Centre for
Digital Wellbeing

The Risk and Challenges presented by Generative AI Tools in Education

Submission

21 July 2023

Table of Contents

Committee Secretary

The House Standing Committee on Employment, Education and Training
PO Box 6021
Parliament House
Canberra ACT 2600

21 July 2023

Submission to the Inquiry into the use of generative artificial intelligence in the Australian education system

Introduction

The Centre for Digital Wellbeing (CDW) welcomes the opportunity to provide a submission to the parliamentary inquiry on the use of generative AI in the Australian Education System.

The Centre is a policy research and design body focusing on technology's impact on overall health and wellbeing, safety, and social cohesion in the Australian community. Our purpose is to facilitate research about the impacts of technology, formulate policy responses and develop initiatives to assist users to better engage in healthy, ethical, and safe digital practices.

CDW is centered around participatory research that uses a human-centered approach to designing policy, particularly focusing on young people, education, and parenting. We develop strong relationships with these beneficiaries which enables us to co-design policy recommendations and solutions to tech-driven initiatives that are centered around lived experience.

The field of AI has accelerated rapidly in the past decade, with advances in the release of Large Language Modules (LLMs) such as ChatGPT. The fundamental issue with AI is that while we have made exponential leaps in the power of such systems to optimise for a wide array of tasks, we have made far fewer gains in the fields of safety and alignment. While these tools provide opportunities in their ability to execute complex utility functions, they have not been programmed to put human values central. At the most rudimentary level, alignment is the attempt to ensure that AI systems work for humans and support human goals, no matter how powerful the technology becomes.

We strongly encourage the House Standing Committee on Employment, Education and Training to consider the risks and challenges that generative artificial intelligence will present to the Australian education system, and draw on the lived expertise of teachers, students, and parents, as well as subject matter experts to form and incorporate legislative and policy solutions that put the wellbeing and safety of

students and teachers first. This aligns to the eSafety Commissioner's Safety by Design approach that advises risk mitigation at the front end and moving thoughtfully to put user rights and safety first[1].

As an advocate for developing systems that promote the safe and considered use of technological advancements in a way that improves overall wellbeing and education outcomes, our submission addresses three of the six areas as per the terms of reference.

For further information on any of the points raised in our submission, please contact CDW on secretariat@digitalwellbeing.org.au or 02 6162 0361.

[1] eSafety Commissioner, Safety by Design. Available at https://www.esafety.gov.au/industry/safety-by-design [Accessed 10 July]

Executive Summary

Generative AI tools have become prevalent in education, but their adoption also brings significant risks and challenges that necessitate thorough consideration. The key areas of concern include data privacy, security, algorithmic transparency, and bias and ethical implications.

This increased use of EdTech during the COVID-19 pandemic posed an extreme risk to the data privacy and security of students. The sensitive student information gathered and in certain instances sold to third parties, made them vulnerable to commercial exploitation and data breaches. Moreover, limited opt-out options and insufficient cybersecurity measures in schools add to the potential risks.

The integration of generative AI models carries added concerns, especially in relation to embedded biases which can further entrench inequalities and discrimination within society. Examples from other countries demonstrate how biased algorithms can unfairly affect grading and other educational processes, especially for marginalized students. Ensuring that systems are trained using diverse and inclusive datasets, continuously monitored and transparent can help mitigate these risks.

Additionally, algorithmic transparency can assist with addressing passive consumption and the decline of critical thinking skills when generative AI systems are over relied upon. Overreliance on AI-generated content may neglect diverse learning styles and impede social interaction, leading to negative impacts on learning outcomes and development of critical and creative thinking skills.

Looking at international practices, the European Union's AI Act stands as a model for effective AI regulation. It categorizes AI systems based on risk levels and imposes corresponding obligations, promoting accountability and transparency. In contrast, the United States is still in the process of developing comprehensive AI legislation and currently relies more on self-regulation by private companies.

To address the risks and challenges associated with generative AI tools in education, a multifaceted approach is recommended. International practices and policies, particularly in the European Union, can serve as models for effective AI regulation. Further, establishing a robust data protection framework, mandating transparent data use policies, and conducting privacy impact assessments, ensuring human oversight, and increasing public awareness around risks are essential steps.

