OXFORD

*Policy and Society*, 2025, **44(1)**, 1–22

**DOI: https://doi.org/10.1093/polsoc/puaf001**
Advance access publication date 3 February 2025

**Introduction**

# Governance of Generative AI

Araz Taeihagh

Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore
Corresponding author: A. Taeihagh, Lee Kuan Yew School of Public Policy, National University of Singapore, 469B Bukit Timah Road, Li Ka Shing Building, Level 2, #02-10 259771, Singapore. Email: spparaz@nus.edu.sg; araz.taeihagh@new.oxon.org

## Abstract

The rapid and widespread diffusion of generative artificial intelligence (AI) has unlocked new capabilities and changed how content and services are created, shared, and consumed. This special issue builds on the 2021 *Policy and Society* special issue on the governance of AI by focusing on the legal, organizational, political, regulatory, and social challenges of governing generative AI. This introductory article lays the foundation for understanding generative AI and underscores its key risks, including hallucination, jailbreaking, data training and validation issues, sensitive information leakage, opacity, control challenges, and design and implementation risks. It then examines the governance challenges of generative AI, such as data governance, intellectual property concerns, bias amplification, privacy violations, misinformation, fraud, societal impacts, power imbalances, limited public engagement, public sector challenges, and the need for international cooperation. The article then highlights a comprehensive framework to govern generative AI, emphasizing the need for adaptive, participatory, and proactive approaches. The articles in this special issue stress the urgency of developing innovative and inclusive approaches to ensure that generative AI development is aligned with societal values. They explore the need for adaptation of data governance and intellectual property laws, propose a complexity-based approach for responsible governance, analyze how the dominance of Big Tech is exacerbated by generative AI developments and how this affects policy processes, highlight the shortcomings of technocratic governance and the need for broader stakeholder participation, propose new regulatory frameworks informed by AI safety research and learning from other industries, and highlight the societal impacts of generative AI.

**Keywords:** governance; artificial intelligence; AI; generative AI; GenAI

In 2021, *Policy and Society* published a special issue on the governance of artificial intelligence (AI) (Taeihagh, 2021). The special issue highlighted the governance challenges related to the rapid development of narrow AI and its increased range of adoption in diverse sectors, highlighting the need for the government to manage the scale and speed of socio-technical transitions occurring because of it (Taeihagh et al., 2021). The 2021 special issue emphasized that, while new applications of AI have the potential to increase quality of life and economic efficiency, they have also triggered changes that are threatening existing markets and social and political orders. This is because the heightened pace of adoption of AI in various domains, such as autonomous vehicles, lethal weapon systems, and robotics, can cause unexpected and unintended consequences that need to be addressed by governments. At the time, the special issue underlined gaps and challenges related to the governance of AI, emerging governance responses, and the need to build policy capacity and address the legal and regulatory challenges of AI.

Since then, there have been significant advancements in generative AI that merit further scrutiny. AI is rapidly penetrating various sectors and is driving some of the most groundbreaking scientific discoveries, as evidenced by the 2024 Nobel Prizes in Physics and Chemistry. Meanwhile, the literature on the governance of AI is—while emerging—still significantly underdeveloped relative to the developments in the field. Building on the success of the previous special issue, this special issue focuses on the governance of generative AI, examining complex and multifaceted challenges related to the development, deployment, and impacts of generative AI. With the successful public release of ChatGPT in November 2022, a new AI spring has begun. Generative AI is transforming industries and has since facilitated content creation and writing, art and synthetic media, data analysis, coding and debugging code, personalized education, product design, game design, scientific discoveries, etc. However, it is important to point out that this technology is not without risks and raises social and ethical concerns around trust, intellectual property rights (IPR), bias, discrimination, and the potential for misuse in creating and spreading deepfakes or disinformation (Abbas & Taeihagh, 2024; Chesterman, 2025; Jaidka et al., 2024). In the longer-term, generative AI can have profound impacts on labor markets and the future of employment, increasing inequality and highlighting the need for reskilling, upskilling, and support from the state (Oder & Béland, 2025). It also affects the creative sector by eroding human creativity and originality, raising questions about the future of creative pursuits (Chesterman, 2025). Additionally, it raises concerns related to individuals, cybersecurity, and national security. Deepfakes can be used for defamation, fraud through impersonation, extortion, non-consensual pornography, and personalized malicious campaigns. Furthermore, generative AI tools can facilitate the automation of malware creation and even the development of weapons of mass destruction and autonomous weapon systems by assisting malicious actors via automation and/or sharing of sensitive information. Lastly, some researchers believe that the current loosely controlled development of these systems can lead to uncontrollable or hostile artificial general intelligence, although others are skeptical about this possibility (Goertzel, 2023; Judge et al., 2025; Russell, 2023).

Given the high levels of technological and economic uncertainty surrounding the development of generative AI, the governance of generative AI is of paramount importance to ensure such systems' ethical and responsible use and to ensure that, while a few companies like OpenAI, Anthropic, Microsoft, and Google benefit from these technologies the most, the risks are not just transferred to society and governments (Khanal et al., 2025).

Over the past few years, there has been a recognition of the need for new approaches to the regulation and governance of AI due to information asymmetry, the pacing problem, ethical concerns, myriads of risks, and fragmentation of efforts (Taeihagh et al., 2021; Tan et al., 2022; Zaidan & Ibrahim, 2024). Scholars have proposed adaptive (Tan & Taeihagh, 2021), hybrid (Radu, 2021), anticipatory (Taeihagh et al., 2021), participatory (Cugurullo & Xu, 2025; Ulnicane, 2025), and complexity-informed approaches (Janssen, 2025) and have criticized the inadequacy of voluntary self-regulatory and co-regulatory efforts due to AI safety risks, potential for conflict of interest, and risk of regulatory capture (Judge et al., 2025; Reuel & Undheim, 2024).

This special issue highlights the complex challenges involved in the governance of generative AI. These include legal, organizational, political, regulatory, and social aspects. This article unpacks generative AI and its risks and highlights unique challenges related to the governance of generative AI that needs more attention. The article then summarizes and highlights the key contribution of the seven articles in this issue. The articles highlight the state-of-the-art thinking regarding the governance of generative AI, identify existing gaps and issues, propose solutions and policy recommendations, and offer some avenues for future exploration.

## Background on Generative AI

The early success of AI was primarily due to a focus on symbolic reasoning and rule-based systems, but machine learning (ML) and neural networks have gained prominence more recently. ML is a subfield of AI that focuses on developing algorithms that learn from vast amounts of data and make decisions without reliance on pre-programmed rules to do so for each task. As ML relies more on the data than pre-programmed rules, it is more difficult to predict the decisions of these systems in advance (Mittelstadt et al., 2016).

Generative AI can be defined as a category of AI systems that create new content (text, images, audio, or video) based on inputs, leveraging ML, particularly generative adversarial networks (GANs),

variational autoencoders (VAEs), large language models (LLMs), and diffusion models. GANs use a two-part architecture of generators and discriminators that compete against each other in an adversarial process to generate outputs (Creswell et al., 2018; Goodfellow et al., 2014, 2020). VAEs learn from a compressed probabilistic latent representation of the inputs to generate new outputs (Diederik P. Kingma & Welling, 2019; Kingma & Welling, 2013). LLMs use transformers, a general-purpose deep learning architecture, and train on vast corpora of input data to generate outputs using self-attention mechanisms to process the data in parallel (Chang et al., 2024; Raiaan et al., 2024; Vaswani et al., 2017). Diffusion models create data by progressively adding noise to data and then learn to iteratively de-noise the data in an organized fashion to generate new data (Ho et al., 2020; Yang et al., 2024).

These systems use enormous datasets to learn patterns and, in turn, use them for generating new outputs. For example, OpenAI's GPT (Generative Pretrained Transformer) can generate text based on prompts provided by the user (Radford, 2018), Stability AI's Stable Diffusion can generate images based on text (Mostaque, 2022), and Google's MusicML (Agostinelli et al., 2023) can generate music from text. More recent trends include the development of multimodal models that process and combine multiple types of data to generate novel outputs (Radford et al., 2021), integration of generative AI with other forms of autonomous systems and robotics (Jabbour & Reddi, 2024), AI agents (Cronin, 2024), and the exploration of new applications of generative AI in science and engineering (Decardi-Nelson et al., 2024).

The accessibility of these systems through open-source libraries, availability of high-quality data, easy access to computational resources through cloud-based platforms, and friendly user interfaces have democratized access to generative AI and has rapidly accelerated their adoption in different sectors. This democratization, while empowering users and facilitating their adoption in new industries, is also increasing the risks and potential for misuse. Therefore, proactive and adaptive governance is necessary to ensure responsible innovation and prevention of harm.

## Key risks of Generative AI systems

Many of the previous AI systems were rule based and addressed predefined tasks, whereas generative AI systems rely more on their training data, work autonomously in creating their responses, and have the capability to create new content, which can result in unexpected outcomes. Some of the unique risks of generative AI systems include the following.

### Hallucination and inaccuracies

Unlike traditional rule-based AI, generative systems typically operate with fewer constraints and can create new content. As such, a key risk of generative AI is when the system produces fabricated or incorrect results, which is highly problematic in high-risk applications such as healthcare, defense, or the judiciary. In May 2023, a highly publicized case involved two lawyers who had submitted a court filing that included hallucinated content without verification, which resulted in legal consequences for the individuals involved (Weiser, 2023).

