Handbook of AI for decision-makers

Tietopolitiikka.fi is a collaborative group established in 2020, consisting of data policy actors from all parliamentary parties. This paper provides background on the current state of AI development and opens a discussion on its political implications, opportunities, and risks for Finland.

Antti Poikola
Jouni Markkanen
Janne Parkkila
Parliamentary Data Policy Collaboration Group

Original Finnish version published on November 14, 2024
English translation published on June 11, 2025

Download the publication in PDF format

Table of contents

Ten Highlights About Artificial Intelligence

1. AI to Improve Service Quality and Enhance Public Services: We must adopt artificial intelligence in the public sector while ensuring citizens' data privacy and transparency in decision-making. The requirements for AI systems in the public sector differ in part from those in the private sector, particularly concerning data traceability. (see Appendix 1: Application Ideas)

2. Cross-Administrative AI Strategy and Governance Model: Coordinated by the Digital Office, the AI strategy will be revised as part of Finland's digital strategy, the Digital Compass. (see 5.1 Updating Finland's AI strategy)

3. Utilizing Finland's Geographical Location: Finland can leverage its clean energy production, secure infrastructure, and stable conditions to attract data centers and AI development activities. (see 4.3 Competition for Computing Power and Semiconductor Industry)

4. Securing Europe's Competitiveness and Sovereignty: The development of artificial intelligence is a key factor in maintaining Europe's vitality. The EU must invest in AI research, innovation, and infrastructure to remain competitive with the United States and China. (see 4 Global Competition in Artificial Intelligence)

5. Harmonization of European Digital Regulation: The EU should focus on aligning recent regulations related to digitalization and data, eliminating overlaps, and harmonizing interpretations. Nationally, legislation must be updated to suit the AI era. For example, updates are needed in legislation concerning automated decision-making and secondary use of data. (see 5.2 Implementing EU Regulations and 5.3 The Need for National Legislation)

6. Preparing for Labor Market Changes: AI will automate routine work, create new jobs, and change the skills required. We must understand this transformation and reform the education system and labor market regulations accordingly. (see 1.1 Labor Market and Skills Policies)

7. Preparing for Changes in the Tax Base and Social Security Structures: AI may reduce the share of wage-based taxes in total tax revenue and concentrate value creation. Taxation and social security systems must be developed to respond to potential changes. (see 1.2 Taxation and 1.3. Social Security)

8. Protecting Democracy: AI influences democratic processes, for example, through decision-making, election interference, and the spread of misinformation. We must develop methods to prevent the misuse of AI that could endanger democracy and ensure that citizens' rights and freedoms are preserved. (see 2.2 Reliability of Information and Democracy)

9. Ethical and Sustainable AI Development: AI must operate ethically, and systems must comply with European fundamental rights and be understandable and transparent. We must ensure human oversight and accountability, especially in healthcare, the justice system, and the economy. (see 3 Ethical Development of Artificial Intelligence)

10. Promoting Global Agreements: Due to AI development, we need global-level agreements related to, for example, AI-assisted warfare, taxation, and AI ethics. Finland should promote the creation of global frameworks for agreement. (see for example 3 Ethical Development of Artificial Intelligence)

Introduction – “we haven’t seen anything yet”

With the rise of AI-driven consumer services like ChatGPT and others rapidly gaining traction, artificial intelligence has entered the mainstream, and interest in it has reached a peak. While this may be another bubble of enthusiasm, real advancements are occurring within this bubble, and these developments have profound and far-reaching societal impacts.

What is discussed when AI becomes a topic in politics? Conversations focus on responsibility and accountability, upheavals in labor markets, efficiency, innovation, technological transformations, education and retraining, the competition for expertise, ecosystems, research, investments, reforming public services, automated decision-making, biases in data, algorithmic transparency, citizens’ trust, deepfakes and political manipulation, the functioning of democracy, international and geopolitical competition, the need for regulation versus overregulation, energy demands and environmental impacts, privacy protection, copyright issues, cybersecurity, taxation, economic sustainability, AI infrastructure, and dozens of other themes.

This discussion paper serves as a guide to the most common and significant political themes related to AI. The development and application of AI in society progress at a faster pace than political decision-making. We must prepare for the extraordinarily rapid and large-scale changes brought by AI in a timely manner, which may often require challenging and lengthy political processes. On the other hand, we must recognize the limits of politics—some aspects of AI development can be guided through policy, while others cannot easily be influenced, though we can still prepare for their consequences.

The increasing use of AI in both the private and public sectors is desirable, and its potential is immense—for instance, simultaneously improving and enhancing public services, driving sustainable economic growth, and increasing well-being. Moreover, AI could help solve some of the world’s most pressing problems, such as climate change and biodiversity loss. However, risks include structural unemployment caused by rapidly changing skill demands, heightened polarization and its risks to democracy, changes in economic structures, growing inequality, and the erosion of privacy.

According to many international assessments, AI will accelerate economic growth in nations. Economic growth stems from long-term productivity increases, which depend on societal structures and how broadly and strategically AI is adopted. A study funded by Google estimates that Finland’s GDP growth due to AI could range between 2% and 8% over the next decade, depending on measures taken. According to the Global AI Index, Finland has the tenth highest potential in the world to benefit from AI. A McKinsey study published in January 2024 suggests that Finland could achieve annual economic growth of 0.7–3.6 percentage points by leveraging AI and other technologies that increase automation. Nobel laureate Acemoglu offers a more moderate global perspective, estimating AI could raise GDP by 0.9–1.1% over the next ten years if investments in AI grow moderately.

On the other hand, the future of AI is rife with both known and unknown threats, such as job retention, the implementation of copyright and privacy protections, and the automation of warfare. AI’s ability to generate convincing and impactful content also creates opportunities for misuse, such as spreading disinformation and manipulation. AI ethics involves considering these various chains of

impact, including how solutions affect people's lives, nature, society, and the economy. Humans cannot outsource their responsibility to machines. From the perspective of AI ethics, it is crucial for humans to remain in control and set goals for societal development.

What is Artificial Intelligence?

Artificial intelligence (AI) refers to computer systems or machines capable of mimicking human cognitive functions to perform various tasks. These capabilities often include reasoning, learning, problem-solving, perception, language comprehension, and creativity.

AI encompasses much more than just ChatGPT. Our understanding of AI, shaped by personal experiences with AI applications and general media coverage, might be limited. AI is a versatile field of technology with countless diverse applications and use cases. In everyday life, AI appears in recommendation algorithms for entertainment services and social media, as well as tools that support productivity and learning, such as text and image generation tools. Businesses and other organizations often use AI behind the scenes for automating decision-making, streamlining processes, and analyzing data. In scientific research, AI algorithms are employed in tasks like developing new drugs, modeling climate change, and simulating physical phenomena. Additionally, AI can be integrated into physical products such as vehicles and household appliances. These are just a few examples of the varied applications of AI.

Our perception of AI evolves over time—as AI applications advance, many of the previously groundbreaking innovations become so commonplace that they are no longer identified as AI. However, all AI systems share a key characteristic: the ability to operate and interact independently with their environment, setting AI apart from other technologies.

Generative AI refers to AI capable of independently creating new content, such as images, music, or text. Today, discussions about generative AI primarily focus on machine learning methods, though other branches of AI also exist and continue to develop. Learning from data distinguishes machine learning from traditional programming. In traditional programming, rules are defined during the programming phase, and the software processes input data based on those predefined rules to produce results. In contrast, machine learning involves a system learning from provided and recurring data. A machine learning algorithm is trained by feeding it large datasets from which it identifies patterns and develops an AI model.

This resulting model can then be used to perform tasks—this is how AI operates. For example, when a question is posed to a language model, the program compares the statistical patterns of the words in the question to the rules within its language model and generates sentences based on these rules.

If the training data for machine learning consists of text, the result is a language model. Similarly, datasets comprising images or motion-related kinetic data result in image or motion models. A foundation model is an overarching concept for large AI models specialized in various data types. Foundation models can also be multimodal, capable of interpreting and generating content such as images, sound, and text. Generative AI solutions are built on foundation models.

Modern AI, underpinned by neural network-based machine learning methods, mimics the functioning of biological neural networks. While neural network technology has been under development since the 1950s, it wasn't until the 2010s that Google's research division succeeded in optimizing the method sufficiently to enable the creation of large, versatile language or image models. A significant driver of recent rapid AI advancements has been improvements in computational infrastructure, particularly parallel computing using graphics processors.

It is only within the last decade that technical developments have allowed for sufficient computational power and storage capacity to enable these breakthroughs. Although technology like Google's is openly available, developing general-purpose foundation models (e.g., ChatGPT) or graphic models (e.g., DALL-E) requires massive amounts of data, which must be standardized for modeling purposes. Additionally, the expertise and computational power needed for these models are largely concentrated in tech giants. As a result, creating these models has initially been feasible only for companies like Google, Microsoft, or Amazon. However, businesses and organizations can utilize tools like ChatGPT by enhancing them with their own data, creating custom ChatGPT extensions for their customers.

The development trajectory of AI has been long, but in recent years, it has clearly crossed a threshold enabling the rapid development of new AI models and applications. As is often the case with technological progress, short-term advancements are overestimated, while long-term progress is underestimated. Self-driving cars haven't quite arrived yet but are likely to become widespread within a decade. Similarly, advancements in kinetic models are expected to bring warehouse and cleaning robots to the market around the same time. A significant transformation of the workforce is on the horizon.

1 Impacts of Artificial Intelligence on Labor Markets, Taxation, and Social Security

AI promises an unprecedented wave of efficiency, innovation, and productivity. However, the risk lies in its benefits being distributed unevenly. AI fundamentally alters how and where value is created, which can weaken the sense of inclusion for many people. Significant shifts in value creation—and consequently labor markets—challenge the structures of taxation and social security. These themes are deeply interconnected and cannot be addressed in isolation.

Harnessing technology requires skills, as well as access to information systems, infrastructure, and data—resources that not everyone possesses. Wealthy nations, companies, and individuals have better opportunities to leverage the benefits of technology, such as in healthcare and education. Already, the greatest benefits of digitalization have been concentrated among a small group of people and businesses. Without proper regulation, AI could accelerate the growth of socioeconomic divides, both globally and within nations and regions.

At its best, AI and automation can simultaneously boost productivity and enhance workplace well-being. AI applications can support human work, and many heavy, dangerous, or cognitively demanding tasks can be fully transferred to AI and robots. However, the rapid pace of change threatens to leave behind some individuals, organizations, and even nations. If implemented without consideration, AI's productivity effects could also be negative, leading to declines in quality without corresponding productivity gains, disrupting social processes, and disregarding the tacit knowledge that underpins organizational operations. How can we create flexibility in labor markets, workplace learning, and educational pathways to ensure a smooth and equitable transition to the new world of work?

The reduction of labor intensity in certain fields and the globalization of digital service businesses are two megatrends enabled by technological development. These trends could significantly impact the distribution of tax revenues across different types of taxes and between nations. How can we redefine tax structures to ensure fairness and promote the overall good of society?

The content and demands of work will undoubtedly change, though it remains uncertain whether automation will reduce the overall demand for human labor. In any case, social security systems face pressure to adapt—either to support people transitioning to new fields or to compensate for the declining significance of work as a source of income. Responsible policy-making and system reforms are essential to avoid creating a future marked by growing inequality and increased societal discontent. How can we build social security mechanisms that smooth the disruptions caused by transitions, fostering inclusion and cohesion?

These questions must be addressed now because AI's role in the economy will only grow. The pace of technological development is likely to outstrip societal and cultural adaptation, underscoring the need for proactive preparation and comprehensive measures in labor, skills, taxation, and social security policies. With AI's advancement, it is imperative to craft holistic policies for these interconnected areas, supporting innovation, encouraging sustainable growth, and ensuring that all members of society benefit from the wealth created by technology. Discussions about tax reform and, more broadly, income redistribution cannot be postponed but must begin promptly to establish fair and sustainable frameworks for future generations.