Recommendations

To address the risks and challenges posed by generative AI tools in education, several measures are recommended by the Centre for Digital Wellbeing.

- That the Government address privacy and data risks in the Privacy Act review by establishing a robust data protection framework that outlines the rights of students in relation to personal data as well as establishing limitations to the collection, use and retention of data of minors.

- That the Government puts in place requirements for educational institutions and AI developers to have clear and transparent data use policies. These policies should outline the types of data collected, the purposes for which the data will be used, how long the data will be retained, and the measures taken to protect data privacy. These policies should be presented to users in a way that is age-appropriate and easy to understand.

- That the Government provides specific powers to the Office of the Australian Information Commissioner to assess AI tools used in education should for their privacy impact. The assessment would identify potential risks to data privacy and outline measures to mitigate them before the tool can be implemented. The Office of the Australian Information Commissioner should be given the power to ban the implementation of AI tools if the risk to a users' data privacy is too high.

- That the Government requires AI developers and educational institutions to implement strong encryption and secure data storage practices. Government should provide support - both technical and financial – to educational institutions to assist with meeting this requirement.

- That the Government strengthens the role of and financial support to the Australian Information Commissioner to assist with the safe implementation and overseeing the use of AI tools in education settings in relation to data privacy.

- That the Government considers the development of comprehensive legislation in relation to Generative AI drawing on international best-practice, like the European Union AI Act. It is crucial that a risk-based approach is taken which includes third party and independent impact and audit assessments, mandatory transparency reports, the human oversight of certain AI systems depending on level of risk, and accountability measures.

- That the Government consider the establishment of a dedicated AI regulatory body assist regulators, policy makers and government in developing, enforcing and implementing legislation, in line with submissions made to "Positioning Australia as a Leader in Digital Economy Regulation" and the Australian Human Rights Commission's Human Rights and Technology Final Report (2021).

- That the Government increases public awareness and education programs to inform users about AI's potential implications and risks, to improve the ability of the users to hold companies and developers accountable and make informed decisions.

Risk and Challenges presented by the use of Generative AI tools

Generative AI Tools pose certain risks and challenges that need to be thoroughly considered. Key risks and challenges associated with generative AI tools in education relate to data privacy, security, algorithmic transparency and bias and ethical concerns.

Data privacy and Security

The speed of generative AI tools uptake into Australian classrooms raises concerns about data privacy and security regarding sensitive student information, parallel to similar challenges faced in recent years with the integration of education technology into the classroom.

The COVID-19 pandemic saw an increase of adoption and implementation of EdTech products into Australian schools to manage online learning in a virtual classroom environment. 89% of these platforms and products put children's rights at risk, monitoring them without their consent and allowing access from or selling the data to third parties.[2]

This access to children's data makes them susceptible to commercial exploitation, exposing them to advertising that is either overt, such as adds in online games, video platforms and in app purchases, or indirect, through sponsored content on influencer platforms[3]. Studies have shown that children under the age of 12 do not understand the pervasive nature of advertising and children 8 years and under cannot differentiate between content and advertising, making them particularly vulnerable to content and microtargeted marketing from brands trying to sell dangerous products such as alcohol, vapes or gambling.[4]

Most schools who adopted the use of EdTech made it compulsory for teachers and students to use the products, rendering it impossible for users to opt-out, limit the amount of data gathered or even limit the use of technology. This raises questions around parents' and children's' ability to provide full consent. Even if there are concerns around the amount of data being collected or the way they are being monitored, there is no recourse available. The impact of this on wellbeing should be considered as studies have shown that if there is a level of discomfort with the way a person is monitored, it can cause severe anxiety, a culture of distrust and symptoms similar to post-traumatic stress disorder.[5]

---

[2] Human Rights Watch. *'How Dare They Peep into My Private Life?' – Children's Rights Violations by Governments that Endorsed Online Learning During the Covid-19 Pandemic*. Available at https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments. [Accessed 11 July]

[3] The Conversation, *How children are being targeted with hidden ads on social media*, November 3, 2021. Available at https://theconversation.com/how-children-are-being-targeted-with-hidden-ads-on-social-media-170502. [Accessed 17 July]

