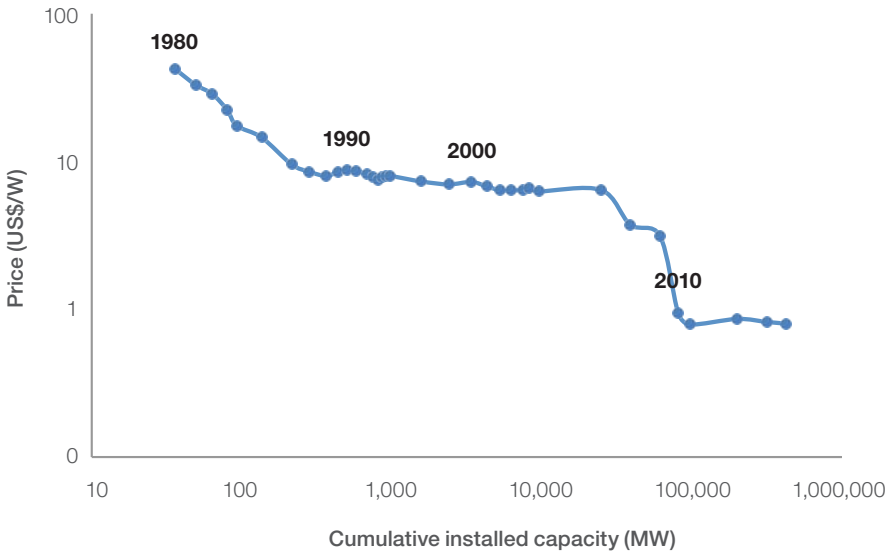
With tight public finances, governments and regulators are having to devise mechanisms for leveraging private finance while seeking to avoid the inflexibility and questions over value for money that have dogged public-private infrastructure finance in the past. It is still unclear how the enormous investment needs for some kinds of infrastructure are going to be met.

The Revolution

Electricity powered the Second and Third Industrial Revolutions, as networks achieved economies of scale by connecting large plants over high-voltage transmission grids to local distribution networks reaching many users. This aggregation of users helped to smooth out much of the local variation in demand, so steady-running base-load plants could be the workhorses of the network, with extra capacity patched in to deal with daily and seasonal peaks. Prohibitively high barriers to entry meant there was little competitive pressure to reduce the significant amount of energy lost as waste heat in the generation, transmission and distribution of electricity.

All of that is now changing. Collapsing prices of photo-voltaic cells make solar panels price-competitive with large-scale generation (Figure 3.3.1). The cost of offshore wind is also dropping fast, with firms such as DONG Energy and Vattenfall bidding prices down as low as €60 per Megawatt hour. Innovation in storage technology is helping with intermittency challenges – from large-scale storage to household battery

Figure 3.3.1: The Falling Price of Photo-Voltaic Modules



Source: Bloomberg New Energy Finance.

Note: Prices are in constant 2015 US$.

units and plugged-in electric vehicles, which will provide an additional buffer. The 4IR is moving electricity networks away from needing to be large-scale, top-down systems.