### Jailbreaking and handling unexpected inputs

Jailbreaking in the context of generative AI refers to the process of manipulating the model to bypass its built-in guardrails (e.g., safety measures, ethical guidelines, and operational constraints) to elicit responses from the model that under normal conditions it would refuse to provide (Niu et al., 2024). Including prompt-level (e.g., filtering, transformation, and optimisation), model-level (e.g., adversarial training, safety finetuning, pruning, alignment checking, etc.), and multi-agents and other defensive mechanisms against jailbreaking can increase the robustness of the system (Peng et al., 2024).

### Data training and validation risks

Generative AI models use vast amounts of data for training, much of which is sourced from the internet without permission. This has raised significant IPR-related questions (Chesterman, 2025). In addition, reliance on unfiltered internet data can result in the incorporation of biases, propagation of misinformation, and infringement of privacy rights (Bender et al., 2021).

These models are heavily influenced by the data used for training them. Therefore, as the data include biases, the models will also exhibit or even amplify those biases in their outputs. For instance, if a model is trained on data that are not demographically representative, it might disproportionately

underperform for those demographics during inference (Rofl et al. 2021; Lim & Taeihagh, 2019). Similarly, if the available training data do not match the deployment setting, this could increase risks and result in poorer performance (Tan et al., 2022). Alternatively, if a benchmark dataset contains systematic labeling errors (Northcutt et al., 2021), or biases are introduced due to annotators' beliefs and backgrounds, these factors will have wide-ranging implications (Sap et al., 2021). Another set of risks for training data arises due to leakage of synthetic and benchmark data into training datasets. This contaminates them, which can reduce model performance, create biases, and compromise the integrity of model evaluation (Xu et al., 2024). Additionally, data poisoning through deliberate manipulation of the training data could introduce vulnerabilities into the system (Dong et al., 2025). Moreover, recent research shows that indiscriminate use of synthetic content from the models in training can eventually cause model collapse (Shumailov et al., 2024).

## Sensitive information risks

The presence of sensitive and personal information in the training data creates risks as the models could be exploited to gain this information. It was announced that the US Department of Energy's National Nuclear Security Administration has begun red teaming with Anthropic to ensure the security of their AI models regarding leakage of sensitive nuclear information (Sabin, 2024).

In addition, the users' level of familiarity with how the system works, or their commitment to protecting their personal data and their organization's sensitive information, can affect the likelihood of them sharing sensitive information with the models. If this information is shared with the generative AI model, it could be used by the providers and, if the data are then used for subsequent training of the models, could eventually create the information risks highlighted earlier.

## Opacity risks

Generative AI models operate as black boxes, as it is unclear how they arrived at their outputs, making it difficult to trace their decisions, ensure accountability, and consequently build trust in them (Taeihagh, 2021). As these models currently learn patterns from billions of parameters (soon to scale to trillions of parameters if the current trajectory of model development continues), they are not easily interpretable by humans because understanding their computational layers is difficult. As a result, it is not clear why they arrive at a certain output (Arrieta et al., 2020). This lack of transparency increases the likelihood of generating unintended outputs such as hallucinations that were discussed earlier, as well as harmful content, as it is difficult to anticipate how the models behave (Bender et al., 2021).

## Control risks

Generative AI systems, particularly with the recent developments in agent-based models, can operate at different levels of autonomy (human-in-the-loop, human-on-the-loop, and fully autonomous) (Firlej & Taeihagh, 2021). This, in turn, results in varying levels of difficulty in regaining control of such systems during malfunctions, necessitating the ability to rapidly regain control or shut down human-on-the-loop and autonomous systems if issues arise. Here, the level of the opacity of the system will affect whether the malfunction can be identified easily. Furthermore, as generative AI systems are now offered to consumers and businesses, aside from the built-in guardrails discussed under jailbreaking risks, the ability to detect and prevent misuse of the systems is important. If the systems have continuous learning capabilities that could result in undesirable behavior, the ability to swiftly control the system becomes more important, for example in the use of Agentic AI or autonomous vehicles with learning capabilities (Tan et al., 2022).

In addition, extensive deployment of the system and use of multimodal models could increase the control risks as they increase the attack surfaces. Deployment of multiple systems that could interact with each other could result in emergent behavior such as in the case of the Facebook chatbot negotiation experiment, wherein the chatbots developed a shorthand language to communicate that was difficult for researchers to follow (Fauzia, 2021; Koff, 2023).

## Design and implementation risks

Design and implementation choices can have significant consequences in terms of the risks of generative AI systems. For instance, access to diverse expertise and involvement of stakeholders can help in reducing unintended consequences (Tan et al., 2021). Similarly, choices in implementation, such as the use of open-source software, can impact the system. Use of external libraries, application programming

**Table 1.** Key risks of generative AI.

| Risk | Description | Example/references |
|---|---|---|
| **Hallucination and inaccuracies** | Generative AI systems can produce fabricated or incorrect results due to their autonomous content creation capabilities. This poses significant risks in critical domains like healthcare, defense, or judiciary. | Legal consequences for lawyers who submitted hallucinated content without verification (Weiser, 2023) |
| **Jailbreaking and unexpected inputs** | Manipulating models to bypass built-in safety measures ("jailbreaking") can elicit disallowed or harmful responses. | Niu et al. (2024); Peng et al. (2024) |
| **Data training and validation risks** | Utilizing vast amounts of unfiltered internet data without permission raises intellectual property concerns and can introduce biases, propagate misinformation, and infringe upon privacy rights. Furthermore, experts are warning of issues such as data poisoning and model collapse due to synthetic data reuse. | Bender et al. (2021); Chesterman (2025); Dong et al. (2025); Lim & Taeihagh (2019); Northcutt et al. (2021); Rofl et al. (2021); Sap et al. (2021); Shumailov et al. (2024); Tan et al. (2022); Xu et al. (2024) |
| **Sensitive information risks** | Models trained on data containing sensitive or personal information can be exploited to extract such data. Moreover, users may share sensitive information, increasing risks if models use this data for further training. | Collaboration to secure AI models against leakage of sensitive nuclear information (Sabin, 2024). |
| **Opacity risks** | The "black box" nature of generative AI models makes it difficult to understand how outputs are produced, hindering traceability, accountability, and trust. Lack of transparency increases the likelihood of unintended outputs like hallucinations and harmful content. | Arrieta et al. (2020); Bender et al. (2021); Taeihagh (2021) |
| **Control risks** | Varying levels of autonomy affect the ability to regain control during malfunctions particularly for Agentic AI. Extensive deployment, multimodal models, and interacting systems increase attack surfaces, potentially leading to unpredictable emergent behaviors. | Facebook chatbot experiment where chatbots developed their own language (Fauzia, 2021; Koff, 2023) |
| **Design and implementation risks** | Design and implementation choices, including stakeholder involvement, significantly impact the risks associated with generative AI systems. Using open-source software and external libraries/APIs can introduce vulnerabilities. | Eiras et al. (2024); Plate et al. (2015); Sconiers-Hasan (2024); Tan et al. (2021) |

Source: Author.

interfaces (APIs), and models can have significant benefits but can also increase risks depending on the level of community engagement and robustness of these external resources (Eiras et al., 2024; Plate et al., 2015; Sconiers-Hasan, 2024). Table 1 summarizes the key risks of generative AI.

# Governing Generative AI

## Governance challenges of generative AI

Generative AI represents a significant advancement in AI, unlocking new capabilities and outputs that could reshape various sectors and foster innovation. As discussed in the previous section, generative AI also presents unique risks, which need to be addressed. Generative AI raises a diverse set of complex governance challenges that extend well beyond engineering of the systems themselves. Addressing these challenges requires urgent attention and a comprehensive approach to safeguard citizens and ensure society benefits from generative AI while mitigating its negative and/or unintended consequences. These governance challenges are discussed in the rest of this subsection.

Data governance and intellectual property concerns arise because models require large amounts of data that are often collected from the internet without permission, raising intellectual property, copyright, and fairness concerns (Chesterman, 2025). Generative AI has created significant legal and ethical concerns for creators, publishers, and consumers. In several high-profile cases, copyright owners have alleged that their work has been included in training datasets without their consent (Coulter, 2024; Gerken, 2023). In addition, there are uncertainties around the copyright of AI-created content, the threshold for copyright for AI co-created content, as well as derivative works (Chesterman, 2025).

The amplification of bias and privacy violations is also a significant governance challenge. As discussed in the previous section, data training and validation practices in generative AI models could propagate or even amplify biases in society related to race, ethnicity, gender, or other protected characteristics, leading to discriminatory outcomes and inequalities through means such as generation of media that reinforce stereotypes (Bender et al., 2021) that require policy responses. Furthermore, widespread deployment of generative AI raises questions about access and equity as it may increase the digital divide and exacerbate social inequalities.