[4] UNICEF. *The children's rights-by-design standard for data use by tech companies, Issue brief no. 5*, November 2020 Good Governance of Children's Data project Office of Global Insight and Policy. [Accessed 12 July]

[5] Vice, *What Constant Surveillance Does to Your Brain*, 14 November 2018. Available at https://www.vice.com/en/article/pa5d9g/what-constant-surveillance-does-to-your-brain. [Accessed 18 July]

Generative AI tools, like some EdTech applications, continuously gather data exposing children to the same risks. Even when data is not sold or accessed, there is still danger in its gathering due to the lack of proper cybersecurity in Australian Schools which are under resourced and lack the critical expertise to address these challenges. They are often slow to react or prevent vulnerabilities in their protection systems and generally have ineffective internal policies specific to cyber security which contributes to poor organizational security culture.[6]

The risks associated with data privacy should be addressed in the Privacy Act review by establishing a robust data protection framework that outlines the rights of students in relation to personal data as well as establishing limitations to the collection, use and retention of data of minors. Other measures that should be explored are:

- Transparent Data Use Policies: Educational institutions and AI developers should be required to have clear and transparent data use policies. These policies should outline the types of data collected, the purposes for which the data will be used, how long the data will be retained, and the measures taken to protect data privacy. These policies should be presented to users in a way that is age – appropriate and easy to understand.

- Privacy Impact Assessment: AI tools used in education should be assessed for their privacy impact by the Office of the Australian Information Commissioner. The assessment would identify potential risks to data privacy and outline measures to mitigate them before the tool can be implemented. The Office of the Australian Information Commissioner should be given the power to ban the implementation of AI tools if the risk to a users' data privacy is too high.

- Encryption and Secure Data Storage: Government can require AI developers and educational institutions to implement strong encryption and secure data storage practices. Government should provide support - both technical and financial – to educational institutions to assist with meeting this requirement.

- Independent Auditing and Oversight: Government should strengthen the role of and financial support to the Australian Information Commissioner to assist with overseeing the use of AI tools in education settings in relation to data privacy.

[6] M Torres, A Mullins, & N Thompson, *'Education Cybersecurity Assessment Tool: A cybersecurity self-assessment tool for the Australian K-12 sector'*, 2022. ACIS 2022 Proceedings. Available at https://aisel.aisnet.org/acis2022/96. [Accessed 12 July]

Algorithmic transparency and critical thinking

In addition to the risk of data capturing practices and lack of appropriate cybersecurity capabilities, algorithmic data-driven profiling and decision-making mechanisms disproportionately affect vulnerable groups, including children. Many generative AI models operate as black boxes[7], making it difficult for children, teachers and parents to understand the underlying technology and how generative AI works. As a result, they may trust and rely on the AI system without critical judgment, potentially accepting biased or false information.

The integration of generative AI tools carries the risk of children, students and teachers becoming passive consumers rather than active thinkers. Instead of actively engaging with problems and coming up with their own solutions, users are presented with answers that are accepted without questioning or exploring alternative ideas. Minister Clare previously mentioned the potential of AI to assist teachers with tasks such as grading, providing them with more time to focus on teaching and mentoring their students.[8] Using AI tools to assist with automating routine tasks can seem innocent, however, it carries the risk of punishing "out of the box" and creative thinking due to only a narrow set of answers being accepted as "correct". The lack of active participation and learning experiences, as well as a decreased need for analysing, reasoning and problem solving will negatively impact learning outcomings, including the development of critical and creative thinking skills.

Further, the black-box nature of some AI models makes it challenging to understand how they arrived at specific outputs. This leads to particular challenges around accountability of developers, concealed biases, discrimination and errors. When decision making processes are hidden, it creates challenges to creating appropriate regulations and attributing responsibility for any negative outcomes, ethical violations or harmful content. The risk lies in developers or organisations avoiding any legal repercussions for the negative impact their generative AI tools cause. Similarly, without insight into the decision-making process, it becomes difficult to assess whether the system is making unbiased choices due to its ability to hide biases and discriminatory patterns. If there is no external oversight, understanding of the process or means of correction, this could lead to unfair treatment of individuals and vulnerable groups.

