ESPAS
European Strategy and
Policy Analysis System

Trends in Artificial Intelligence and Big Data

Trends in Artificial Intelligence and Big Data

SUMMARY

This paper addresses the present state of play and future trends, uncertainties and possible disruptions of Artificial Intelligence (AI) and Big Data in the following areas:

Political: *AI is* biased, difficult to scrutinise and to estimate its power, and – more so when embodied in autonomous systems – potentially dangerous. Policy responses are accountability, transparency, safety and control, and public debate. These should be based on ethics. *AI will lead to* better governance, more debate, new policy actors and processes, a contest over centralisation, and the EU catching up. *AI might* progress in a revolutionary or evolutionary mode, lead to new political divisions, and change democracy. AI might be misused as a "superior orders" defence. *What if* data analysis changes or replaces democracy?

Socio-economic: *Big Data is* changing the role of data, is often dependent on sensitive information, is handicapped in the short term but better in the long term due to data protection, and its industry is in danger of monopolisation. *AI* lowers the cost of prediction, replaces human prediction and human labour and causes social problems, increased nudging and misuse of the term AI. *AI will lead to* more data, economic growth and more job market distortions. *AI might* lead to new industry giants, a request for more privacy, new state solutions, yet unknown jobs, AI taxes and increased state ownership. *What if* new economic ideologies emerge, singularity strikes or AI monopolies are broken up?

Geopolitical: *AI is* increasing the power competition between the US and China and gives both more power. Europe tries to create businesses and find its strengths. All are investing in military solutions and the west has a slight disadvantage here. *AI will lead to* a shakeup of the international system, hierarchies and networks becoming more powerful, and real-life deception being more difficult. *AI might* lead to China becoming the most powerful power overall and in AI. The future of AI R&D and the success of Europe's broad approach is uncertain. *What if* there are two digital worlds, China becomes a data-privacy defender, and AI become targets?

Technological: *Superintelligent AI is* invested in and researched, challenged by philosophy, and possible this century. *It might* imitate the brain, be assembled together or be designed by other AIs. An intelligence explosion or a conscious AI could be possible, and might be the last invention of humanity. *It would* require long term funding, need to overcome many technical hurdles, be dangerous due to its intellect, possibly be contained with collective intelligence, and maybe have humans lose their jobs, safety or purpose.

Key questions for policy-makers: What makes European AI distinctive? What areas can and should we prioritise, if any? What should be regulated? *How could and should the EU* foster AI development, avoid monopolisation, provide data pools, use high data standards, link researchers and corporations, balance fundamental with applied AI research and private with state funding, boost applications, compensate for job loss, keep AIs away from dangerous actors, support EU foreign policy (neighbourhood, FPI, democracy and human rights, aid and development, economic freedom), improve our lives with AI, change the geopolitical AI race, deal with autonomous weapons and superintelligent AI and organise Foresight?

In this Briefing

- Main trends, key uncertainties and possible disruptions of AI
- Questions for policy-makers

Author: Leopold Schmertzing

Introduction

This ideas paper looks at the political, economic, social, geopolitical and technological present and **possible future** status of Artificial Intelligence (AI) and Big Data to then pose a number of questions to policy makers. The future analysis is divided into trends, uncertainties and possible disruptions to anticipate likely, plausible and possible future events.

This analysis is meant to enrich the interinstitutional debate provided by the ESPAS network and to **improve the resilience** of EU's AI strategies and policies.

Defining and comparing AI and Big Data

According to the European Commission,

> ➤ Artificial Intelligence "refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."

> ➤ Big Data "refers to large amounts of data produced very quickly by a high number of diverse sources."

Defining is political agenda setting: Is AI a philosophical concept or a scientific discipline? Is AI a part of robotics or Big Data, or vice versa? The new draft AI definition by the European Commission's High-Level Group on AI includes Big Data and Robotics in its AI definition. This paper also incorporates Big Data in its mentions of AI, if no differentiation is made.

Technology and (Historic) Change

Powerful new technologies can interfere with the status quo in politics, society, economy, science, and the international order. This is not automatic. It is in our hands to steer technology in the right direction. Several past inventions that are comparable to AI and Big Data show the power of technological change:

> ➤ Niall Ferguson compared the **digital revolution** to the information explosion due to the **printing press**, which he deemed responsible for the Reformation, the scientific revolution and the Enlightenment.

> ➤ Sebastian Thrun, adjunct Professor at Stanford and founder of Google X, stated that **AI** "is to the brain what the **steam engine** has been to the human muscle." The steam engine was a major cause for railroads, industrialisation, colonialism and socialism.

> ➤ The Economist and many others compared **data** with **oil**, the arguably most important commodity for the twentieth century. It led to the rise of middle eastern autocracies, petrochemics and plastic, and much of modern capitalism, including tycoons and competition law.

> ➤ All of these comparisons are also politically sensitive and have flaws. Especially the comparison of data and commodities sparks debates. As a conference panellist was recently quoted: "Data is not oil. **Data is people**."

Political Lens: Between Laissez-Faire and the 5-Year Plan?

State of Play

AI and Big Data have already proven to be powerful forces for good and bad, but both technologies pose inherent **policy challenges** that make some contemplate regulation:

The (self-) learning process of AI is opaque, as is Big Data's ability to piece together conclusions from disparate sources. Sometimes these processes behave like a black-box, where input and output correlations are difficult to understand. This means their actions are **no longer open to scrutiny** by humans to the extent they were used to with "normal" machines. The report by the European Group on Ethics states that "it is impossible to understand how exactly AlphaGo managed to beat the human Go World champion."