Another concern is the spread of misinformation and disinformation. Generative AI's ability to rapidly create realistic synthetic media (e.g., text, audio, and visual content) creates significant challenges in terms of misinformation and disinformation (Jaidka et al., 2024; Ng & Taeihagh, 2021). Furthermore, the ability to use agentic generative AI as well as personalize the content further facilitates automation of the creation and dissemination of the content, increasing its potency. This could erode social capital and public trust. Beyond technical fixes, this requires increasing stakeholder engagement to build digital information resilience.

The potential for fraud and cybercrime is heightened as malicious actors can deploy synthetic media generated by generative AI, raising concerns for individuals' safety, national security, and cybersecurity (Mavikumbure et al., 2024; Pasupuleti et al., 2023). The proliferation of deepfakes (Abbas & Taeihagh, 2024) by malicious actors can damage reputations; spread falsehoods; facilitate fraud through phishing attacks, extortion, and non-consensual pornography and impersonation; and manipulate public opinion through personalized campaigns at scale. Automation facilitated by generative AI can facilitate the development and deployment of malicious code and the creation of large-scale and autonomous cyberattacks.

The societal impacts and the future-of-work challenges due to generative AI—such as the impact on the creative industries, human expression, and the wider labor force due to automation (Frey & Osborne, 2023)—require proactive engagement from governments to ensure that adequate protections are offered to affected workers (Oder & Béland, 2025). Furthermore, due to the black box and opaque nature of generative AI, questions around transparency, accountability, and liability need to be addressed so that the risks are not just transferred to society (Taeihagh, 2021). The traditional regulatory frameworks for data protection, intellectual property, and the like were often defined based on human agency and are unable to adequately address the complexities of generative AI.

Power imbalances pose another governance challenge. Big Tech's dominance over generative AI development and deployment can reinforce existing inequalities. The informational advantages and technical capabilities of a few dominant firms (Khanal et al., 2025; Zhang et al., 2024) can lead to an increased ability to shape policy discussions and regulatory frameworks to their benefit. Stronger governance measures, including antitrust policies, transparency requirements, and public engagement, are needed to promote inclusivity, prevent regulatory capture, and ensure that AI governance remains balanced and representative of diverse interests and perspectives. Voluntary guidelines alone are insufficient when faced with entrenched power dynamics (Khanal et al., 2025).

There is also a narrow focus and limited public engagement in relation to the development and deployment of generative AI. Even though generative AI will impact the whole of society, public participation and engagement are limited (Ulnicane, 2025). Decisions around the development and deployment of the technology are often made in a technocratic fashion (Cugurullo & Xu, 2025). This lack of democratic engagement could lead to developments that have a narrow focus on risk management without aligning the purpose and direction of the developments with wider societal values and considering the views of people who will be affected (Judge et al., 2025; Oder & Béland, 2025).

The public sector faces a variety of challenges in regard to the governance of generative AI. However, it must play an important role in such governance, not just in the development and enforcement

of comprehensive laws and regulations but also in promoting awareness, collaboration, and partnerships among stakeholders to ensure responsible adoption of the technology (Taeihagh, 2023). The public sector itself faces significant challenges due to technical skills shortages, information asymmetry, and institutional resistance. The pacing problem and the rapid technological development and emergence of new capabilities and applications relative to legislative developments create regulatory gaps (Pande et al. 2023) that have resulted in reactive rather than anticipatory and adaptive governance (Tan & Taeihagh, 2021). The difficulties in attracting and retaining AI talent due to the higher salaries and attractiveness of working in the private sector also compound these difficulties. In addition, the general-purpose nature of generative AI and its applicability across various sectors and industries makes it difficult to define the scope of regulation and apply it consistently. This problem has often resulted in vague laws that lack the specificity required of effective regulation (Guihot et al., 2017; Larsson, 2020).

Finally, there is a need for international cooperation. The global nature of AI developments demands international collaboration on standard setting, the development of guidelines and best practices, and addressing transboundary concerns given different cultural norms, priorities, and legal frameworks in different jurisdictions. This issue is particularly important for mitigating the risks of the deployment of AI in autonomous weapon systems (Firlej & Taeihagh, 2021) and preventing a race to the bottom given the US–China AI rivalry (Zhang et al., 2024). A collaborative international effort is needed to effectively govern generative AI.

By focusing on the governance of generative AI, policymakers, private companies, researchers, and the public can work together to build trust in the technology and ensure that it is used responsibly, moving beyond limited quick "governance fixes" (Ulnicane, 2025) and embracing a more comprehensive and systemic approach that prioritizes social values and public good (Janssen, 2025).

## Governance of generative AI: a comprehensive framework for the future

Addressing the risks and challenges highlighted earlier underscores the urgent need for a comprehensive approach to the governance of AI. A fundamental shift in how governments approach AI needs to occur. Governments need to be more proactive and adaptive in governing AI and make the process more participatory. This requires capacity building within the public sector, regulatory innovations, and taking bold actions to protect citizens and democratic values in the face of the rapid technological disruptions wrought by AI advancements and technological convergence in areas such as robotics, energy, manufacturing, and biotechnology.

The governance of generative AI requires an adaptive approach that facilitates iteratively responding to new risks and challenges swiftly and ensuring that regulations do not become obsolete due to the pace of advancements (Tan & Taeihagh, 2021). The use of regulatory sandboxes and pilot programs could provide a safe and controlled environment for experimenting with generative AI, thus helping regulators understand practical challenges and adapt regulations iteratively, in addition to traditional audit and compliance mechanisms (Philipsen et al. 2021). While self-regulation is by no means adequate for governance of generative AI (Judge et al., 2025), it can play an important role in promoting collaboration within the industry and establishing codes of conduct that can be complementary to regulations. In addition, promotion of safer AI architectures as well as research into the development of AI for monitoring, and enforcement of regulations need further attention.

### *Improving data governance and IPR*

Given the risks and challenges of data governance, developing novel frameworks for data sharing and data protection is crucial to balance the needs for open data for innovation and public safety with individual privacy. Beyond mandatory disclosure requirements related to training data and model capabilities, an exploration of novel concepts regarding data ownership and investigation of the role of blockchain technology for data management and privacy are important. With the increased use of generative AI and recent success of multimodal models, questions on ethics and the impacts of the use of these models and biometrics in various applications need more attention (Ng et al., 2023). While some applications such as teleoperation and training of robots and autonomous systems (Cheng et al., 2024; Ding et al., 2024) are groundbreaking and should be encouraged, the ease of use of these models opens the door for abuse and misuse as well.

Similarly, a novel approach to copyright and IPR is needed to ensure the rights of creatives while promoting innovation as well as addressing new concerns around AI-generated content, AI-cogenerated content, and derivatives [see Chesterman (2025) in this issue].

### *Reducing bias and privacy violations*

Governments need to encourage responsible innovation and ethical AI developments given that tackling bias and privacy concerns in generative AI systems requires a concerted effort in addressing various aspects of data collection, training, and validation of models. Focuses need to include improving practices around data curation, ensuring the use of diverse and representative datasets, and development of tools and procedures for detecting and addressing biases during training and deployment. Having diverse teams, engaging stakeholders, and developing workflows for swiftly addressing the concerns of those who have been negatively affected are of outmost importance for building trust.

Within organizations, red teaming, impact assessment, and internal auditing need to become routine. Furthermore, independent regulators should conduct audits of the models, particularly if the impact of the model or scale of its deployment is significant.

Another area that needs urgent attention is the development of standards, ethical guidelines, certification schemes, and codes of conduct. Much like the rest of the engineering professions, computer science as a discipline needs to rapidly develop such procedures to ensure responsible and ethical use of technology. While such efforts are underway by professional organizations, the rate of development and adoption needs to significantly increase.

### *Improving risk management and cybersecurity*

The development of technical solutions can help in addressing the spread of misinformation and disinformation, as well as fraud and cybercrime. These include techniques to identify synthetic media, watermarking, and provenance tracking (Abbas & Taeihagh, 2024; Jaidka et al., 2024). With the increased use of digital platforms and generative AI in the coming years, the frequency and spread of cyberattacks and mis/disinformation will undoubtedly increase. This requires developing digital resilience in platforms through the development of public–private partnerships and among the population through media literacy and educational campaigns to highlight the new capabilities and risks of AI.

Generative AI is already facilitating fraud and cybercrime, and it is possible that in the near term, more people will fall victim to such crimes. At the moment, the speed of the response to such incidents on the part of banks and police needs to significantly improve. Due to the nature of such crimes, often funds are quickly transferred from mule accounts to other jurisdictions, which requires fast reaction from police and banks to block such activities and/or claw back the funds. New cross-border mechanisms need to be established so that the transfer of funds does not render their retrieval impossible.

### *Reducing negative societal impacts*

While there are questions about whether the rate of developments in generative AI can be maintained, in the longer run, profound societal changes due to AI adoption are inevitable. This requires preparation for the socioeconomic impacts of automation due to labor market disruptions, concentration of wealth, and income disparities. Another area that needs significant attention is the study of the environmental impact of generative AI, keeping in mind the high computational demands of training and operating these systems.