Robodebt is a recent example of the negative impact that the use of machine learning technologies can have on individual wellbeing when there is a lack of human oversight in the implementation and use of AI-like tools[9]. The program trained to detect and pursue discrepancies in income to ensure greater compliance in payments received by welfare recipients. It was meant to be cost-effective and efficient with the goal of optimising for the detection of non-compliance, but due to the absence of human

[7] The Conversation, *What is a black box? A computer scientist explains what it means when the inner workings of Ais are hidden*, 22 May 2023. Available at https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when-the-inner-workings-of-ais-are-hidden-203888. [Accessed 11 July]

[8] Innovation Australia, *Inquiry to probe generative AI use in schools, higher ed,* 25 May 2023. Available at https://www.innovationaus.com/inquiry-to-probe-ai-use-in-schools-higher-ed/#:~:text=A%20parliamentary%20inquiry%20will%20examine,technology%20is%20here%20to%20stay. [Accessed 15 July]

[9] Pursuit, The Flawed Algorithm at the Heart of Robodebt, 10 July 2023. Available at https://pursuit.unimelb.edu.au/articles/the-flawed-algorithm-at-the-heart-of-robodebt. [Accessed 20 July]

supervision the program made a series of decisions that in the case of Jarrad Madgwick led to the loss of life.[10]

The mandatory disclosure of information about algorithms (training data used, limitations identified during development, decision-making logic, etc.) they employ can assist with countering these challenges. Further measures CDW urges the Government to explore are:

- Audit and Impact Assessments: Organisations developing AI tools should be required to conduct regular audits or impact assessments of their AI systems to address potential biases or unintended consequences. In order to ensure accountability, these assessments should be subject to external scrutiny.

- Transparency Reports: Legislation should require organisations to publish transparency reports detailing how their AI systems are used, including any cases of significant impact on individuals or society. These reports would provide insights into how algorithms function in practice and how they affect users.

- User Access to Data: Legislation can grant users the right to access and understand the data collected about them and the algorithms used to make decisions based on that data. Users should have the opportunity to challenge or correct inaccurate information.

- Third Party evaluations: Legislation should mandate third party evaluations of AI tools and systems to ensure they meet certain criteria for transparency, fairness, and accountability. These evaluations should be conducted by independent auditors and regulatory bodies.

- Human review and Oversight: Legislation should require that critical decisions made by AI systems to have a human-in-the-loop component, where a human reviewer can examine and validate the system's outputs before final decisions are implemented.

- Accountability measures: Legislation should establish accountability measures for organisations that deploy AI systems that lead to biased, discriminatory or harmful outcomes to help deter irresponsible AI development and use.

- Public Awareness and Education: Government should also support public awareness campaigns and educational programs to inform the general public about AI technology and its potential implications to improve the ability of the public to hold organisations accountable and make informed decisions.

[10] The Age, *Kathleen Madgwich tells Robodebt royal commission about her son Jarred and the damage the scheme caused*, 10 March 2023. Available at https://www.abc.net.au/news/2023-03-10/qld-robodebt-scheme-government-royal-commission-fraud/102027838. [Accessed 17 July]

Overreliance and reduced human interaction

Generative AI tools such as ChatGPT have the potential to provide benefits to schools, teachers and students such as assisting with the development of study plans and automating administrative tasks, allowing more focus on personalized instruction, guidance and mentorship. However, excessive dependence on AI tools in education carries the risk of diminishing the role of human teachers and interpersonal interactions.

It is important that the integration of Generative AI tools is not rushed due to its potential to assist with some challenges the education sector in Australia currently faces. In particular, AI is unable to address in full the core issues currently faced by the education sector in Australia, such as teacher shortages. In 2023, Victoria saw over 900 vacancies advertised in government schools[11], South Australian teachers reported working over 50 hours per week[12] and in Tasmania, no government schools are receiving appropriate funding[13] in order to achieve the minimum NAPLAN standards. While the National Teacher Workforce Action Plan aims to attract more people to the profession, the Plan does not include significant pay increases for teachers or the provision for more funding to government schools across the board. If we are to improve the state of education in Australia and ensure quality teachers are retained, there needs to be greater nationwide financial investment and value of the teaching profession.

The need for education to have tools and methods that will ease the pressures of the profession are valid and necessary. However, using generative AI tools to ease current pressures on teachers without the implementation of rigorous, risk-based legislative approach to guide schools, will pose very real and potentially significant risks to the safety, intellectual development and wellbeing of students.