Biases and errors – introduced via the data pools AIs learn from – can become engrained in an AI, leading to "online ads that show men higher-paying jobs; delivery services that skip poor neighbourhoods; facial recognition systems that fail people of colour; recruitment tools that invisibly filter out women", according to Powles and Nissenbaum. This is not necessarily the AI's fault, but the designers', and less a technical rather than a social problem. Proof of this can be found in platforms without AI: for example, guest applications with typically African-American names have a 16% lower chance of being accepted in Airbnb.

This makes it **difficult to know how powerful** an AI is, both in terms of the ability to do its job, and in terms of its implications. Policy makers and experts advocating regulation are often ridiculed as exaggerating, but there are plenty of examples of AI specialists being surprised by the power or weakness of the AIs they have built. The probably biggest case for this is Facebook. The company seemed genuinely surprised by the power of its newsfeed-AI that had quickly polarised its users and had facilitated political earthquakes.

AIs possess two forms of power. One is comparable to the power of the brain: AI algorithms can find novel solutions to all kinds of intellectual problems. AIs and Big Data can, however, also be part of an **autonomous system** (quasi-synonyms are robot, cyber-physical system, advanced mechatronic or embodied AI). These are physical systems guided by AI that also include sensors and mechanics and therefore become part of the physical world. Examples are autonomous cars, cleaning robots or autonomous weapons. For now they are, by many, considered the most dangerous. Autonomous systems might be not only opaque, hard to scrutinise, probably biased and difficult to estimate in their power, their real-world decision-making independence and physical lethality is also deeply unnerving to many.

Most current **policy responses** to these issues fall into two camps. One camp has a laissez-faire approach assuming that the AI market (and angered consumers) would control companies to not get too unsafe. The US has mostly followed that approach. Meanwhile, authoritarian states such as China have strengthened control over most important aspects of AI research and development (more on the US and China in the geopolitical vignette). The EU seems to have decided to take the (probably best, but most complicated) middle road: preparing policy options that are sufficient to guarantee beneficial AI, but not to - as MEP Giménez Barbat stated - "give in to the temptation to legislate on non-existent problems." This leaves many policy options:

➤ **Accountability:** This includes the democratic oversight over government-AI, different forms of human oversight, verification and control depending on the sensitivity of AI, the disclosure of AI decisions affecting humans, the setting up of appeal mechanisms (e.g. GDPR establishes the right for individuals to know and challenge automated decisions), control mechanisms ensuring data privacy of AI data pools, and registering/certifying AIs with certain tasks or capabilities.

➤ **Transparency:** In an effort to minimise the black-box problem, which has created much of the political and social backlash, researchers (e.g. MIT, Google) are working to translate the decisions and the reasoning of AIs to humans. This needs human input that could make AI more expensive, but it would also make AI more efficient and safer. Another option would be a "right for explanation" or making developers keep historical records of issues and what went in and out of the AI.

➤ **Safety and control:** This comes in many facets like for example ensuring consistent decision making of AI, denying hacking, reducing possible errors and risks, in case something goes wrong. Another facet is limiting AIs power through e.g. competition law and taxation (see next chapter). Some also call for banning or severely limiting R&D in fields of AI such as autonomous weapons, superintelligent AI and offensive cyber capabilities.

➤ **Public debate:** This includes providing sufficient education about AI, paying for AI training for natural and social scientists and policy makers, facilitating science translation, and creating incentives for AI experts to work for NGOs, government and in politics.

Policy should be based on **ethical AI principles**. These are, however, no technical solutions. The often-used example of an autonomous car deciding on how to crash is an inherently political question, as there is no agreement in the world on what ethical behaviour entails. Balancing different moral approaches is therefore a political task that cannot be relegated to entities like companies, programmers or AIs themselves (a list of ethical guidelines by expert groups is provided at the back).

Main Trends

AI will have the potential to **improve government** and governance by increasing effectiveness and oversight. For example, police use predictive tools to assess e.g. where to patrol, or what policeman are likely to engage in misconduct.

Just like social-media, freedom-of-internet and data-privacy debates today, AI is very likely to become a **more central topic** in future political discourses. Most likely this will be due to affecting the daily lives of ordinary citizens, causing political scandals, job loss (see next chapter), perceived loss of control, and societal divisions. This can lead to political crisis and violence, especially if AI is seen as oppressive, business-controlled or technocratic.

AI will give rise to **new policy actors** and will rebalance the power of existing ones. New expertise will become essential, new policy entrepreneurs will emerge. Technology companies seem to have a mixed track record as European policy actors. Some, like Google, which runs over 2000 commercial AI processes, are again leading the industry, and have already partnered up to contribute to the policy debate. New powerful government actors that are central in verification and control might emerge too.

The **policy cycle** or process will change too, with new ways to come up with, formulate, decide and implement policy due to AI. AI will help in finding policy gaps and political strategies, and in assessing probable policy impacts.

There will be more of the already ongoing political battle between **top-down** and centralising, and **bottom up** and distributive AI development and usage. Especially authoritarian countries want to centralise political power in AIs as a way to keep control of an ever-more complex society. Europe's strength and values lie in bottom up AI, and it could make a "brand" of it. It could, for example, export GDPR and other regulation, provide AI for SMEs, government watch-dogs and news corporations, and invest and become a market leader in less data-hungry, decentral or blockchain-based AI.

The **EU will lag behind** in AI for some more time, because it has a more complicated task than others. On the other hand, with a resilient and free economy, a balanced regulatory system, an interested public, intact societies and world class research it will be well-placed in the medium term.