Governments must prepare for and adapt to significant job displacements and ensure social stability during the transition period, deploying the learning from past technological transitions. Given the significant failures and limitations of universal basic income experiments (Giles, 2024; Hiilamo, 2020; Miller et al., 2024; Talgo, 2024; Vivalt et al., 2024), options such as windfall taxes and capital gain taxes on AI-generated wealth need to be seriously explored to provide support for those impacted by AI automation and to ensure that the benefits of AI are broadly shared by society rather than just by Big Tech. Governments need to prepare for significant disruptions that may render certain skills and professions obsolete. The education system's priorities need to be shifted toward skills that are less likely to be automated, as well as to reskilling/upskilling individuals to adapt to changes. AI literacy should become foundational in education systems to ensure a broad understanding within society of how these technologies are affecting our lives. This education will ultimately help maintain human agency and

dignity. Mandating efficiency improvements and mitigation strategies is necessary for addressing the environmental impact of generative AI.

### *Reducing power imbalances*

Given the concentration of power in Big Tech, it is important to empower small and medium-size enterprises as well as universities to make meaningful contributions to AI development. Training and running generative AI models is capital intensive, and Big Tech recruits some of the best AI talent, which further entrenches their power and influence in AI development and in fact the policy process as well (Khanal et al., 2025). Furthermore, Big Tech is adept at employing scaremongering tactics with the aim of advocating for regulations that would hinder the progress of open-source AI developments (Davidson, 2023). Beyond establishing strong regulatory and oversight frameworks that promote transparency and accountability on the part of Big Tech, governments must facilitate the creation of AI development ecosystems that support small and medium-size enterprises and universities, promote the development of open-source AI communities, and offer financial support.

### *Enhancing public engagement*

Generative AI governance should address the needs of those who will be affected by the technology and not just those powerful entities that are driving the development of generative AI. Rather than prioritizing quick technical solutions (governance fixes), they should take into account deeper societal considerations (Ulnicane, 2025). Policymakers should use participatory methods and seek continuous input from the public regarding policy decisions and the direction and purpose of technological developments rather than just focusing on risk mitigation strategies. Bringing together government, industry, civil society, academia, and the public should result in transparent discussions around generative AI developments and ensure consensus-building and collaboration so that these governance frameworks reflect societal values and priorities.

### *Public sector enhancements*

The public sector must develop the technical, organizational, and policy capacities needed for understanding the impact of generative AI on different aspects of the substance and process of policymaking. Public officials need to be educated about the benefits, challenges, and implications of AI development and deployment, and cross-sector taskforces need to be established to rapidly address issues such as ensuring equitable access and preventing exacerbation of the digital divide.

The potential role of generative AI in public service delivery and decision-making should be thoroughly examined. Necessary reforms should be conducted to facilitate the attraction and retainment of AI talent within the government at the highest levels of decision-making, as well as relevant regulatory agencies. The necessary resources need to be made available for overseeing AI developments, reducing the gap in technical know-how, and improving the efficiency and effectiveness of public sector service delivery through use of AI.

### *Increasing international cooperation*

The establishment of norms and common standards, best practices, and mechanisms for addressing cross-border regulatory challenges and global risks and avoiding regulatory fragmentation is essential for ensuring effective governance of generative AI. Given the global AI race, pursuing diplomatic efforts and international agreements—particularly regarding AI safety, banning the use of fully autonomous weapon systems and preventing the use of AI to trigger nuclear weapons and develop weapons of mass destruction—can help avoid escalations and increase global stability (Egan & Kine, 2024; Firlej & Taeihagh, 2021; Zhang et al., 2024), as well as help governments focus on addressing power imbalances in AI development (Khanal et al., 2025). Ulnicane (2025) suggests moving toward a polycentric system of governance and highlights a number of early efforts toward international coordination on the governance of generative AI, while emphasizing that more needs to be done to increase the inclusivity of such efforts. In addition, in the short term, establishment of Intergovernmental Panel on Climate Change-like efforts can help bring scientific consensus on the risks of AI, while, eventually, the establishment of an intergovernmental organization such as the International Atomic Energy Agency with authority based on legally binding agreements and treaties is necessary. See Table 2 for the summary of generative AI governance strategies.

**Table 2.** Governance Strategies for generative AI.

| Area | Key actions and recommendations |
|---|---|
| **Improving data governance and intellectual property rights** | (a) Develop novel data frameworks to balance innovation, public safety, and individual privacy.<br>(b) Implement mandatory disclosure of training data and model capabilities.<br>(c) Explore new concepts of data ownership, such as utilizing blockchains.<br>(d) Address ethical concerns of multimodal models and biometrics (Ng et al., 2023).<br>(e) Recognize potential for misuse due to ease of use.<br>(f) Develop novel approaches to copyright and IPR to protect creatives (Chesterman, 2025). |
| **Reducing bias and privacy violations** | (a) Promote responsible innovation and ethical AI development.<br>(b) Improve data curation practices using diverse and representative datasets.<br>(c) Develop tools and procedures to detect and address biases during training and deployment.<br>(d) Assemble diverse teams and engage stakeholders to build trust.<br>(e) Implement swift workflows to address concerns of negatively affected individuals.<br>(f) Institutionalize red teaming, impact assessments, and internal auditing.<br>(g) Conduct independent audits for models with high impact or large-scale deployment.<br>(h) Develop and rapidly adopt standards, ethical guidelines, certification schemes, and codes of conduct in computer science and accelerate efforts by professional organizations. |
| **Improving risk management and cybersecurity** | (a) Develop technical solutions to identify synthetic media, such as watermarking and provenance tracking (Abbas & Taeihagh, 2024; Jaidka et al., 2024).<br>(b) Enhance digital resilience through public–private partnerships.<br>(c) Promote media literacy and educational campaigns about AI capabilities and risks.<br>(d) Prepare for increased cyberattacks and the spread of misinformation and disinformation.<br>(e) Improve response times of banks and police to AI-facilitated fraud and cybercrime.<br>(f) Establish cross-border mechanisms to quickly block illicit activities and recover funds. |
| **Reducing negative societal impacts** | (a) Prepare for impacts of automation, including job displacement and income disparities.<br>(b) Study and address the environmental impact of generative AI. Mandate efficiency improvements and mitigation strategies for environmental concerns.<br>(c) Recognize limitations of universal basic income experiments (Giles, 2024; Hiilamo, 2020; Talgo, 2024) and explore options like windfall and capital gain taxes on AI-generated wealth to support affected individuals (Miller et al., 2024; Vivalt et al., 2024).<br>(d) Shift education priorities to skills less likely to be automated. Emphasize reskilling and upskilling.<br>(e) Integrate AI literacy as foundational in education to maintain human agency and dignity. |
| **Reducing power imbalances** | (a) Empower small and medium-sized enterprises (SMEs) and universities to contribute to AI development.<br>(b) Promote open-source AI initiatives and communities and offer financial support.<br>(c) Recognize Big Tech's influence (Khanal et al., 2025) and establish strong regulatory oversight for Big Tech to ensure transparency and accountability. Counteract scaremongering tactics by Big Tech that hinder open-source AI progress (Davidson, 2023). |
| **Enhancing public engagement** | (a) Use participatory methods to involve the public in AI policy decisions and ensure that governance frameworks reflect societal values through consensus-building and collaboration.<br>(b) Address deeper societal considerations beyond quick "governance fixes" (Ulnicane, 2025).<br>(c) Seek continuous public input on the direction and purpose of technological developments.<br>(d) Foster transparent discussions among government, industry, civil society, academia, and public. |

(continued)

**Table 2.** (Continued)

| Area | Key actions and recommendations |
|---|---|
| **Public sector enhancements** | (a) Educate officials about AI benefits, challenges, and implications. Develop capacities to understand AI's impact on policymaking. Examine AI's role in public service delivery and decision-making. Establish taskforces to ensure equitable access and prevent digital divide.<br>(b) Reform policies to attract and retain AI talent within government and regulatory agencies.<br>(c) Allocate resources to oversee AI developments and improve public service efficiency with AI. |
| **Increasing international cooperation** | (a) Develop mechanisms to address cross-border regulatory challenges and global risks; avoid regulatory fragmentation. Establish global norms and standards for effective AI governance.<br>(b) Pursue diplomatic efforts and international agreements on AI safety, including banning fully autonomous weapon systems and preventing AI-triggered nuclear weapons and weapon of mass destruction (WMDs) (Firlej & Taeihagh, 2021; Egan & Kine, 2024;Zhang et al., 2024).<br>(c) Move toward polycentric governance with increased inclusivity (Ulnicane, 2025).<br>(d) Initiate efforts like Intergovernmental Panel on Climate Change (IPCC) to build scientific consensus on AI risks and establish an International Atomic Energy Agency (IAEA) like international organization with authority through legally binding agreements and treaties. |

Source: Author, based on numerous sources.