1.1 Labor Market and Skills Policies

AI and automation exert pressure on labor markets and professional skill requirements. Just as automation fundamentally transformed manual labor and its markets, AI is set to revolutionize office and administrative work. At the same time, technological innovations create new jobs and entirely new professions while reshaping how work is performed. In the long term, the transformation of the workforce will be radical, and even in the short term, the changes will be significant. Job descriptions, skill requirements, operating models, and service structures are evolving—affecting all sectors and organizational levels.

The gradual change has already begun in so-called white-collar work. In many cases, AI has simplified routine tasks, allowing employees to focus on creative and strategic work. For example, in programming, basic coding tasks have become heavily AI-supported, enabling programmers to concentrate on managing larger systems. On the other hand, some tasks are becoming democratized,

as seen in communications, where some content creation previously handled by specialists is now done using AI tools. AI is initially transforming knowledge work, and many manual labor sectors are not under the same immediate pressure. However, the advent of robotics will soon bring disruptions to routine physical work, as versatile and affordable robots enter the market. For instance, Tesla's upcoming multipurpose humanoid robots are expected to cost less than a car, around $20,000.

According to the World Economic Forum's (WEF) The Future of Jobs 2023 report, automation and AI are expected to replace a significant portion of existing jobs, potentially leading to more job losses than gains. The report suggests that up to 23% of all workers could either lose their jobs or need retraining. Additionally, over 85% of organizations view the adoption of new technologies as the most critical driver of change in the coming years. While AI's productivity benefits and their impacts on labor markets are challenging to predict, alternative perspectives exist. For instance, economist Daron Acemoglu has criticized overly optimistic estimates of AI's productivity benefits, arguing they may be more limited than anticipated.

The expansion of automation and investment in education must go hand in hand. According to the WEF report, 82% of organizations are investing in employee training, while 80% are accelerating automation. Countries and companies that invest in reskilling their workforce stand to benefit from automation by creating new, higher-productivity jobs. In particular, the technology sector is expected to see a significant increase in new job opportunities. For example, Amazon's use of robotics in warehouses has reduced the need for manual labor while increasing demand for technical specialists.

Emerging jobs, such as AI trainers, robotics maintenance technicians, and experts addressing ethical issues, highlight the need to redesign education systems. Alongside technological skills, social skills, critical thinking, problem-solving, and lifelong learning capabilities must also be emphasized. Ensuring broad access to retraining and continuous learning is essential to help people adapt to changing labor markets.

In Finland, labor market regulations and organizations must significantly enhance their readiness for change. Rigid labor markets and adversarial positions can hinder smooth transitions to new roles and slow the implementation of necessary retraining programs.

1.2 Taxation

Technological development raises critical questions about how to create a taxation system that accounts for the unique characteristics of the digital economy and ensures the sustainable financing of public services. Traditional tax models no longer align with the realities of a digitized and automated economy, where value is increasingly created immaterially and across borders.

Technology companies leveraging the power of advanced AI systems already generate significant profits with minimal workforces. The trend of decreasing labor intensity is accelerating, and it has even been speculated that we may soon see the first "one-person unicorn" startup—valued at over a billion dollars but employing only its founder, with all other operations automated.

Currently, national tax bases rely heavily on labor and consumption taxes. A common fear is that tax revenues will plummet if robots and AI replace human labor. Historically, machines and automation have driven productivity growth, which, in turn, has raised wages and strengthened tax bases. The issue is not so much “robots taking jobs” but “robots boosting sales and creating jobs.” For instance, Finnish factories have often increased their workforce alongside the introduction of new robots because higher production volumes, improved quality, and faster production have boosted sales.

However, the productivity leap brought by AI could be qualitatively different, potentially shrinking the tax base for income taxes. Recent productivity gains have already led to a decline in the share of wages in national income while entrepreneurial and capital income shares have risen.

One proposed solution is a robot tax, which would target the use of robots and automation systems in workplaces. The goal is to offset income tax losses and fund social programs, such as reskilling workers for new opportunities. However, implementing such a tax at a local level would be shortsighted. A robot tax would be akin to taxing other productivity-enhancing tools like computers, mobile phones, or excavators. If robots were heavily taxed or banned, production would simply shift to countries where they are not. Defining what qualifies as a “robot” for tax purposes is also challenging, as most labor-saving technologies do not fit the traditional definition but still exert similar pressures on structural unemployment. The EU Parliament rejected the idea of a robot tax in 2017.

A better approach is to reduce labor taxes to avoid “penalizing” human work compared to the returns of automation. Lower labor taxes would also help attract foreign workers. At the same time, tax revenue must be maintained by taxing the output and value added by both robots and humans—corporate profits, capital income, and value-added taxes (VAT) in appropriate proportions. While international tax competition remains a challenge, there is hope in global agreements on minimum taxes. The OECD (with around 140 countries) has established minimum tax rules, and the EU has already implemented a directive to enforce them across its member states.

The rise of remote work introduces new considerations. For example, in knowledge work, it may be worth exploring the possibility of waiving income taxes for workers who wish to remain in their home countries while working for Finnish companies. Such measures could improve worker retention and help companies secure critical talent. The EU could play a key role in simplifying and facilitating tax agreements to make such arrangements feasible across borders.

In addition to reduced labor intensity, another significant trend is the internationalization of digital business. Digital service companies, such as social media platforms, search engines, and gaming firms, operate globally but pay corporate taxes primarily based on their physical locations. Digital services also displace local businesses—for instance, global digital platforms have taken over local media advertising.

The concentration of tax revenues from digital companies in a few countries has increased pressure elsewhere to implement tax mechanisms granting consumption-location countries greater tax rights, even if the companies providing the services lack a physical presence there. Digital services taxes aim to ensure that digital companies contribute to tax revenues in countries where they have a significant user base, addressing profit-shifting by multinational tech firms and creating a level playing field between traditional and digital businesses.

Implementing digital services taxes requires reliable data collection and reporting on service usage locations. For consumer services, this raises privacy concerns, as usage and location data would need to be collected solely for tax purposes.

A global infrastructure for tax collection (similar to SEPA for payments) would be needed to avoid excessive administrative burdens and ensure simplicity for businesses operating in multiple countries under different tax regimes. Digital services taxes must not lead to double taxation, where digital companies pay taxes in both their home countries and the countries where they provide services. National or regional digital service taxes could also strain international relations, as countries hosting large tech companies may retaliate with trade sanctions or other measures. A coordinated international model through the OECD is a better alternative.

AI's role in economic automation is also emerging. For instance, fully automated stock funds already exist, where algorithms manage trading without human intervention. This suggests a future scenario where AI can autonomously purchase services, products, or information beyond the stock market. Economic dynamics could shift significantly as AI creates new value chains rapidly and continuously while optimizing its operations. These AI-driven transactions would likely use cryptocurrencies, managed and governed through blockchain-based decentralized autonomous organizations (DAOs). The legal status of DAOs and the regulation of autonomous algorithmic commerce, including taxation issues, should be proactively addressed.

1.3 Social security

Historically, improvements in labor productivity have not reduced the overall demand for human labor. Over the past century, labor productivity has increased approximately 15-fold, yet employment levels have remained relatively stable as new types of work have emerged. However, the changes brought by AI are occurring so rapidly that when considering the structure of social security, it is prudent to prepare for the possibility that history may not repeat itself. The proliferation of AI could lead to significant structural unemployment, as the remaining and newly created jobs may demand higher skills than those displaced by automation. In this scenario, displaced workers might be unable to transition to other jobs, resulting in a net increase in unemployment—a departure from the historical trend where productivity growth has always created new opportunities.

If automation and robotics were to significantly reduce the need for traditional human labor, fewer people would earn their livelihood from paid employment, leading to a decline in income tax revenues. Preparing for this scenario requires a fundamental reassessment of social security structures and funding sources. For example, in the United States, the rise of AI has sparked discussions about transitioning to a universal basic income (UBI) or a negative income tax-based social security model. Financing such models would necessitate emphasizing non-labor-related tax sources and harmonizing international taxation to prevent harmful tax competition.

If the overall demand for human labor does not decrease but instead changes in nature, the pressure on social security systems would be less severe. In this case, the focus would be on ensuring that the population's skills align with the evolving demands of the labor market. Even in this scenario, social security systems need to be reformed to include more flexible support mechanisms that help workers transition to new jobs and industries.

Social security solutions and education also have a feedback loop with the adoption of automation. Adequate levels of basic security and investments in lifelong learning reduce the supply of cheap labor for undesirable routine tasks, creating incentives to automate such jobs.

2 Other Societal Issues Related to Artificial Intelligence

Artificial intelligence is a general-purpose technology with an enormous range of applications. This wide spectrum of applications means there are also numerous societal questions related to AI. The public discussion has particularly highlighted the opportunities AI brings for advancing data-driven decision-making, the risks it poses to democracy through opinion manipulation, cybersecurity concerns, copyright issues, environmental impacts, and the use of AI in warfare.

As tools evolve, we are increasingly able to model global wicked problems and seek more effective solutions to them. Even global decision-making and democracy can rest on a stronger foundation when we have access to reliable forecasting models that can be used to guide decisions. On the other hand, AI's ability to manipulate content—and thereby influence human thinking—poses a significant risk to democracy. With minimal resources, AI tools can be used to generate vast amounts of credible images, text, audio, and video content, and distribute them widely across various online platforms. AI also enables easy mass customization of messages in ways that speak directly to each individual.

AI's role in cybersecurity is dual in nature: while it introduces new means of defense, it also creates new vulnerabilities.

The use of AI in content creation raises concerns about copyright enforcement, patent protection, and the ability to verify the authenticity and originality of works. These challenges are tied both to the training data used by AI models and to the content generated by AI, as well as how such content is treated. Legislation must not become stuck in the past in ways that hinder AI research or the development and application of new revenue models based on it. However, the rights of content creators must still be ensured in the future.

Training and using large AI models also have significant environmental impacts due to the computing capacity they require—and the resulting energy consumption. Sustainable AI development demands substantial investments in data center energy efficiency and the production of new clean energy to fully meet the growing energy demand.

Risks related to security and militarization are also serious concerns as AI is applied in military contexts and cybersecurity. Warfare has already become highly technological, and AI is a central component of modern weapons systems. Actors with access to new technologies may gain a significant advantage over even traditionally well-armed forces. International cooperation and the development of regulatory frameworks are essential to maintaining global security.

2.1 AI Enables Data-Driven Decision-Making

Never before have we had such an opportunity to create a unified situational picture of many interconnected and complex issues. AI enables the creation of situational awareness by efficiently

and in real-time processing vast and fragmented datasets. For example, AI was used to create a digital twin of global weather systems, significantly improving weather forecasting capabilities almost instantly.

Developing data collection and utilization plays a crucial role in addressing society's increasingly complex challenges. Efficient data collection allows for the analysis of broader datasets, providing more accurate insights to support decision-making. Utilizing AI and machine learning for data analysis enables unprecedentedly rapid and multidimensional data processing, opening doors to deeper understanding and more proactive decision-making.

For instance, open tax data and a real-time income registry can allow quick responses to economic conditions, such as implementing fiscal stimuli or tightening measures. Real-time data on tax payments also reflects citizens' economic activity and can support data-driven policymaking while ensuring privacy. For example, the real-time analysis of automatically reported household deduction data could inform rapid rule adjustments if needed.

Generative AI Transforms Weather and Climate Forecasting (NVIDIA, 3/2024)

NVIDIA's Earth-2 model is revolutionizing weather and climate change forecasting and modeling. It enables the creation of accurate simulations that allow for unprecedented precision and speed in predicting weather phenomena, such as storms and floods. Earth-2 can produce forecasts 1,000 times faster and 3,000 times more energy-efficiently than traditional models. This can improve preparedness for extreme weather events and global disaster readiness.