Key Uncertainties

The need for future regulation will depend a lot on the **technological progress** of AI. Some experts believe that the advances in machine learning are plateauing and that AI will only develop slowly and incrementally from now on. Others see much more change coming, even revolutionary jumps like superintelligent AIs that are able to be employed in many fields at the same time (see the technology chapter). A more powerful AI would also need new forms of control. One theoretical way would be to align AI values with ours. For value-aligned AI to become a reality, humanity would need to better understand its own and the AIs intelligence, values, goals and modes of learning.

- What will be the political division in 2030? State versus market might be replaced by community versus algorithm.
- Which positions on AI will become left and right-wing mainstream and populist positions? For now they are incoherent.
- Will mainstream politicians catch up with fringe parties', populists' and nationalists' use of social media and Big Data analytics? Like with previous media revolutions, non-status quo powers seem better at exploiting them.

How will **political belief systems** and myths adapt to AI? The advents of general purpose technologies have had profound impacts on ideologies before (e.g. the internet's effect on liberalism or nationalism). Liberalism's emphasis on freedom could be challenged by the good life in an AI-led world. Even humanism's focus on humans could be challenged by an appreciation for AI.

Most importantly, will polities be able to find a way to harness the positive power of AI without becoming dependent on it? Maybe a future resilient deliberative democracy with a new generation in charge will integrate and embrace the power of AI. Maybe bogged down democracies will be undermined by AIs superior way of decision making. Paula Boddington stated that 'The quintessentially bad excuse of the twentieth century was, "I was only following orders," the quintessentially bad excuse of the twenty-first century may yet be, "I was only following the **dictates of a machine** (or algorithm)"'.

Possible Disruptions

Political parties could be even more sidelined than today, should a direct, AI-controlled line between politician and the electorate exist. Even political leaders themselves could fall victim to a more decentralised real-time direct democracy simplified by Big Data and regulated by AI.

And what if…

- the rights to the AI leading this decentral democracy, with its vast knowledge and power, would be in private sector possession?
- political deliberation would be dominated by bots with superhuman abilities to persuade you of a political position?

Imagine you have a very powerful AI version of Siri that has a friendly relationship with, and listens to most of your citizens. Wouldn't that AI know which policy would satisfy the biggest group of people? One rather scary idea is that the perfection of data analysis might **replace democracy** organically. Instead of voicing your opinion directly at elections or in direct democracy, you voice it indirectly, by leaving data traces. Your public and private communication and the changes in your daily routine might tell more about your political position and wishes than you could verbalise.

Economic and Social Lens: The Impact of Data and Prediction

State of Play

AI and Big Data has **already changed** the economy. They made available new consumer goods (e.g. Alexa) and services (e.g. traffic or translation apps). They have also advanced the productivity of entire markets, such as finance, health care or manufacturing. AI interacts with other high-tech innovations such as cognitive sciences, biotech or nanotech. Finally, AI led robotics assist or replace humans in difficult, dirty, dull or dangerous work.

In the wake of Big Data, the **role of data** has changed dramatically:

> *"From power grids and water systems to hospitals, public transportation, and road networks, the growth of real-time data is remarkable for its volume and criticality. Where once data primarily drove successful business operations, today it is a vital element in the smooth operation of all aspects of daily life for consumers, governments, and businesses alike."*

There are **downsides** too. As with oil, digging up, sending, trading and storing of data has stirred heavy political and economic battles, many of those on old fault lines.

The **amount of data** has increased drastically: more than 5 billion consumers interact with data every day. Every minute people exchange 527.760 snapchat and 46,740 Instagram photos, send 456.000 tweets, watch 4 million YouTube videos, send 156 million emails and book 45,788 Uber trips. Gartner predicts that in 2021 there would be one million new Internet-of-things (ioT) devices going online every hour of the day. Cisco predicts that in 2021, 94 per cent of global workloads will be processed in the cloud.

The most important data is often **critical and sensitive**, but it is also the life blood for machine learning. Health statistics and x-rays, recorded individual movements or political discussions could be the basis for AIs helping doctors, urban planners and politicians. Unfortunately, they can also be used against individuals and democracy as a whole.

As we see in the next chapter, non-democratic countries have an advantage in collecting this data without asking. Against such a competition, European **data protection** is perceived as an economic handicap. As German Chancelor Merkel admitted last year: "The fact that we have a difficult relationship to data ... has set us back." What to do? Providing safe and anonymised government-run datapools and incentivising the opening of business data would be one way. A way out of centralised datapools that create privacy concerns is federated learning, where AI learning takes place in the periphery – a mobile phone or an autonomous car – and improvements are conveyed to the larger federation for distribution to the rest of the collective. There are also new AI learning approaches that need less data, like quantum computing. This way, strong data protection might make European AIs more competitive, as less data used means less costs.

Monopolisation and "**eDistortions**" are other issues in the current data industry. There are old monopolistic tendencies based on size, power, attraction to experts and unregulated conduct. And there are new ones: the more your service (e.g. social platform) is used, the more data the AI can use to learn. This improves the service, which attracts more people. If the service builds on communities, this effect is even stronger, because bigger networks attract more newcomers, as they have more people to interact with. This tendency for monopolisation risks market distortions. Services deteriorate, as customers cannot easily change. In exchange for ostensible free services, data is collected far under its market price. Social costs are shifted onto others. Solutions to monopolisation range from classical measures such as anti-trust regulation, data and security auditing and transparency measures (such as disclosing the worth of stored data) to informing and helping the media, academics and users to become more aware of problems.

