



AI ASIA PACIFIC  
INSTITUTE

# Policy Brief: ChatGPT and Other Generative AI Systems





## About the AI ASIA PACIFIC INSTITUTE

The AI Asia Pacific Institute focuses on addressing the social, legal, and ethical risks associated with artificial intelligence in order to unlock its potential for creating a sustainable world.



## Acknowledgements

We are grateful for the insightful expertise provided by Dr. Stuart Russell, Dr. Toby Walsh, Dr. Pedro Domingos, and Dr. Luciano Floridi.

Thank you to Joseph Negrine (intern of the AI Asia Pacific Institute) for his significant contribution in drafting this document and to Dr. Wendy Bonython (Associate Professor of Law at Bond University) for her valuable contribution to the editing process.

### Copyright

© 2023 AI Asia Pacific Institute. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

### Important disclaimer

The reader is advised and needs to be aware that the information contained in this publication may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, the AI Asia Pacific Institute (including its members and advisors) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

After engaging in a series of discussions with experts in the field of artificial intelligence ('AI'), the AI Asia Pacific Institute previously disseminated a briefing paper that delves into the risks and prospects associated with ChatGPT. Building upon the foundation laid by the briefing paper, this Policy Brief explores nascent methodologies applicable to the regulation of generative AI systems. Generative AI systems are models that *generate* new output—whether text, audio, or visual material—based on data they have been trained on. ChatGPT is one example of a generative AI system. Other examples include Alibaba's 'Tongyi Qianwen', Google's 'Bard', and Microsoft's 'Vall-E'.

## An Overview of the Applicable Regulations

This part outlines current and proposed regulatory frameworks governing generative AI systems in China, the European Union ('EU') and the United States ('US'), recognising their pioneering role in these advancements.

### China

China is a central figure in the development of AI, and is expected to play a leading role in shaping the contours of the 'Fourth Industrial Revolution'. China has adopted a top-down approach to AI regulation, characterized by strong government intervention and control. The Chinese government has issued comprehensive national strategies and plans for AI development, outlining specific goals, targets, and policy frameworks. In contrast, Western countries generally follow a more decentralised approach featuring a combination of government regulations, industry self-regulation, and collaborative initiatives. On a brief comparative analysis, the regulatory developments emerging from China place a greater emphasis on social responsibility and group and community relations, with relatively less focus on individualistic rights. The regulatory framework includes the 'Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)' (*'Draft Measures for Generative AI Services'*); *Provisions on the Administration of Deep Synthesis of Internet Information Services* (entered into force on 10 January 2023); and the *Personal Information Protection Law* (entered into force on 1 November 2021). An official translation of the *Draft Measures for Generative AI Services* is not available at the time of writing. Therefore, this Policy Brief relies on Stanford University's [April 2023 translation](#). Crucially, a holistic analysis of China's high-level regulatory developments requires policymakers to assess the structural, cultural, and political context that shapes its approach to and development of AI. As is the case with all jurisdictions, a nuanced understanding of the domestic and international interests is paramount for policymakers.

### The European Union

The EU's regulatory framework includes the General Data Protection Regulation ('GDPR') and the [proposed 'AI Act'](#) (*'Draft EU AI Act'*). Under the [16 May 2023 compromise amendments](#) to the *Draft EU AI Act*, generative AI systems will be considered 'foundation models', as they are 'designed to optimize for generality and versatility of output' and 'trained on a broad range of data sources and large amounts of data' (para 60(e)). The explanatory paragraph (60g) states that this is intended to subject generative AI systems to specific requirements that distinguish such systems from 'high-risk' AI systems. Accordingly, article 28b requires providers of foundation models to ensure their systems comply with EU regulations prior to becoming available on the market by documenting risk assessment and mitigation; incorporating appropriate data sources and examining for biases. The *Draft EU AI Act* has also received criticism from the AI ecosystem on the argument that it imposes excessive regulatory burdens and stifles innovation, contributing to Europe's further disadvantage in AI. The *Draft AI Act* is still undergoing review and revision, and it is expected that the final version will address some of the concerns raised during the consultation process.

## The United States

The US does not have comprehensive federal regulations specifically focused on AI. However, various initiatives and efforts are underway to address AI-related concerns. This mainly consists of Agency Guidance: federal agencies such as the Federal Trade Commission ('FTC'), the National Institute of Standards and Technology ('NIST'), and the Food and Drug Administration ('FDA') have all issued guidance and recommendations related to AI. For instance, the FTC provides guidelines on consumer protection and privacy concerns associated with AI technologies, while NIST offers technical standards and best practices to promote trustworthy AI development and deployment and has proposed the [AI Risk Management Framework](#), which is intended to be voluntary. Supplementing the Agency Guidance is the [Blueprint for an AI Bill of Rights \(2022\)](#) (the 'Blueprint'). The Blueprint—which is not binding—is intended to support the development of policies and practices to protect the civil rights and promote democratic values in the building, deployment, and governance of automated systems.