While AI can tailor content based on user data, overreliance can lead to a one-size-fits all approach to education with no consideration for diverse learning styles and needs of students. Relying solely on AI-generated content may not address individual learning preferences effectively leaving certain students who might already be struggling more, behind. Further, limited social interaction between students and educators has the risk to lead to a decrease in the use of collaborative learning methods and interpersonal skill development. Studies have found that a lack of academic social interaction has led to learning and self-regulation difficulties,[14] as well as reduced motivation and interest in the learning process due to a

[11] The Age, *Workload, lack of respect among key causes of why teachers leave profession, research finds*, 26 April 2023. Available at https://www.abc.net.au/news/2023-04-26/teaching-crisis-studied-in-national-research-by-federation-uni/102254252. [Accessed 9 July]
[12] Australian Education Union, South Australian Branch, *Fix the Crisis*. Available at https://www.fixthecrisis.com.au/. [Accessed 9 July]
[13] Australian Education Union, Tasmanian Branch, *The real state of Tasmanian Education*. Available at https://aeutas.sgp1.digitaloceanspaces.com/2023/02/Real-State-of-Tasmanian-Education.pdf. [Accessed 9 July]
[14] T P Ivanec, *The Lack of Academic Social Interactions and Students' Learning Difficulties during the COVID-19 Faculty Lockdowns in Croatia: The Mediating Role of the Perceived Sense of Life Disruption Caused by the Pandemic and the Adjustment to Online Studying*, 2022. Available at https://www.mdpi.com/2076-0760/11/2/42. [Accessed 17 July]

lack in novelty and engagement ultimately contributing worsening learning outcomes, particularly for students with specific needs.

Bias and ethical concerns

Generative AI models use machine learning models (MLM) which are trained using large datasets to perform a programmed function. If the training data predominantly represents a specific demographic, culture, or perspective, the generated output may reflect those biases and further entrench and exacerbate inequalities and stereotypes present in our societies.[15]

Many examples of bias have already eventuated in the application of AI tools. In 2018, Amazon scrapped the use of its automated recruitment tool due issues with bias against women. The program used artificial intelligence to score job candidates using resumes submitted to the company over a period of 10 years. The male dominance in software development and other technical jobs present in the training data caused the tool to teach itself that male candidates were preferable over female ones.[16] For students, this could mean that challenges historically faced by marginalised students further intensify when a demographic identifier such as race, intersects with other statuses such as language, income, etc. If gaps/biases in the training data are not addressed, they further amplify current socio-economic, racial and ethnic disparities.[17]

Often these tools are tested in educational institutions that only have a small number of multicultural and First Nations students and/or students from low-income backgrounds. In 2020, the UK experienced the risks associated with using AI in educational settings firsthand. Due to the COVID-19 pandemic, the British Government cancelled the advanced-level qualification exams. Instead, teachers were asked to provide an estimate of result they expected their students to achieve. Those estimated scores were weighted using an algorithm that relayed upon the historic performance of individual secondary schools. It was later found that students from less-advantages schools were more likely to have their grades decreased and students from richer schools had a higher likelihood of having their grades raised, essentially reinforcing economic and societal bias present in the U.K education system.[18]

The issue spans across many areas other than grading and, while there are advocacy groups working for equity and fairness in AI development, the problem remains largely unaddressed. The lack of legislation

---

[15] UNESCO, *Artificial Intelligence: examples of ethical dilemmas*, 21 April 2023. Available at https://www.unesco.org/en/artificial-intelligence/recommendation-.ethics/cases#:~:text=But%20there%20are%20many%20ethical,and%20privacy%20of%20court%20users. [Accessed 10 July]

[16] Reuters, *Amazon scrapes secret AI recruiting tool that showed bias against women*, 11 October 2018. Available at https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. [Accessed 11 July]

[17] N Gaskins, *Interrogating Algorithmic Bias: From Speculative Fiction to Liberatory Design,* 2022. Available at https://link.springer.com/article/10.1007/s11528-022-00783-0#:~:text=Section%202%3A%20Algorithmic%20Bias%20in,AI%20and%20machine%20learning%20systems. [Accessed 10 July]

[18] Axios, *How an AI grading system ignited a national controversy in the UK,* 20 August 2020. Available at https://www.axios.com/2020/08/19/england-exams-algorithm-grading. [Accessed 11 July]

and guidelines for educational institutions leaves many disadvantages students at risk of further discrimination. It is therefore crucial that generative AI tools and models are trained on diverse and inclusive datasets and regularly monitored to mitigate bias.

