

Kirsi Hantula

EPISTEMIC RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE

A democratic society's antidote
to the concentration of epistemic power
in AI development

Sitra memorandum

© Sitra 2026

Epistemic rights in the age of artificial intelligence – A democratic society’s antidote to the concentration of epistemic power in AI development

Authors: Kirsi Hantula and a working group consisting of Jukka Vahti, Tiina Härkönen, Heikki Lauha, Rosa-Maria Mäkelä, Tarmo Toikkanen, Hannu-Pekka Ikäheimo, and Antti Lehtinen

Layout: Grano Oy

ISBN 978-952-347-470-3 (PDF) www.sitra.fi

ISSN 2737-1034 www.sitra.fi

Sitra’s memos are background materials produced to support our futures work.

Sisällys

Preface	4
Executive summary	6
Tiivistelmä	8
Sammanfattning	10
1. Epistemic inequality is a pressing democratic issue in digitalising societies	12
2. Epistemic power in generative AI development is increasingly concentrated in the hands of tech giants	16
3. The societal consequences of epistemic inequality in AI development	20
4. Epistemic rights as the foundation of more democratic AI development	24
Access to information	24
Competence and skills	26
Participation	29
Freedom from coercion or undue influence	30
5. Afterword: Three steps for applying epistemic rights in practice	32
References	34

Preface

The rise of social media was once compared to the invention of the printing press and the emergence of newspapers. Now, the same narrative is being repeated about rapidly advancing AI applications.

According to an old piece of wisdom, technological disruptions tend to be overestimated in the short term and underestimated in the long term. With social media, we already live in the medium-term phase, and some conclusions can be drawn. As with earlier communication technology transformations, social media has brought more people into the sphere of information dissemination and production. At the same time, it has created a host of new societal challenges.

Not least of these problems is the concentration of power in the hands of a few international giants. At the height of the social media hype, hundreds of millions of users around the world – myself included – carelessly and almost without thinking handed over their personal data to Facebook and other social media companies. This gave rise to an entirely new business logic: the data economy, in which individuals' personal data, the tracking of behaviour, along with data generated by communities, became commodities. Now, many of these same companies are central players in the AI transformation, as they possess vast amounts of data needed to train AI models and the resources required to develop them.

The AI transition is progressing while many of the problems highlighted by the rise of social media remain largely unresolved. However, mistakes need not be repeated. If the rise of social media and the data economy taught us anything, it is that technologies permeating society should fall much more firmly within the scope of democratic decision-making. In the context of AI, this means that, for example, the limits of acceptable AI use in public services should be subject to public discussion and consideration.

This memo continues Sitra's earlier work on digital power and the impact of digitalisation on societal power relations. Its purpose is to reflect on how different actors in society could use the perspective of knowledge rights in AI-driven Finland to build a more democratic, rightsrespecting and innovative information society, while also developing their own operations. At the same time, the memo provides background for Sitra's experiment which was launched in autumn 2025: the world's first citizencreated AI code of conduct.



We also want to remind actors across society that digitalisation – and AI development as part of it – is not a ‘law of nature’. We have many ways to influence the direction of development – if we choose to.

Hannu Ikäheimo

Director, Democracy Innovations Programme

Sitra

Executive summary

The rapid development of artificial intelligence (AI) in recent years has raised justified concerns about the concentration of power in international AI development in the hands of a small group of tech companies that possess the largest technological and financial resources. The fear is that these companies can dictate the direction of AI development without broad consultation with other stakeholders.

This memorandum addresses AI development from the perspective of epistemic, or informational, power. The concentration of epistemic power in AI development is as serious a problem for democratic societies as the concentration of economic and technological power. This is because technology giants make many decisions when developing AI, decisions that have far-reaching ecological, ethical, social, or societal impacts.

To ensure that these decisions are made more transparently in the future and considering the overall benefit of society, democratic societies should dismantle the epistemic inequality that has entrenched itself in AI development and strengthen citizens' AI agency. This memorandum presents a framework for epistemic rights, which provides democratic societies and various organisations with ways to begin this work. The original framework was developed by Professor Emeritus of Communication Hannu Nieminen, but in this memorandum, the framework has been adapted to better fit the context of AI development.

The framework for epistemic rights helps to concretely illustrate how the epistemic power accumulated by large tech companies weakens the AI understanding of other societal actors and narrows citizens' opportunities to influence the decisions made in connection with AI development. At the same time, the examination of epistemic rights helps various societal actors – such as public administration, educational institutions, media, or companies adopting AI – to see how they can strengthen and protect epistemic rights in the emerging AI era through their own actions.

Democratic societies should dismantle the epistemic inequality that has entrenched itself in AI development and strengthen citizens' AI agency.



The framework for epistemic rights consists of four rights:

- 1. Access to information:** Everyone should have equal access to relevant and truthful information about AI development and use.
- 2. Competence and skills:** Everyone should be guaranteed equal opportunities to understand the possibilities and challenges involved in AI development. People must be able to protect personal or societal interests based on this understanding.
- 3. Participation:** Everyone should be given equal opportunities to influence the public discussion and decision-making regarding the value base, direction, and impacts of AI development.
- 4. Freedom from coercion and undue influence:** Everyone should be guaranteed equal opportunities to form opinions about emerging AI, free from coercion or manipulation by those who hold the greatest epistemic power.

Tiivistelmä

Viime vuosien nopea tekoälykehitys on herättänyt perusteltua huolta, että valta kansainvälisessä tekoälykehityksessä uhkaa keskittyä pienelle joukolle teknologiayrityksiä, joilla on suurimmat teknologiset ja taloudelliset resurssit. Pelkona on, että nämä yritykset pystyvät määräämään tekoälyn kehityksen suunnan, muilta liiemmästi kyselemättä.

Tässä muistiossa tekoälykehitystä käsitellään tiedollisen vallan näkökulmasta. Tiedollisen vallan keskittyminen tekoälykehityksessä on demokraattisille yhteiskunnille yhtä vakava ongelma kuin taloudellisen ja teknologisen vallan keskittyminenkin. Tämä johtuu siitä, että teknologiajätit tekevät tekoälyä kehittäessään monia päätöksiä, joilla on kauaskantoisia ekologisia, eettisiä, sosiaalisia tai yhteiskunnallisia vaikutuksia.

Jotta näitä päätöksiä tehtäisiin jatkossa läpinäkyvämmiin ja yhteiskunnan kokonaisetua puntaroiden, demokraattisten yhteiskuntien tulisi purkaa tekoälyn kehittämiseen pesiytynyttä tiedollista epätasa-arvoa ja vahvistaa kansalaisten tekoälytoimijuutta. Tässä muistiossa esitellään tiedollisten oikeuksien kehikko, joka tarjoaa demokraattisille yhteiskunnille ja erilaisille organisaatioille keinoja päästä työssä alkuun. Alkuperäisen kehikon on laatinut viestinnän emeritusprofessori Hannu Nieminen, mutta kehikkoa on tässä muistiossa muokattu tekoälykehityksen kontekstiin sopivammaksi.

Tiedollisten oikeuksien kehikko auttaa hahmottamaan konkreettisesti, millä tavoin suurille teknologiayrityksille kasautunut tiedollinen valta heikentää yhteiskunnan muiden toimijoiden tekoäly-ymmärrystä ja kaventaa kansalaisten mahdollisuuksia vaikuttaa tekoälyn kehittämisen yhteydessä tehtäviin päätöksiin. Samalla tiedollisten oikeuksien tarkastelu auttaa erilaisia yhteiskunnallisia toimijoita – esimerkiksi julkista hallintoa, oppilaitoksia, mediaa tai tekoälyä käyttöön ottavia yrityksiä – näkemään, miten ne voivat omilla toimillaan vahvistaa ja suojella tiedollisia oikeuksia kehittyvän tekoälyn aikakaudella.