US regulation is lagging relative to China and the EU, with no actual or proposed binding federal legislation to date. There have been ongoing discussions and legislative proposals in the U.S. Congress regarding AI regulation. These efforts seek to establish federal laws that address a wide range of AI-related issues, including privacy, bias, transparency, and accountability. Several bills have been introduced, but as of now, no comprehensive federal AI legislation has been enacted. While this absence of federal legislation may be thought to encourage innovation, an alternative possibility is that clarifying the regulatory framework—or lack thereof—can increase investment in AI due to greater legal certainty.

## Regulating Risks Posed by Generative AI Systems

Of paramount importance is the proper categorisation of the risks posed by generative AI systems, distinguishing between the short-term and long-term risks. This Policy Brief focuses on the short-term risks outlined in the [Briefing Paper](#), which are (i) impersonation and disinformation; (ii) privacy and security; (iii) bias and discrimination; and (iv) intellectual property infringement. The regulatory strategies used, or proposed, to mitigate these risks are discussed in turn.

### 1 Impersonation and Disinformation

Generative AI systems have the capacity to mimic individuals, resulting in the proliferation of advanced disinformation and fraudulent activities. This section considers two epistemic threats. The first threat concerns disinformation created by malicious agents, like phishing messages.<sup>1</sup> The second threat concerns factual failures and reasoning errors generated by AI systems, referred to as 'hallucinations'.<sup>2</sup> These threats can be thought of as human-made disinformation and AI-generated misinformation, respectively.

In regard to human-made disinformation, regulations can be imposed on both the providers and users of generative AI systems. The EU's [2022 Strengthened Code of Practice on Disinformation](#) provides a set of guidelines for signatories to better self-regulate. Commitment 15 provides that signatories can mitigate human-made disinformation by warning users of the AI systems and proactively detecting such content. This standard mirrors article 52(1) of the *Draft EU AI Act*, which provides that, unless it is 'obvious from the circumstances and the context of use', providers must inform natural persons they are interacting with their AI systems. To prevent the distribution of manipulative content, article 52(3) provides that users of AI systems who create deepfakes<sup>3</sup> must disclose that they have been artificially generated. A [report by the European Parliamentary Research Service](#) notes that this labelling obligation 'could be a first step towards mitigating potential negative impacts', but is insufficient to address other issues. The *Draft EU AI Act* neither contains guidelines for disclosure nor includes sanctions for non-compliance (article 71). Further, it is unclear how actors sharing deepfakes anonymously would be held accountable.

China's '[Provisions on the Administration of Deep Synthesis of Internet-based Information Services](#)', which came into force on 10 January 2023, are instructive here.<sup>4</sup> These provisions prohibit persons from using generative AI systems to produce or disseminate false information (article 6); require the authentication of persons using deepfake technologies (article 9); and establish mechanisms for refuting rumours created by such systems (article 11).

Policymakers should consider whether different sociopolitical conditions would affect the feasibility of these regulations within their own countries. Because Chinese internet users have a greater 'digital fingerprint', as they are often required to link their online accounts to their [government-ID-linked phone numbers](#), malicious agents may struggle to conceal their identities online. There may be competing policy objectives, as suggested by measure 15.2 of the 2022 Strengthened Code of Practice on Disinformation, which notes that policies to detect and sanction the impermissible use of deepfake technologies must be trustworthy and 'respect the rights of end-users'.


With respect to AI-generated misinformation, regulators ought to maintain a balanced approach. In a [blog post concerning GPT-4](#), OpenAI acknowledged that 'new risk surfaces' will emerge as models become more powerful and their fields of content expand. It is conceivable that the complete eradication of hallucinations may be an unattainable objective. Greater attention is required in the EU to address this issue. As has been [observed elsewhere](#), the original proposal of the *Draft EU AI Act* did not foresee the proliferation of generative AI systems. The *GDPR* addresses this gap to a limited extent. In circumstances where the hallucination provides false information about a particular individual, a claim may arise under article 5 of the *GDPR*, which provides that personal data must be accurate, and reasonable steps must be taken to rectify or erase inaccurate data. In addition to the aforementioned scenario, the absence of provisions governing hallucinations presents a potential peril whereby individuals may excessively depend on systems without engaging in discerning assessment of their outputs.