Further, it is important that during the implementation and development of any legislation, the Federal government works in close collaboration with State governments to develop strong guidance for educational institutions that will help establish policies that are in accordance with legislation and that puts the safety, wellbeing and learning outcomes of students, teachers and administrators at front of mind.

International practices and policies in relation to the use of Generative AI

Most countries, like Australia, are relying on existing law and guidelines for best practice in AI while considering regulatory measures. The European Union however is set to introduce the first comprehensive AI legislation in the Western world and is often looked towards for a best practice approach towards AI regulation.

European Union

While many countries are working on introducing AI legislation, the European Union is considered at the forefront of developing the most comprehensive law to regulate AI. The AI Act, first proposed in April 2021 and likely to be adopted by the end of 2023, is part of the EU's broader digital strategy, Europe's Digital Decade[19]. Its purpose is not to halt development of AI, but rather ensure that system that pose a risk to safety and the fundamental rights of European Union citizens do not become operational. As such, the Act regulates AI on the basis of four risk categories: unacceptable risk, a high risk, limited or minimal risk.[20]

The legislation sets out each risk level in detail and provides requirements for AI development depending on risk classification. AI systems that are considered a threat to people – such as cognitive behavioural manipulation, social scoring and real-time and remote biometric identification systems - present an unacceptable risk and will be banned. High-risk AI systems are subject to a strict set of obligations before they can be made accessible to the public including risk assessments, data set requirements, traceability of results, clear information for users, appropriate human oversight and high level of robustness, security and accuracy. Applications with limited risk have specific transparency obligations to make the user aware it is interacting with an AI system. Finally, minimal-risk systems are allowed free use.[21]

When an AI system is considered high risk, it must undergo a rigorous process before entering the market. During the developing a high-risk AI system, AI Impact Assessment and Codes of Conduct should be used to guide the development overseen by multidisciplinary teams. Afterwards, the system would be required to undergo an approved conformity assessment and maintain compliance with the AI requirements for the entire duration of its application. For particular systems, an external body will be involved in the assessment. If there are any changes to the system required, this step has to be repeated until all

[19] European Commission, *Europe's Digital Decade: digital targets for 2030.* Available at https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en. [Accessed 17 July]

[20] European Commission, *Shaping Europe's digital future - Regulatory framework proposal on artificial intelligence,* 20 June 2023. Available at https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. [Accessed 17 July]

[21] European Commission, *Shaping Europe's digital future - Regulatory framework proposal on artificial intelligence,* 20 June 2023. Available at https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. [Accessed 17 July]

requirements are satisfied. Experts have recommended to make the external body involvement mandatory for all high-risk systems.[22]

Once compliance assessments are completed, all high-risk AI systems will be logged in a dedicated EU database and a declaration of conformity must be signed by the company. When the system has become publicly available, relevant authorities will monitor and review its application and the company is required to have a post-market monitoring system implemented.[23] This approach acknowledges the rapid development of systems and ensures ongoing quality and risk management.

In order to enforce the legislation, a new enforcement body at the Union level - the European Artificial Intelligence Board (EAIB) - would be established which can hand out fines for non-compliance. Member States will be required to established national supervisory bodies similar to its national data protection authorities under the GDPR.[24]

The Act further aims to define AI in a deliberately broad way to future proof its application. The definition proposed is: "'a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate output such as predictions, recommendations, or decisions influencing physical or virtual environments."[25] Similarly, the Act tries to apply an open standard to its classification norms for high-risk applications. However, this carries the risk of differences of opinion about their interpretation which would require courts to make a final decision.[26]

Until the AI Act is adopted, the EU provides some rules around automated decision making in the GDPR which protects citizens from being subjected to decisions based solely on automated processes.[27]

---

[22] M Kop, *EU Artificial Intelligence Act: The European Approach to AI,* Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021. Available at https://futurium.ec.europa.eu/sites/default/files/2021-10/Kop_EU%20Artificial%20Intelligence%20Act%20-%20The%20European%20Approach%20to%20AI_21092021_0.pdf.