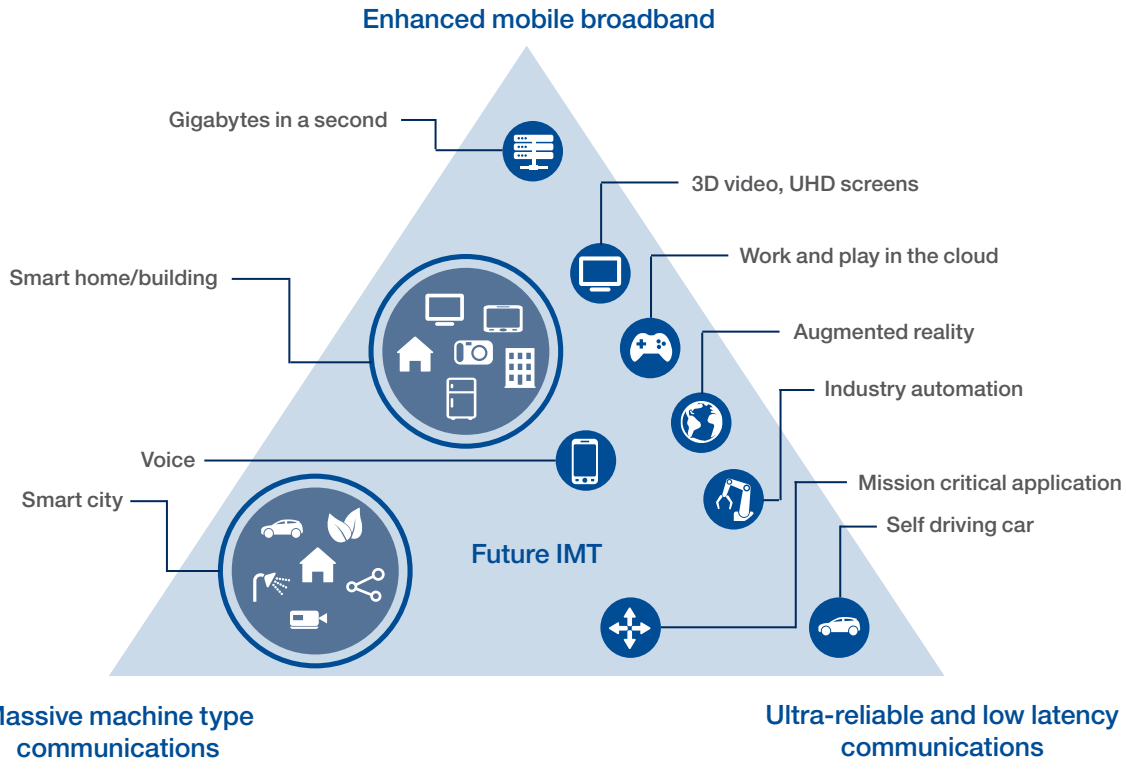
Technological innovations will increasingly offer households and firms the possibility of going "off-grid" entirely – but even if they increasingly generate their own power, most are still likely to want to remain connected to the high-voltage networks that are the backbone of today's electricity supply systems. Indeed, the rising use of solar, wind and tide power – with their associated intermittency issues and their greater need to tap the energy storage possibilities of hydropower in mountainous regions – will increase the appeal of high-voltage connections over long distances. But the growing scope for businesses and homes to supply and store their own electricity will make electricity networks multi-scale and less "lumpy" in terms of their capital requirements.

Beyond supply and storage, technology is improving efficiency by integrating supply and demand. Until very recently, energy suppliers and network operators have had to rely on crude methods to forecast demand for electricity. Big data, pervasive sensors and the Internet of Things are making it easier for users to monitor and control their energy demand, and for grids to predict and manage energy supply. In a world of prosumers and distributed suppliers, the challenges are how to synchronize supply and demand and pay for resilience.

Water could also transition from centralized networks towards more distributed systems. New materials and sensor technologies allow treatment at the household or community level, creating opportunities to harvest rainwater and directly reuse waste water. For the time being, economies of scale still favour large, centralized plants in existing urban areas: they also allow utilities to monitor water quality centrally and address failures quickly. Relying on localized water storage would also create challenges in prolonged periods of drought. But centralized networks are costly to create, and the balance of costs and benefits is beginning to tip in favour of distributed water systems if cities can be planned for these systems from the outset.

