Australian National University

Response to the Australian Government Department of Industry, Science and Resources
Safe and Responsible AI in Australia Discussion Paper

July 2023

INTRODUCTION

The Humanising Machine Intelligence Program (HMI), a grand challenge initiative supported by the Australian National University, welcomes the opportunity to respond to the Discussion Paper, "Safe and Responsible AI in Australia". HMI is a multi-disciplinary research team with expertise that spans the humanities, social sciences, and computer science, supporting cross-cutting collaborations on AI. Drawing on our collective understanding of AI, derived from deep engagement with existing literature and active production of new research in this area, we adopt a socio-technical approach to AI systems.[1]

Although AI has been around for decades, recent developments and expansive AI integration across personal, professional, and civic life have garnered public attention. Government responses recognise that AI holds both promise and peril and work towards fostering the former while mitigating the latter. HMI shares this commitment and offers research-based insights towards those ends.

Our response includes recommendations to expand and adjust the focus of the Discussion Paper where appropriate. Specific recommendations revolve around problem orientation, the scope of regulatory issues related to AI, and practical considerations not-yet-addressed in the Paper's proposals. We structure our submission across three interrelated pillars:

1. Reorienting key assumptions about risk, trust, and bias in AI systems
2. Considering actionable regulatory targets
3. Cultivating critical awareness of regulatory responses to support domestic initiatives

Through these pillars, we emphasise the structural nature of AI in society and the corresponding need for holistic approaches that attend to how human, economic, environmental, infrastructural, and legislative realities are intertwined factors in the production, deployment, and regulation of AI, machine learning, and other data-intensive technologies.

THREE PILLARS FOR SAFE AND RESPONSIBLE AI

1. Reorienting Key Assumptions about Risk, Trust, and Bias in AI Systems

The way a society acts towards AI — its development, deployment, and regulation — is shaped by how that technology is framed and understood. It is therefore vital to identify core assumptions about AI as they operate in policies and practices, and to critically interrogate the validity of existing assumptions so they may be reconfigured as necessary.

To this end, we take as our starting point the implicit assumptions about AI that presently underlie the Discussion Paper, focusing on areas that may benefit from rethinking. These include, in particular, issues of risk, trust, and bias.

[1] The U.S. Office of Science and Technology (OSTP), National Institute of Standards and Technology (NIST) and the National AI Advisory Committee have all endorsed this approach to AI.

1.1 Risk and Trust: Finding the Balance

Implied in the Discussion Paper, and in popular discourse, is a mutual exclusivity between risk and trust, with a preference to increase trust and trustworthiness in AI systems. We contend instead that healthy and informed scepticism — rather than high levels of trust — can be instrumental for risk reduction, while focusing attention on earned trust within the social institutions that implement AI.

For context, research shows that the Australian public has a relatively low level of trust in AI (Gillespie et al. 2023, p.14) and maintains concerns about AI-related risks (Biddle 2023). These technologies are not necessarily at issue for Australians; it is their application by powerful institutions and related impacts on democratic integrity, personal privacy, economic security, and fundamental rights. We do not think that sensitivities to these risks are problems to resolve. Instead, they are resources for cultivating an engaged citizenry that holds government and industry accountable. A sceptical public asks tough questions of central institutions about how AI implementation will impact on their daily lives — questions that would have saved significant hardship had they been, for example, posed prior to the "Robodebt" scheme.

Understanding the reasoning behind postures such as scepticism, pushback, and defiance also offers valuable information for enhancing regulatory responses (see Braithwaite, 1995, 2014). The challenge is balancing healthy scepticism and critical engagement with AI systems against a base level of institutional trust, especially when trying to encourage AI for social benefit. Striking this balance means involving Australian communities — especially those who are at disproportionate risk of negative impact — from the earliest stages of AI-related developments (including law, policy, regulation, and deployment). It also means:

- Building mechanisms of transparency and accountability so that Australians can better understand how AI systems operate in terms of the data they use and how those systems impact people individually and collectively;
- Offering easy pathways for reporting when AI systems impinge on rights, threaten safety, or enact harm; and
- Tying reporting tools to legal and regulatory consequences for private and public sector actors and organisations.

Such approaches can empower Australians to advocate for themselves and their communities, bolstering potential opportunities wrought by AI and other data-intensive technologies.

1.2 Bias: From Fairness to Reparative Justice

This Discussion Paper, like many regulatory efforts, conveys concerns about bias. The goal is presented as one of bias neutralisation via fairness. Research demonstrates that this approach does not, and cannot, attend to or address historical inequities represented and reinforced in model outputs. Shifting the standard from 'fairness' to 'justice' begins with a baseline imperative towards redress, affirmatively accounting for and rectifying the systemic inequalities that permeate existing datasets, data sources, and the outputs generated therein. AI and its regulation present not only a responsibility to address systemic and historical inequalities, but also an opportunity to do so as these technologies reshape our worlds (Davis et al. 2021; So et al. 2022).