2.2 Reliability of Information and Democracy

Various actors actively exploit AI's capabilities to manipulate information to suit their purposes. The same technologies and tools are used for lawful and acceptable communication, such as targeted marketing, but also for criminal and manipulative purposes. We may not always recognize whether we are encountering tailored, mass-produced persuasive content or deceptive, false information created with malicious intent.

For instance, creating convincing deepfake content that imitates individuals or otherwise fabricates false narratives is becoming easier, while identifying such content is increasingly challenging as technology advances. This increases the risk of various scams and can fundamentally affect information dissemination and public opinion formation. Deepfake use during election campaigns has already impacted democratic elections in Europe and elsewhere.

Deepfakes have existed for years, such as falsified speeches by Barack Obama in 2017. However, the technology continues to improve, and the tools are now widely accessible. AI-generated ultra-realistic images, videos, and audio are advancing rapidly. The latest solutions already enable almost real-time deepfake creation using minimal data, such as cloning a person's voice from a three-second audio sample.

Previously, the Finnish language provided some protection against email scams, but advancements in language models have made Finnish-language scams highly convincing and introduced new elements, such as video and audio.

Automated content creation is transforming journalism. The competition for speed in news production is intensifying, increasing the risk that even traditionally reliable sources may inadvertently spread and legitimize fake news. A weakened media landscape, combined with increased production of falsehoods, distorts public discourse and the quality of information. False or inaccurate information can have severe consequences for social stability, public health, and the economy.

State actors also utilize artificial content. AI-assisted disinformation campaigns are becoming a significant component of warfare. For example, Russia has actively used AI-generated content to justify its aggression in Ukraine. AI has also made influence attempts on foreign decision-makers more sophisticated and subtle. NATO has explicitly addressed these challenges in its preparedness strategies.

Social media platforms' profit models pose additional challenges. These platforms are designed to maximize user engagement to increase advertising revenue, creating algorithms that present individuals with a personalized, biased view of the internet. This polarization is further exacerbated by prioritizing divisive content. Such algorithms are not aligned with societal well-being, leading to fragmented information bubbles and eroding public discourse while reinforcing prejudices against opposing viewpoints.

However, algorithms that drive discussions need not necessarily increase polarization or discord. Algorithms can be adjusted to optimize constructive dialogue and reward users for meaningful interaction. Research and experience with such "pro-social algorithms" exist, for instance, in Taiwan. These algorithms aim to foster interaction and dialogue between people with differing views. They emphasize content that challenges users' preconceptions and presents diverse perspectives constructively.

The ability of technology to shape political discussions and influence citizens highlights the need for policies that protect democracy and ensure access to reliable information. The creation and dissemination of false and misleading information must be actively countered, as its widespread presence poses a security threat to society and individuals.

Countermeasures include a diverse and critical media landscape, ethical journalism, interdisciplinary collaboration, the availability and preservation of credible information, and promoting critical thinking and media literacy among citizens. These countermeasures must be supported and developed. Citizens need to be educated to navigate the digital environment, with digital and media literacy cultivated from early childhood.

The Rise of Virtual Influencers (John Herrman, Intelligencer, 2023)
At the beginning of 2024, computer-generated social media influencers began appearing on platforms. These synthetic personas offer companies a cost-effective alternative to traditional influencers for advertising campaigns. While replacing human influencers entirely is unlikely, virtual influencers may compete in campaigns where a simple face is sufficient.

2.3 Intellectual Property Rights

We must address the questions and practices related to copyrights and data privacy concerning the materials used to train AI. While AI has primarily been used in digital production so far, advancements in robotics and 3D printing will soon bring AI into the realm of physical world modification. However, legislative changes may not be necessary if interpretation practices can be updated.

Human involvement in content creation will likely continue to hold special value. In copyright law, the creator's contribution is key. The open question remains how human-created content can be distinguished from content generated by generative AI and what hybrid forms may exist between entirely AI-generated and human-generated works. The assumption has been that content available online is human-made and thus subject to permissions unless otherwise indicated. From a copyright perspective, it is essential to develop ways to identify the human role in content creation, even if a strict distinction between human and AI-produced content may not always be meaningful.

Currently, copyrights protect human creative work expressed as literary or artistic creations. To qualify for copyright protection, a work must be the original result of its creator. The fulfillment of these requirements is assessed case by case, based on the "threshold of originality," which varies for different types of works. In Finland, only a natural person can hold copyright to a work. Purely AI-created images, music, or videos are not protected by copyright. However, a creator using AI as a tool in the creative process may hold copyright if their contribution is sufficiently independent and original.

A major paradigm shift in intellectual property rights concerns the idea of machines having agency and, consequently, authorship. This idea remains distant. In the coming years, focus will instead be on practical development, guidance through case law, and especially on determining human creativity and originality in assessing the threshold of originality.

The EU's competitiveness in AI depends on a well-functioning copyright system. This system can be strengthened by developing the technical infrastructure for managing copyright data. It relies on interoperable and trustworthy services based on reliable copyright information. New technologies, such as the EU's blockchain-based infrastructure (EBSI), can support copyright enforcement and creators' livelihoods. Finland has led this EU-level copyright data infrastructure effort since its presidency, focusing on technical solutions for use cases such as training language models in collaboration with other EU countries. Finland must also influence the new Commission to create a clear and balanced roadmap for this development and invest in measures tailored to the needs of various sectors.

2.4 Challenges of Cybersecurity

The discussion around AI-related security can be divided into three areas: risks leading to the misuse of AI, the use of AI to facilitate crimes, and traditional cybersecurity risks that affect all IT systems, including AI systems. It is also important to recognize that machine learning and AI have long been used in cybersecurity solutions, such as analytics and management services, as well as in quickly identifying threats.

AI systems are vulnerable to security breaches and misuse, where AI can be manipulated to act incorrectly, for example, through distorted input data. Additionally, cybersecurity threats include the risk that advanced AI systems might behave in unforeseen or incomprehensible ways.

The exploitation of AI systems enables the automation of activities aimed at compromising cybersecurity, accelerating and enhancing serious actions such as data breaches and sabotage. For instance, self-modifying and replicating malware can be created and distributed, posing a significant threat as such malware evolves rapidly, evading detection and countermeasures.

AI-powered software bots already search for vulnerabilities in IT systems. For example, the Vastaamo data breach was carried out using a software bot that attempted to log into systems with default usernames and passwords. AI can also be used to develop automated attack software that identifies vulnerabilities in other software quickly and effectively.

Ensuring the security of software products and management services throughout the supply chain is essential. The complexity of AI systems and the integration of various components can expose the entire supply chain to cybersecurity risks, regardless of where the issue originates.

The continuous updating and maintenance of AI systems are crucial because the cyber threats they face evolve rapidly, requiring adaptation to new conditions. Risk management strategies related to cybersecurity must be modern and reasonably easy to update to respond effectively to rapidly changing threats and ensure robust security management. The EU AI Regulation's fundamental rights impact assessments and the implementation of risk management measures based on risk assessments contribute to efforts to ensure the cybersecurity of AI systems.

2.5 Sustainability and the Environment

AI's impact on society goes far beyond its technological achievements, offering significant opportunities to advance sustainability and address global environmental challenges. Applications in combating climate change, managing natural resources, and protecting the environment enable more efficient and precise actions for the benefit of the planet. However, leveraging AI to achieve sustainability goals requires a responsible approach, and it is essential to recognize its significant negative environmental impacts, such as high energy consumption and water usage.

Data centers' electricity consumption is projected to double from 2022 to 2026. In 2022, data centers consumed 460 TWh of electricity, equivalent to 2% of global electricity usage, and this could rise to 650–1,050 TWh, comparable to Germany's annual electricity consumption.

Although legislation and technological advances, such as the European Union's Energy Efficiency Directive, may curb consumption growth, the Jevons paradox is evident: improvements in energy efficiency can lead to unexpected increases in overall consumption. This occurs because as energy efficiency reduces the energy costs of individual services or products, demand for those services grows, potentially resulting in greater overall energy consumption. For example, more efficient data centers might lower the cost of computational services, increasing the use of AI and other energy-intensive applications, ultimately raising energy consumption instead of reducing it. Addressing this challenge requires a systemic approach beyond mere technological improvements.

On the other hand, AI can support efforts against climate change, sustainable resource management, and environmental protection. Climate change impacts can be modeled more accurately, leading to the development of more effective mitigation strategies. AI can optimize renewable energy production and demand management, reducing the reliance on fossil fuels and lowering carbon emissions. By analyzing vast amounts of data and predicting future trends, AI helps optimize the use of water, land, and forests. This can reduce waste, support biodiversity, and promote the sustainable use of natural resources.

In environmental protection and species conservation, AI systems such as drones and AI-based monitoring systems enable real-time surveillance of protected areas. These tools provide critical information to prevent poaching and protect endangered species and their habitats. For example, damage to freshwater mussel populations could have been prevented with basic technology, where a machine's GPS could have simply blocked access to restricted areas.

In waste reduction and recycling, AI enhances efficiency by automating sorting processes and optimizing recycling chains. This can reduce landfill waste, support circular economy principles, and promote sustainable resource use.

Moreover, AI opens new opportunities for businesses to integrate sustainability more deeply into their operations. Analytics and predictive models enable companies to develop environmentally friendly and socially responsible practices.

Sustainable AI development requires attention to energy efficiency, material recycling, and expanding global carbon-free energy production. Training and using AI consume vast amounts of energy, highlighting the need for more energy-efficient processors and algorithms. Expanding renewable energy production is essential to meet AI's growing energy demands without increasing emissions. Simultaneously, circular economy principles, such as extending device lifespans and improving resource efficiency, must be promoted.

Open and interdisciplinary collaboration is crucial for sharing best practices and developing comprehensive solutions to environmental challenges. When used responsibly, AI can be a key player in building a greener and more sustainable future for everyone.

2.6 Warfare and Artificial Intelligence

AI is transforming warfare by enabling autonomous weapon systems capable of independently navigating to targets, identifying them, and attacking without human intervention. For example, a German defense AI startup provides technology for sensor fusion and collaboration between unmanned aerial vehicles and aircraft. Such weapons allow faster responses, are often cheaper to use, and reduce risks to human soldiers' lives.

AI might also reduce human errors, accidental discharges, or intentional misconduct, such as war crimes. However, at current technological levels, the risk of errors may increase if devices operate without continuous human oversight. The core question remains how reliably and predictably the system functions and what rules govern its operation. A middle ground involves semi-autonomous

systems, where AI performs target identification and/or designation, but a human ultimately authorizes the weapon's launch.

History offers examples where human judgment has averted catastrophe. In 1983, the Soviet Union's missile warning system falsely reported a massive incoming missile strike. A single officer chose to deviate from protocol and waited for visual confirmation before retaliating. Without this decision, nuclear escalation could have ensued. Such scenarios underscore the need to account for error possibilities when evaluating AI systems.

On future battlefields, the critical issue may not be the capabilities of individual weapons but the information exchange between different units, an area where AI could provide significant advantages. The political debate around AI-driven autonomous weapon systems has grown increasingly complex as military drones become easily and cheaply accessible to non-state actors. The UN has called for a global ban on autonomous weapons by 2026, and while most nations support this, not all have agreed to the proposal.

AI also has potential in intelligence, analysis, and conflict prevention. When combined with large datasets, AI can generate near real-time situational awareness of the battlefield. In conflict prevention, AI can analyze extensive historical, social, and political data to identify early signs of potential conflicts, enabling preemptive or peacebuilding measures. For instance, AI could monitor changes in economic, political, and social indicators in a target country to anticipate violent behavior. AI is also highly promising for simulations, such as war games, negotiation scenarios, and exploring peace mediation options.