Regarding communications, the 4IR will continue to shift the balance between mobile and fixed networks. To improve mobile broadband, 5G technologies are envisaged to provide much faster data transfer (>1 Gigabyte per second) and reduced end-to-end latency (sub-1ms). By consolidating existing layers of technology, such as 2G, 3G, 4G and Wi-Fi, 5G will also improve coverage and 'always-on' reliability – it is an ensemble of different technologies, rather than a single type of new technology. Although the experience of those previous technologies suggests that new uses

Figure 3.3.2: Usage Scenarios for Mobile Technologies



Source: ITU 2015.

for 5G will emerge after deployment, two key roles are already anticipated for 5G: providing gigabit connectivity for businesses and consumers for a range of content, applications and services (the top of the pyramid); and enabling ultra-reliable, low latency machine-to-machine (M2M) communication (the bottom of the pyramid), which will help to achieve objectives in other infrastructure systems, such as easing congestion (Figure 3.3.2).

Governments are facing a difficult decision about whether to be first movers in rolling out 5G or wait to learn lessons from first movers, in the expectation that costs will decrease. For now, the bandwidth of fibre-optic cables remains hard to beat – but it is also expensive in towns and cities: 80% of the costs are attached not to the technology itself but to the labour-intensive process of digging trenches and laying ducts. Uncertainty about future technological development can inhibit investment: is it better to dig trenches for cables or wait for 5G? The same dilemma applies to other types of infrastructure – for example, in the time it takes to roll out smart metres, new and better metres are being developed.

While improving some infrastructure assets, the 4IR promises to ease pressure on others by finding alternative ways to deliver the same functionality. For example, meeting in virtual reality is becoming an increasingly acceptable substitute for physical business travel, while drones may substitute for delivery vans in cities. Satellite technologies will help to fill the gaps in digital connectivity where fixed or terrestrial mobile technologies are not cost-effective. Where energy companies once defined themselves by their physical infrastructure assets, they increasingly see themselves as being in the business of providing specific services such as heating and lighting. As the 4IR creates new ways to deliver services, it may begin to challenge whether infrastructure should be seen as a special category at all.

The Risks

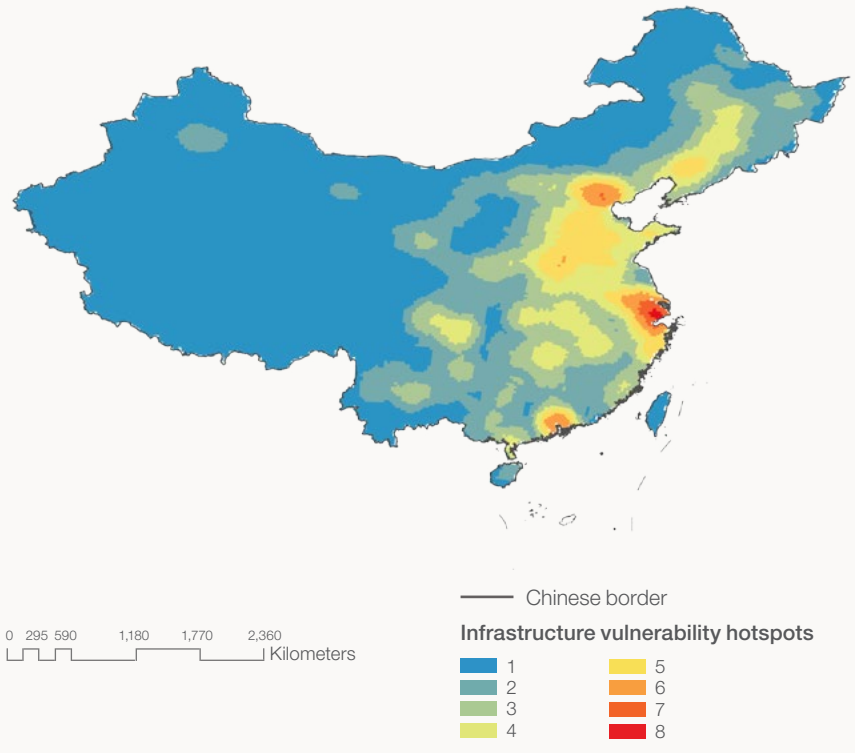
In theory, greater connectivity brings intrinsic resilience: electricity networks with more supply points, for example, should be less prone to failure. However, as different infrastructure networks become more interdependent, there is also growing scope for systemic failures to cascade across networks and affect society in multiple ways. In particular, electricity networks are now assuming an increasingly central role in many areas of life, such as road transportation and heating (taking over from gas and liquid fuels).