## Overview of the special issue articles

The seven articles in this special issue tackle the complex legal, organizational, political, regulatory, and social challenges of governing generative AI. A central theme is the inadequacy of existing governance frameworks for addressing the unique challenges that generative AI presents. In light of the rapid advancements, the authors collectively highlight the urgent need for innovative governance approaches that are adaptive, inclusive, and aligned with societal values. Chesterman (2025) examines the intricate IPR challenges of generative AI in terms of the inputs and outputs of the models. He highlights how the traditional legal frameworks fall short, which resonates with Janssen's (2025) critique of traditional information technology (IT) governance models and proposed complex adaptive system (CAS) perspective for the governance of generative AI. Extending this discussion, Khanal et al. (2025) examine how Big Tech's power is increasing and they are acting as "super policy entrepreneurs," leveraging their influence to shape policymaking, calling for a re-evaluation of public policy theories to account for this influence. Ulnicane (2025) further critiques the technocratic nature of generative AI governance and advocates for democratic participation and responsible innovation. Judge et al. (2025) echo the need for new regulatory frameworks as we are facing AI systems with emergent behaviors, which aligns with the model Janssen (2025) suggests. Cugurullo and Xu (2025) explore the social implications of the use of generative AI and city brains and advocate for participatory anticipatory governance, echoing Ulnicane's (2025) call for inclusive and participatory governance, a theme that is also highlighted in the work of Oder and Béland (2025) examining the impacts of generative AI on low-skilled workers. Below is a summary of the articles in the special issue on the governance of generative AI.

### Good models borrow, great models steal: Intellectual property rights and generative AI [Chesterman (2025)]

The rapid advancement of generative AI has raised significant questions regarding IPR and is impacting knowledge workers and creatives. This article addresses the following two questions: Should creators or owners of data that are scraped be compensated for their use in the training of AI models? Who, if anyone, should own the output of generative AI models?

Chesterman (2025) first examines the training of generative AI models. The author highlights the ethical and legal challenges arising from the use of data without permission or compensation for creators. He argues that our current legal frameworks are stressed due to such wholesale use of copyrighted material for training AI models, citing the lawsuit brough by Getty Images against Stability AI (Coulter,

**Table 3.** The input and output challenges of generative AI in relation to IPR and potential solutions [based on Chesterman (2025)].

| | Challenges | Solutions |
|---|---|---|
| Inputs (training data) | Permissionless and uncompensated data scraping for training models | Encouraging models that respect copyright and compensate creators<br>Regulation mandating the disclosure of the sources of training data |
| Outputs (AI-generated content) | Ambiguity surrounding the ownership of AI-generated content | Developing legal frameworks for the ownership of AI-generated content<br>Promoting licensing and provenance of AI-generated content |

Source: Author, based on Chesterman (2025).

2024). The article also highlights different approaches adopted by jurisdictions such as Singapore and the European Union, which have introduced exceptions for "computational data analysis" and "text and data mining" to balance the interests of creators and AI developers [Chesterman (2025) citing Copyright Act (2021), ss. 243–244] and the EU Directive on Copyright in the Digital Single Market (EU Directive, 2019). Then, the article examines the question of ownership in AI-generated outputs, highlighting that IPRs were designed for protecting human creativity and that most jurisdictions do not grant copyright protection to works solely produced by AI. The author examines the ambiguity surrounding AI-assisted content and the difficulty of establishing the threshold of human involvement required for works to be granted copyright protection, highlighting the UK's model of "computer-generated" works. This model assigns limited rights to the person who made the arrangement for the creation of the content as a potential model to be followed (Copyright, Designs and Patents Act, 1988).

The author argues that failing to address IPR questions will have serious consequences, as unrestricted use of copyrighted material for training of the AI models reduces the incentives for humans to take on creative professions and erodes the value of human intellectual contributions, while intellectual property protections that are too stringent could stifle innovation and limit the benefits of generative AI in reducing the cost of content creation and dissemination. The author draws learnings from the music industry and issues of piracy with the rise of Napster and how legal actions and the rise of legitimate and licensed platforms helped the industry to adapt. He argues that the rise of a new breed of AI models that respect the copyrights of human creators and are transparent regarding the source of their training data can balance the interests of the AI developers and creators. Table 3 shows the key input and output challenges of generative AI in terms of IPR and potential solutions to them, highlighting the importance of transparency.

## Responsible governance of generative AI: conceptualizing AI governance as a complex adaptive system (Janssen, 2025)

With the increased adoption of generative AI in organizations and their rapid evolution, Janssen (2025) argues that traditional IT governance strategies are often deterministic and focus narrowly on the technology itself without adequately considering the context in which the technology operates. The article addresses the following overarching question regarding the governance of generative AI: How can generative AI be governed responsibly within organizations considering its complex and evolving nature? To address this question, the article explores why traditional IT governance approaches are insufficient for effective governance of generative AI and how alternative frameworks and strategies could improve governance of generative AI.

Janssen critiques different streams of IT governance (centralization/decentralization, contingency strategy, aligning IT with organizational objectives, managing risk/return trade-offs, control mechanisms and procedures, human involvement in AI decisions, and complying with regulations) and how they fail to account for the dynamic and emergent properties of generative AI. The author highlights the opacity and unpredictability of generative AI outputs and points out issues such as lack of transparency, data quality, bias, and hallucinations, along with the rapid pace of technological advancements, necessitating the need for responsible generative AI governance that considers social values and citizens'

**Table 4.** Traditional IT governance vs. CAS-based AI governance [based on Janssen (2025)].

| | Traditional IT governance | CAS-based AI governance |
|---|---|---|
| Focus | Internal to the organization, technology focused | Socio-technical system as a whole |
| Accountability | Predictive accountability with clearly defined roles and responsibilities | Joint accountability among all stakeholders |
| Governance approach | Control-oriented strategy | Adaptation-oriented strategy |

Source: Author, based on Janssen (2025); Li et al. (2021).
Abbreviations: CAS, complex adaptive system.

expectations. The author proposes adopting a CAS lens (Holland, 2006) for conceptualizing AI governance to help better understand and manage the evolving nature of it, as an alternative to traditional AI governance strategies. The author argues for viewing AI as part of a socio-technical system where people, policies, processes, and technologies dynamically interact and co-evolve. The CAS lens highlights that governance should recognize the interdependencies and interactions within the system. It should try to influence the behavior of different agents within the system (developers, users, organizations, and society) to achieve systemwide optimal outcomes.

Given the "responsibility gap" stemming from the complexity and unpredictability of generative AI, assigning responsibility for specific outcomes in generative AI is difficult. The author proposes moving away from "predictive accountability" toward "joint accountability" and a collaborative approach and use of feedback loops in the CAS model to share responsibility among all stakeholders for ensuring responsible use of generative AI. Janssen (2025) posits that there is no shortage of governance strategies. It is important to select appropriate combinations of policy instruments into policy packages to ensure their effectiveness. He argues that governance mechanisms should guide rather than control AI and advocates for establishing clear accountability, robust risk assessment, open communication, and continuous learning, as well as adaptation-oriented strategies to influence socio-technical systems' evolution (Li et al., 2021). Table 4 contrasts traditional IT governance and the CAS model for governance of generative AI.

## Why and how is the power of Big Tech increasing in the policy process? The case of generative AI (Khanal et al., 2025)

With the rapid rate of digitalization, large technology companies (Big Tech)—benefiting from network effects—have accumulated significant wealth and power through their control over digital platforms and infrastructure. With access to top talent, data, and compute power, the current surge of generative AI has acted as a catalyst, further increasing and consolidating their dominance in various domains. Khanal et al. (2025) explore how Big Tech's dominance in the context of generative AI is increasing and reshaping the policymaking process. To examine this overarching issue, they explore how the nature of the relationship between Big Tech and governments is changing, how Big Tech uses its resources to influence the policy landscape, and how the emergence of generative AI is accelerating this process.

Khanal et al. (2025) use and extend Kingdon's multiple streams framework (Kingdon, 1984) to study why and how the power of Big Tech is increasing in the policy process. By introducing the "Big Tech-centric technology stream," they highlight how Big Tech companies exert influence on all streams and distinguish it from the traditional innovation-centric technology stream, which focuses on technological advancement for the betterment of society (Goyal & Howlett, 2018). The authors posit that, in this new technology stream, the primary objective of Big Tech is maintaining and increasing dominance by shaping the political and regulatory environment, and the expansion and acceleration of the diffusion of technologies that serve its interests, even if it results in stifling innovation, highlighting that when Big Tech gets involved, the technology stream moves away from an innovation-centric model to the one dominated by the self-interest of Big Tech (Khanal et al., 2025). They illustrate that Big Tech, in the problem stream, impacts the epistemic community through their control of the platforms, provision of research funding, and using the media to frame the problems in a favorable way. In the policy stream, they impact instrument constituencies through the provision of digital solutions and influence

**Table 5.** Differences between the traditional approach of technology companies and Big Tech [based on Khanal et al. (2025)].

| | Innovation-centric approach | Big Tech-centric approach |
|---|---|---|
| Actors | Traditional innovators and technology companies | Big Tech companies |
| Focus | Focused on the promotion of innovation in their domain | Primarily use technology to expand power and market dominance, serving self-interests, even at the expense of stifling innovation |
| Characteristics | Confined to the technology stream<br>Limited involvement in policy outside of the technology domain<br>Aim of advancing technology for societal benefits | Not confined to the technology stream<br>Exerting significant influence in all streams and stages of the policy process<br>Aim to align and reshape the policy process with their interests and expand their dominance |

Source: Author, based on Khanal et al. (2025).

policy choices by promoting technological solutionism (Morozov, 2013). In the politics stream, Big Tech—through lobbying, political donations, and revolving doors—exerts a significant influence over political decisions, despite the growing concerns on the part of the public over privacy, competition, and fake news. Table 5 highlights the differences between the traditional approach of the technology companies and Big Tech.