AI in the War in Ukraine (Jason McGee, TechInformed, 2/24)
The war in Ukraine has already demonstrated practical applications of AI, particularly in military operations analysis and target designation. For example, Palantir Technologies' AI systems have been effectively used to identify and destroy Russian tanks. AI enables the collection and analysis of vast amounts of data, accelerating battlefield decision-making and enhancing the operational effectiveness of Ukrainian forces.

AI in Fighter Jets (Emma Roth, The Verge, 4/2024)
The U.S. Air Force has utilized AI to pilot a fighter jet. An AI-controlled aircraft participated in its first air combat test against a human pilot in September 2023. This test was part of a program to develop autonomous systems for piloting military aircraft and managing combat scenarios. During the test, the AI piloted the aircraft without requiring human intervention or safety overrides. Testing is set to continue throughout 2024.

3 Ethical Development of Artificial Intelligence

Throughout history, technological advancements, such as the widespread adoption of cars or centralized electricity production, have created new opportunities while bringing risks and challenges that require ethical scrutiny alongside technological innovations. AI parallels these earlier revolutions by reshaping daily life, transforming organizational practices, and accelerating societal change.

The rapid development of foundational models has raised various issues, including biases, injustices, lack of transparency, and risks of misuse. Addressing these challenges requires effective regulation, clear guidelines, and standards. Establishing international frameworks for agreeing on general principles for AI governance is also critically important.

The ethical development and implementation of AI should be a broader societal process that extends beyond technology developers and policymakers. Engaging citizens in this conversation is particularly important, as it can not only increase trust in AI but also bring valuable perspectives to its ethical development. Participatory democratic methods, such as deliberative citizen panels and democracy technology platforms, offer practical tools for incorporating citizens' voices into AI development. Systematic use of these methods can help build genuinely human-centered and ethical AI that enjoys broad societal acceptance and trust.

Ethics, a branch of philosophy, seeks to answer questions about how people should act in different situations and what values and norms should guide behavior. It addresses what we aim to promote and what we seek to prevent. AI ethics, a field of applied ethics, examines the effects of AI systems from the perspective of values and norms. Does the technology promote the things we want to uphold, or does it threaten them? How can we mitigate potential adverse effects?

Fundamental and human rights apply in the realm of AI as well. Public administration must adhere to the principles of good governance, and the legal protections of those involved must be ensured when AI is used in decision-making. Justice and harm avoidance should be pursued regardless of whether systems involve AI, and data protection must be upheld when processing personal information.

What is new is that AI systems can learn while making decisions. Although AI systems are highly capable and can operate largely independently, the ethical principle remains that humans set goals, and humans are always accountable for the actions of AI systems. Responsibility must also be clearly assigned. In public administration, this involves official accountability, but similar clarity about responsible parties must be demanded in the private sector, especially in high-risk applications.

All technology affects society, but AI's ability to act autonomously in its environment introduces new social and societal impacts compared to other technologies. For example, advanced AI combined with robotics enables machines to operate independently in the physical world, as seen with self-driving cars. Even more rapidly developing are AI agents operating in the digital world, capable of performing tasks such as conducting transactions. As AI applications grow more capable, their societal and social impacts become increasingly comprehensive.

As we navigate the crossroads of AI's opportunities and challenges, we must do so guided by ethical principles, fairness, and responsibility. It is essential to ensure that decision-makers have the skills to use data ethically and responsibly. This requires active participation and collaboration among all stakeholders—technology developers, researchers, policymakers, users, and civil society. AI must be approached beyond its technological aspects to ensure that its development respects human rights and benefits society broadly. Long-term perspectives are also critical; short-term gains may lead to undesirable outcomes in the future.

Aligning AI's operations with human needs aims to ensure that systems act ethically and safely. This involves effective programming and oversight mechanisms and AI's ability to understand and independently implement ethical principles. Ethical decision-making challenges must be addressed before foundational models or other AI technologies broadly surpass human capabilities in cognitive tasks. Otherwise, it will be difficult to keep AI development under human control.

Superalignment Project for AI Governance (Jan Leike ja Ilya Sutskever, OpenAI, 7/2024)
OpenAI has launched the Superalignment team to address the challenge of managing superintelligent AI systems. The goal is to develop AI capable of ensuring that future systems smarter than humans act in alignment with human intentions by 2029. Achieving this requires technical breakthroughs, such as developing automated monitoring and testing of models. OpenAI is dedicating 20% of its computational capacity over the next four years to solving this issue. (Neither of the article's authors currently works for OpenAI.)

3.1 Data and Bias

Bias in datasets refers to the fact that the data is not entirely representative, fair, or objective, which can lead to misleading or distorted results when used for analysis or modeling.

In machine learning, algorithms independently structure data and draw conclusions from it. The reliability and safety of AI systems depend heavily on how and with what data the algorithms are trained. The rapid growth of data volumes has been a key factor in enabling recent advancements in AI technologies, and the amount of data continues to increase.

However, data is not objective. The very existence of data means that someone has already decided which signals to collect and how to collect them. Thus, data is always incomplete and messy. Data is also political, reflecting historical opinions and biases. While data biases have always existed, AI amplifies their negative effects by replicating and reinforcing the biases in its training data, which can lead to discriminatory or biased outcomes. Some technical biases, such as a dataset's temporal coverage, can be tested and measured, but understanding how an AI based on data portrays a group or culture often cannot be easily discerned by simply analyzing the data.

The quality of data and its suitability for the intended purpose are among the most important factors for the quality, responsibility, and ethics of AI solutions. Data quality also affects its evidentiary value—the more critical the decision, the stricter the criteria for the data should be. Additionally, traceability of data is a central aspect of responsible AI solutions: the source of the data and the

reliability of the facts it contains must always be carefully assessed. Traceability helps ensure that the data used is trustworthy and its origins known, which is particularly important when making critical decisions. Understanding both the possibilities and limitations of data is a cornerstone of responsible and ethical AI development and use.

Many large datasets used to train foundational models currently fail to withstand critical scrutiny regarding bias. These datasets often consist of information accumulated from various projects since the early days of the internet. In the future, if we want AI to provide high-quality assistance in various areas, external criteria for data quality must be established. For example, if AI is to give answers inclusive of different cultures, those cultures need to be well represented in the data.

AI developers and providers must pay greater attention to the sources and collection methods of data and develop practices to identify and correct biases. This is essential to ensure that AI systems do not exacerbate but rather improve fairness and equality in society. However, fairness and non-discrimination cannot be resolved purely through mathematical methods because these are fundamentally philosophical, cultural, and legal questions. Therefore, AI development must be a multidisciplinary effort and cannot rest solely on the shoulders of programmers.

A significant source of bias is data degeneration, where AI uses previously AI-generated data as training material. This can quickly degrade the quality of the AI model and reinforce any initial biases. With large amounts of AI-assisted content now being produced online, filtering such material out of training data is becoming increasingly challenging.

On a political level, decisions need to be made about the extent of public control over data quality and the level of transparency requirements for data. For instance, OpenAI, the company behind ChatGPT, currently provides no visibility into the data used to train its models, despite millions of people using the tool daily. Policies must especially address the social biases inherent in data related to individuals and groups, which have no technical solutions. Examples include stereotypes tied to certain groups or the impact on unrepresented individuals. Without regulation and guidance, these issues are often overlooked in the fast pace of technological development and business contexts due to their complexity.

The Dutch Algorithm Scandal as a Warning for Europ (Melissa Heikkilä, Politico, 3/2022)
The Dutch tax authority's algorithm scandal, where AI falsely accused families of subsidy fraud, caused enormous human and economic suffering. Many families, especially from minority groups, fell into massive debt, and in some cases, children were placed in foster care. This case highlights the risks of using AI in public decision-making and sparks discussions in Europe about regulating algorithms and their human consequences.

3.2 Fairness

Fundamental and human rights, as well as legally defined anti-discrimination principles, must apply regardless of whether a system uses AI or not. The use of AI systems to manipulate or control people is therefore condemnable. Discriminatory practices in automated processes can easily spread to

affect large groups. However, discrimination exists even without algorithms, and machine learning can also be used to detect hidden biases.

When comparing humans and machines as decision-makers, it is essential to recognize that "technically" good decisions alone are insufficient. Machines can systematically apply decision-making criteria and scale the process efficiently. If uniformity is the goal, machines excel. However, human decision-making processes often involve interpersonal elements that are difficult to automate without losing something essential.

Special attention must be paid to equality, and procedures must be fair and traceable back to the origins of the decision-making chain. Citizens must trust that the principles guiding AI-supported systems are clear and well-understood. Moreover, systems must be designed to allow post hoc corrections if issues such as unequal treatment arise.

Trust is a cornerstone of fairness, and Finland, as a global leader in trust, must maintain this standing in the age of AI. Fairness also includes a material dimension, where the harms and benefits generated by AI should be balanced among citizens, businesses, NGOs, and the state. Ultimately, fairness depends on trust and the balance between the benefits and harms of AI.

3.3 Explainability and Responsibility

Transparency in AI decision-making is a prerequisite for sustainable democracy. The complexity of AI models can make their operations opaque, creating a "black box." Users often struggle to understand how and why models arrive at specific decisions or responses. Additionally, AI models are not static; they can develop new operational rules based on new data, which may be ethically sustainable or unsustainable. These challenges raise concerns, particularly given AI's role in critical decision-making areas such as healthcare, the justice system, and financial services.

The use of AI in decision-making does not currently eliminate human responsibility. Whether in public services, medical diagnoses, or any other processes traditionally handled by humans, accountability must remain with people. Building user trust requires transparency and comprehensibility in system operations.

Ensuring consumer protection, legal rights, and non-discrimination necessitates that individuals affected by automated decisions can appeal incorrect decisions and influence future decision-making processes. Automated decisions must be explainable to those they affect, directly or indirectly. Furthermore, affected individuals should have the normal ability to challenge and seek changes to AI-made decisions. Without explanations for decision backgrounds and a clearly accountable party—who must be human—erroneous decisions cannot be properly contested.

It is necessary to develop methods and standards to improve the transparency of AI systems, making it possible to understand how and why specific decisions are made. Evaluating language models involves assessing training data (how and where it was sourced), the explainability of their responses (how the AI reached its conclusion), and the quality of their responses.

3.4 Autonomy

AI systems must not manipulate, oppress, coerce, mislead, condition, or patronize people. While AI can guide individuals toward the most suitable services or provide helpful information for evaluating products and services, our society is built on the premise of autonomous individuals capable of making decisions and participating in democracy, markets, and civic activities.

Human autonomy, along with the realization of fundamental and human rights, must be considered throughout the design and use lifecycle of AI. AI systems should be designed to complement and enhance human cognitive, social, and cultural skills while adhering to human-centered principles. In work processes and operational management, authentic human oversight and control must be ensured.

The right to one's data—or informational self-determination—is a human right in the digital age. Individuals should determine who can access and use their data. Data is a production factor in the information society, much like labor and capital were a century ago. Strengthening individuals' ability to utilize and control their data is essential. Permissions to use data must be easily revocable, and data must be deletable and transferable, ensuring human autonomy and market functionality. Individuals should have access to essential services from both private and public sectors without having to surrender unnecessary personal data. Additionally, the quality or cost of such services should not be significantly inferior.

3.5 Avoiding Harm

AI systems must not cause or exacerbate harm or otherwise negatively impact people, animals, or the environment. Both physical and mental integrity must be safeguarded. AI systems and their operational environments should be safe and technically reliable. Special attention should be given to vulnerable individuals, who should, where possible, also be included in the development and implementation of AI systems. Separate solutions must address the challenges faced by underrepresented groups. No one should be discriminated against based on personal characteristics.

Human oversight is necessary to monitor AI learning processes and ensure adherence to non-discrimination and widely accepted ethical codes and values as the system evolves.

Privacy protection must also be ensured. AI systems can collect, analyze, and store unprecedented amounts of personal information, raising significant concerns about individual rights and freedoms. Special attention must be paid to data privacy in high-risk fields such as healthcare and justice systems, where issues are inherently personal. Mechanisms must be developed and implemented to protect individual privacy while maximizing the benefits of AI.