Systemic risks can come from many directions – whether these are cyberattacks or software glitches, solar storms or even just unexpectedly

Box 3.3.1: Mapping Infrastructure Vulnerability to Natural Hazards

An "infrastructure criticality hotspot" is defined as a geographical location where there is a concentration of critical infrastructure, measured according to the number of customers directly or indirectly dependent upon it. In the map of China below, red spots indicate where the highest numbers of people and businesses would be affected if a natural disaster caused infrastructure failure. According to this research, from the Environmental Change Institute at the University of Oxford, China's top infrastructure hotspots are Beijing, Tianjin, Jiangsu, Shanghai and Zhejiang.

Given the scale of China's manufacturing production and its role in the global supply chain, the business impacts of natural disasters could be astronomical: flooding in the more economically developed coastal provinces already accounts for more than 60% of the country's losses due to flooding.[1] The Oxford study finds that severe flooding events could disrupt infrastructure (rail, aviation, shipping and water) services for an average of 103 million people, while drought could affect an average of 6 million electricity users.



Source: Hu et al. 2016
Note: http://www.mwr.gov.cn/zwzc/hygb/zgshzhgb/201311/t20131104_515863.html

widespread and persistent clouds – and the increased complexity bring brought about by the 4IR makes the severity of those risks very difficult to estimate (Box 3.3.1). Society is increasingly dependent on information and communication technology networks in particular, and these have their own dependencies and vulnerabilities. In a 20th-century electricity network, it is possible to analyse the consequences of any given sub-station failing. That becomes impossible when every household is supplying and storing electricity and constantly adapting how much it uses based on price signals: we may suspect that our networks are acceptably resilient, but we cannot model them accurately enough to be sure.

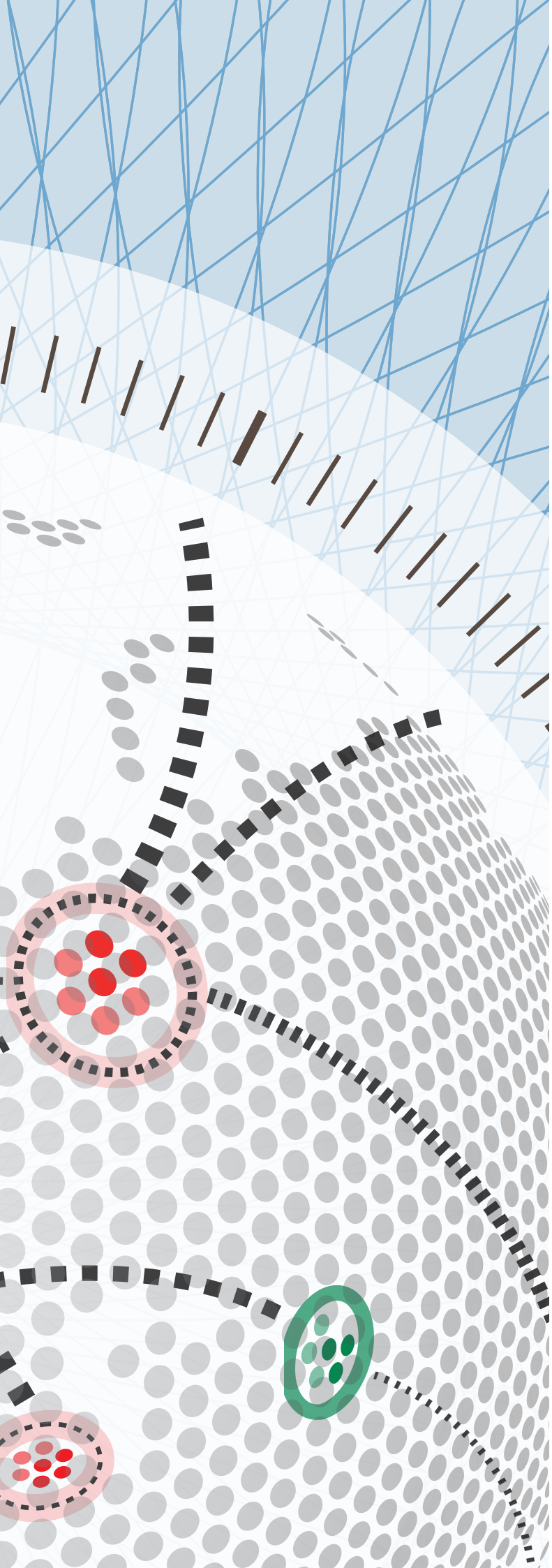
Because the 4IR intensifies networks' reliance on each other, there is a need for information sharing – utility providers tend to understand their own systems well, while often being more or less in the dark about the resilience of the systems to which they are connected. However, concerns about commercial confidentiality and security increase the challenge of developing protocols for information sharing that would help dependent customers to understand their risks. Not only infrastructure providers but also businesses need to understand risks and resilience more fully: analysis of supply chain risk tends to focus more on physical sites than the infrastructure networks that sustain those sites and move goods and services between them.