Steam engines blew up regularly before people knew how to deal with them, and so might AI. Like the steam engine it could also foster a completely new economic paradigm. In a way, this already happens through AI's role as a Prediction Machine. AI is a General Purpose Technology that,

according to Ajay Agrawal, is just making the step into the mainstream of the economy. Just like in 1995, when the internet radically lowered costs of arithmetic, AI lowers the cost of **predicting** (generating needed information out of existing data). Due to the power of AI, firms will convert other processes (e.g. human prediction and expertise) into AI prediction processes. For example, AIs learn through observation to predict human behaviour when driving or choosing people for a job. As AI prediction becomes cheaper and replaces other processes, data – the input for prediction – increases in worth. So too does human judgement (from which AIs learn), actions and implementations (e.g. factories) and feedback data, although AI will improve in these functions as well, over time.

Prediction is an important part of **decision making** and sometimes, by getting better, it changes strategies (e.g. Amazon being so good at predicting what customers want that they send goods before they get ordered). Finally, Agrawal explains the discrepancy between current AI capability and the high expectations of countries and companies with the expectation that prediction accuracy will rise over the next years. Prioritising prediction improvement creates the need for trade-offs, e.g. in data privacy or other research.

The current effect of AIs on **jobs** is unclear and blends into two bigger questions of the effect of automation on job loss:

- Is the classical theory of creative destruction – that more jobs are created due to innovation and optimisation than lost in the long run – not correct anymore? A recent EPRS study on the economic and social trends to 2035 stated that recent automation led to more job loss than additional employment.
- Is the decline in manufacturing jobs due to globalisation and outsourcing or automation? It seems to be both, which explains why the US was hit so hard. There factories left to China, and so did R&D and industrial knowledge. Many of the four million workers who lost their jobs since 2007 do not have the skills to work in an automated economy, which causes a rise in illnesses, suicides and alcoholism, and furthers the opioid epidemic that killed eight Americans every hour in 2017. The US might have more structural problems than Europe, but the next wave of AI and robotics is just around the corner.

Another social downside of AIs is the increased and improved **nudging**. According to Richard Thaler, a nudge is a small feature of the environment that encourages a desired behaviour**.** Their usage has a long history and seen an increase before the advent of AI, but now that we are constantly on the phone and everything is connected to the internet we constantly get nudged. Even worse, AI refines these nudges to be more effective and less visible. Fake news for example spread so quickly because AIs find ever better ways to encourage customers to click on a link.

AI is a **polarising buzzword** that provokes the misallocation of R&D capital. All kinds of analysis pretends to be AI driven, although most of the work is still done by humans. Companies also negate and avoid AI use in sectors like health care, where it could lead to quick successes, due to the bad press.

Main Trends

The **amount of data** subject to AI analysis will grow by a factor of 50 to 2025, mostly because "each connected person will have at least one data interaction every 18 seconds. Many of these interactions are because of the billions of IoT devices connected across the globe" backed by G5 networks. We are moving to a world where everything –all traces on earth, including of machines - will be collected and transformed into digital data. It will further fuel Big Data analytics and AI learning.

AI and automation will **boost economic growth**, but rapid creative destruction may impose limits. A PwC study found AI "could contribute up to 14 per cent to global GDP by 2030, equivalent to around USD15 trillion at today's values". An Accenture study projects an average additional growth of 1,7 per cent of GDP per year by 2035 due to AI. US defence officials estimate that 26 per cent of

Chinese economic growth will be due to AI by 2030. Others warn that GDP figures are unsuitable for distinguishing beneficial from harmful AI policy.

AI together with demographics and inequality will probably **replace even more jobs**. This PwC analysis found in the recent EPRS study (p. 61) projects a huge short-term impact on financial and retail business, and a bigger medium term one on transportation and manufacturing:
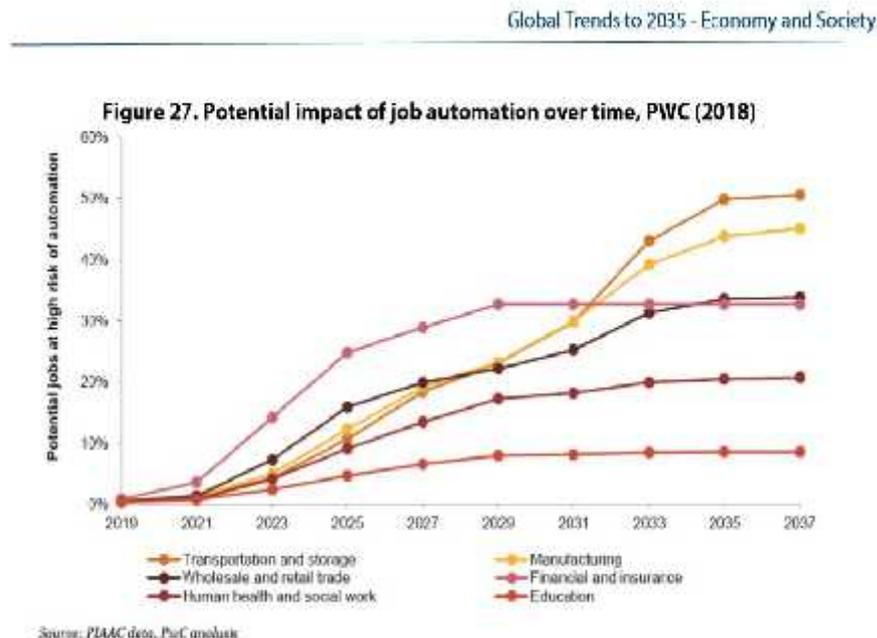
Between 2017 and 2022 it is expected that a quarter of all US malls close. This means millions of mainly female and part–time retail workers lose their jobs. The 2018 Future of Jobs Report by the WEF states that by 2022 42 per cent of work tasks will be done by machines (not necessarily AIs), from 29 per cent today. Despite this, they project a medium term global economic boost by 2022. In the long term, however, Bain projects that by 2030 the "rapid spread of automation may eliminate as many as 20% to 25% of current jobs – equivalent to 40 million displaced workers – and depress wage growth for many more workers. The benefits of automation will likely flow to about 20% of workers—primarily highly compensated, highly skilled workers—as well as to the owners of capital." The EPRS study summarised that automation may lead to favouring skilled workers to unskilled ones, capital to labour, and economic superstars to all others. It however also states that "in the long run (…) job creation should prevail." Another article looked at the many job prognoses and stated: "we have no idea how many jobs will be lost to the march of technological progress".