Tiedollisten oikeuksien kehikko rakentuu neljästä oikeudesta:

- 1. Pääsy tietoon:** Kaikilla ihmisillä tulee olla yhdenvertainen pääsy tekoälyn kehitystä ja käyttöä koskevaan olennaiseen ja totuudenmukaiseen tietoon.
- 2. Osaaminen ja valmiudet:** Kaikille ihmisille tulee taata yhdenvertaiset valmiudet ymmärtää tekoälyn kehitykseen sisältyviä mahdollisuuksia ja haasteita. Ihmisten on kyettävä varjelemaan henkilökohtaista tai yhteiskunnan etua tämän ymmärryksen pohjalta.
- 3. Osallistuminen:** Kaikille ihmisille tulee suoda yhdenvertainen mahdollisuus vaikuttaa tekoälykehityksen arvopohjaa, suuntaa ja vaikutuksia koskevaan julkiseen keskusteluun ja päätöksentekoon.
- 4. Ei pakkoa tai painostusta:** Kaikille ihmisille tulee turvata yhdenvertaiset mahdollisuudet muodostaa näkemyksiä kehittyvästä tekoälystä, ilman suurinta tiedollista valtaa käyttävien pakottamista tai manipulointia.

Sammanfattning

Den snabba utvecklingen av artificiell intelligens (AI) under de senaste åren har väckt berättigad oro över maktkoncentrationen inom internationell utveckling av AI till en liten grupp företag som innehar de största teknologiska och ekonomiska resurserna. Rädslan är att dessa företag kan bestämma riktningen för utvecklingen av AI utan att rådfråga andra i större utsträckning.

I detta memorandum behandlas AI-utveckling ur perspektivet av kunskapsrelaterade maktförhållanden. Koncentrationen av kunskapsrelaterad makt inom utvecklingen av AI är ett lika allvarligt problem för demokratiska samhällen som koncentrationen av ekonomisk och teknologisk makt. Detta beror på att teknologijättarna fattar många beslut när de utvecklar AI, beslut som har långtgående ekologiska, etiska, sociala eller samhälleliga konsekvenser.

För att dessa beslut i framtiden ska fattas mer transparent och med hänsyn till samhällets helhetsintresse, bör demokratiska samhällen bryta den kunskapsmässiga ojämlikhet som har rotat sig i AI-utvecklingen och stärka medborgarnas AI-aktörskap. Till stöd för detta presenteras i memorandumet en ram för rättigheter relaterade till kunskap, som erbjuder demokratiska samhällen och olika organisationer verktyg för att komma i gång med arbetet. Den ursprungliga ramen har utarbetats av emeritusprofessor i kommunikation Hannu Nieminen, men i detta memorandum har ramen anpassats för att bättre passa AI-utvecklingens kontext.

Ramen för kunskapsrelaterade rättigheter hjälper till att konkretisera på vilket sätt den makt som har ackumulerats hos stora teknologiföretag försvagar andra aktörers förståelse av AI och begränsar medborgarnas möjligheter att påverka de beslut som fattas i samband med utvecklingen av AI. Samtidigt hjälper utvärderingen av kunskapsrättigheter bland olika samhällsaktörer – till exempel offentlig förvaltning, utbildningsinstitutioner, media eller företag som använder AI – att se hur de genom sina egna åtgärder kan stärka och skydda kunskapsrättigheter i en tid av artificiell intelligens.



Ramen för kunskapsrättigheter består av fyra rättigheter:

- 1. Tillgång till information:** Alla människor ska ha lika tillgång till relevant och sanningsenlig information om utveckling och användning av AI.
- 2. Kompetens och färdigheter:** Alla människor ska garanteras lika möjligheter att förstå de möjligheter och utmaningar som utvecklingen av AI innebär. Människor måste kunna skydda personliga eller samhällsintressen baserat på denna förståelse.
- 3. Deltagande:** Alla människor ska ges lika möjlighet att påverka den offentliga diskussionen och beslutsfattandet om AI-utvecklingens värdegrund, riktning och konsekvenser.
- 4. Inget tvång eller påtryckning:** Alla människor ska garanteras lika möjligheter att bilda sig åsikter om AI, fria från tvång eller manipulation från de som innehar den största makten angående kunskap.

1. Epistemic inequality is a pressing democratic issue in digitalising societies

Public discussions about the relationship between digital technologies and democracy typically highlight how digital technologies have eroded the foundations of democracy. Attention is drawn, for example, to recommendation algorithms that have reshaped media consumption. It is equally common to discuss how the rise of social media has contributed to the difficulties faced by traditional news media. Following recent leaps in AI, public debate has also raised concerns about AI-generated disinformation and the use of AI for election manipulation.

This memo does not treat democracy and digital technological development as being inherently opposed. Instead, it focuses on a challenge common to both: *the production of epistemic inequality*.

Both democratic decisionmaking and digital technology development require broad and deep understanding of complex issues. In practice, this knowledge often accumulates among a relatively small group – referred to here as the *epistemic elite*. This group possesses the information and expertise that shape situational awareness, guide strategic choices, and inform decisions. Most citizens do not share this depth of knowledge. As a result, epistemic gaps emerge between the elite and everyone else. Over time, this can erode trust in decisionmaking and undermine the legitimacy of decisions.

Epistemic inequality in democratic decisionmaking has been discussed previously. In Finland, for example, concerns have been raised about whether civil servants or experts – who hold significant information relevant to policymaking – exert excessive influence over political decisions. There has been discussion of increasing expert power and of how the language used by experts, as well as the ways of producing and presenting knowledge that they favour, prevent others from participating in the collection and interpretation of knowledge (Ylöstalo 2020).

By contrast, little attention has been paid to epistemic inequality. At first glance, the claim that epistemic gaps constitute a problem in the development of digital technologies may even sound strange. Why, after all, should we be concerned about epistemic inequality at a time when information, in its various forms, is more abundant than ever?

The answer is simple: the shift toward a modern, networked information society has not eliminated epistemic inequality. Rather, developments in digital technologies appear to have reinforced it.

The rapid development of information and communication technology has indeed brought ever larger segments of the population within reach of modern information dissemination and advanced tools for data processing and production. Yet, it has also destabilised the information environment and empowered a handful of tech giants capable of defining the pace of technological development.

The biggest technology companies can collect vast amounts of digital data and turn it into commercially valuable knowledge. Their research and development capacities dwarf those available elsewhere. As a result, economic, technological and epistemic power becomes concentrated, enabling them to shape public debate and its tone.

For digitalising societies, this constitutes a major democratic challenge. A cornerstone of Finnish and European democracy is the legal right of citizens to influence decisions that affect them. This should apply equally to choices made about digital technologies, particularly when they shape daily life or societal structures.

Epistemic inequality embedded in digital technology development threatens this right. When citizens, researchers, policymakers, civil society and digital service users lack adequate access to meaningful information or lack the ability to interpret it, deliberation about important decisions taken in the context of technological development is weakened.

This also increases the risk that technology companies make shortsighted decisions that fail to account for longterm ecological, social, or societal consequences.

In the following chapters of this memorandum, the topic is examined in greater depth. The aim is to demonstrate the kinds of mechanisms that generate and perpetuate epistemic inequality in the development of digital technologies. At the same time, ways of dismantling epistemic gaps are explored.

Figure 1. Epistemic rights as a foundation for strengthening people's epistemic agency**ACCESS TO INFORMATION**

Everyone should have equal access to relevant and truthful information about AI development and use.

COMPETENCE AND SKILLS

Everyone should be guaranteed equal opportunities to understand the possibilities and challenges involved in AI development. People must be able to protect personal or societal interests based on this understanding.

PARTICIPATION

Everyone should be given equal opportunities to influence the public discussion and decision-making regarding the value base, direction, and impacts of AI development.