It must be noted that the providers of AI systems are not *entirely* blameworthy for hallucinations in all circumstances. Users may unintentionally pose a question in a manner that causes the system to 'untether' from factual training data. For example, when asked to provide 'at least five examples, together with quotes from relevant newspaper articles' of sexual harassment by American law professors, ChatGPT falsely listed the (real) George Washington University Law School professor, [Jonathan Turley](#). The response (falsely) claimed that Turley was a member of Georgetown University Law Center, citing a [\(non-existent\) Washington Post article](#) that supposedly reported the harassment occurred during a class trip to Alaska (which Turley clarified had never happened before).

As indicated within the recommendations section, regulatory measures aimed at mitigating the incidence and associated hazards of hallucinations may necessitate comprehensive coverage of both the development and utilization of generative AI systems.

## 2 Privacy and Security

Models trained on personal data can generate highly realistic and identifiable information, creating risks for privacy and security. This concern transcends the realm of individual privacy and security, encompassing broader considerations. In April 2023, [Samsung banned their employees from using ChatGPT](#) following concerns that internal sensitive code that had been uploaded could be provided to other users.



The *GDPR* is more relevant to privacy and security issues than the *Draft EU AI Act*. The *GDPR* requires consent from individuals before collecting personal data (article 7); provides individuals with the rights to access their personal data (article 15) and delete it (article 17); and contains measures to protect personal data from unauthorised access, use, or disclosure (article 21). Since [March 2023](#), OpenAI and the ‘Garante’—Italy’s data protection authority—have contested the lawfulness of ChatGPT’s data collection processes. This reveals the regulatory challenges that generative AI systems pose to the *GDPR*. Specifically, the Garante called for [a range of measures concerning data processing](#), which sought to resolve four identified breaches of the *GDPR*. Firstly, OpenAI did not initially prevent minors from accessing ChatGPT. Secondly, ChatGPT can generate inaccurate information about people (ie hallucinate). Whether the *failure to explain* how personal data is processed, and *failure to explain* responses that utilise personal data are, breaches of the *GDPR* is a contested issue. The ‘right to explanation’ is not mentioned in the *GDPR*’s articles, and the relevant Recital ([Recital 71](#)) has [no legal force](#). Nevertheless, article 5 of the *GDPR* concerns erasure or rectification of inaccurate personal data. Thirdly, users were not provided with an explanation as to how their data was being collected. Article 17 of the *GDPR* provides for the ‘right to be forgotten’ such that an individual can request to have their data removed from the model. It is worth noting that an absolute ‘right to be forgotten’ may be unattainable. There are concerns that [personal data becomes embedded in generative AI systems](#), making it ‘nearly impossible to remove all traces of an individual’s personal information’. Lastly, the large amounts of personal information being collected to train future iterations of ChatGPT could not be justified under any of the six bases in article 6(1) of the *GDPR*.

The US Blueprint lists ‘data privacy’ as a guiding principle. The Blueprint calls for privacy protection by default, with an ongoing review for privacy risks; minimising data collection by confining it to situations where it is ‘strictly necessary to [achieve] the [system provider’s] identified goals’; and ensuring best practices are followed to prevent data leaks beyond the consented use case. Further, the Blueprint outlines rights for the peoples whose data is being collected, including the rights to access that data; know who has access to that data; correct the data where necessary; and request the deletion of their data. Certain domains (eg health, employment, education) are identified as deserving of enhanced data protection. This includes an ethical review of sensitive data that may limit opportunities or access to services; auditing data quality to ensure it is not inaccurate; limiting the extent to which sensitive data can be shared, sole or made public; and additional reporting requirements where necessary.

China’s *Draft Measures for Generative AI Services* contains similar measures. The consent of data subjects is required for personal data used in the pre-training and optimisation of generative AI systems (article 7(3)). Further, providers must ‘ensure the data’s veracity, accuracy, objectivity, and diversity’ (article 7(4)). Once this data is collected, providers have an obligation to protect it and treat it appropriately. Pursuant to the *lex generalis* article 11, providers must not ‘illegally preserve input information from which it is possible to deduce the identity of users, ... conduct profiling on the basis of information input by users and their usage details, and ... provide information input by users to others’. Last, the ‘right to be forgotten’ is recognised in article 13, which provides that providers must ‘promptly handle individual requests concerning revision, deletion, or masking of their personal information’.