Key Uncertainties

Who will be the new powerful economic players in 2030? Will the big **Tech Giants** (e.g. Google, Amazon, Facebook, Netflix, Tencent, Baidu) that own most AI still benefit from their advantages in ten years?

The **request for privacy** of part of the European population includes demands for data protection, but might be a broader trend of feeling overburdened and oversaturated by technology's access and power, and fearing a loss of autonomy. Meanwhile the technological progress makes privacy impossible to uphold. The constant analysing and nudging by AI could lead to social and individual harm and strong reactions.

Initially, the industrial revolution often meant misery for workers, but then resulted in the expansion of state functions, in terms of education, social security or democratic participation. Will Europe come up with new **governance solutions** to address the AI revolution before a new round of

economic misery? Ideas such as universal basic income, time-banking and digital democracy might be first attempts.

Will opportunities or challenges dominate in terms of **job opportunities** due to AI? Humans are especially bad at anticipating things they have never seen, these are the so-called unknown-unknowns. Therefore, we have a hard time imagining tomorrows jobs, as many people in the 1980s would have had a hard time understanding many of today's jobs.

Will a major economic power decide to tax automated services by 2030? Would it be feasible in such a highly connected environment? Would it be fair or productive? The European Parliament has voted against it in 2017, but is there a way around it in a world with less jobs, and therefore less taxes and smaller welfare for more unemployed? According to the New York Times, "South Korea, the most robotized country in the world, instituted a robot tax of sorts in 2018 when it reduced the tax deduction on business investments in automation". This could be a way forward, as many current tax systems favor automation.

Will AIs be mainly **publicly or privately-owned**? Past general purpose technologies started as inventions of one and ended in the hands of another. In addition, governments decided to shoulder the research and development costs of revolutionary advances for purposes of national security or in the expectation of overwhelming economic or social returns at certain times.

Possible Disruptions

Will there be **new economic ideologies** in twenty of thirty years that represent the AI age, such as Marxism and Capitalism rose due to industrialisation? How will people define themselves, will labels like 'consumer' and 'worker' still be useful then?

What happens to the economy in a **singularity event** (an exponential growth in intelligence in AI, see next chapter)? Contrary to expectations, growth might remain limited or even slow down, as "economic growth may be constrained not by what we do well but rather by what is essential and yet hard to improve".

What if Facebook further falls into the "curse of bigness", it is misuses its monopoly power? What if normal antitrust measures such as fines do not work, and a future US government decides to break Facebook up, just like it did with Standard Oil? Can data giants with one or two major platform be broken up at all? What would be the ripple effects?

Geopolitical Lens: AI Policy and Grand-Strategy in 2030

State of Play

AI and geopolitics have an interesting relationship: First, AI, like space exploration during the cold war, is used as a projection space for a power competition by the two main players, China and the US. Second, like dreadnoughts and nuclear weapons before, AI pledges to increase geopolitical power both in terms of influence and military power.

In his book "AI Superpowers", Kai-Fu Lee tells the story of how official **China** was shocked when a US AI beat the best Asian players in Go. China also realised that AI gave it the chance to become leader in an important technology by 2030 and at the same time stabilise regime control. So President Xi made the Internet, Big Data and AI core parts of his modernisation effort, betting on its effect on the economy and security. With its closed internet, control over companies, and the buildup of AI-managed surveillance, China is developing a counter model to the vulnerable, decentralized but powerful original (western) digital revolution.

China has some **advantages**: With no real data privacy and the authoritarian urge to collect all possible data, China possesses a huge data reservoir for AI and Big Data to learn from. Computing power is centralised, and China is capable of organising huge state projects. Due to the close links between SOEs, private companies and the state, China can choose the best actor for each problem. It shows results: for example, according to its own head of research, Alibaba's Siri-like AI customer phone service has overtaken Google duplex. It already manages conversations with interruptions, nonlinear meanings, and implicit intent, and is already used million times a day. On the other hand, President Xi repeats old mistakes like stuffing SOEs with old communist party cadre. The authoritarian climate cannot be favourable for researchers, who depend on criticism and a free exchange of ideas.

Daniel Fiott notes that "not only did China release its own national development plan for AI technologies in July 2017, it is developing a **military** doctrine designed to overmatch the US's military technological prowess." The West meanwhile still allows Chinese companies to buy themselves into digital firms and the supply chain, thereby probably weakening defences of critical infrastructure and giving Chinese intelligence insights and sources for data collection. Only recently did the US commerce department consider export controls for AI and other advanced technology.