2. Considering Actionable Regulatory Targets

Section 2 of the Discussion Paper draws attention to tensions that emerge in relation to balancing the opportunities and challenges of AI, given the stated potential of AI to drive productivity growth. These possible economic benefits do not necessarily ensure positive social outcomes or their equitable distribution. Accordingly, the prospect of productivity gains from AI should not deter the development of regulatory targets or actioning them now. We highlight several considerations for AI regulatory frameworks and policy settings in the short term.

2.1 Future Proofing

We submit that regulatory actions seeking to balance opportunities and challenges should be historically informed, immediately relevant, and adaptable for the future. To this latter point, regulatory action must be attentive to technological conditions, while underpinned by broader principles that will carry forward into new and evolving socio-technical landscapes.
AI is a rapidly developing field that presents many unknowns when it comes to its societal effects and trajectories of social change. Regulatory responses should address the technologies of the present while accommodating future developments and new discoveries (as also recognised in the draft EU AI Act, see Annex A). For example, the next generation of Large Language Models (LLMs) may well be able to attribute sources for the text they produce, which will improve reliability and trust. However, issues of intellectual property, privacy, and data provenance still require critical attention in the present, albeit in potentially altered forms compared to earlier innovations. Regulating AI is thus best approached through solid and consistent principles, implemented through adaptable frameworks that can adjust to technological change.

2.2 Transparency with Propriety

Businesses have a vested interest in maintaining ownership over their intellectual property, which encourages opacity. Yet, as the Discussion Paper suggests, transparency is an important dimension of technology regulation. Clear rules about transparency, applied equally across industries, can generate necessary insights into AI systems and how to audit them, without compromising a competitive business environment.

Australia should have disclosure requirements for AI algorithms and systems as a concrete regulatory step. Common among several global recommendations (see EU 2023, Mądry 2023, Montgomery 2023), these requirements are widely applicable, pose a relatively low barrier for organisations to adopt, and can accommodate a diverse set of deployment environments. They do, however, require appropriate technical specification on what should be disclosed, negotiated levels of detail when there are relevant commercial interests, consideration of incentives for innovation, and assessment of appropriate measures for monitoring and comparing risks.

These regulatory responses require collaboration with experts who can provide independent technical advice to design and evaluate transparency protocols and requirements. Take, for example, LLMs, which often produce factually incorrect statements and provide both logically and factually incorrect explanations; the content that they produce generally should not be trusted (Antoniak et al 2023). Disclosure of LLMs aid in facilitating user education, help avoid the downstream harm of LLMs' potential influence and provide people with warnings about simulated feelings, personality, and relationships. In sum, disclosure can have spillover benefits. As such, these requirements should be necessary costs associated with development and deployment of AI.

2.3 AI Prohibitions

The Discussion Paper poses questions about prohibiting the use of AI in certain situations in Australia, acknowledging that some states have banned ChatGPT in schools. It also highlights how the European Commission's proposed AI Act has scope for banning AI technologies in contexts where risks are deemed unacceptable. When AI are implemented into social systems, they necessarily inherit the histories and dynamics of those spheres. In domains that are highly consequential and carry strong histories of inequality (e.g., criminal justice, policing), AI tools are likely to entrench and amplify existing patterns. In these circumstances, and others in which inequality reproduction and other social ills prove intractable, it is vital that AI prohibitions can be enforced, including the ability to dismantle them when harms are identified.

We support the prohibition of AI types classified as unacceptable risk. In addition to the categories endorsed by the draft EU AI Act, the U.S. Blueprint for an AI Bill of Rights states that "monitoring should not be used in education, work, housing, or in other contexts where the use of such surveillance technologies is likely to limit rights, opportunities, or access" (White House OSTP, 2022). Such issues warrant careful attention and the prospect of prohibition if risk mitigation strategies do not counteract harms (discussed further in the next section). Enforceable prohibitions are not only important for counteracting specific AI harms, but also in demonstrating consequences for using AI in ways that undermine positive societal outcomes.

3. Cultivating Critical Awareness of Regulatory Responses to Support Domestic Initiatives

The Discussion Paper acknowledges both general (laws that apply across industries) and sector-specific regulations, international developments, and complexities underpinning governance environments. It considers how multiple actors already regulate and shape the development, deployment, and use of AI technologies. While a range of regulatory approaches and tools are both available and necessary to encourage safe and responsible AI, risk management is a central feature of proposed responses in the Discussion Paper. We suggest a wider appreciation and critical assessment of regulatory responses to support the development of domestic initiatives.

3.1 Broadening the Recognition of Different AI Risks and Harms

We appreciate the increasing recognition of AI-related risks, both domestically and globally. Addressing AI-related risks and harms, however, should involve dynamic, iterative, and informed modes of assessment with clear overarching governance principles to guide action and decision-making. This requires attention to the significant technical variation subsumed under the "AI" umbrella and attending to AI in its many and diverse forms. We point here to other governing entities that specify relevant AI risks vis-à-vis a range of technical conditions and recommend a similar approach in Australia (see discussion of the EU AI Act and NIST framework in Annex A).