### 3 Bias and Disinformation

Generative AI systems can reflect biases present in their training data, entrenching discriminatory narratives. One of the highlighted issues presented by experts is the importance of improving the transparency of AI systems. However, gaining complete insight into how generative AI systems are trained may be an unattainable objective. The combination of complex training algorithms, proprietary considerations, large-scale data requirements, iterative processes, and the evolving nature of research and development pose challenges that can limit transparency and make it difficult to fully understand the intricacies of the training process.

The European Commission has recently [proposed auditing requirements](#) for very large online platforms ('VLOPs') and search engines ('VLOSEs'), which would be inserted into the [Digital Services Act](#). The current proposal would impose an obligation on VLOPs and VLOSEs to provide vetted researchers with privileged access to data; subject themselves to annual independent audits; and publish reports on content moderation, risk assessments and risk mitigation. At the time of writing, ChatGPT and other generative AI systems are not included in the list of 17 VLOPs or 2 VLOSEs. However, paragraph 29 of the proposal indicates that these systems may be subject to requirements if they are used by the listed VLOPs and VLOSEs. In addition, the [newly proposed article 28b](#) of the *Draft EU AI Act* provides that foundation models—which include generative AI systems like ChatGPT—must contain 'only datasets that are subject to appropriate data governance measures'. Providers of such models must therefore take appropriate measures to examine the suitability of the data sources and mitigate possible biases.

In China, generative AI systems must 'respect social virtue and good public custom'. Thus, article 4(2) of the *Draft Measures for Generative AI Services* provides that measures must be taken to prevent discrimination 'on the basis of race, ethnicity, religious belief, nationality, region, sex, age, or profession'. This must occur throughout the lifespan of the AI system, including during the stages of algorithm design, selecting training data, model generation and optimisation, and service provision.

Lastly, the US Blueprint lists 'algorithmic discrimination protections' as a guiding principle. The Blueprint calls for a 'proactive assessment of equity' during a system's design phase; 'representative and robust data' that mitigates potential biases; proactive testing to guard against discrimination via proxy data; and disparity assessment and mitigation.

#### 4 Intellectual Property Infringement

Finally, generative AI systems raise questions about intellectual property rights and the implications of generating copyrighted works. Providers have already faced liability risks, with class action suits having been brought against companies like [Github](#) for their code-generating AI and [Midjourney](#) for their art-generating tool.

Similar regulations have been proposed across China and the US. In China, article 4(3) of the *Draft Measures for Generative AI Services* provides that generative AI systems must 'respect intellectual property rights and commercial ethics'. Article 7(2) accordingly prohibits the use of data that infringes intellectual property rights from being used as pre-training or optimisation material for generative AI systems. In the US, the proposed [AI Risk Management Framework](#) states that training data subject to copyright should align with relevant intellectual property right laws.

Meanwhile, [article 28b\(4\)\(c\)](#) of the *Draft EU AI Act* imposes an obligation on providers of generative AI systems to publicly disclose the training data used that is protected under copyright law. This transparency provision was [favoured over a blanket prohibition](#) on the use of copyrighted material as training data.

Ultimately, this is a balancing act between protecting individuals' (intellectual property) rights on one hand and fostering innovation on the other. The application of copyright law to machines—as opposed to the human creative process—raises a [fundamental question](#) moving forward: should copyright law protect an artist's creative expression or their style more broadly? This discourse has been further convoluted by the [recent position adopted in Japan](#), wherein copyrighted materials incorporated within AI datasets are exempted from the purview of copyright law, except in cases where such utilisation would unreasonably jeopardize the rights of the copyright owner.



## Conclusion and Recommendations

The regulatory framework surrounding AI is characterised by rapid evolution. As of the present writing, the United States Government is actively soliciting public input for forthcoming AI regulations, while divergent opinions persist among experts within the AI community regarding the potential suspension of training activities for systems surpassing the capabilities of GPT-4. This section presents a conclusive overview of the proposed methodologies aimed at regulating the aforementioned risks associated with AI. Of paramount importance is the proper categorization of these risks, distinguishing between the short-term and long-term risks posed by generative AI systems. As mentioned already, this Policy Brief places high emphasis on the short-term risks of the technology.

### *Short-Term Risks*

Short-term risks pertain to the immediate hazards currently posed by generative AI systems. This Policy Brief has examined four such risks.

#### **1. Impersonation and Disinformation**

- 1.1. For users, inform them that they are interacting with generative AI systems.
- 1.2. For users, inform them of the AI system's limitations, which improves their ability to critically evaluate its output.
- 1.3. For users, establish platforms where they can refute the validity of harmful generated content.
- 1.4. For users, require identity verification before enabling access to generative AI systems.
- 1.5. For providers, require detection procedures to identify malicious generated content.
- 1.6. For providers, require content produced by generative AI systems to be labelled as generated.
- 1.7. For providers, require 'stress testing' of models to test the vulnerability to hallucinations.