3.6 Respecting Languages and Cultural Contexts

AI tools are often globally operational, while norms and values are culture-specific and vary across regions. International regulation of AI requires finding common ethical denominators and ensuring that AI operates within these frameworks worldwide.

Large foundational models still do not sufficiently cover even many mid-sized languages in their applications. It is crucial to ensure that AI applications also accommodate less common languages and their cultural contexts. Estimates suggest that 70% of training data for language models is in English, meaning cultural emphases primarily come from English-speaking countries, particularly the United States. The lack of data from diverse cultural environments and languages reduces the diversity of AI applications, widens the digital divide, and limits the equitable global distribution of technological benefits.

The position of low-resource languages, such as Finnish, can be improved through curated and openly accessible training datasets, as well as targeted language models like Finland's Poro and the newer Viking model for Scandinavian languages.

4 Global Competition in Artificial Intelligence

The concentration of control over the most advanced AI technologies in the hands of a few companies or nation-states poses the risk of power centralization. Therefore, when developing technology, the goal of strategic autonomy must be considered, at least at the EU level.

Many nations and companies are striving for leadership in AI development. The geopolitical competition between the United States and China is particularly significant, often likened to the space race between the U.S. and the Soviet Union.

The U.S. justifies its geopolitical competition with internal and external security concerns. This has led to numerous measures restricting Chinese participation and cooperation in AI development, extending beyond AI to critical and emerging technologies, trade relations, and education. Similarly, China has its strategic initiatives aimed at excelling in the AI race. Currently, China leads in scientific publications on AI, while the U.S. boasts the most attractive market environment for AI companies.

Among companies, the race for AI technology is increasingly concentrated within established tech giants. Dominant market positions enable these companies to amass vast amounts of data, continuously invest in expertise through acquisitions, and expand computational capacity. All major tech companies are either developing their own foundational models or partnering with AI model developers. For example, Microsoft has a licensing agreement with OpenAI, Apple with Perplexity, Amazon and Google with Claude, and Google has also developed its own models. European players like France's Mistral have licensed their products to Microsoft. This has led to a consolidation trend where successful AI companies are acquired or tied into the ecosystems of tech giants through quasi-mergers.

Broadly speaking, advancements in AI capabilities have primarily resulted from the scaling of foundational models. Larger models generally perform significantly better in many tasks than smaller ones. Training larger models requires exponentially more computational power and data, leading to similarly exponential cost increases. For instance, training OpenAI's GPT-3.5, released in 2022, is estimated to have cost around $10 million. Training GPT-4 cost $100 million, and the still-unreleased GPT-5 is expected to exceed $1 billion.

The competition between nations and companies intertwines because AI companies are highly concentrated globally, primarily in the United States. By early 2024, there were approximately 70,000 AI companies worldwide, with a quarter based in the U.S. The U.S. has doubled its number of AI companies over the past seven years, maintaining its position as the leading destination for private capital investments. The U.S. also has 4.5 times as many AI unicorns (startups valued at over $1 billion) as China. This dominance suggests that U.S. companies will continue to wield significant influence over the development of international agreements and regulations.

The next major challenge in AI lies in scalability, emphasizing the availability of resources such as expertise, data, computational capacity, and, relatedly, microchips and energy. In global competition, control over these critical resources gives certain players power over others.

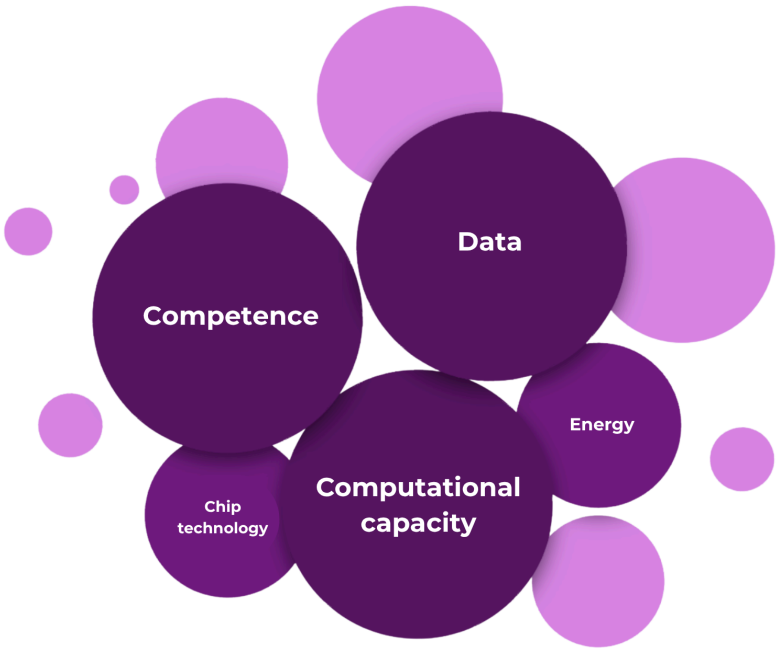

Figure 1: The most important resources for developing artificial intelligence are expertise, data, and computing capacity. Computing capacity, in turn, requires chip technology and clean energy. Together, these are critical resources over which there is intense global competition for control.

4.1 Competition for Expertise and Innovation

AI development and applications, chip technology advancements, data center construction, energy production, and the surrounding infrastructure all require skilled professionals, developers, and researchers. There is fierce international competition for experts in these fields, with expertise in many specialized areas already concentrated in a few global hubs. This trend of concentration reinforces itself, as investments attract talent, and talent attracts investments.

It is vital to develop AI expertise in Finland and attract talent from abroad. To do so, Finland must simplify the residence permit process, similar to Canada, provide services in English, and ensure employment opportunities, fair working conditions, and competitive salaries for newcomers and their families. Finland conducts highly advanced AI research at the European level, and its strengths in attracting international talent include a high education level, functional society and infrastructure, and beautiful nature.

As AI becomes increasingly integrated into the workplace, new skills such as understanding AI fundamentals, data handling, and collaborating with automated systems will be necessary. All Finns should have the opportunity to learn about AI and its applications. AI education should be incorporated at all levels, from early childhood to higher education. Adult and re-skilling programs should also be prioritized to enable those already in the workforce to update their skills. Keeping

pace with rapid changes requires constant updates to curricula, teacher training, and adequate educational funding.

Beyond developing expertise for AI development, individuals also need skills to navigate AI in everyday life. As machines surpass human capabilities in many tasks, critical thinking, creativity, systems thinking, and learning-to-learn skills must become central in school curricula. Understanding the connections between things and the implications of AI applications is essential for leveraging these tools wisely and protecting against risks. In a world where videos, photos, or even familiar voices on the phone may not be genuine, citizens must be equipped with awareness and source criticism.

In addition to individual expertise and talent attraction, it is crucial to invest in research and innovation within companies. Supporting the practical application of research will help establish AI-developing and AI-utilizing companies in Finland. Deep expertise is created and strengthened within communities.

Accelerating Finland's AI ecosystem also requires funding. Research, development, and innovation (RDI) funding should emphasize AI applications. Strengthening collaboration between funding organizations, both nationally and internationally, is essential. At the national level, closer cooperation between organizations like the Academy of Finland and Business Finland would be beneficial. Additionally, developing the expertise and collaboration of public RDI funding bodies is key to ensuring the most impactful allocation of resources.

Small and medium-sized enterprises (SMEs) must also enhance their readiness to incorporate AI into their products and services. The AI Finland initiative, launched by the Technology Industries of Finland, aims to promote AI adoption among SMEs.

4.2 Competition for Data

Developing AI increasingly requires more diverse and high-quality data—discussions on internet forums alone are insufficient for applications such as healthcare AI. Data has become an even more critical currency in the AI development market. Those with the most high-quality data can develop the best AI systems.

Training large foundational models has already utilized the entirety of the open internet, including copyrighted materials without permission. A critical challenge now is: where will new training data come from? Major players are competing for training data with significant financial investments.

For individual companies, large foundational models do not offer a competitive advantage as they are public goods accessible to all. The real value arises when foundational models are combined with an organization's proprietary data. These models can be fine-tuned with closed and internal datasets, enabling the AI to better address the specific needs of the company.

AI training often requires human intervention. Training typically involves humans evaluating the outputs generated by AI and correcting its mistakes. Tech companies have employed underpaid labor in poor working conditions for this mentally taxing task. Finland has also faced criticism for a trial where prisoners were used to train AI. Beyond unethical working conditions, cultural questions arise:

for example, what AI is taught to deem acceptable or unacceptable may reflect the values of different countries and companies. In U.S. culture, depicting violence may be more acceptable than nudity, unlike in Finnish culture.

As human-generated data is limited, the AI industry is exploring the use of synthetic data for model training. This “artificial” data can supplement real data and help reduce biases, but it also carries challenges, such as the risk of degraded model quality. So far, synthetic data has not resolved the issue of training data shortages, leaving data holders as highly sought-after partners in the near future.

In addition to synthetic data, another possibility is developing more diverse learning methods in which AI models generate their own training data, for instance, by simulating various scenarios and problem-solving situations in virtual environments. In such environments, models can experiment, play, and even improvise to acquire new skills and concepts in a manner more akin to human learning.

4.3 Competition for Computing Power and Semiconductor Industry

The rapid development of AI has triggered a global race for the infrastructure required by AI, such as semiconductor technology, computing power, data centers, and clean energy production. For example, in 2024, technology companies invested an amount equivalent to Finland’s GDP in constructing new data centers.

European unity is crucial in developing this infrastructure. Competing against nations like the U.S. and China requires avoiding fragmentation within Europe on essential infrastructure projects. It’s vital to identify projects executed collaboratively, leveraging the resources of all of Europe, as well as areas under the responsibility of nation-states and companies.

Finland currently hosts Europe’s most efficient supercomputer for demanding scientific computing, Lumi, whose further development would benefit all of Europe. The supercomputer also serves as an economical platform for training AI models. Decisions on developing European high-performance computing (HPC) infrastructure, including Lumi, are made under the EuroHPC regulation, which is currently being updated. In addition to Lumi, the quantum computer located in Otaniemi offers possibilities for combined use with the supercomputer—a globally rare combination.

The upcoming legislative work on digital infrastructure, to be prepared during the 2024 parliamentary term, provides Finland with an opportunity to actively influence the development of Europe’s shared computing power and infrastructure. Expanding EuroHPC capacity as part of the forthcoming Cloud and AI Development Act is a critical step. It’s important to note that resources needed for training and deploying AI models differ: while Lumi can train large AI models, deploying them requires computing power elsewhere.

Growing computing power for AI is intrinsically tied to the semiconductor industry. Semiconductor development is a key factor in geopolitical competition, as advancements in semiconductors are critical for the future of AI. The more powerful and abundant semiconductors are, the greater the

computing capacity becomes. Many developers of large foundational models have announced multi-billion-dollar investments in AI model processing chips.

Manufacturing semiconductors suited for advanced AI computations is highly specialized and conducted in limited cleanroom environments, primarily in Taiwan. For Europe, the exclusive production of EUV lithography machines (extreme ultraviolet lithography) by the Dutch company ASML is significant. These machines are essential for semiconductor production and are supported by an extensive network of subcontractors. Overall, semiconductor design and manufacturing form a single global production network that no single country or company can fully control. Taiwan possesses critical expertise, but even Taiwan cannot independently produce chips from start to finish. The same applies to the U.S.

Currently, the manufacturing of processors for AI computations is dominated by the American company Nvidia, which controls an estimated 80% of the market. Nvidia's processors were initially designed for rendering video game graphics but are particularly well-suited for the parallel computing demands of AI. Many AI libraries and software are optimized for Nvidia's hardware, and its CUDA programming environment has gained widespread adoption, enabling efficient development and optimization of AI applications.