FREEDOM FROM COERCION AND UNDUE INFLUENCE

Everyone should be guaranteed equal opportunities to form opinions about emerging AI, free from coercion or manipulation by those who hold the greatest epistemic power.

Within the framework of epistemic rights, the key measures for strengthening citizens' epistemic agency have been condensed into four epistemic rights. Source: Nieminen 2024.

The starting point for analysis is Professor Emeritus Hannu Nieminen's (2024) epistemic rights framework. The framework argues that in modern democracies, information and understanding should be treated as fundamental civic rights. Without them, equal participation in democratic decisionmaking is not possible.

To strengthen the foundation of democracy, institutions such as the media, public administration and educational institutions must enhance people's abilities to evaluate and process information and expand opportunities for public deliberation about important societal matters.

This memo applies the epistemic rights framework to challenges in developing and deploying new digital technologies – particularly generative AI. The aim is to help organisations producing or using information to strengthen individuals' epistemic rights during rapid digitalisation.

The framework can also support companies developing new technologies by helping them assess the 'democratic resilience' of their products and ideas. For decision-makers or civil servants, the framework of epistemic rights can, in turn, serve as a tool for more strategic and humancentred decision-making, for example in situations where digital services aimed at citizens are introduced or developed.

INFO BOX 1

CORE CONCEPTS IN GENERATIVE AI DEVELOPMENT

Generative artificial intelligence refers to advanced language models that produce new content in the form of written text, audio, images, or videos.

Advanced language models are also often referred to as 'large' or 'broad' language models. These models have been trained on vast datasets, in practice on data mined from the internet. Their training has made use of multilayered neural networks. As a result of the training methods employed, advanced language models can perform complex tasks that require sophisticated analysis and interpretation of speech, text, or visual material.

A chatbot is a computer program designed to communicate with humans either through speech or via written text. With the development of generative artificial intelligence, chatbots have evolved rapidly, as new large language models have begun to be used in their construction. Thanks to advanced language models, chatbots are now able to communicate with their users far more accurately than before, with linguistic fluency and in a manner that convincingly imitates human-to-human communication.

Sources: Sitra's Dictionary, undated; Tietotekniikan termi talkoot, 2025; Tietotekniikan termitalkoot, 2018.

2. Epistemic power in generative AI development is increasingly concentrated in the hands of tech giants

At the end of 2024, the Organisation for Economic Cooperation and Development (OECD) published an overview examining the risks and benefits of AI development (OECD 2024). One significant risk they identified was the concentration of power in AI development in the hands of a small number of influential technology companies. The OECD is not alone in this assessment. The same issue has recently been highlighted, for example, in the World Economic Forum's annual global risk assessment (Torkington 2024).

In these and other similar analyses, the concentration of power in AI development is typically examined from an economic and technological perspective. They emphasise that the capital and resourceintensive nature of developing stateoftheart AI models threatens to concentrate power in giant technology companies with the greatest financial and technological muscle. What receives less attention, however, is that the same factors may also lead to a concentration of epistemic power. This is because, in AI development, epistemic power derives not only from deep expertise and AI competence, but also from control over the key technological and material resources required for developing AI models, as well as from the strategic use of strong market positions and deep customer insight.

The world's largest and most influential technology companies have a clear advantage in all these areas. Put simply, these companies have the largest financial resources, which enable them to secure access to the crucial resources needed to build AI models, such as advanced chip technology and computing power, data, and toplevel expertise.

In the development of generative artificial intelligence, this dynamic is most clearly visible, for example, in the development of the most advanced large language models that underlie many conversational chatbot applications such as ChatGPT. The development of these language models is led by a few growth companies that triggered the generative AI boom, together with a handful of large technology giants. In this memorandum, particular attention is nevertheless paid to the latter, as the technology

giants have demonstrated the strongest capacity to use economic and technological power to reinforce their own epistemic power.

The largest technology companies have managed to remain at the forefront of language model development because their vast technological and financial resources allow them to adapt most rapidly to the changing demands and conditions of the operating environment. This ability to respond effectively is their key competitive advantage, enabling them to consolidate their epistemic power.

At the beginning of the current AI boom, in the early years of the 2020s, this capability helped the largest technology companies to successfully navigate the intense competition among actors developing language models for the critical resources required to build them. At this stage, investment in generative artificial intelligence surged dramatically worldwide, and numerous new AI laboratories emerged, some of which have since closed or been absorbed by larger competitors. At the same time, the development of increasingly advanced language models required ever greater amounts of data and computing power. As a result, demand grew sharply for all the key resources needed to build the most advanced language models – advanced chip technology, cloud computing infrastructure, and training data mined from the internet – while the supply of these resources was insufficient.

The large technology companies, however, were able, thanks to their formidable financial and technological muscle, to secure access to critical resources and to continue developing their most advanced language models. As a consequence, their epistemic power in the development of generative artificial intelligence was strengthened.

A clear example of how technology giants have been able to reinforce their epistemic power by relying on their own financial and technological resources is cloud computing. Alphabet (Google's parent company), Microsoft, and Amazon together control roughly two thirds of the global cloud services market. Owing to their market position, they did not have to worry about the development of their own language models slowing down due to shortages in computing capacity, even during the most intense phase of competition over resources.

At the same time, market leadership in cloud services has enabled Alphabet, Microsoft, and Amazon to use their computing capacity as a means of creating dependency relationships favourable to them in the development of generative artificial intelligence. This was most strikingly evident in the way these giants used computing power as a bargaining chip when entering into multibillioneuro partnership agreements with smaller startup companies developing advanced language models. In return for their investments, they provided startups with muchneeded computing capacity, but demanded agreements in exchange, guaranteeing them access

to many of the smaller companies' technological and research innovations (Hallamaa 2024).

The strengths of tech giants in the next phase of generative AI development

Since then, the shortage of resources needed to develop advanced language models has clearly eased. Factors contributing to this include innovations related to the development of the most advanced language models, which have improved the efficiency of the use of the data, chips and computing power required in language model development (see e.g. Ng et al. 2025). It is therefore likely that the development of the most advanced language models is gradually moving into a new phase. In this phase, the leading companies will be those capable of creating, on top of advanced language models, consumer services that function as increasingly adaptable, generalpurpose tools for a variety of use cases, or that integrate generative AI broadly across different services. (Ruokonen 2025)

Achieving such market leadership requires partly new kinds of expertise and different capabilities than in the early phase of the generative AI boom. It requires the ability to combine cuttingedge research with systematic and continuous product development, with the goal of building services that function intuitively, are technically reliable, and successfully meet growing consumer expectations. It requires the capability to integrate advanced generative AI smoothly into different workflows and service processes.

In this phase of the generative AI race, the epistemic power of the largest technology companies appears to be strengthened by at least two significant factors. First, large technology companies with the greatest financial and technological resources can strengthen their position at the forefront of generative AI development by developing their own AI competence over the long term – comprehensively and without sparing expense. This is evidenced by their published investment plans for the coming years, according to which the companies intend to spend hundreds of billions of euros on the development of the most advanced language models and the infrastructure supporting them (Bloomberg Intelligence 2024). This is also reflected in the sheer number of AI developers employed by large technology companies. By one estimate, Amazon, Alphabet and Meta (which owns Facebook among others) together employ over 33,000 AI experts (Olson 2023).

The same tech giants are also able to increase their understanding of AI by engaging in active research cooperation with universities. All these companies engage in various forms of cooperation with academic research

groups studying AI, and this development has not been viewed entirely positively in the research community. Fears have been voiced that such collaborations concentrate toplevel AI research among an even smaller number of large technology companies and elite universities (Khanal, Zhang and Taeihagh 2024). At the same time, concerns have been raised about whether the influence of tech giants in the scientific community is becoming excessive (Färber and Tampakis 2023; Abdalla and Abdalla 2020; Ahmed, Wahed and Thompson 2023).