One reason for this was that the **United States** had its own "Sputnik moment" when it realised it would be challenged on a field it has called its own until recently. The US department of defense under President Obama – the traditional state investor in high-tech – did not finalise its third offset strategy, a massive R&D investment scheme, and President Trumps administration has not picked up on it until recently. It seems the current leadership is skeptical of novel technological solutions to warfighting. Nevertheless, the new administration has not stopped running research and the buildup of the Defense Innovation Unit Experimental (DIUx) which is meant to finance silicon valley start-ups of interest to the military. Secretary of Defense Mattis also made it clear that his strategy was focused towards China, which gave the AI strategies more of a direction. On the whole-of-government level, the US is streamlining its national security effort in AI by creating a "National Security Commission on Artificial Intelligence" uniting congress and agency heads.

US AI policy looks well on its way. However, it seems many in the US still trust **Silicon Valley** with its talent and financing to lead the way. Maybe that will be enough, but one needs to bear in mand that former technological advances like personal or mobile computing or the internet were based on state-funded inventions. There also is a political component that should not be underestimated: The US was the economic winner of automation and social media, but also lost many jobs because of it (see the previous chapter). Therefore, President Trump and many politicians in both parties might not see helping the Californian IT industry as their priority.

The **European Union** focuses on two goals of its own: catching up on transforming its good research into viable products and services, and trying to figure out where the strengths of its AI approach are (experts mentioned business-to-business AI, robotics, industrial AI). Many member states and the European Institutions (e.g. the European Parliament's 2017 Delvaux Report on Robotics) came up with AI strategies. Last year, France and Germany set up the decentralised Joint European Disruptive Initiative (JEDI). Two large groups of European scientists founded CERN-like scientific projects called CLAIRE and ELLIS to structure the catching up effort of European business and research entities in AI. In terms of financing military research, the EU also makes up leeway. Recently it made commitments to use 5% of the European Defence Fund for disruptive technologies, and some of the 34 PESCO projects have an AI component.

On the warfighting level the "ethics gap" – the weakening of western military power due to ethical considerations - might look bigger than it really is, argues RAND. Still, as stated in the ESPAS ideas paper on Warfare, "western democracies are more at ease with the AI-augmentation of humans, the teaming of unmanned systems with manned ones (e.g. formations of drones and planes, where the pilot decides for all), and AI in an advisory instead of a command role. The US and most European militaries has already decided on having an AI-policy of a **'human in the loop'** in all operations that involve harming someone, but as always, the devil is in the detail."

Main Trends

According to the Belfer Center, the "impact of technologies such as autonomous robots on war and international relations could rival that of nuclear weapons". Deciding whether to field and use such weapons will be a hard choice. Improvements in AI might "shake up the balance of international power by making it easier for smaller nations and organisations to threaten big powers like the US." It is fairly cheap and quick to couple AIs with physical or cyber weapons. On the other hand, only a few states seem to be on the forefront of AI, and most are big (except maybe Israel).

Again quoting the ESPAS Warfare paper, "the West has to be aware that preferring AI in a supportive role to purely AI-controlled weapons might be a **price it pays in combat losses** at some point. It would be a price that the West might be willing to pay for keeping to its ideals. This is also not new: Soviet weaponry cared less about the survivability of their crews, so long as they were efficient, while Western countries prioritised safety. This sometimes put the West in weaker positions, but it was deemed worth it."

Offensive cyber capabilities will increase rapidly due to AI. Earlier this year, technology giant IBM unveiled a proof-of-concept called DeepLocker, the first-ever **malware based on AI**. It is more or less invisible, can be employed with unprecedented accuracy and provides high anonymity to attackers.

While there will be an overall shift away from **hierarchies** towards **networks**, AI and Big Data will make both more powerful. Centralised military command and control will gain in oversight in rapidly changing environments due to AI. Meanwhile, AI will also facilitate decentral communication and learning.

In addition, AI supported analyst networks, using all available sensors to their advantage will make **camouflage and deception** much more difficult. "Already today, it seems that intelligence services struggle to build fake identities due to the missing data trail, military units give away their positions when soldiers post photos online, and collaborative or AI-guided commercial satellite image analysis makes it difficult to hide any larger piece of equipment in the field."

Key Uncertainties

China is becoming more powerful in all aspects. In terms of **military power**, China has the capabilities and the will to catch up. However, to surpass "the West" China would need to undertake its full transformation into a interventionist power. In addition, no one is sure in how far China's current strengths - centralised control over priorities, investment and the private sector, a long-term plan and a focus on big ticket technical solutions such as AI and space technology – are useful in

innovating its military in the future. Military strength also depends on economic strength, and China's economy seems less stable in the recent years. Finally, what if the fixation of China (just like Russia) to beat the US is making them invest inefficiently both in military and economic research?

Will China overtake the US in its **AI capability**? In the short term no, quoting Eric Schmidt, CEO of Google: "US will maintain its lead over the People's Republic of China for the next five years". Business leaders are split (yes, no argument) on this, so are academics. In the long run, after these five years, Schmidt expects China to "pass us [the US] extremely quickly." But how would an economic downturn or political instability affect AI policy?

What is the next magic formula for **AI innovation**? Nobody knows how good companies really are in the battle over future technological change. It might be that Silicon Valley with its "let's try it" approach, its connections to money and the best and brightest, has only been good for a certain period. Just like US government-led innovation has been working best from the Manhattan project to NASA's moon landing.

It seems that **Europe** follows a broad approach, highlighting a human-centred focus, high ethical standards and broad investment. It is not clear what the rewards or penalties for such a broad strategy are.