Competition in designing processors and producing chips for AI computations is intensifying. Traditional processor manufacturers such as AMD and Intel, along with AI-specialized chip designers like Groq and SambaNova, are striving to challenge Nvidia's market position. Major tech companies, including Google, Meta, and Microsoft, are also developing their own AI chips for specific needs.

Semiconductors have become a focal point in global politics. The U.S. government has imposed export restrictions on the most advanced chips, particularly to China. Despite this, U.S. chip manufacturers have sold billions of dollars' worth of chips to China through creative means to bypass the restrictions.

Finland has also taken steps in the semiconductor industry. Starting chip production in Tampere appears likely, supported by €40 million in funding from the European Chips Act. However, the chips potentially manufactured in Finland will differ significantly from the most sought-after and competitive AI chips.

Finland's Semiconductor Strategy (Teknologiateollisuus, 4/2024)
Semiconductors are essential for modern technology, from smartphones to quantum computers. Finland aims to triple the turnover of its semiconductor industry by 2035, leveraging strengths in chip design, MEMS technology, photonics, and quantum technologies. The strategy focuses on expanding research and development, increasing the workforce, and attracting international investments. The goal is to create 20,000 new jobs and generate economic impacts of €90–180 billion. Finland seeks to become a leading player in advanced semiconductor technologies, emphasizing sustainable development and the strategic autonomy of the European Union.

4.5 Energy

Data centers used for training and running AI models require vast amounts of energy for cooling and computation. Energy resource management and availability are becoming increasingly critical factors. Nations with access to affordable and sustainable energy sources may gain a competitive edge in AI development.

In this context, the Nordic countries are excellently positioned due to their cool climate, stable ground conditions, and politically stable environments. Norway and Sweden are leaders in producing zero-emission and cost-effective energy, but Finland also has opportunities to invest in renewable energy to attract companies to establish data centers within its borders. Successful initiatives could help Finland, together with other Nordic countries, create a northern hub for AI computation.

Energy consumption in AI development occurs in three stages: collecting training data, training models, and deploying them. Data processing and storage take place in data centers that operate powerful supercomputers. The two largest energy needs in data centers are electricity to run the supercomputers and the energy required for their cooling. Electricity for operations can come from either renewable or fossil fuel sources, and cooling solutions vary widely. Finland's cool climate makes it easier to cool data centers, and the heat generated by machines is often redirected to district heating systems, providing heat to local households. In contrast, in many other countries, data centers rely purely on mechanical cooling powered by electricity, consuming large amounts of water, which is already scarce in hotter regions.

The number of data centers will continue to grow. Predictions for their future energy needs are staggering, as the computing power required for AI development continues to increase. Finland's geographic location, functioning infrastructure, and stable operating environment position it as a potential key player in energy production and as a prime location for data centers.

Recent developments in AI highlight the enormous energy demands of training models. Current estimates suggest that the energy consumption of AI-focused data centers rivals that of a small nation's annual electricity use, and this need is projected to grow significantly if current trends persist.

Electricity alone is not the only critical factor. Alongside energy availability, cooling solutions are equally vital. Today's tech giants—Microsoft, Google, and Meta—consume massive amounts of water to cool their data centers, an approach that is unsustainable in the long term. Investments in clean energy are being made even within the AI sector itself.

4.6 Experimentation Culture and Risk-Taking

Adopting AI in the public sector requires embracing a culture of experimentation and increasing risk tolerance. Encouraging innovative solutions and accepting a degree of uncertainty and mistakes—inevitable when utilizing new technologies—is essential. This requires a cultural shift that sees learning from mistakes and continuous improvement as central to organizational development. At the same time, successful solutions must be identified and widely disseminated across society, crossing organizational boundaries. This necessitates building models for inter-organizational

knowledge sharing and creating incentives for sharing data, applications, and co-developing solutions.

The public sector must be prepared to take calculated risks and support the development of new operational models to address the complex challenges of modern society. Risk-taking incentives could include shares in products or companies developed with public resources.

The responsibility of individual officials for decisions must be humanized by creating a clear channel for obtaining preliminary opinions on applications or solutions. For example, the approval processes brought by the EU AI Act should be made flexible and quick to ensure that emerging ideas can be implemented as swiftly and safely as possible.

Financing innovative companies and ideas is a central challenge in Europe, where obtaining funding is often slower and more bureaucratic compared to the United States. One significant difference is the proportion of private venture capital relative to GDP, which is much higher in the U.S., where public funding plays a smaller role. This funding disparity puts Europe at a disadvantage in global competition, as innovative startups often lack sufficient capital. However, U.S. venture capital also plays a significant role in scaling European AI companies. A possible solution to this challenge could be greater utilization of pension funds. In the U.S., pension funds are a significant source of venture capital, but in Europe, their potential remains untapped due to regulations and risk aversion. By reforming regulations and encouraging pension funds to invest in innovations, Europe could unlock significant new funding sources to support the growth of new companies and technological development.

Lowering the barriers for public and private entities to utilize AI would be beneficial by providing support for identifying and developing useful use cases and creating solutions that instill confidence in adopting applications. A common service channel could be established to guide questions related to the AI Act and other AI application-related legislation. This service channel could be implemented nationally, as done in Austria, or through Nordic cooperation.

5 Leadership and Legislation in AI Development

AI development and utilization are progressing rapidly worldwide, bringing new opportunities and threats. Political leadership plays a crucial role in leveraging AI's potential and preparing for its risks. Finland's long history in AI development provides a strong foundation for continued success in harnessing AI's capabilities.

Finland's public administration has systematically produced information on AI for seven years. The 2017 Finnish AI Strategy was among the first national AI strategies and has since been supplemented by the 2022 AI 4.0 program and the 2024 recommendations of the Technology Negotiation Council. These initiatives have addressed AI's general impacts and outlined national goals and trends across various administrative sectors. However, Finland lacks an active, comprehensive operational model for systematically anticipating AI's opportunities and challenges. Additionally, the level of ambition must be significantly raised. For example, Sweden has already published an AI strategy addressing generative AI in November 2024.

Achieving AI's potential requires national measures to examine AI adoption across different sectors relative to the current state. Each sector must develop roadmaps to systematically enhance the use of AI and automation.

A significant portion of AI-related legislation comes from the EU and must be considered in all national regulations. Finland should avoid overlapping legislation while addressing the need to update existing laws to align with the AI era and create a clear, supportive regulatory environment.

5.1 Updating Finland's AI Strategy

Finland must create a clear, cross-ministerial strategic vision for AI utilization, governed parliamentary across electoral terms. This vision should be integrated into the Digital Decade plan and coordinated through the Digital Compass Office.

Key elements of this vision should include:

- **Target Stat**e: Define an ambitious yet achievable position for Finland in AI development and utilization, leveraging its strengths.
- **Concrete Goal**s: Outline Finland's specific objectives in AI adoption.
- **Actions**: Detail plans for widespread societal adoption of AI and automation.
- **Governance Model**: Establish leadership structures to maximize AI's benefits nationwide.
- **Resourcing**: Identify responsibilities and allocate resources to implement these actions.

AI adoption in the public sector must be efficient, responsible, and impactful. Effective resource allocation requires a comprehensive, cross-sectoral assessment to prioritize initiatives with the highest societal and economic benefits.

Finland's strategy should align with Europe's broader AI ambitions. Finland can strengthen its position as an AI leader by ensuring Europe's status as a global AI powerhouse. Preserving cultural diversity and ensuring resilience require the EU to raise its ambition level and enhance its strategic autonomy in AI. Competitive and reliable European foundation models are essential for building new applications and commercial services. However, as AI advances, foundation models themselves will become commodities, with most value arising from how organizations fine-tune these models using proprietary data and integrate them into human-AI interfaces and production ecosystems. Finland, as part of Europe, must consider its contributions to global AI ecosystems.

5.2 Implementing EU Regulations

Europe has primarily focused on regulatory issues in AI development, which risks leaving Europeans in a spectator role. Europe's key challenge will be creating a unified and robust market environment that supports the establishment and growth of AI startups and directs private capital flows to Europe instead of the United States. A positive European example is France's AI ecosystem, which has developed through collaboration between private capital, public investment, and university structures. On the other hand, Europe's emphasis on human rights and ethics can be seen as a strength, representing an intrinsic value for preserving humanity in the AI era. At the same time, this approach could provide a competitive advantage if the market environment becomes functional.

Finland must invest in influencing EU legislation and ensuring consistent implementation of regulations across EU member states. The EU has recently adopted a new AI Act, complemented by other digital regulations, such as the General Data Protection Regulation (GDPR), the Data Act, and sector-specific regulations like the European Health Data Space (EHDS). These frameworks guide AI development in Europe. It is now crucial to influence the implementation of European regulations, ensuring their application is straightforward and unambiguous, with access to preliminary opinions and guidance. Implementation must focus on enforcement measures, standardization, technical specifications, and above all, consistency in interpreting regulations. This requires evaluating the effectiveness of different regulations, cross-checking overlaps, eliminating redundancies, and correcting regulations if necessary. How the European Commission establishes the EU AI Office and succeeds in recruiting the right experts will significantly impact the success of the AI Act in achieving its goals.

The EU AI Act also requires Finland to establish a regulatory testing environment, commonly referred to as a regulatory sandbox. Sandboxes offer companies the opportunity to test and develop new products and services in a supervised environment with specific regulatory exemptions. The goal is to advance both technology and regulation simultaneously. Regulatory sandboxes enable companies to experiment and test new ideas before market entry while allowing authorities to gain valuable insights into the impacts of AI technologies and necessary regulatory actions. Implementing the sandbox should not be delayed. Many EU countries have already piloted their sandboxes, providing lessons to learn from, while bold, independent choices could allow Finland to attract AI developers beyond its borders.

During the current parliamentary term, regulation is led by Commissioner Henna Virkkunen. Ensuring that other skilled Finns are placed in the right forums and networks is the most effective way to promote Finland's key interests. Finland should proactively identify development-oriented partner countries and build a shared agenda with them. For instance, delegated acts within the AI Act and other digital regulations will still undergo political processes, and Finland, together with its partners, must ensure their successful implementation. Suitable partner countries include those in the LUMI supercomputer consortium and members of the digitally advanced D9+ group.

Meta and Apple Withhold Latest AI Models from Europe
Ben Lovejoy, 9to5Mac, 7/2024, "After Apple, Meta also withholding future AI models from EU countries"

Meta and Apple have delayed releasing their latest AI models in European markets, citing regulatory uncertainties in the EU. This raises concerns that stringent regulatory requirements hinder innovation in the region. Meta is withholding its multimodal Llama AI model from release in the EU until regulatory clarity is achieved. This could slow AI development and weaken Europe's position in the global technology market.

Regulation advocates argue that the restrictions are working as intended, reducing the likelihood that technology companies will use end-users as guinea pigs for incomplete AI products rushed to market.

5.3 The Need for National Legislation

AI thrives on data, which is why data mobility requires special attention at the national level. For instance, the Finnish Secondary Use Act effectively prevents the real-time use of data by anyone other than registry holders, making available data at least a year old. This is unsustainable in the AI era.

Furthermore, legislation on automated decision-making needs updating as part of the working group implementing the EU AI Act. Currently, municipalities are unable to utilize automated decision-making effectively because the existing legislation does not allow it if the appeal authority is different from the entity making the automated decision. Most appeals concerning municipal decisions are directed to other authorities.

At present, general legislation on administrative automated decision-making is tied to rule-based automation and does not yet account for the use of generative AI models in decision-making. The rights of AI systems to make decisions must be evaluated on a risk-based approach. For example, language models could be used for certain decisions, accepting a predefined level of errors and their impacts. AI-based decision-making cannot currently be made entirely error-free, but neither can decisions made by humans. The key is to achieve fewer errors than the existing solution. At the same time, it is crucial to ensure that responsibility for AI-made decisions is always assigned to a person or organization. Even if AI can reduce the number of errors, it is fundamentally important to have a clearly accountable party for any harm caused and for resolving difficult ethical questions.