Alongside broad AI competence, the second cornerstone of the epistemic power of the largest technology companies in the new phase of the generative AI race appears to be the large portfolios of digital services that the companies have been building long before the breakthrough of generative AI. Many of the digital services owned by tech giants – such as search engines, online stores and social media services – already have extremely large numbers of users. When companies bring generative AI into their services, they continuously accumulate large amounts of realtime, valuable data about the use of generative AI.

The data helps tech giants to identify users' expectations of generative AI and new needs, as well as to spot opportunities to expand their generative AI offering into new, complementary markets, for example healthcare, education and defence (Landi 2024; Stone 2025; Requiros 2025). In this way, large technology companies are able to exploit their global ecosystem of digital services as an important engine for the development and growth of generative AI.

3. The societal consequences of epistemic inequality in AI development

The concentration of epistemic power in the hands of tech giants constitutes a new thorny problem for democratic societies, which is underlined by the rapid development and spread of generative AI.

As generative AI becomes interwoven with society's functions and services, the decisions made in its development will have a variety of social, ecological or societal consequences. For this reason, it is in the public interest that these decisions be made by a sufficiently diverse group, weighing different viewpoints broadly and transparently.

The problem can be illustrated with conversational chatbots built on advanced language models.

The use of chatbots in various customer service contexts is now spreading rapidly. From the perspective of European societies, the change is not only about whether customer service can be made more efficient by using chat technology. It is also about the fact that conversational chat interfaces will, in the near future, become an important gateway to the internet and to information, while at the same time becoming interwoven into various service pathways.

Is the rapid proliferation of these interfaces an entirely desirable development, or should we consider the possibility that conversational interfaces could be used for deliberate deception, the concealment of accurate information, or the theft of personal data (cf. e.g. Lu 2025; Montreal AI Ethics Institute 2025)? Or, as a society, do we lose something else of importance if the different viewpoints and values associated with the use of chat technology are not carefully weighed together?

In recent times, for example, Google has already integrated a chatbottype interface into its search engine. Meta has brought new generative AI features to Facebook, Instagram and WhatsApp, and Apple is connecting generative AI to its Siri assistant. Likewise, the use of conversational AI in various software used in knowledge-based work has rapidly become widespread. Next, a new generation of AI assistants is emerging capable of independently handling a wide range of tasks on behalf of users in digital environments.

An evergrowing body of research, however, shows that the current types of generative AI language models remain in many ways prone to producing erroneous or skewed information. Studies have already shown

that shortcomings in language model training methods can leave value biases in the models, which distort the answers provided by the AI services built on them (Buschek and Thorp, undated). It has also been observed that language models sometimes hallucinate, that is, produce entirely fabricated answers. Moreover, their operational logic itself could lead to language models subtly misleading their users (Wong 2024).

Chatbot services that provide a virtual friend, a partner, or a sexual companion have also entered the market. In extreme cases, the use of these services has been observed to lead to emotional dependency, and in the worst cases, with tragic consequences (El Atillah 2023; Montgomery 2024). Their data protection is also often extremely weak, and their business models predatory (Caltrider, Rykov and MacDonald 2024; Kim 2024).

The EU's AI Act imposes some new obligations on providers of chatbot services, but legislation cannot address all societally significant concerns. Societies must find alternative ways to respond to the changes that, for example, the sharp rise in the popularity of chatbot 'friends' and 'partners' may bring to social relationships (see e.g. Muldoon 2024; Hill 2025).

Questions related to chat technology, however, are only one example among many of the complex social, ethical and ecological issues associated with the development and deployment of generative AI, issues about which broad public discussion or joint deliberation has not yet taken place, even though the development of generative AI is now proceeding in great leaps.

In order for generative AI to be developed in the coming years in a more balanced way and with broader considerations, democratic societies should now begin, in a determined manner, to create new ways of working and practices that address the epistemic inequality that has taken root in the development of generative AI.

A key guiding principle in developing these practices should be that the perspectives or values of the large technology companies with the greatest epistemic power in generative AI development do not become overly dominant in the public debate, nor should decisions about the direction of this technology's development be left largely in their hands.

Now is the right time to develop new operating models, when many of the norms, values, standards, and operating models related to the development of language models – and even parts of the international legislation concerning their development and use – are only just taking shape.

INFO BOX 2

THE CONCENTRATION OF EPISTEMIC POWER PRODUCES NEW SOCIETAL, ETHICAL AND ECOLOGICAL CHALLENGES

When technology companies with the greatest epistemic power develop generative AI, they can make fairly independent decisions that have far-reaching effects on people's daily lives or on society.

Copyrights. Building advanced language models requires vast amounts of training data, which is mined from the internet. This material has been shown to contain, among other things, copyrighted content and sometimes even sensitive personal information, such as personal health data (King and Meinhardt 2024; Leffer 2023). The EU's AI Act requires developers of advanced language models to disclose more transparently than before what material is contained in the training data used by their language models. With regard to the use of copyrighted material, however, international legal interpretations of the extent to which languagemodel developers have the right to use such material are still taking shape. Several significant lawsuits concerning copyright are pending against companies developing the most advanced language models (see e.g. Allyn 2025).

Biases and discriminatory attitudes contained in training data. Training data used to build advanced language models, originating from the internet, reflects biased and discriminatory attitudes embedded in societies. The languagemodel training methods currently in use cannot completely prevent these attitudes from flowing into the language models and subsequently into the services built on them (see e.g. Hoffman et al. 2024). This should be taken into account both when assessing the reliability of the content produced by these language models and when considering the deployment of services that utilise advanced language models in new contexts.

Challenges in the use of chatbots. A great deal of hope is being placed on the development of chat technology that utilises generative AI, and the use of chatbots in a variety of contexts is rapidly becoming more common. Chatbased services are already being actively developed and piloted, for example, in the mental health and education sectors (Taleva 2025; Elliott 2024; Chadwick 2025).

Because the field is still young and developing very rapidly, more research and understanding would be needed, for example, about how users' relationships to chat interfaces change over long periods of time. Research to date has already found indications that longerterm use of chatbots as learning support may not improve schoolchildren's learning outcomes, even though it often does so in the short term (see e.g. Araujo and Bol 2024; Wu and Yu 2023). Recent studies have also shown that current chatbots still have serious limitations in therapeutic use (see e.g. Moore et al. 2025).



The significant energy consumption of AI and other environmental impacts. Building the most advanced language models and providing services built on them consume vast amounts of computing power and electricity. Cooling data centres also requires large amounts of water. Awareness of the growing ecological footprint of generative AI has been dawning gradually, but there should still be more active public discussion about curbing the sector's carbon dioxide emissions and environmental burden, and ways should be sought to reverse the negative trend. At the same time, closer examination should be made of whether, for example, the data centre construction boom created by global AI development might bring with it risks related to regional energy availability.

Is it possible that the placement of a new data centre in an area leads to local energy availability problems during major peak consumption periods? If that were to happen, it could, at worst, prevent other energy-intensive investment projects from coming to the area – which might have a greater impact, for example, on local employment or the economy (Hanhivaara 2025; Velkova 2024).

There should also be a discussion about the types of applications for which generative AI is genuinely worth employing. For instance, is it sensible to use AI that consumes natural resources merely for marketing new products?

Criminal and malicious use of advanced language models. The most advanced language models and the services built on them can also be used for malicious or criminal purposes, aiming to undermine the stability, democracy, security of societies or individuals' fundamental rights (see e.g. NCSC 2024).