Possible Disruptions

China has already isolated its internet from the rest of the world. Meanwhile its AI and social media services face resistance in other countries due to data issues and fear of Chinese mischief. With western products being replaced in China, and Chinese digital products not selling in advanced economies, will there be two incompatible **digital worlds** and markets in the future?

What if, in a turn around, China becomes a **champion of data privacy**? Which could the causes for such a change be? Can authoritarian states be trusted to guard data rights?

What if another country makes a **dash for AI leadership**? India lags behind, but AI promises to make the country more transparent and efficient, something many strive for. Japan has a good research base and sees a way to replace its ageing workforce in AI.

What is the chance that we see **AIs fighting each other** in the real world or the cyber dimension? The goal of war is to make an enemy submits to your will. This means AIs and robots will not be the ultimate targets, but humans and the things they cherish. However, AIs are vulnerable to **hacking**: governments, terrorists, activists and businesses might try to introduce intentional bias into AIs, leading to vulnerabilities.

Technical Lens: Superhuman and Conscious AI?

State of Play

According to Nick Bostrom, any intellect that greatly exceeds the cognitive performance of humans in all domains is superintelligent. Similarly, Mark Tegmark defines a **Artificial General Intelligence** (AGI) as the ability to accomplish any cognitive task at least as well as humans.

While many policy makers see the question of AGI as science fiction, **huge investments** are made into researching it. For example, DeepMind – developers of the Go-champion AI AlphaGo and bought by Google for 500 million USD – spends up to 200 million USD each year to come closer to that goal. OpenAI, funded with an Endowment of 1 billion USD, has the same goal. Since this research is not required to be transparent, it is likely that states such as the US, Chinese and probably others are also already working on such programmes. The biggest project by the European Union is the Human Brain Project, an effort to construct a virtual human brain, although this is not exactly the same as building an AGI.

Since around 2014, a few US and UK institutions such as the Singularity University, the Future of Humanity Institute (Oxford), the Future of Life Institute, the Centre for the Study of Existential Risk (Cambridge), and the Center for Human-Compatible AI (Berkley) began to look at the broader implications of developing a general AI and how to mitigate **existential risk** when building and running it.

AI experts are divided over questions such as: (1) Will AI ever be capable of general intelligence similar or surpassing that of a human, and if yes, when? (2) Will AI ever be conscious or self-aware as a life form? (3) What effect would a yes on one of the questions have on humanity?

There are a couple of **philosophical challenges** to the concept of a general AI (we get to the technical ones later). What is intelligence and how important is it? What are theoretical limits to it? The "no free lunch" theorem, for example, stipulates that no one solution or algorithm can be best at all problems.

Still, recent advances in AIs has made researchers such as Stuart Russell and others from the beneficial-AI movement believe that an AGI was a real possibility **this century**. Others believe in preparation, as at least it cannot be ruled out.

Precise predictions about when a breakthrough could happen are close to meaningless. In the last years AI advanced much quicker than even experts thought and the AI community has a notorious history of **over- and underestimating** its advances. In 2015, for example, a huge majority of experts believed that it would take another 15 years for an AI to beat a Go-master. Two years later, the world best player was beaten repeatedly by the above mentioned AI AlphaGo.

What are ways to develop an AGI? One could try to imitate the functions of a **human brain**. The world's largest neuromorphic computer that operates like a brain (SpiNNaker) was activated recently at the University of Manchester. As part of the Human Brain Project it "will enable scientists to create detailed brain models." Today it only manages to mimic less than 1 per cent of the human brain, but it probably already has the capability to mimic that of a mouse.

Another way to reach the goal would be to combine different **general capabilities**. Today, AIs are good at generalised tasks through automated learning. They need less and less feedback from humans and start to combine different skills, such as haptics and vision. They quickly become good at difficult video games by finding creative solutions to advance. Text analysis made another big jump to reach more human-like speaking and understanding, including the meaning of language in social contexts. Deep learning AIs from OpenAI and Google are close to human capability in guessing missing words and reading comprehension.

A third, more hypothetical, way would be to specialise in building AIs that are good at building AIs. This idea is based on the idea of the **singularity** (also called intelligence explosion): Since the design

of machines is an intellectual activity, an intelligent enough machine could design an even better machine. AlphaZero, Microsoft and other are making headway in a precursor: automated AI development.

Even a superintelligent AI would not automatically possess **consciousness**. Some say consciousness is needed to deal with and relay vast amounts of conflicting unconscious knowledge, others say it is essential in determining a personal relationship to objects and ideas and in giving meaning. Without going deeper into this, it is neither sure, if consciousness is a prerequisite to intelligence, if an AI could or should possess it, and if we would understand an AIs consciousness, if it did possess one.

Main Trends

For some more years, humans will have the advantage in "critical thinking, problem solving, managing human interactions, and above all else, expressing creativity." Humanity's "future role involves embracing these last bastions of human exclusivity and becoming more human". But AI makes **fast inroads**, even in being creative (whatever that means). AIs make art, surprise their mentors with creative digital evolutions, and transform conceptual sketches into real world pictures.

AI is advancing rapidly, but the chance is high that a mix of too high expectations and stalling progress in new AI frontiers such as AGI – see the trough of disillusionment - will raise doubts and shift capital to other advanced fields for a time.