The authors introduce the concept of "super policy entrepreneurs" for Big Tech, after demonstrating that they are now operating in all the streams and stages of the policy process. This calls for re-evaluation of policy theories and governance frameworks to account for the growing role of Big Tech in the policy process and necessitates that scholars, policymakers, and society at large critically examine Big Tech's expanding role and ability to influence agendas, formulate solutions, and shape political discourse. The authors conclude by highlighting the need for stronger governance mechanisms to ensure that democratic processes and societal well-being are not undermined by the growing influence of Big Tech.

## Governance fix? Power and politics in controversies about governing generative AI (Ulnicane, 2025)

Polycentric governance (Carlisle & Gruby, 2019; Ostrom, 2017) is a governance system in which multiple independent, overlapping, and autonomous centers of decision-making operate. Ulnicane (2025) suggests that the current governance of generative AI exhibits such polycentricity as organizations such as the organisation for economic co-operation and development (OECD), G7, EU, and the UK government showcase cooperation, interdependence, and competition in efforts regarding the governance of AI. The paper examines the current policy proposals put forward on the governance of generative AI and the current framings and controversies surrounding it.

Ulnicane (2025) examines the efforts in the G7 Hiroshima process, OECD reports, and the UK AI Safety Summit and the links of these initiatives to other organizations. She highlights that these efforts are predominantly led by developed countries and raises concerns over issues of inclusion, equality, and representation from less developed countries. She argues that the dominant framing around risk management—particularly existential risks—has diverted attention from considerations of how generative AI can be directed to address societal needs.

The author highlights the current technocratic approach to generative AI governance and the concomitant limited public participation, drawing attention to the paradox that it is widely used by the public. This suggests that while the use of generative AI is democratized, its governance is imbalanced and dominated by a handful of developed nations and Big Tech (Khanal et al., 2025).

Inspired by the concept of "technological fix," whereby technical solutions are applied to societal problems without addressing their underlying issues, Ulnicane (2025) conceptualizes the current approach to generative AI governance as a "governance fix." A governance fix involves the application of a technocratic and expert-driven governance approach to manage risks with limited deliberation and

**Table 6.** Generative AI governance approaches [based on Ulnicane (2025)].

| | Current approach | Proposed approach |
|---|---|---|
| Governance | Technocratic governance fix focused on risk management | Democratic and responsible innovation for addressing societal challenges |
| Participation | Led by experts with limited public participation | Broad and inclusive participation with active involvement of the public and diverse stakeholders |

Source: Author, based on Ulnicane (2025).

democratic engagement. The author advocates for responsible innovation and fostering democratic participation and stakeholder engagement in the co-production of governance strategies, embracing the political elements in polycentric governance to better address societal challenges and ensure that AI developments are aligned with societal values. Table 6 summarizes the current and proposed generative AI governance approaches as discussed by Ulnicane (2025).

## When code isn't law: rethinking regulation for artificial intelligence (Judge et al., 2025)

Lessig (1999) popularized the phrase "code is law," suggesting that the rules embedded in software systems (code) can govern behavior similar to how laws govern behavior. However, Judge et al. (2025) argue that in generative AI, this principle is no longer valid as LLMs and deep learning systems exhibit emergent behaviors that make their operations opaque and unpredictable. They examine the Federal Aviation Administration (FAA) and Nuclear Regulatory Commission (NRC) regulatory frameworks in the US, highlighting that using standards and audit mechanisms, it can be tested whether engineering systems comply with specifications in their design and code in these domains. However, they argue that current AI systems cannot be similarly regulated, since their behavior emerges from training data and inference rather than explicit programming.

Judge et al. (2025) posit that the unpredictability of these AI systems raises significant challenges in the domains of safety, accountability, and alignment with human values, which can ultimately result in increased inequality, job loss, or even loss of human control over superintelligent systems, This is further exacerbated by the current AI arms race (Zhang et al., 2024). This necessitates alignment (Gabriel, 2020) and control of AI systems (Firlej & Taeihagh, 2021; Russell, 2019).

The authors question the use of voluntary guidelines for ensuring AI safety and highlight that by drawing lessons from established regulatory practices such as in the FAA and NRC and integrating AI safety research, novel regulatory frameworks for AI can be developed that can address AI risks and foster responsible innovation. The authors propose establishing a specialized regulatory body for overseeing AI development throughout its lifecycle, with "teeth" that can enforce regulations, respond swiftly, and conduct audits as need be. It should have the ability to monitor AI systems in real time and recall or deactivate the systems if necessary. They highlight the need for use of formal mathematical verification methods to ensure appropriate AI behavior and move toward provable safety guarantees. Table 7 highlights the challenges of regulating current generative AI systems.

## When AIs become oracles: Generative artificial intelligence, anticipatory urban governance, and the future of cities (Cugurullo & Xu, 2025)

City brains are large-scale AI systems that have been introduced to urban platforms to aid in managing various aspects of cities, such as transport, public safety, and environmental monitoring (Caprotti & Liu, 2022; Cugurullo, 2021). This paper examines how generative AI and the use of LLMs as part of city brains influence anticipatory governance in urban policy. The authors critically examine the use of generative AI in the management of urban services and use of its predictive capabilities. They study the risks and challenges associated with the use of such black box and opaque systems, focusing on transparency, accountability, and the marginalization of stakeholders through the reduction of their agency.

Cugurullo and Xu (2025) examine the use of generative AI in a technocratic anticipatory urban governance setting by studying the implementation of a city brain project in Haidian, a district in Beijing, China. Haidian collects data extensively, including over 14000 closed-circuit television and 20000 sensors, which is then fed to AI systems including an LLM called WuDao that is used for predicting urban

**Table 7.** Challenges of regulating generative AI systems based on Judge et al. (2025).

| | Traditional engineered systems | Generative AI systems |
|---|---|---|
| Characteristics | Often deterministic based on explicit code and design | Emergent behavior from complex training processes on vast datasets |
| | Transparent and auditable components | Opaque black box models |
| | Verifiable, predictable, traceable, and thus correctable | Difficult to verify, unpredictable, untraceable |
| Regulatory challenges and approach | Code is law | Code is not law: emergent behaviors make traditional regulation inadequate |
| | Compliance ensured through design specifications, audits, and testing (e.g., FAA, NRC) | Need new regulatory frameworks drawing lessons from regulating traditional engineered systems while incorporating AI safety research |
| | Traceability allows for correcting sources of failure | Rules cannot be encoded directly, and misbehaviours cannot be traced easily |
| | Established practices for ensuring safety and reliability exist | Proposed consolidated oversight of AI lifecycle by a single authority, formal verification, independent monitoring with ability to deactivate or recall systems, and establishment of red lines |

Source: Author, based on Judge et al. (2025).

issues (e.g., environmental risks, congestion, health, security, etc.), which are used by policymakers to take pre-emptive actions (Cugurullo & Xu, 2025).

While proponents of city brains highlight their potential for increasing efficiency and effectiveness in urban management, the authors point out challenges resulting from the opacity of the algorithms (at times even to the developers) and lack of transparency that complicate the validation of the predictions of the system for use in policymaking. They highlight varying accuracy rates of the system between 60% and 90%, raising concerns about the reliability of the policies based on these results. The authors posit that the technocratic nature of this anticipatory approach restricts public engagement and citizen participation in governance and that the algorithmic predictions can prioritize certain futures serving the interests of powerful stakeholders rather than the public.

The authors explore the epistemological foundations of technocratic anticipatory governance, emphasizing its focus on rationality, interest in avoiding future shocks, and deterministic nature. They argue that these epistemological foundations result in the integration of AI in anticipatory governance as a means of generating calculable predictions, which can lead to "posthuman governance," where human agency is diminished (Cugurullo et al., 2023).

They highlight the need to move from technocratic anticipatory governance to participatory models of anticipatory urban governance for achieving transparency and inclusivity. These, in turn, will serve the public and uphold democratic principles. They advocate a human-centered approach to urban governance and the use of strategies that promote public engagement and involve citizens in developing plausible futures rather than relying only on AI predictions. Table 8 compares technocratic anticipatory governance with participatory anticipatory governance.

## Artificial intelligence, emotional labor, and the quest for sociological and political imagination among low-skilled workers (Oder & Béland, 2025)

Oder and Béland, 2025 focus on the impact of generative AI on labor markets. Through a single in-depth case study, they examine the effects of generative AI on the daily routines of low-skilled workers in call centers in Vienna, Austria. The article examines how generative AI affects the nature of the emotional labor required by the call center workers and their perception regarding the disruption caused by generative AI on their work. The study also explores their coping mechanisms and their social and political imaginaries.