Such use cannot be completely prevented, but it can be made more difficult if the companies building advanced language models systematically trace and prevent malicious use of their models and actively communicate their observations to the various actors in the internet ecosystem (see e.g. Open AI 2025). Democratic societies, for their part, can establish new rules and operating procedures to counter threats arising from the illegal use of the most advanced language models, and can especially seek to protect their most vulnerable population groups from AI-assisted abuses.

4. Epistemic rights as the foundation of more democratic AI development

When new operating models begin to be developed, the framework of epistemic rights presented at the beginning of this memo could provide a foundation for this work. The framework could form a shared valuebased starting point for different societal actors, based on a common understanding that in rapidly digitalising societies, people's agency and opportunities for influence must be protected more vigorously than before.

For this reason, this chapter presents an expanded version of the epistemic rights framework, adapted to the context of developing advanced language models and generative AI.

In preparing the framework, an effort has also been made to consider how different types of societal actors can improve epistemic equality through very different kinds of actions.

For example, public administration, the media, educational institutions and companies developing generative AI have different roles and tasks in society, and the resources available to them vary significantly.

Still, all of them can have an important role in defending and strengthening epistemic rights.

The updated framework gives different actors the opportunity to focus their attention precisely on those epistemic rights whose protection aligns most naturally with their own daytoday activities.

Access to information

Everyone should have equal access to relevant and truthful information about AI development and use.

Influential technology companies that develop the largest language models and generative AI have been criticised for sharing increasingly scarce information with outsiders about many of the important decisions they make in the development of their language models (Nolan 2023).

Although reporting on the functioning and development of language models has increased somewhat in recent years, certain aspects of their

development remain underreported and opacity persists (Bommasani et al. 2024a).

The EU's AI Act, which entered into force in 2024, improves access to information. However, the Act has been criticised for requiring less public reporting from companies that develop the most advanced language models than, for example, the European Parliament had hoped for earlier in the legislative process. The AI Act obliges companies that build advanced language models to provide information to the EU and the member states, as well as to organisations that build services on the language models developed by these companies. For users of services built on language models – citizens – the Act only tries to safeguard access to information by requiring developers of advanced language models to publish 'a sufficiently detailed summary' of the training data collected to train the models (Bommasani et al. 2024b).

In addition to transparency reporting, companies developing language models could provide more information about their models and the services built on them in other contexts as well, taking into account different information needs and usage situations. Ordinary service users may also have very different reasons, in different situations, for wanting to know about the functioning of a language model. They may want to better understand how a model uses and combines data, or they may want to learn more about how language models work or evaluate the accuracy of the content produced by a service built on a language model (cf. e.g. Hantula and Korkman 2022).

At a more general level, the key is often ensuring that enough information is produced and presented in the right form so that different groups can assess how much a language model can be trusted across various contexts and purposes (Bansal et al. 2022). Achieving this goal requires companies developing advanced language models to improve information access and adopt more diverse and humancentric ways of sharing information and creating understanding of how language models function (Liao and Wortman Vaughan 2024).

In addition to the companies developing the largest language models, organisations that adopt AI services built on language models should also create new practices to ensure that service users and other external parties receive sufficient information about the use of AI. Public organisations could set an example by committing to publishing information about how and based on what principles they use AI. Such transparency-promoting activity would naturally build on the new obligations that the EU's AI Act imposes on public administration. According to the Act, public sector actors must conduct a fundamental rights impact assessment for all their AI procurements.

Good practices for disseminating information on publicsector AI systems already exist – for example, the public AI register maintained by the City of Helsinki, which contains information on all AI systems used by the City. In producing information about AI systems, special attention should be paid to ensuring that the information provided is understandable, accessible and widely available.

In addition to improving citizens' access to information, it would be important to create practices that better safeguard the access of civil society organisations and independent researchers to essential information on languagemodel development and the use of generative AI. In the research community investigating the safety and reliability of language models, concern has recently grown about the future of independent research. This is because companies developing advanced language models have not guaranteed independent researchers the same conditions as, for example, have been provided in traditional software development (A Safe Harbour for Independent AI Evaluation 2024; Tiku 2024).

Competence and skills

Everyone should be guaranteed equal opportunities to understand the possibilities and challenges involved in AI development. People must be able to protect personal or societal interests based on this understanding.

The increasing prevalence of generative AI creates needs for renewal in education and training. Studies show that students and teachers hold several misconceptions about AI (e.g. Mertala and Fagerlund 2024). Knowledge related to AI as a phenomenon, as well as AI pedagogy, remains unfamiliar to most professionals in the education sector. According to research, three out of four early childhood and basic education teachers have not used AI applications at all (TanhuaPiiroinen et al. 2025).

A key objective of Finland's national digitalisation strategy, the Digital Compass, is for Finland to become a digitally literate and capable country by 2030. Finland aims to remain a country where citizens have the knowledge, skills and capacities to participate actively and express themselves safely and creatively in a digital society. For this reason, safeguarding the digital literacy of Finns also requires a strong contribution from the school system. AI literacy and digital information literacy are new civic skills of the digital age. Digital information literacy is the ability to find, interpret, analyse, manage, understand, create and disseminate information safely and appropriately in different digital environments (Faktabaari 2022). In the era of generative AI, new kinds of digital

environments include chat interfaces and digital assistants built on advanced language models.

AI education in schools has so far been strongly rooted in the tradition of media and technology education. The aim has been to understand the operational logic of AI, its application possibilities, and the individual's relationship to technology. In contrast, the societal impacts of AI, and its relationship to democracy and citizens' agency, have been examined only superficially.

In the future, democracy and humanrights education in schools could provide an effective means of highlighting the relationship between AI and democracy, while strengthening citizens' epistemic agency and capacity to act. This would, however, require a new pedagogical approach and necessitate the renewal of both media and technology education, and democracy and humanrights education.

Research shows that citizens' societal participation already diverges and becomes stratified from basic education onwards (e.g. KestiläKekkonen et al. 2022). Children and young people could be given more equal opportunities to understand technological development from perspectives beyond the purely technical by introducing content dealing with the societal impacts of AI, for example as part of civics or religious and ethics curricula (Juurola and Korhonen 2025). This would simultaneously improve citizens' abilities to participate in discussion and decisionmaking related to AI.

Examples for the renewal of technology education and democracy and humanrights education could be drawn from Scotland. In 2022, the country launched an AI education programme for primaryschool pupils as a continuation of an earlier AI strategy emphasising humancentric AI (Children's Parliament 2024).

As in Scotland, the renewal of Finland's media and technology education and democracy and humanrights education should be based on a coherent national vision. In March 2025, the Finnish National Agency for Education and the Ministry of Education and Culture published materials on the use of AI, aimed at supporting early childhood education and school providers in developing AIrelated competence and promoting responsible, safe and innovative use of AI (Finnish National Agency for Education 2025). The recommendations included in the materials address AI literacy, sustainable development and ethical principles. However, the link between AI and individual agency and the functioning of a democratic society receives very little attention. In the future, such AI guidelines could encourage actors in the education sector to engage in a more active societal dialogue about the direction of AI development and thus promote civicoriented AI education.

In addition to the education sector, the media also has its own role in developing people's understanding of AI and their agency. The media should strive to ensure that citizens maintain sufficient understanding of both the positive and negative societal impacts of generative AI development, as well as of alternative developmental trajectories. For example, in 2023, Associated Press (AP), one of the largest news agencies in the United States, added a section to its style guide defining how AI is to be handled journalistically (Meir 2023). One aim of AP's new guidelines was to ensure that AI technologies are not given forms of agency in reporting that actually belong to the companies and individuals developing them.

Alongside the media, other organisations communicating about AI can also create internal guidelines on the concepts and terms they use, so as to avoid anthropomorphising AI or reinforcing misconceptions about its capabilities (see e.g. Robbins and Bryson 2023; Kambhampati et al. 2025).