#### **2. Privacy and Security**

- 2.1 For users, enable age-gating to protect the personal data of minors.
- 2.2 For users, provide notice and require their consent to collect and use data for training purposes.
- 2.3 For providers, minimise the amount of personal data collected for training purposes.
- 2.4 For providers, allow users to opt-out of having their personal data used for training purposes.
- 2.5 For providers, establish mechanisms that allow users to delete their personal data from existing datasets.

#### **3. Bias and Discrimination**

- 3.1 For providers, establish [bias mitigation measures](#) throughout the lifespan of their systems, including the following stages: algorithm design, selecting training data, model generation and optimisation, and service provision.
- 3.2 For providers, exercise enforcement on the disclosure of data training models.
- 3.3 For providers, require regular independent audits and reporting on risk assessment and mitigation measures.

#### **4. Intellectual Property Infringement**

- 4.1 For providers, prohibit the use of training data that infringes intellectual property law. Alternatively, require reporting on copyrighted data that is used for training purposes.

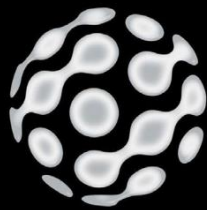
## Long-Term Risks

The longer-term risk that generative AI systems *may* pose is in accelerating society's arrival at Artificial General Intelligence ('AGI'). AGI [refers to](#) the ability of an AI system to perform a variety of tasks in different contexts and environments. The concept of AGI exists in contrast to our current, 'narrow' AI systems, which perform tasks in specified contexts. Within the AI community, there exists a division of perspectives regarding the fundamental inquiries surrounding the feasibility of AGI and, if indeed attainable, the prospective timeframe for its realization.<sup>5</sup>

Assuming that AGI proves viable, apprehensions arise regarding its potential to profoundly and adversely impact humanity, potentially engendering existential perils. These concerns demand equitable recognition and warrant further scrutiny. In this vein, there are recommendations that policymakers increase investment in ['safety and alignment research'](#) so that our understanding of AI's long-term risks keep pace with technological developments. Nonetheless, it is of utmost significance to uphold the differentiation between short-term and long-term risks, thereby averting potential diversions from the pivotal role that regulation may play in offering remedies to the aforementioned risks.

## Endnotes

- 1 One prominent example has been the ‘hundreds’ of fake profiles of Ukrainian President, Volodymyr Zelenskyy, that attempt to trick individuals into sending them money: Peter Suci, ‘There Are Now Hundreds Of Volodymyr Zelenskyy Impersonators On Social Media’, Forbes (online, 16 March 2022) <<https://www.forbes.com/sites/petersuci/2022/03/16/there-are-now-hundreds-of-volodymyr-zelenskyy-impersonators-on-social-media/?sh=3d28dcbe4a6e>>.
- 2 This term has not been universally approved of, largely on the basis that it frames the problem in a manner that appears to anthropomorphize AI systems rather than depict the issue as one of ‘untethered’ text generation: Ben Zimmer, ‘“Hallucination”: When Chatbots (and People) See What Isn’t There’, Wall Street Journal (online, 20 April 2023) <<https://www.wsj.com/articles/hallucination-when-chatbots-and-people-see-what-isnt-there-91c6c88b>>.
- 3 Deepfakes are AI-generated audio or visual content. They need not be manipulative, and have recognisably beneficial purposes in areas like education and entertainment. However, deepfake technology can be used maliciously. A 2019 report by Deeptech Labs found that the vast majority of deepfake videos (96%) were non-consensual pornography: see Deeptech Labs, The State of Deepfakes (Report, 2019) 1, 6.
- 4 Again, there is no official English translation of these regulations, so the Stanford University translation is relied on: Rogier Creemers and Graham Webster, ‘Translation: Internet Information Service Deep Synthesis Management Provisions (Draft for Comment) – Jan. 2022’, DigiChina Stanford (online, 4 February 2022) <<https://digichina.stanford.edu/work/translation-internet-information-service-deep-synthesis-management-provisions-draft-for-comment-jan-2022/>>.
- 5 See, eg, Toby Walsh, ‘The Singularity May Never Be Near’ (2017) 38(3) AI Magazine 58. Cf Sébastien Bubeck et al, ‘Sparks of Artificial General Intelligence: Early experiments with GPT-4’ [2023].



**AI ASIA PACIFIC**  
INSTITUTE

[contact@aiasiapacific.org](mailto:contact@aiasiapacific.org)  
[aiasiapacific.org](http://aiasiapacific.org)