Through interviews and qualitative research, the authors identify that call center workers perform significant emotional labor to meet organizational expectations when interacting with consumers and

**Table 8.** Technocratic vs. participatory anticipatory governance [based on Cugurullo & Xu [2025]].

| | Technocratic anticipatory governance | Participatory anticipatory governance |
|---|---|---|
| Governance | Top-down and technocratic | Bottom-up involving multiple stakeholders |
| Role of technology | Technology is a tool to calculate and control future events | Technology facilitates public engagement and supports foresight |
| Episte-mological basis | The future is deterministic and calculable | Multiple plausible futures exist that should be explored collectively |
| Citizen engagement | Limited | Broad |

Source: Author, based on Cugurullo and Xu (2025).

the introduction of chatbots and advanced interactive voice response (IVR) systems has intensified their emotional labor. This is because these tools often fail to resolve consumer issues effectively and the call center workers have to deal with angry and/frustrated consumers who have been transferred to them, which increases call center workers' stress and requires them to manage their emotions even more effectively.

Through the case study, Oder and Béland (2025) establish that only a minority of employees recognized the long-term impact of AI on their jobs and a majority were preoccupied with the immediate challenges of the work and the frustration/anger of consumers caused by the inefficiency of the AI and IVR systems The authors highlight that none of the employees recognized the political dimension of the introduction of AI and AI-driven change and did not consider collective action against it. Thus, they suggest that the increased emotional labor of the workers in the call centers is linked to their low levels of sociological and political imagination due to the intensification of the demands of their jobs via the introduction of flawed AI and IVR systems.

To address these issues, Oder and Béland (2025) recommend (a) fostering awareness among workers through union and government programs, (b) inclusive policy development that engages stakeholders to ensure that their concerns are heard and addressed, (c) developing legal frameworks that ensure that workers have co-determination rights when AI systems are introduced in the workplace, (d) developing reskilling and upskilling programs that consider the specific needs and constraints of workers, and (e) not individualizing the problem and providing state-funded support.

## Acknowledgements

The special issue editor is grateful for the support provided by the Lee Kuan Yew School of Public Policy, National University of Singapore (NUS), the Centre for Trusted Internet and Community (CTIC) at the NUS, and NUS Deputy President Professor Chen Tsuhan. The special issue editor would like to thank the editors of *Policy and Society* and all the anonymous reviewers for their constructive feedback and support for this and other articles in the special issue on the governance of generative AI. Special thanks are extended to Shaleen Khanal and Hongzhou Zhang, and the events team and research office at the Lee Kuan Yew School of Public Policy, for their support in various aspects of organising this special issue and the accompanying workshop held on 6–7 October 2023.

## Funding

This research/project was supported by the Ministry of Education, Singapore, under its MOE AcRF TIER 3 Grant (Award Number: MOE-MOET32022-0001), and the National Research Foundation Singapore under its AI Singapore Programme (Award Number: AISG3-GV-2021-002). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore, or the National Research Foundation, Singapore.

## Conflict of interest

None declared.

## Declaration on the use of generative AI

During the preparation of this work, the author used generative AI to improve the writing of this document, after which the author reviewed and further edited the content as needed.

## References


Abbas, F., & Taeihagh, A. (2024). Unmasking Deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. *Expert Systems with Applications*, 252(Part B), 124260. https://doi.org/10.1016/j.eswa.2024.124260

Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., & Sharifi, M. (2023). MusicLm: Generating music from text. *arXiv:2301.11325*.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., and Chatila, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58(June 2020), 82–115. 10.1016/j.inffus.2019.12.012.

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, Association for Computing Machinery, NY, USA, pp. 610–623.

Caprotti, F., & Liu, D. (2022). Platform urbanism and the Chinese smart city: the co-production and territorialisation of Hangzhou City Brain. *GeoJournal*, 87(3), 1559–1573. https://doi.org/10.1007/s10708-020-10320-2

Carlisle, K., & Gruby, R. L. (2019). Polycentric systems of governance: A theoretical model for the commons. *Policy Studies Journal*, 47(4), 927–952. https://doi.org/10.1111/psj.12212

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., & Ye, W. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 1–45. https://doi.org/10.1145/3641289

Cheng, X., Li, J., Yang, S., Yang, G., & Wang, X. (2024). Open-television: Teleoperation with immersive active visual feedback. arXiv:2407.01512.

Chesterman, S. (2025). Good models borrow, great models steal: intellectual property rights and generative AI. *Policy and Society*, 44(1), 23–37. https://doi.org/10.1093/polsoc/puae006

Copyright Act (2021). Singapore.

Copyright, Designs and Patents Act (1988). United Kingdom.

Coulter, M. (2024). Aiming for fairness: an exploration into getty images v. stability ai and its importance in the landscape of modern copyright law. *DePaul Journal of Art, Technology and Intellectual Property Law*, 34(1), 124–142.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202

Cronin, I. (2024). Autonomous AI agents: Decision-making, data, and algorithms. In *Understanding Generative AI Business Applications*. Apress.

Cugurullo, F. (2021). *Frankenstein Urbanism: Eco, Smart and Autonomous Cities, Artificial Intelligence and the End of the City*. Routledge.

Cugurullo, F., Caprotti, F., Cook, M., Karvonen, A., McGuirk, P., & Marvin, S. (2023). *Artificial Intelligence and the City: Urbanistic Perspectives on AI*. Routledge.

Cugurullo, F., & Xu, Y. (2025). When AIs become oracles: generative artificial intelligence, anticipatory urban governance, and the future of cities. *Policy and Society*, 44(1), 98–115. https://doi.org/10.1093/polsoc/puae025

Davidson, J. (2023). Google Brain founder says big tech is lying about AI extinction danger. Financial Review. https://www.afr.com/technology/google-brain-founder-says-big-tech-is-lying-about-ai-human-extinction-danger-20231027-p5efnz

Decardi-Nelson, B., Alshehri, A. S., Ajagekar, A., and You, F. (2024). Generative AI and process systems engineering: The next frontier. *Computers and Chemical Engineering* 187, 108723. 10.1016/j.compchemeng.2024.108723

Ding, R., Qin, Y., Zhu, J., Jia, C., Yang, S., Yang, R., Qi, X., & Wang, X. (2024). Bunny-visionpro: Real-time bimanual dexterous teleoperation for imitation learning. *arXiv:2407.03162*.

Dong, T., Xue, M., Chen, G., Holland, R., Meng, Y., Li, S., Liu, Z., & Zhu, H. (2025). *The Philosopher's Stone: Trojaning Plugins of Large Language Models*. In the 32nd Annual Network and Distributed System Security Symposium (NDSS), 24–28 February 2025 in San Diego, California.

Egan, L., & Kine, P.. (2024). Biden's final meeting with Xi Jinping reaps agreement on AI and nukes, Politico. https://www.politico.com/news/2024/11/16/biden-xi-jinping-ai-00190025 (Accessed November 2024).

Eiras, F., Petrov, A., Vidgen, B., Schroeder, C., Pizzati, F., Elkins, K., Mukhopadhyay, S., Bibi, A., Purewal, A., Botos, C., & Steibel, F. (2024). Risks and opportunities of open-source generative AI. arXiv:2405.08597.

EU Directive. (2019). Directive (EU). 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC 2019 (EU).

Fauzia, M. (2021). Fact check: Facebook didn't pull the plug on two chatbots because they created a language. USA Today. https://www.usatoday.com/story/news/factcheck/2021/07/28/fact-check-facebook-chatbots-werent-shut-down-creating-language/8040006002/ (accessed November 2024)

Firlej, M., & Taeihagh, A. (2021). Regulating human control over autonomous systems. *Regulation & Governance*, 15(4), 1071–1091. https://doi.org/10.1111/rego.12344

Frey, C. B., & Osborne, M. (2023). Generative AI and the future of work: a reappraisal. *Brown Journal of World Affairs*, 30, 161.

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, 30(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Gerken, T. (2023). New York Times sues Microsoft and OpenAI for 'billions'. BBC. https://www.bbc.com/news/technology-67826601 (accessed October 2024)

Giles, C. (2024). Universal basic income: the bad idea that never quite dies. Financial Times. https://www.ft.com/content/27057ff2-e9b6-4630-a6ea-201e0f6d72d9 (Accessed November 2024).

Goertzel, B. (2023). Generative ai vs. agi: The cognitive strengths and weaknesses of modern llms. *arXiv:2309.10371*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Goyal, N., & Howlett, M. (2018). Technology and instrument constituencies as agents of innovation: Sustainability transitions and the fovernance of urban transport. *Energies*, *11*(5), 1198. https://doi.org/10.3390/en11051198

Guihot, M., Matthew, A. F., & Suzor, N. P.. (2017). Nudging robots: Innovative solutions to regulate artificial intelligence. *Vanderbilt Journal of Entertainment and Technology Law*, *20*, 385.

Hiilamo, H. (2020). The basic income experiment in Finland yields surprising results. The University of Helsinki. https://www.helsinki.fi/en/news/fair-society/basic-income-experiment-finland-yields-surprising-results (Accessed November 2024).

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, *33*, 6840–6851.

Holland, J. H.. (2006). Studying complex adaptive systems. *Journal of Systems Science and Complexity*, *19*(1), 1–8. https://doi.org/10.1007/s11424-006-0001-z

Jabbour, J., & Reddi, V. J. (2024). Generative AI agents in autonomous machines: A safety perspective. arXiv: 2410.15489.

Jaidka, K., Chen, T., Chesterman, S., Hsu, W., Kan, M. Y., Kankanhalli, M., Lee, M. L., Seres, G., Sim, T., Taeihagh, A., Tung, A., Xiao, X., & Yue, A. (2024). Misinformation, disinformation, and generative AI: Implications for perception and policy. In *Digital Government: Research and Practice*. Association for Computing Machinery.