In addition to educational institutions and media actors, companies and other organisations that adopt AI-based systems can also support people's agency and understanding of AI. The EU's AI Act already guides them in this direction. The Act requires organisations using AI to ensure that their staff possesses sufficient AI literacy (Toikkanen and Toivanen 2025).

Strengthening broad and critical AI literacy will be particularly important in the future in fields where AI systems become tightly interwoven with information production or decisionmaking (see e.g. Nelimarkka, Mervaala and Laaksonen 2025). It is also essential that organisations using AI employ a sufficient number of people, in different roles, who have the ability to examine the impacts of AI systems from various societal perspectives and who have the opportunity to influence their organisation's decisions about AI. For this reason, organisations adopting AI should develop internal processes and working practices that ensure that they assess their significant AI policies comprehensively, including any possible social, ecological or societal impacts.

One helpful tool for considering different perspectives might be the international DEDA assessment tool, which allows people representing different functions and skills within an organisation to jointly reflect on the ethical tensions and choices in data projects. A Finnish version of the tool has recently been published (Repair Working Group 2023).

Developing employees' AI competencies, however, does not remove organisations' responsibility to develop ethically sustainable AI solutions that take into account all kinds of groups. The responsibility for developing such AI systems and for their safe deployment always lies with the system's developer and the organisation that implements it. It must not depend on the vigilance of individual employees.

Participation

Everyone should be given equal opportunities to influence the public discussion and decision-making regarding the value base, direction, and impacts of AI development.

The technology sector generally suffers from too little diversity in its workforce, and the development of the largest language models is no exception (Ville 2023). For example, in the important reinforcement learning phase of training large language models, the enormous model is often trained by only a very small and demographically homogeneous group (Buschek and Thorp, undated). It has been shown that the attitudes of this group can influence the training of language models, and as a result, models may develop, for example, religious or political biases (Perez et al. 2023).

In addition to the reinforcement learning phase, there are other stages in the development of advanced language models in which the choices made by those developing the models have a significant impact on the final outcome. These include, for example, selecting training data, cleaning data, and the redteaming phase prior to the publication of the model, during which efforts are made to identify potential problems in its behaviour.

At the same time, companies developing advanced language models suffer from a more general challenge associated with AI development: the problem of biased data. Training data for language models originating from the internet is always biased to some extent. As a result, the completed model reproduces sexist or racist attitudes present in societies unless the problem is recognised and addressed sufficiently during its development (see e.g. Chatterjee 2023; Valonen 2022).

To reduce the risk that advanced language models or the digital services using them inadvertently reproduce strongly skewed viewpoints or racist or sexist attitudes, companies developing language models should ensure that a sufficiently diverse group is involved in all important stages of model development and evaluation. It is also important to take into account the perspectives and concerns of civil society actors and citizens in these phases.

For this reason, the toolbox of companies developing advanced language models could, in the future, include, for example, deliberative citizens' panels and new types of digital participation platforms designed to facilitate respectful and attentive citizen dialogue. These methods of so-called participatory democracy have been developed to reduce gaps in societal participation. Their use has already been tested in isolated instances in the development of the largest language models, but in the future their use should be consolidated as part of the development of generative AI, and it should be ensured that the methods are not used merely for symbolic participation (Anthropic 2023).

In addition to the companies developing the largest language models, political decisionmakers and public administration could strengthen citizens' participation and influence in AI development.

These actors could ensure that citizens have opportunities to influence broad visions – such as national or sector-specific plans – regarding the use of AI or desired directions for its development. Citizens should also have the opportunity to influence policies that guide and manage the societal changes caused by AI development.

Tools for such citizen participation would include, in part, the same participatory democracy methods that companies developing advanced language models could also use. For example, Belgium has already organised a deliberative citizens' panel that produced recommendations on the most important societal aspects of AI development. The formation of shared positions in the panel was supported by a group of experts whose task was to ensure that all Belgian panellists had sufficient capacities to deliberate in a multifaceted way on the different viewpoints, benefits and risks associated with AI development (Bürgerrat 2024).

Freedom from coercion or undue influence

Everyone should be guaranteed equal opportunities to form opinions about emerging AI, free from coercion or manipulation by those who hold the greatest epistemic power.

In recent years, the growing lobbying power of large technology companies has caused concern, particularly among civil society actors (see e.g. Corporate Europe Observatory 2025a). The large financial and technological resources of tech giants give them substantial means that they can channel into systematic lobbying. As evidence of this, the lobbying expenses of the largest technology companies in the European Union – such as Alphabet, Apple, Facebook and Microsoft – grew by 16.5 percent between 2021 and 2023, reaching 113 million euros (Djurickovic 2023). In practice, technology company lobbyists often provide, for example, detailed proposals for legislative texts to advance the interests of the entities they represent.

In Europe in recent years, large technology companies have sought to influence policymakers and authorities who have been preparing the EU's AI Act and its implementation (see e.g. Corporate Europe Observatory 2025b). The outcome of legislative work may become skewed if large companies have disproportionately greater influence than other actors – such as civil society organisations – by leveraging time, expertise and other resources to influence the preparation of legislation. If different

parties and viewpoints are not heard equally or if all parties do not have equal resources to participate or influence, the principles of democratic decisionmaking are jeopardised (Djurickovic 2023).

Civil society organisations have demanded that the excessive representation of companies in various EU advisory groups, lobbying meetings and political processes should be prevented. They and others have also called for reforms and clarifications to the EU Transparency Register, which in its current form, in their view, does not provide a sufficiently comprehensive or consistent picture of the lobbying activities of different actors within EU bodies (European Court of Auditors 2024; Corporate Europe Observatory 2024).

Civil society organisations should also be ensured sufficient financial resources to participate in the preparation of legislation or in the important subsequent phase of setting technical standards.

For example, in the standardisation work related to the EU's AI Act, it has already been observed that civil society organisations' participation in creating standards is often hindered by financial barriers. To enable civil society organisations to participate more actively in the drafting of AI standards, the funding channels created to support participation in standardisation work should be further developed (Galvagna 2023).

In addition to direct lobbying, systematic attention should in the future also be paid to the more indirect ways in which large technology companies may attempt to exert influence through the AI services they control. From the perspective of the development of generative AI, concerns in the coming years include rapidly evolving AI assistants whose communication is intentionally designed to mimic human interaction.

Research already shows that the more convincingly these interfaces imitate emotional reactions or produce opinions that feel persuasive, the more likely users are to perceive them as having their own 'personality'. Human likeness may also lead users to trust chat interfaces more blindly or lead them to believe that they can form an interactive, social relationship with the interface (Cornelius and Leidner 2021; van Wynsberghe 2022).

Companies or other actors using advanced chat technology may attempt to deliberately manipulate chatbot users, and democratic societies should not be overly naive in assuming that this will not occur. In a global political environment where choices made in technology and service development may be increasingly shaped by overtly political motives, it is plausible that humanlike, emotionally persuasive chat technology could be used to advance political or societal agendas (cf. Jackson 2024; Laforge 2025). For this reason, democratic societies should now engage in broader discussion about how the most vulnerable can be protected from manipulative and deceptive uses of generative AI technology.

5. Afterword: Three steps for applying epistemic rights in practice

The outline of epistemic rights related to AI development presented in this memo is the first attempt to apply the concept of epistemic rights within the context of technological development. It should therefore be regarded primarily as a starting point for discussion, from which understanding and analysis can deepen and evolve over time.

Even in its current form, however, the outline offers many kinds of organisations the opportunity to leverage epistemic rights as drivers for their own activities. Using the framework helps ensure that perspectives on epistemic rights are considered broadly, diversely and at the appropriate time.

The following three steps provide a starting point for those interested in exploring epistemic rights.