[23] M Kop, *EU Artificial Intelligence Act: The European Approach to AI,* Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021. Available at https://futurium.ec.europa.eu/sites/default/files/2021-10/Kop_EU%20Artificial%20Intelligence%20Act%20-%20The%20European%20Approach%20to%20AI_21092021_0.pdf.

[24] M Kop, *EU Artificial Intelligence Act: The European Approach to AI,* Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021. Available at https://futurium.ec.europa.eu/sites/default/files/2021-10/Kop_EU%20Artificial%20Intelligence%20Act%20-%20The%20European%20Approach%20to%20AI_21092021_0.pdf.

[25] King & Wood Malleson, *Developments in the Regulations of Artificial Intelligence,* 19 April 2023. Available at https://www.kwm.com/global/en/insights/latest-thinking/developments-in-the-regulation-of-artificial-intelligence.html. [Accessed 16 July]

[26] M Kop, *EU Artificial Intelligence Act: The European Approach to AI,* Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021. Available at https://futurium.ec.europa.eu/sites/default/files/2021-10/Kop_EU%20Artificial%20Intelligence%20Act%20-%20The%20European%20Approach%20to%20AI_21092021_0.pdf.

[27] King & Wood Malleson, *Developments in the Regulations of Artificial Intelligence,* 19 April 2023. Available at https://www.kwm.com/global/en/insights/latest-thinking/developments-in-the-regulation-of-artificial-intelligence.html. [Accessed 16 July]

United States

In October 2022, the Biden Administration announced a Blueprint for an AI Bill of Rights focussed on protecting civil rights and democratic values.[28] The document shares a non-binding roadmap for responsible AI use through 5 core principles that should guide the development and implementation of AI systems. However, the Blueprint has received criticism as it lacks the ability for enforcement of the guidelines and relies on the private sector to self-regulate from a consumer rights-based approach.[29]

While there are current efforts to draft AI specific legislation in the U.S. with Senate Majority Leader Chuck Schumer leading a congressional effort to draft AI regulation and U.S. Representative Ritchie Torres filing the AI disclosure Act of 2023, there is currently no federal legislation other than the American Data Privacy and Protection Act and the National Artificial Intelligence Initiative Act of 2020. Companies are required to self-regulate to ensure any systems operate within the boundaries of relevant frameworks such as the AI Risk Management Framework from the U.S. National Institute of Standards and Technology. [30] On a state level, regulators have been more active with at least 17 states introducing bills or resolutions related to the use of Artificial Intelligence.[31]

While legislation is still wanting, other efforts have been made to mitigate the potential misuse or unintended negative consequences of AI use. For example, the US National Institute of Standards and Technology launched an initiative involving workshops and discussions around the development of federal standards for trustworthy AI systems.[32] Nevertheless, experts remain concerned with the current overreliance on self-regulation by private companies.

---

[28] The White House, *Blueprint for an AI Bill of Rights.* Available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/. [Accessed 19 July]

[29] Brookings, *Opportunities and blind spots in the White House's blueprints for an AI Bill of Rights,* 19 December 2022. Available at https://www.brookings.edu/articles/opportunities-and-blind-spots-in-the-white-houses-blueprint-for-an-ai-bill-of-rights/. [Accessed 20 July]

[30] IAPP, *What's next for potential global AI regulation, best practices,* 23 April 2023. Available at https://iapp.org/news/a/iapp-gps-2023-whats-next-for-potential-global-ai-regulations-best-practices-for-governing-automated-systems/. [Accessed 21 July]

[31] National Conference of State Legislatures, *Legislation Related to Artificial Intelligence,* 26 August 2022. Available at https://www.ncsl.org/technology-and-communication/legislation-related-to-artificial-intelligence. [Accessed 20 July]

[32] National Conference of State Legislatures, *Legislation Related to Artificial Intelligence,* 26 August 2022. Available at https://www.ncsl.org/technology-and-communication/legislation-related-to-artificial-intelligence. [Accessed 20 July]