Janssen, M. (2025). Responsible governance of generative AI. Conceptualizing AI governance as a complex adaptive systems. *Policy and Society*, 44(1), 38–51. https://doi.org/10.1093/polsoc/puae040

Judge, B., Nitzberg, M., & Russell, S. (2025). When code isn't law: rethinking regulation for artificial intelligence. *Policy and Society*, 44(1), 85–97. https://doi.org/10.1093/polsoc/puae020

Khanal, S., Zhang, H., & Taeihagh, A. (2025). Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. *Policy and Society*, 44(1), 52–69. https://doi.org/10.1093/polsoc/puae012

Kingdon, J. (1984). *Agendas, Alternatives, and Public Policies*. Pearson.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv:1312.6114

Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, *12*(4), 307–392. http://dx.doi.org/10.1561/2200000056

Koff, D. (2023). Why Facebook shut down its AI, Bultin.com https://builtin.com/artificial-intelligence/facebook-shuts-down-ai (accessed November 2024)

Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*, *7*(3), 437–451. 10.1017/als.2020.19

Lessig, L. (1999). *Code and Other Laws of Cyberspace*. NY.

Li, Y., Taeihagh, A., de Jong, M., & Klinke, A. (2021). Toward a commonly shared public policy perspective for analyzing risk coping strategies. *Risk Analysis*, *41*(3), 519–532. https://doi.org/10.1111/risa.13505

Lim, H. S. M., & Taeihagh, A. (2019). Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities. *Sustainability*, *11*(20), 5791. https://doi.org/10.3390/su11205791

Mavikumbure, H. S., Cobilean, V., Wickramasinghe, C. S., Drake, D., & Manic, M. (2024). *Generative AI in cyber security of cyber physical systems: Benefits and threats*. In 16th International Conference on Human System Interaction (HSI) (pp. 1–8). Paris, France, IEEE.

Miller, S., Rhodes, E., Bartik, A. W., Broockman, D. E., Krause, P. K., & Vivalt, E. (2024). Does income affect health? Evidence from a randomized controlled trial of a guaranteed income. No. w32711. National Bureau of Economic Research.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

Morozov, E. (2013). *To Save Everything, Click Here: The folly of technological solutionism*. Public Affairs.

Mostaque, E. (2022). Stable diffusion public release. *Stability AI*. https://stability.ai/news/stable-diffusion-public-release (accessed October 2024)

Ng, L. H., Lim, A. C., Lim, A. X., and Taeihagh, A. (2023). Digital ethics for biometric applications in a smart city. *Digital Government: Research and Practice*, *4*(4), 1–6 10.1145/3630261.

Ng, L. H. X., & Taeihagh, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy and Internet*, *13*(560), 585. https://doi.org/10.1002/poi3.268

Niu, Z., Ren, H., Gao, X., Hua, G., & Jin, R. (2024). Jailbreaking attack against multimodal large language model. arXiv:2402.02309.

Northcutt, C. G., Athalye, A., & Mueller, J. (2021). Pervasive label errors in test sets destabilize machine learning benchmarks" arXiv:2103.14749.

Oder, N., & Béland, D. (2025). Artificial intelligence, emotional labor, and the quest for sociological and political imagination among low-skilled workers. *Policy and Society*.

Ostrom, E. (2017). Polycentric systems for coping with collective action and global environmental change. *Global Justice* (pp. 423–430). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315254210-18/polycentric-systems-coping-collective-action-global-environmental-change-elinor-ostrom

Pande, D., & Taeihagh, A. (2023). Navigating the governance challenges of disruptive technologies: insights from regulation of autonomous systems in Singapore. *Journal of Economic Policy Reform*, *26*(3), 298–319. https://doi.org/10.1080/17487870.2023.2197599

Pasupuleti, R., Vadapalli, R., & Mader, C. (2023). *Cyber Security Issues and Challenges Related to Generative AI and ChatGPT*. In Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–5). Abu Dhabi, United Arab Emirates, IEEE.

Peng, B., Bi, Z., Niu, Q., Liu, M., Feng, P., Wang, T., Yan, L. K., Wen, Y., Zhang, Y., & Yin, C. H. (2024). Jailbreaking and mitigation of vulnerabilities in large language models. arXiv:2410.15236.

Philipsen, S., Stamhuis, E. F., & de Jong, M. (2021). Legal enclaves as a test environment for innovative products: Toward legally resilient experimentation policies 1. *Regulation & governance*, *15*(4), 1128–1143.

Plate, H., Ponta, S. E., & Sabetta, A. (2015). *September. Impact assessment for vulnerabilities in open-source software libraries*. In 2015 IEEE International Conference on Software Maintenance and Evolution (pp. 411–420). IEEE.

Radford, A. (2018). Improving language understanding by generative pre-training. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed October 2024)

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., & Krueger, G. (2021). *Learning transferable visual models from natural language supervision*. In International conference on machine learning (pp. 8748–8763). PMLR.

Radu, R. (2021). Steering the governance of artificial intelligence: national strategies in perspective. *Policy and Society*, 40(2), 178–193. https://doi.org/10.1080/14494035.2021.1929728

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., Ahmad, J., Ali, M. E., & Azam, S., (2024). *A review on large Language Models: Architectures, applications, taxonomies, open issues and challenges*. In IEEE Access.

Reuel, A., & Undheim, T. A. (2024). Generative AI needs adaptive governance. arXiv:2406.04554.

Rolf, E., Worledge, T. T., Recht, B., & Jordan, M. (2021). *Representation matters: Assessing the importance of subgroup allocations in training data*. In International Conference on Machine Learning (pp. 9040–9051). PMLR.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.

Russell, S. (2023). Written testimony. Hearing before the Committee on the Judiciary, United States Senate. https://www.judiciary.senate.gov/download/2023-07-26-testimony-russell

Sabin, S. (2024). Anthropic, feds test whether AI will share sensitive nuke info, Axios. https://www.axios.com/2024/11/14/anthropic-claude-nuclear-information-safety (accessed November 2024)

Sap, M., Swayamdipta, S., Vianna, L., Zhou, X., Choi, Y., & Smith, N. A. (2021). Annotators with attitudes: How annotator beliefs and identities bias toxic language detection. arXiv:2111.07997.

Sconiers-Hasan, M. (2024). Application Programming Interface (API) Vulnerabilities and Risks, Software Engineering Institute, Carnegie Mellon University. CMU/SEI-2024-SR-004. https://kilthub.cmu.edu/ndownloader/files/47218084 (accessed November 2024)

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot N, Anderson R, Gal Y (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

Taeihagh, A. (2023). Addressing policy challenges of disruptive technologies. *Journal of Economic Policy Reform*, 26(3), 239–249. https://doi.org/10.1080/17487870.2023.2238867

Taeihagh, A., Ramesh, M., & Howlett, M. (2021). Assessing the regulatory challenges of emerging disruptive technologies. *Regulation & Governance*, 15, 1009–1019. https://doi.org/10.1111/rego.12392

Talgo, C. (2024). Universal basic income is a moral hazard. Newsweek https://www.newsweek.com/universal-basic-income-moral-hazard-opinion-1863775 (Accessed November 2024).

Tan, S., Joty, S., Baxter, K., Taeihagh, A., Bennett, G. A., & Kan, M. Y. (2021). Reliability Testing for Natural Language Processing Systems. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 1, 4153–4169.

Tan, S., Taeihagh, A., & Baxter, K. (2022). The risks of machine learning systems. arXiv:2204.09852.

Tan, S. Y., & Taeihagh, A. (2021). Adaptive governance of autonomous vehicles: Accelerating the adoption of disruptive technologies in Singapore. *Government Information Quarterly*, 38(2), 101546. https://doi.org/10.1016/j.giq.2020.101546

Ulnicane, I.. (2025). Governance fix? Power and politics in controversies about governing generative AI. *Policy and Society*, 44(1), 70–84. https://doi.org/10.1093/polsoc/puae022

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

Vivalt, E., Rhodes, E., Bartik, A. W., Broockman, D. E., & Miller, S. (2024). The employment effects of a guaranteed income: Experimental evidence from two US states. No. w32719. National Bureau of Economic Research.

Weiser, B. (2023). ChatGPT lawyers are ordered to consider seeking forgiveness, The New York Times. https://www.nytimes.com/2023/06/22/nyregion/lawyers-chatgpt-schwartz-loduca.html (Accessed November 2024).

Xu, C., Guan, S., Greene, D., & Kechadi, M. (2024). Benchmark data contamination of large language models: A survey. arXiv:2406.04244.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M. H. (2024). Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4), 105. https://doi.org/10.1145/3626235

Zaidan, E., & Ibrahim, I. A. (2024). AI governance in a complex and rapidly changing regulatory landscape: A global perspective. *Humanities and Social Sciences Communications*, *11*, 1121. https://doi.org/10.1057/s41599-024-03560-x

Zhang, H., Khanal, S., & Taeihagh, A. (2024). Public-private powerplays in generative AI era: Balancing big tech regulation amidst global AI race. *Digital Government: Research and Practice*. https://doi.org/10.1145/3664824