1. Integrate epistemic rights into existing processes and functions

Begin by examining your organisation's processes and practices and assessing in which contexts the use of the epistemic rights framework would be most essential and impactful. The framework can be used, for example, to conduct a democracybased review of ethical principles or other key guidelines. The framework may also provide an important perspective for strategy work or assist in assessing how sustainable a new AIbased solution is from the standpoint of democracy. Using the epistemic rights framework to support the responsibility work of organisations that use AI can, in turn, encourage reflection on how to strengthen the positive societal impacts – often referred to as the 'social handprint' – of their operations.

2. Create partnerships boldly across traditional boundaries

In public debate, AI development often appears as an almost inevitable trajectory – one that citizens or even most organisations cannot influence. At the same time, people harbour uncertainties and concerns about this development, and Finnish trust in AI is at a low level in an international comparison (Gillespie et al. 2023).

The epistemic rights framework provides a way to create a counterforce to this development. It gives organisations that have not

traditionally appeared as active actors or changemakers in public discussions on technology development the opportunity to direct their resources and capabilities toward strengthening more democratic and equal AI development. Simultaneously, the framework provides these organisations with a shared valuebased foundation. On this basis, it becomes possible to form partnerships that span traditional sectoral boundaries and other silos.

Different societal actors can also use the epistemic rights framework as a common foundation for preparing a roadmap outlining actions that strengthen epistemic equality.

3. Identify the torchbearers

The logic of epistemic rights will not take root in society unless it gains the support of torchbearers who strategically use their strengths to spread epistemic rights. It is therefore important that organisations whose actions to strengthen epistemic rights can produce more lasting and systemic effects identify their own influence, as well as the leverage points where their actions have a particularly significant impact.

Torchbearers of epistemic rights may include, for example, organisations that conduct – or financially support – applied research and development. They can direct resources to development work and experiments in which practices that strengthen epistemic rights are created for different contexts or sectors.

Key actors may also include publicadministration organisations. They can integrate epistemicrights perspectives into their guidelines, recommendations, policies or decisions that relate, for example, to democracy work, AI education or the responsible use of AI. Public administration can further set an example for other actors by adopting practices that strengthen epistemic rights, improve transparency in the use of AI, or enhance citizens' opportunities to influence decisions regarding AI use.

References

- Ahmed, N., Wahed, M. & Thompson, N. (2023) 'The growing influence of industry in AI research', *Science*, 379, pp. 884–886.
- Allyn, B. (2025) 'The New York Times takes OpenAI to court. ChatGPT's future could be on the line', NPR, 14 January 2025. (Accessed: 12 February 2025).
- Anthropic (2023) *Collective Constitutional AI: Aligning a Language Model with Public Input*, 17 October 2023. (Accessed: 4 October 2024).
- Araujo, T. & Bol, N. (2024) 'From speaking like a person to being personal: The effects of personalized, regular interactions with conversational agents', *Computers in Human Behavior: Artificial Humans*, 2(1).
- A Safe Harbour for Independent AI Evaluation (2024) AI researchers' open letter, March 2024.
- Bansal, G. et al. (2022) 'Workshop on Trust and Reliance in AI-Human Teams (TRAIT)', CHI EA '22 Extended Abstracts, Article 116.
- Bloomberg Intelligence (2024) *Big tech 2025 capex may hit \$200 billion as gen-AI demand booms*, 4 October 2024.
- Bommasani, R. et al. (2024a) 'The 2024 Foundation Model Transparency Index', *Transactions on Machine Learning Research*, December 2024.
- Bommasani, R. et al. (2024b) *Foundation Models under the EU AI Act*. Stanford CRFM, 1 August 2024. (Accessed: 13 November 2024).
- Buschek, C. & Thorp, J. (n.d.) *Models All the Way Down*. Knowing Machines. (Accessed: 13 September 2024).
- Bürgerrat (2024) 'Artificial intelligence discussed with wit', 26 May 2024. (Accessed: 13 September 2024).
- Caltrider, J., Rykov, M. & MacDonald, Z. (2024) 'Romantic AI Chatbots Don't Have Your Privacy at Heart', Mozilla Foundation, 14 February 2024. (Accessed: 12 February 2025).
- Calvagna, C. (2023) *Inclusive AI governance*. Ada Lovelace Institute, 30 March 2023. (Accessed: 4 October 2024).
- Chadwick, L. (2025) 'Estonia announces a new AI initiative with OpenAI to roll out chatbots in schools', *Euronews*, 25 February 2025. (Accessed: 31 March 2025).
- Chatterjee, M. (2023) 'Bias in AI is real. But it doesn't have to exist', *Politico*, 14 July 2023. (Accessed: 31 March 2025).
- Children's Parliament (2024) *Exploring Children's Rights and AI. Stage 2 Summary Report*.
- Cornelius, S. & Leidner, D. (2021) 'Acceptance of Anthropomorphic Technology: A Literature Review', *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 6422–6431.
- Corporate Europe Observatory, LobbyControl, Transparency International & The Good Lobby (2024) *Open Letter to EU Commission President Ursula von der Leyen*, 15 July 2024. (Accessed: 13 November 2024).

Corporate Europe Observatory (2025a) *Coded for Privileged Access: How Big Tech weakens rules of advanced AI*, 30 April 2025. (Accessed: 21 May 2025).

Corporate Europe Observatory (2025b) *Bias baked in: How Big Tech sets its own AI standards*, 9 January 2025. (Accessed: 21 May 2025).

Djurickovic, T. (2023) 'The impact of big tech: Increased lobbying expenses in the EU', *Tech.eu*, 12 September 2023. (Accessed: 13 November 2024).

El Atillah, I. (2023) 'Man ends his life after an AI chatbot "encouraged" him to sacrifice himself to stop climate change', *Euronews*, 31 March 2023. (Accessed: 19 September 2024).

Elliott, D. (2024) 'This AI tutor could make humans 10 times smarter, its creator says', *World Economic Forum*, 29 July 2024. (Accessed: 31 March 2025).

European Court of Auditors (2024) *Special report 05/2024: EU Transparency Register – provides useful but limited information on lobbying activities*. (Accessed: 13 November 2024).

Faktabaari (2022) *Digitaalinen informaatiolukutaito -opas*. Edited by K. Kivinen. Avoin Yhteiskunta ry / Faktabaari EDU.

Färber, M. & Tampakis, L. (2024) 'Analyzing the impact of companies on AI research based on publications', *Scientometrics*, 129, pp. 31–63.

Gillespie, N., Lockey, S., Curtis, C., Pool, J. & Akbari, A. (2023) *Trust in Artificial Intelligence: A Global Study*. University of Queensland and KPMG Australia.

Hallamaa, T. (2024) 'Automaattinen tietojenkäsittely keskittyy valtaviin datakeskuksiin', *Yle*, 27 May 2024. (Accessed: 13 November 2024).

Hanhivaara, J. (2025) 'Asiantuntija kriittinen Suomeen rakennettavista datakeskuksista', *Yle*, 23 February 2025. (Accessed: 28 February 2025).

Hantula, K. & Korkman, O. (2022) *Embracing user unpredictability*. Solita.

Hill, K. (2025) 'They asked ChatGPT questions. The answers sent them spiraling', *The New York Times*, 13 June 2025. (Accessed: 15 June 2025).

Hoffman, V. et al. (2024) 'AI generates covertly racist decisions about people based on their dialect', *Nature.com*, 28 August 2024. (Accessed: 28 February 2025).

Jackson, D. (2024) 'An Eight-Year Struggle for Accountability in the US Ends in Defeat', *Tech Policy Press*, 18 November 2024. (Accessed: 6 March 2025).

Juurola, L. & Korhonen, T. (2025) *Demokratiakasvatus, lasten oikeudet ja tekoäly*, University of Helsinki Innokas network publications. (Accessed: 20 May 2025).