AI researchers are working to quickly **overcoming barriers** to AGI:

- AI has a **tendency to forget** useful knowledge that is not relevant for some time. When AIs become better at deciding what knowledge to keep and what not, and when to use different streams of learning, this might get them closer to a AGI.
- Similarly, **deep learning** is still complex, costly and needs a lot of human input. The mentioned unsupervised and federated learning (Google, cars) and generative adversarial networks (GAN), where AIs learn from each other, seem promising.
- **Quantum computing** is in its infancy, but could jumpstart a general AI with its enormous processing power. More than that, a quantum computer has the ability to work much like our brain, which makes it perfect for machine learning.

It was said before that currently, autonomous systems or embodied AIs are more dangerous than their purely digital siblings. Max Tegmark argues, however, that the AI itself - the closer it comes to an AGI - becomes the actual **danger**, not the physical robot. We should know best: there are many stronger animals than us on earth, but we control them all, due to our bigger brain.

Key Uncertainties

Irving Good, the founder of the singularity concept, noted: "the first ultra-intelligent machine would be the **last invention** that man need ever make." No one is sure what might happen then. In the film "She" for example, the AI just becomes disinterested in human problems, because it has become to intelligent to care about petty human issues. But would there be just one AGI, as self-improvement would make any other superfluous, or would there be many AGIs?

No one is sure where a superintelligent AGI would leave **humans**. Steven Pinker writes "As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come." Daniel Kahneman disagrees, he thinks robots will not only be better in statistics and wiser, but will also have more emotional intelligence than humans.

Science Fiction authors and philosophers have had many ideas of how a **world after singularity** looks like. They normally differ on the questions if humans will still exist, are in control, are safe, happy, and if the AGI has consciousness. Many of the worst scenarios are based on classical alignment issues such as the AI pursuing a given objective brilliantly to the destruction of mankind.

Possible Disruptions

What if AIs could prevent humans from **breaking the law**? Nowadays laws can be violated if you are willing to risk punishment. In the future, a car might not let you speed, your clothes would prohibit you from beating someone, and a pistol might not let you fire. AIs can already learn from eye movement to spot lying. This sounds good, but are there not cases where one might need to break a rule? Context can justify a misdeed - for example in self-defence. The violation of rules oftentimes starts a political discussion and lead to change. So is there a right to break a rule?

Radical new ways to communicate and **collaborate more efficiently might** balance out the power of AI. Technical solutions range from existing augmented-reality to far future ideas such as connecting brains directly. Those would produce more powerful groups that might be less susceptible to AI analysis, nudging and disinformation. Transhumanists are a growing group believing in the importance of the enhancement of the human body through technology.

Imagine, in 20 years, there is a superintelligent, friendly, conscious AI which is a source of pride to the world and fulfils all our wishes. Would this be a **paternalistic world**? The difficult question goes to the core of the human condition: What are we to do, if we are not needed anymore? What then is the purpose of humanity?

Questions for Policy Makers

1. What should be the AI narrative of the European Union? What makes EU AI distinctive?
2. What are the fields of AI in which the EU could possibly become a world leader?
3. What should or could be the areas of AI that the EU should prioritise, if any at all?
4. What are the opportunities and costs of state-led or private-public-partnership prioritising and research?
5. What should the AI priorities of the next EU legislature be?
6. What should be regulated in advance – because, for example, the risk or opportunity is so clear or big – and what not?
7. What can be better dealt with through market self-regulation or self-regulation of industries (e.g. industry/expert standards or codes)?

8. How could and should the EU:
   a. Foster the development of a European AI ecosystem?
   b. Avoid the monopolisation of AI providers?
   c. Provide better (access to) data pools?
   d. Balance the need for data with data protection?
   e. Use its high data standards as an advantage?
   f. Balance fundamental research and applied development, and what would the right proportion of state and private funding for both be?
   g. Improve synergies between researchers and corporations?
   h. Boost the development of AI applications?
   i. Compensate for job loss due to automatisation and prepare for future job loss?
   j. Keep AIs out of the hands of dangerous actors, nationally and internationally?
   k. Use AI to improve our individual and communal lives?
   l. Use AI to support a EU foreign policy, e.g. neighbourhood policy, our foreign policy instruments, such as election observation or security sector reform, our commitment to democracy and human rights, humanitarian aid and development, and to promoting economic freedom globally?
   m. Change the AI race to its and global benefit? Is there, for example, a way for Europe be a bridge between the US and China to avoid further escalation?
   n. Go about the R&D of autonomous weapons?
   o. Implement AGI risk and research policies?
   p. Better organise its Foresight capabilities?

MAIN REFERENCES

Further Reading:

Tegmark, Max. *Life 3.0: Being human in the age of artificial intelligence*. Knopf, 2017.

Lee, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin, 2018.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. *Prediction Machines: The simple economics of artificial intelligence*. Harvard Business Press, 2018.

European Union:

Report with recommendations to the Commission on Civil Law Rules on Robotics, JURI committee, European Parliament, 27 January 2017.

Global Trends to 2035: Geo-politics and international power, EPRS, European Parliament, September 2017

Global Trends to 2035: Economy and Society, EPRS, European Parliament, November 2018

A definition of Artificial Intelligence: main capabilities and scientific disciplines, European Commission's High-Level Expert Group on Artificial Intelligence, 18 December 2018

Reillon V, Understanding artificial intelligence, EPRS, 11 January 2018

Ethical Guidelines:

Draft Ethics guidelines for trustworthy AI, European Commission's High-Level Expert Group on Artificial Intelligence, 18 December 2018

Asilomar AI Principles, 2017

Ethically Aligned Design, Institute of Electrical and Electronics Engineers (IEEE), 2017/2018

Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Atomium-European Institute for Science, Media and Democracy, December 2018