A risk-based legislative framework should be developed for verifying work performed by AI, as it is unsustainable to require humans to manually review large volumes of data. AI performs such tasks better, and in addition to random sampling, only exceptional cases should be flagged for human review. Humans should not be tasked with monotonous data analysis, where they would likely make significantly more errors than AI.

In Finland, data usage rights are largely determined by direct legislation or licensing. Therefore, it is essential to have straightforward ways to license data usage and to revise legislation related to data utilization to be compatible with the AI era.

The implementation of regulation should identify the most critical legislation for collaborative development and ensure its enforcement, for example, through concurrent funding for various stakeholders. The government program offers a tool in the form of a co-development budget, which, if successfully applied, could lead to significant improvements over the current situation. When drafting national regulations, compatibility with the EU AI Act must be ensured.

6 Future Directions in AI Development

With the advent of ChatGPT, the use of AI has started to become mainstream, making AI accessible to everyone worldwide simultaneously. Large foundational models are pushing the boundaries of current applications and deepening our understanding of machine learning and intelligence. However, large foundational models are just one component of the AI ecosystem and a step toward even smarter technologies. Many researchers and leading AI companies state that the next significant advances in AI development will occur outside foundational models.

In addition to large foundational models, smaller, task-specific models are being developed, which are more efficient and better suited for specific purposes. For example, Google and OpenAI are developing user-friendly platforms that allow users to customize language models to their needs without programming skills.

Current AI services are often closed platforms provided by commercial operators, but the market also includes open-source models, such as Meta Llama and Mistral, which challenge these platforms. In principle, open models can be utilized by anyone with sufficiently powerful hardware. However, large tech companies have a competitive advantage as they can provide the necessary computational resources and standardize industry development according to their priorities.

Technological boundaries are constantly being pushed. Today, advanced foundational models and AI-based assistants are already part of many people's daily lives. AI technologies are also accelerating the development of augmented and virtual reality to become even more immersive. The time is not far off when integrations between humans and machines will not just be laboratory experiments but part of everyday life. These new human-machine interaction technologies will offer richer and more immersive experiences for work, learning, entertainment, and social interaction, reshaping our perception of reality.

AI development is often described on a continuum from narrow AI to artificial general intelligence (AGI). Narrow AI refers to AI designed and trained to perform a specific task without human intervention. General AI refers to a theoretical form of AI capable of understanding, learning, and applying knowledge across all areas of intellectual activity as well as or better than humans. As we move from narrow AI toward more general AI, the autonomy, adaptability, and ability of AI to influence its environment increase. Current AI applications, such as algorithms providing movie recommendations, speech recognition systems, and chatbots, represent narrow AI. They operate in well-defined environments with clearly specified tasks.

6.1 Small and Local AI Models

Many players, such as Google, Mistral, Meta, and Apple, are developing small AI models alongside foundational models. The market is shifting toward smaller models for economic reasons since the operating costs of large models are high, and most use cases do not require the capabilities of the largest models. Inference costs, or operating costs, refer to the resources (computational power, energy, time) needed to use an AI model to generate results. These costs depend on the model's size

and complexity, the hardware used, and the required processing speed. Operating costs differ from the training costs of a model, which are significant but one-time.

Companies develop their own smaller models, which can be trained with the organization's proprietary data to perform specific tasks exceptionally well. This also ensures data protection for organizations, as training data and interactions with AI do not leave the local environment. High-security environments such as healthcare, banking, insurance, and government operations particularly benefit from this development.

The advancement of small, resource-efficient models enables AI to operate locally on phones and user devices and more broadly in various physical devices, such as surveillance cameras and traffic lights. This trend, known as edge computing, involves data processing being performed as close as possible to the physical location where the data is collected rather than being sent to large data centers over the network. Edge computing and centralized computing can also work together. When a small, device-based AI model encounters a task beyond its capacity, it sends a request to a larger AI model in the cloud for processing.

Small and local models are expected to become part of the operating systems of phones and computers. AI can optimize resource management, provide predictive suggestions, and automate routine tasks. As part of the operating system, AI models primarily use local device data. Models are trained before installation and continue learning through interaction with users. AI models operating on smartphones enable personalized AI agents for users. For example, Google, Apple, and Microsoft have already introduced device-based AI models that enhance the capabilities of operating systems and individual applications.

6.2 Open and Closed Models

Access to AI today typically comes through subscription-based or usage-based services. Well-known cloud services, such as Microsoft, Google, and Amazon, offer various AI tools ranging from language translation to image recognition. Recently, large foundational models like OpenAI's ChatGPT have entered the market. These company-hosted services are closed in the sense that the AI system's source code and training materials are only known to the service provider.

A significant portion of today's computer systems relies on open-source software. The principle of open-source software is that the source code is available to anyone and can be used as they see fit. Some open-source licenses allow free commercial use, but there are exceptions. AI models are also available under open-source licenses.

Open-license AI models can be used by anyone with a sufficiently powerful computer to run the models. However, not all openly available AI models are entirely open—they may lack information about training materials or include restrictive license terms. For example, Meta has released its Llama model for public use. Similarly, the French company Mistral has made its model publicly available, but its training data is not disclosed. Finnish company Silo.ai (now owned by AMD) has documented its training data for the Poro and Viking language models with academic precision. The Hugging Face community plays a significant role in connecting developers and users of open AI models, offering a wide range of models for text, image, and audio processing.

Open models have the advantage of helping the AI community develop better models, compare their performance, and create new services using AI models. Open licensing allows for crowdsourcing development efforts and provides opportunities for entirely new innovations. Open models enable anyone to further develop and apply the models for desired purposes, for better or worse.

While open models provide opportunities for collaborative development and increased transparency, their impact on market structures and competition remains limited. Many large tech companies continue to benefit most from open models, as they provide significant computational resources and cloud services often required to utilize open models. Additionally, large companies can set industry standards by supporting specific open-source projects, thereby influencing the technical solutions and practices favored in the market. This power dynamic raises questions about whether openness truly democratizes the AI ecosystem or ultimately benefits large corporations and entrenches concentrated market power.

6.3 Interaction Between AI and Humans

Interaction between humans and computers is becoming increasingly intuitive, as AI-powered voice and text-based interfaces enable natural language communication with machines. This simplifies and enhances the user experience, making technology more accessible to users of all ages and backgrounds. More broadly, these advancements are transforming how we communicate, learn new skills, and work, bringing both new opportunities and challenges.

One challenge in AI development is its limited ability to understand and interpret human emotions and social signals. Diversifying the training data for AI systems from text to images, video, and audio has opened possibilities for developing emotion recognition. For example, OpenAI's GPT-4o model, released in May 2024, can somewhat recognize expressions and emotions. In the future, we can expect AI systems to better interpret tones of voice, expressions, gestures, and other non-verbal communication forms, enabling deeper and more meaningful interaction with humans. However, emotion recognition is highly complex and error-prone because emotions are so ambiguous, culturally specific, and context-dependent that even humans struggle to interpret them correctly. This field risks veering into pseudoscience, raising concerns among privacy and human rights activists. China has been a visible example of the problems associated with using emotion recognition technology. The country's authorities have utilized the technology to monitor minorities like the Uyghurs, attempting to identify "suspicious" behavior and anxiety. Organizations like Access Now and European Digital Rights have called for a complete ban on emotion recognition. The EU AI Act prohibits the use of emotion recognition in policing, border control, workplaces, and schools.

A significant development in human-AI interaction is the ability of advanced AI-based assistants to provide tailored assistance and advice. By learning users' preferences and needs, they can support various daily activities, ranging from healthcare advice to personal financial management and education.

In the future, we may have a large array of specialized AIs that are aware of the physical world we observe and can assist us from the digital realm. Current early examples include AI assistants from Microsoft, Google, and Apple, which allow users to automate various processes and manage information comprehensively within these companies' application ecosystems. The next step in

development is already visible in the integration of AI applications with other applications and their operation. These semi-autonomous AI assistants are called agents. What distinguishes an AI agent from a conventional AI program (e.g., chatbots) is the ability to perform actions such as calling APIs. For example, a search agent can respond to our conversations and automatically find solutions. Such autonomous AI agents, capable of using various tools extensively, can be introduced to perform human tasks, either alongside people or as “independent workers.”

AI-enabled reform of learning and education is another significant area. By offering personalized learning experiences that match an individual’s learning style and pace, AI can help overcome the limitations of traditional education and provide more effective and meaningful learning experiences.

Additionally, AI opens the door to new forms of interaction. For example, virtual assistants and chatbots can serve as moderators and facilitators in conversations, improving the quality of meetings and promoting collaboration among participants. Similarly, AI can act as a tool to enable scalable and efficient interaction between customers and companies. AI can, for example, analyze large amounts of customer communication, such as emails, chat conversations, or survey responses, and provide insights to product managers on how to improve products or services.

As AI advances, social challenges will grow. For instance, how can an AI-based therapy service designed to be as engaging as possible align with an individual’s best interests? This challenge must be addressed as we face developments that significantly exacerbate existing social challenges.

6.4 The Blurring Boundaries Between the Digital and Physical Worlds

AI also has a physical dimension, most visibly seen in robotics. Robotics refers to the automation of various processes where machines and software take over tasks previously requiring human effort. Robotics has existed even before modern AI, but AI development is propelling the field forward significantly. An example of this is the integration of motion models into robots, enabling them to learn by experimenting in virtual environments, watching videos, and mimicking humans.

In the future, robots are expected to become more intelligent, enhancing their range of capabilities. For instance, a robot named Figure learns to perform complex tasks by observing humans. This model is already being tested on BMW’s production line, and its second-generation version can handle multiple tasks with voice commands. The startup Covarian has introduced a method for communicating with robots via images and text, simplifying their training compared to earlier methods. Among tech giants, Nvidia has developed a methodology based on large foundational models that allows AI to design algorithms for training robots.

Another trend bridging the physical and digital worlds is the development of brain-computer interfaces (BCIs) and cybernetic or biotechnological integrations. Examples of future possibilities include real-time health-monitoring implants, neuroprosthetics that restore mobility or senses to the disabled, and thought-controlled devices that enhance our work efficiency and creative expression. For example, Apple has developed and patented technology that enables measuring biological signals, such as brainwave activity, through AirPods via the ear. Another example of a brain-computer interface is the solution by the U.S.-based neurotechnology company Synchron, which enabled an

ALS patient paralyzed in their limbs to use virtual glasses to control hand movements through a combination of eye tracking and thought control using technology implanted in the brain cortex.

Neurosurveillance and neurohacking have been highlighted as significant risks. Technology already exists to monitor employees' focus by tracking brainwave patterns. Duke University professor Nita Farahany has warned that we are at a critical crossroads defining how freely human brains will be exposed to surveillance and hacking in the future. Farahany advocates for declaring mental privacy a human right.

While connecting brains to computers may not become commonplace just yet, the proliferation of technically robust and lightweight virtual reality glasses is imminent. With VR glasses, computer-generated augmented reality will become a daily layer of the physical world for many. This virtual environment, partially blended with the physical world and partially independent, is called the metaverse. Developers of the metaverse envision a future where we gather in the metaverse for experiences that feel almost as real as those in the physical world. However, industrial use cases, such as remotely conducted machinery maintenance, are likely to materialize much sooner than this vision.

The everyday adoption of augmented reality applications may further exacerbate social fragmentation, as people may indeed begin to see the world differently. At the same time, the increasing realism of virtual reality is introducing entirely new challenges. For example, AI-powered virtual girlfriends or boyfriends have led people to fall in love with digital entities. These challenges are expected to grow significantly as next-generation AI models enable the creation of virtual individuals that look, sound, and feel real, integrating seamlessly into our lives.