Kestilä-Kekkonen, E. et al. (2022) *Nuoret kansalaiset: Tilastollinen tutkimus yhdeksäsluokkalaisten kansalaispätevydestä*. Ministry of Justice Publications 2022:11. (Accessed: 12 February 2025).

Kambhampati, S. et al. (2025) *Stop Anthropomorphizing Intermediate Tokens as Reasoning/Thinking Traces!*, arXiv preprint.

Khanal, S., Zhang, H. & Taeihagh, A. (2024) 'Why and how is the power of Big Tech increasing in the policy process? The case of generative AI', *Policy and Society*.

Kim, A. (2024) 'Is AI Companionship the Next Frontier in Digital Entertainment?', ARK Invest, 18 June 2024. (Accessed: 12 February 2025).

King, J. & Meinhart, C. (2024) *Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World*. Stanford University white paper.

- Laforge, G. (2025) 'Big Tech's Foreign Policy Takeover', *Tech Policy Press*, 2 May 2025. (Accessed: 6 June 2025).
- Landi, H. (2024) 'Google's healthcare AI ambitions', *Fierce Healthcare*, 19 March 2024. (Accessed: 17 January 2025).
- Leffer, L. (2023) 'Your Personal Information Is Probably Being Used to Train Generative AI Models', *Scientific American*, 19 October 2023.
- Liao, Q.V. & Wortman Vaughan, J. (2024) 'AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap', *Harvard Data Science Review*, 31 May 2024. (Accessed: 20 September 2024).
- Lu, D. (2025) 'We tried out DeepSeek. It worked well, until we asked it about Tiananmen Square and Taiwan', *The Guardian*, 28 January 2025. (Accessed: 3 February 2025).
- Meir, N. (2023) 'AI guidance, terms added to AP Stylebook', *Associated Press*, 16 August 2023. (Accessed: 20 September 2024).
- Mertala, P. & Fagerlund, J. (2024) 'Finnish 5th and 6th graders' misconceptions about artificial intelligence', *International Journal of ChildComputer Interaction*, 39, Article 100630.
- Montgomery, B. (2024) 'Mother says AI chatbot led her son to kill himself', *The Guardian*, 23 October 2024. (Accessed: 13 November 2024).
- Montreal AI Ethics Institute (2025) *The AI Ethics Brief #166*, 2025. (Accessed: 11 June 2025).
- Moore, J. et al. (2025) *Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers*, arXiv preprint.
- Muldoon, J. (2024) "Maybe we can role-play something fun": When an AI companion wants something more', *BBC*, 9 October 2024. (Accessed: 12 February 2025).
- Nelimarkka, M., Mervaala, E. & Laaksonen, SM. (2025) *Rajapinta guidelines for research ethics with generative artificial intelligence*, Rajapinta ry.
- Nieminen, H. (2024) 'Why we need epistemic rights?', in Aslama Horowitz et al. (eds.) *Epistemic Rights in the Era of Digital Disruption*. Cham: Palgrave Macmillan.
- Ng, E. et al. (2025) 'DeepSeek: The Chinese AI app that has the world talking', *BBC*, 4 February 2025. (Accessed: 28 February 2025).
- Nolan, M. (2023) 'Llama and ChatGPT Are Not OpenSource', *IEEE Spectrum*, 27 July 2023. (Accessed: 13 November 2024).
- NCSC – National Cyber Security Centre (2024) *The nearterm impact of AI on the cyber threat*, 24 January 2024. (Accessed: 12 February 2025).
- OECD (2024) *Assessing potential future artificial intelligence risks, benefits and policy imperatives*, OECD AI Papers No. 27. Paris: OECD Publishing.
- Olson, P. (2023) 'Tech's Armies Are Huge, Yet Struggling to Innovate', *Bloomberg*, 26 March 2023. (Accessed: 13 November 2024).
- OpenAI (2025) *Disrupting malicious uses of our models: an update*, 21 February 2025.
- Opetushallitus (2025) *Tekoäly varhaiskasvatuksessa ja koulutuksessa – lainsäädäntö ja suositukset*. (Accessed: 20 May 2025).

Perez, E. et al. (2023) 'Discovering Language Model Behaviors with ModelWritten Evaluations', *Findings of the ACL 2023*, pp. 13387–13434.

Repair-työryhmä (2023) 'Helsingin yliopiston johtama hanke toi tekoälyn eettisen arvioinnin työkalun Suomeen', *University of Helsinki*, 25 October 2023. (Accessed: 13 November 2024).

Requiroso, K. (2025) 'Pentagon Advances Generative AI in Military Operations Amid USChina Tech Race', *eWeek*, 22 April 2025. (Accessed: 20 May 2025).

Robbins, S.A. & Bryson, J.J. (2023) 'Do We Collaborate With What We Design?', *Topics in Cognitive Science*, 1–20.

Ruokonen, M. (2025) *Applying AI in key European industries*. Sitra, 27 March 2025.

Sitra (n.d.) Generatiivisen tekoälyn määritelmä Sitran tulevaisuussanastossa.

Stone, A. (2025) 'How Higher Ed Institutions Are Using BuiltIn Generative AI Tools', *EdTech*, 6 February 2025. (Accessed: 13 March 2025).

Taleva, K. (2025) 'Näin mielenterveyden avuksi kehitetty tekoäly toimii – MTV Uutiset testasi', *MTV3*, 28 January 2025. (Accessed: 28 February 2025).

TanhuaPiironen, E. et al. (2025) *Esi- ja perusopetuksen digitalisaation tilannekuva 2025*. Publications of the Ministry of Education and Culture 2025:13.

Tietotekniikan termitalkoot (2018) 'Chattibotin määritelmä', Sanastokeskus ry.

Tietotekniikan termitalkoot (2025) 'Suuren tai laajan kielimallin määritelmä', Sanastokeskus ry.

Tiku, N. (2024) 'Top AI researchers say OpenAI, Meta and more hinder independent evaluations', *The Washington Post*, 5 March 2024. (Accessed: 20 September 2024).

Toikkanen, T. & Toivanen, M. (2025) 'Tekoälyasetus tulee – onko lukutaito kunnossa?', Sitra, 3 February 2025. (Accessed: 13 March 2025).

Torkington, S. (2024) 'These are the 3 biggest emerging risks the world is facing', *World Economic Forum*. (Accessed: 12 December 2024).

Valonen, T. (2022) 'Tekoälyjärjestelmät täytyy arvioida yhdenvertaisuuden näkökulmasta', *Yhdenvertaisvaltuutettu*, 27 September 2022. (Accessed: 16 September 2024).

Velkova, J. (2024) 'Dismantling public values, one data center at the time', *Reimagining Public Values in Algorithmic Futures Consortium*, 19 February 2024. (Accessed: 13 March 2025).

Ville, L. (2023) *System update: Addressing the gender gap in tech*. Fawcett Society, 30 October 2023.

van Wynsberghe, A. (2022) 'Social robots and the risks to reciprocity', *AI & Society*, 37, pp. 479–485.

Wong, M. (2024) 'Chatbots are primed to warp reality', *The Atlantic*, 30 August 2024. (Accessed: 12 September 2024).

Wu, R. & Yu, Z. (2024) 'Do AI chatbots improve students' learning outcomes? Evidence from a metaanalysis', *British Journal of Educational Technology*, 55(1), pp. 10–33.

Ylöstalo, H. (2020) 'Asiantuntijavalta politiikassa', *Politiikasta*, 24 October 2024. (Accessed: 24 October 2024).

SITRa

SITRA MEMORANDUM, APRIL 2026

Sitra's memoranda are content produced to serve
as background for our futures work.

ISBN 978-952-347-470-3 (PDF) www.sitra.fi

SITRA.FI

Itämerenkatu 11-13
PO Box 160
00181 Helsinki
P. +358 294 618 991
LinkedIn @Sitra