Governance of Infrastructure Networks in the 4IR

Like infrastructure networks themselves, arrangements for their governance have evolved incrementally and mostly siloed by sector – not least because ownership arrangements can be so different, ranging from highly competitive privatized markets (e.g. in mobile phone provision) through regulated monopolies, public-private partnerships, state-owned enterprises and direct public provision.[1] Governments are increasingly recognizing that this fragmented approach is becoming unfit for purpose in the 4IR. As networks become interconnected – for example, as digital technologies enable the routing of vehicles and the management of electricity and water demand – a "system-of-systems" approach to governance is needed. That requires appropriate sharing of information among network operators, and also requires regulators adopting common principles across networks.
Just as network operators and businesses need to better understand and manage systemic risks, governments and regulators need to take a wider view. Examples of new governance structures that recognize the need for a more integrated approach include the National Infrastructure Commission in the United Kingdom, Infrastructure Australia, and the National Infrastructure Unit in New Zealand. These new entities are having to navigate tensions between taking a national-level strategic approach to articulating needs for infrastructure to support growth and productivity and creating space for competition and innovation.

While the 4IR is creating complex new challenges for planners and regulators, it is also providing powerful new tools for monitoring and analysing system performance at hitherto unprecedented spatial and temporal scales – and testing resilience through simulation. Modelling exercises in a virtual environment will never give infallible results, but in itself the exercise of constructing and testing models can help to expose vulnerabilities in system resilience. Alongside their traditional role of minimizing the harmful effects of natural monopolies, infrastructure regulators in the 4IR should be paying more attention to systemic risks, building technical capabilities and standards for information sharing and stress testing.

Chapter 3.3 was contributed by Jim Hall, Oxford Martin School, University of Oxford.



Endnotes

[1] OECD 2015.

References

Hu, X, Hall, J.W., Shi, P. and Lim, W-H. 2016. "The spatial exposure of the Chinese infrastructure system to flooding and drought hazards". *Natural Hazards* 80 (2): 1083–118. doi:10.1007/s11069-015-2012-3

ITU (International Telecommunication Union). 2015. "IMT vision: Framework and overall objectives of the future development of IMT for 2020 and beyond". Recommendation ITU-R M.2083. http://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf

OECD (Organisation for Economic Co-operation and Development). 2015. *Towards a Framework for the Governance of Infrastructure*. Paris: OECD. https://www.oecd.org/gov/budgeting/Towards-a-Framework-for-the-Governance-of-Infrastructure.pdf

173003129 (02/17) TCL

ZURICH®