South Korea in Turmoil Over Spread of Deepfake Pornography – Authorities' Actions Fall Short for Victims (BBC, 9/2024)
A major scandal has erupted in South Korea over the proliferation of deepfake pornography, where AI is used to manipulate images of young women and even minors in sexually degrading ways. These images are shared in Telegram groups, and the number of cases has risen significantly. While authorities have launched investigations, many victims and activists are outraged by the lack of adequate oversight on platforms like Telegram.

6.5 From Narrow AI to General AI

Foundational models represent a significant leap in the realm of narrow AI, particularly in understanding and generating natural language. While these models are not general AI, their ability to process and produce language is remarkably versatile even in this early stage of development. The capabilities of foundational models are expanding beyond natural language processing into other areas, and it is anticipated that future AIs will be able to grasp broader contexts and better understand the world. For instance, graphical models have already incorporated basic mechanics, enabling objects in AI-generated videos to adhere to principles like gravity and ensuring that humans always have five fingers. However, the ability to generate human-like text or photorealistic videos

does not equate to genuine intelligence or consciousness. The rapid development of creative AI lays the groundwork for future research toward more versatile and advanced AI systems.

One of the most significant abilities of the human mind is abstract thinking—the capacity to understand and process ideas that are not directly observable or concrete. AI is expected to advance in the realm of abstract thinking, enabling it to solve complex problems and develop creative solutions in ways increasingly similar to human thought processes.

General AI is anticipated to apply its learned knowledge broadly across different fields and situations. It would possess the ability for creative thinking, skilled decision-making, and independently performing various tasks. General AI would be capable of acquiring knowledge from diverse sources, making logical inferences based on incomplete information, adapting learned concepts to new circumstances and environments, and autonomously correcting its malfunctions—all without human intervention. Achieving this requires advanced mechanisms for self-assessment and optimization, enabling AI systems to continually improve and adapt. For now, general AI remains a subject of research and speculation.

The progression from narrow AI to general AI is described through tier models. For example, OpenAI's roadmap for achieving general AI advances through five levels: 1) conversational AI capable of engaging in natural discussions, 2) "thinkers" solving problems with doctoral-level expertise, 3) autonomous agents operating independently for extended periods, 4) "innovators" generating new ideas and solutions, and 5) "organizations" functioning as entire entities. OpenAI estimates it is at level one and approaching level two. Similarly, Google's DeepMind has proposed its own tier model, progressing from non-expert performance to superhuman capabilities exceeding human skills. The goal is to establish a shared standard among AI developers to clarify what general AI entails, allowing progress to be measured and compared across different actors.

Experts disagree on the timeline for achieving general AI. Some predict that general AI systems may become mainstream within this decade, while others believe progress will be significantly slower.

Various entities are striving to develop general AI, each with its own objectives. The closer we get to general AI, the greater the societal impact of the developed technology. It is essential to establish clear global frameworks for AI development and use. The ultimate goal of developing general AI should be to create systems that serve as partners to humans—complementing and expanding human abilities in transformative ways. This requires a deep understanding of human behavior, culture, and social norms to ensure that general AI operates harmoniously within human communities. The journey toward general AI is both a technological and philosophical exploration, challenging us to contemplate the nature of human intelligence and its potential augmentation through AI.

Rapid Progress in Handling Large Data Volumes (Demis Hassabis, Google Blog, 2/2024)
The capabilities of AI models to manage large volumes of data are advancing rapidly. For example, OpenAI's GPT-3.5, released in November 2022, could process 4,096 tokens simultaneously. In February 2024, Google's Gemini 1.5 Pro AI model was released, capable of accurately handling up to one million tokens. Practically, this means a user could input, for example, a 402-page transcription of the Apollo 11 mission to the language model, and it would be able to analyze, categorize, and summarize the vast amount of information. The model could draw conclusions from discussions, events, and details across different parts of the document.

Significant Leap in AI Reasoning Capabilities (OpenAI, 9/2024)
OpenAI introduced the o1 model, designed to solve more complex problems with increased accuracy and efficiency. The model employs a new foundational model methodology that enables step-by-step problem-solving, significantly improving answer quality, particularly in science, mathematics, and coding. This development is noteworthy because the advancement of reasoning capabilities is considered a critical milestone on the path toward general AI.

Appendix 1: Application Ideas

Many tasks can be partially or fully transferred to AI solutions. Below is a list of example solutions that can leverage AI to improve the efficiency and quality of public sector operations. Additional potential use cases for public administration can be found in Futurice's catalog of over 60 use cases.

Imaging	AI applications are already used in examining X-rays. AI solutions can analyze image data significantly faster and more accurately than humans, identifying potential tumors or other anomalies quickly and precisely. For example, in ultrasound examinations, AI-assisted solutions can more efficiently identify fetal measurements, blood flows, and other characteristics, analyzing dozens of images per second. This frees up time for human interaction in patient situations.
Recording Patient Data	AI has significant potential in recording patient data, which currently consumes a substantial portion of physicians' working time. Voice-recognition AI can document information directly during a medical procedure, asking for confirmation and additional details at the end to ensure all necessary data is recorded.
AI Support for Physicians	AI can serve as a support tool for physicians, providing access to required information in seconds and assisting in making diagnoses far more effectively than current medical information services. It can compare new symptoms and patient reports with all legally collected information and correlate it with research data. In pilot projects conducted years ago, computers identified tumors that doctors had missed, enabling earlier detection of illnesses, leading to better treatment outcomes and lower costs for society.
Assessing Care Needs	AI solutions can support comprehensive pre-assessments based on individuals' data and additional information they provide. This enables directing people to appropriate services without requiring human intervention.
Anonymous and Capable Advisory Services	Anonymous advisory services can lower barriers to seeking help when people have sensitive issues to discuss. AI ensures privacy by providing services without human listeners, encouraging individuals to seek help when needed.
treamlining Licensing and Application Processes	Various licensing and application processes can be largely or fully automated with the use of tools. Intelligent forms ask for necessary information in the correct format, with much of the data retrievable from public sector databases. Interfaces can also be enhanced by enabling forms and applications to be completed through an AI interview, improving service levels and reducing queues in different agencies.
Ensuring the Legality of Procurement and Avoiding Overlaps	AI can assist in procurement by ensuring compliance with municipal procurement guidelines and existing legislation. For example, the Hilma service includes the AI-powered "Procurement Wizard" tool. Sharing data nationally could reveal collaboration opportunities between actors and enhance data interoperability.

Avoiding and Correcting Overlapping and Conflicting Regulations	AI applications trained on all public sector guidelines can check drafts for potential overlaps. The same concept can identify issues with existing regulations and suggest corrections.
Enhancing Report Preparation	AI solutions can draft preliminary data-based reports, which can be refined as needed by human effort. AI also allows the creation of continuously updated, text-based, and easily readable reports based on real-time data.
Automating Appointment Scheduling	AI can automate appointment scheduling via websites, chatbots, and AI-powered phone services. This saves staff time for higher-level tasks and genuine human assistance.
Simplifying Access to Decision-Making Data	AI systems can process decision-making data, making it easier to find and understand information for anyone interested in governance. The tool can also enable novel analysis of decision-making processes, such as identifying overlapping processes across cities, regions, and national levels, uncovering new collaboration opportunities.
Designing Transportation Networks	With sufficient foundational data and a functional platform, AI can compare scenarios to minimize temporary or permanent disruptions during project implementation. AI solutions can analyze numerous scenarios that traditional traffic planning methods cannot handle.
Coordinating Repair Projects	Public and private sectors often operate in the same areas, and aligning development projects can significantly reduce overall costs. AI tools can facilitate synchronization of projects among various stakeholders.
Simplifying Service Discovery	Compiling information about various services through AI systems makes it easier for users to find the right service. For example, AI could help a user navigate Finland's social support system and pre-fill application forms. Another example might involve consolidating necessary information for event planning into one channel, streamlining the permit process.
Multilingual Service Access	AI translators, which are constantly improving, can offer native-language information to even the smallest linguistic minorities. AI applications can translate any language into another in real-time, enhancing public service accessibility in various ways. Face-to-face interactions and online services can utilize AI to ensure smooth communication in the user's preferred language.
Simplifying Bureaucratic Language	Language models can clarify bureaucratic jargon into plain language, create summaries, and automatically check the readability of materials.
Public Services with Voice Interfaces	AI can train voice-controlled interfaces for any online system, enabling access to public digital services for those unable to use computers or smartphones, provided identification and reliability challenges are resolved.
Monitoring and Quality Assurance	AI systems can assist in quality assurance for accounting and other numerical data. They can identify anomalies in large datasets, making it easier to detect and correct errors.

Analyzing Research Data	AI systems can help summarize and interpret research data, regardless of its format. They can also be used in longitudinal studies to interpret previously collected data, especially older open-ended responses, more effectively than any previous technology.
Energy Consumption Management	AI can optimize energy distribution, reduce unnecessary consumption, analyze real-time data to identify spikes, and predict future needs for better energy management and cost savings. AI can also optimize smart grids and integrate renewable energy sources, supporting more sustainable energy use.
Improving Legislative Processes	AI can enhance legislative processes by providing faster and more accurate analysis of large volumes of legal and administrative data. For instance, AI can assist in drafting impact assessments for legislative proposals by analyzing similar past cases and identifying potential issues.
Accessibility and Different Population Groups	AI can improve the accessibility and user experience of digital services. For instance, chatbots can provide 24/7 support, AI can describe environments or read texts aloud for visually impaired users, and interfaces can be personalized to meet user needs, while automated translations improve linguistic accessibility.
AI as an Enabler of Human-Centeredness	AI enables tailoring services to individual needs. In healthcare, AI can provide predictive and personalized care by monitoring patients' health in real-time and forecasting future risks. Similarly, in education, AI can adapt learning materials and resources to the learner's needs by analyzing learning styles and tracking progress in real time.
Managing Large Data Sets	Public authorities' information obligations can be reduced by creating user-friendly information retrieval channels capable of processing extensive datasets. For example, AI solutions for understanding Finnish law will be available in 2025, with Edita launching the Finlex AI service and making Finlex data available for AI training. Similarly, all decision-making data at the national, municipal, and regional levels could be compiled into accessible chatbots, providing understandable answers in seconds.
AI for Expert Assistance	AI tools like ChatGPT, Microsoft Copilot, and others can significantly improve productivity in expert work, depending on the task. Routine tasks like information retrieval, presentation creation, and text production can be automated and accelerated, reducing workload and improving quality, allowing experts to focus on more creative and critical tasks.
Anticipation	AI can efficiently anticipate service needs in the public sector by analyzing large datasets on population behavior and existing services. The analysis can identify trends, forecast challenges, and optimize resource use. For example, AI can predict peaks in healthcare demand or the need for educational resources regionally.

Targeting	AI can improve the targeting of public services by analyzing data on individual behavior and past service usage. This enables tailored service proposals that meet individual needs. In social services, AI can identify individuals who would benefit from special support, or in healthcare, those at risk for certain diseases, allowing for proactive interventions. Developing such applications requires a focus on individual rights and data security.
Digital Twins for Anticipation	AI can simulate digital twins that model their real-world counterparts. By simulating potential events for the twin at an accelerated pace, service needs can be identified based on probabilities. For instance, a healthcare client's digital twin could simulate health developments to identify and prevent future issues.
Simplifying Transactions	AI agents managing different applications are becoming more common. For the public sector, this could mean seamless transactions via agents, creating a "one-stop-shop" experience.
Meeting Minutes and Task Assignments	AI can assist in creating high-quality minutes, identifying tasks, and directly assigning them. In the future, meeting devices could recognize participants and automatically record tasks into work management systems.
Real-Time Economy	AI and digitalization can automate and streamline tax payments, such as VAT, reducing the administrative burden on taxpayers and minimizing tax gaps (e.g., some countries have VAT gaps exceeding 10%, whereas Finland's is very small). An example of this development is the EU's VAT in the Digital Age directive proposal..