We recognise that full knowledge of technological conditions can be challenging due to the 'black boxed' nature of many AI and ML systems, with inner workings that remain hidden or obscured. However, most 'black boxes' have at least partial windows, or avenues by which partial windows can be opened. Significant scholarly attention focuses on opening technological black boxes and rendering them observable. These efforts, which require extensive skill and intensive labour, are essential for risk and harm assessments that achieve both accuracy and breadth. They must be supported and incentivised for researchers in the field.

Increased technical transparency serves, and will be served by, careful consideration of social circumstances that may not be immediately obvious when assessing risks and harms. The Discussion Paper accounts for levels of risk (as is increasingly common in proposed frameworks), but does not account for important social factors, such as:

1. Aggregated risk, which might not be visible in specific applications or individual situations but nonetheless emerges across a larger population, at scale, or over time;
2. The variety of harms enabled by AI (for an overview, see Shelby et al., 2023), some of which are not captured by formal risk assessments or can be inherent to AI and its development (e.g., model training labour, which is not a focus of the Discussion Paper);
3. Strong governance objectives and principles, as any effective regulatory approach requires the identification of clear public-interest aims, mechanisms for community consultation, appropriate checks and balances, the delineation of enforcement powers, and the ability to apply meaningful sanctions.

3.2 More Robust Regulatory Approaches Informed by Evidence-Informed Approaches

While identifying AI risks is an important step in developing a regulatory framework, the Discussion Paper's articulation is limited. Evidence-informed approaches and concepts, studied

and refined over decades by regulatory governance scholars, demonstrate that voluntary or self-regulatory approaches have limited impact if there are not strong incentives for compliance and binding forms of co-regulation (see, e.g., Drahos, 2017). These concerns are particularly relevant for AI, where industry-led voluntary measures and self-regulation are commonplace but often reflect corporate interests rather than public interests.

We thus recommend adopting more explicitly human-centred values as the underpinning norms for safe and responsible AI in Australia. This approach would mean prioritising social benefits for the diverse groups that make up our society, not simply promises of innovation, efficiency, or market gains. Regulation would entail developing mechanisms that would incentivise and require AI adopters and developers to show how the use of these technologies support positive societal outcomes. The Department of Industry, Science and Resources is well positioned to help in steering private and public actors to support the pursuit of safe and responsible AI for all Australian residents.

In addition to developing clear governance expectations, there are many existing mechanisms to leverage, which are domestic and transnational in nature. Strategies can — and should — vary by context, drawing on technical standards (e.g., ISO/IEC JTC 1/SC 42 standard series), certification programs (e.g., CSIRO Responsible AI Pattern Catalogue), domain-specific authorities already adapting their rules and practices to account for the impact of AI (e.g., ACMA, TGA), and models and practices from other jurisdictions. Incorporating new measures, some of which have been proposed elsewhere, could include, but are not limited to:

- Strengthening human rights protections in relevant legislation;
- Creating processes to enable greater public oversight and an Ombudsman to ensure authorities are informed about AI-related harms;
- Developing AI safety toolkits to support testing frameworks and responsive risk assessment;
- Tracking case studies (contemporary and historical) on AI-related challenges, including lessons from regulatory successes and failures (see Annex B for example topics); and
- Investing in and delivering socio-technical training to enhance critical AI literacy and regulators' capabilities.

These individual strategies require coordination, harmonisation, and interdisciplinary expertise, not simply risk-based frameworks, to guide them. Australia's AI Ethics Principles offer a foundation, but they will not materialise without stronger forms of governance. Fortunately, there are evidence-informed regulatory concepts and practices that can be adapted to enhance Australia's AI governance landscape so that it supports safe and responsible AI.

ANNEX A – Regulatory Definitions and Identified Challenges Specific to Types of AI

In June 2023, the members of the European Parliament commenced negotiating positions on the EU AI Act (2021). Section 5.2.1 of the EU AI Act states that the definition of AI should be as technology neutral and as future proof as possible. Annex I of the proposed EU AI Act (2021) recognises the diversity of AI technologies — past, present, and state-of-the-art. Different AI technologies can pose challenges ranging from traditional and well-studied to ill-posed and uncertain.

Title I in the EU AI Act (2021) defines an “artificial intelligence system" as generating content, predictions, recommendations, or decisions. The decision-making aspects of AI distinguish it from traditional considerations of privacy, liability, responsibility, and regulation. This lies in contrast to technology that is primarily used as instruments by human decision-makers.

A risk management framework has been compiled by NIST. This U.S. organisation had its origins in controlling standards of weights and measures, perhaps reflecting on the balancing act necessary with AI's benefits and risks. Appendix B of the document by Tabassi et al (2023) outlines the uniqueness of AI challenges. Compared to traditional software, AI specific risks are new or increased. Particularly applicable to machine learning-based variants of AI that are trained on data, these touch upon questions of bias, trust, data dependency, scale, black box nature, unpredictability, and reproducibility. These systems also make it difficult to apply traditional software-oriented strategies like testing, documentation, and verification.

Not all the of categories of AI technologies in Annex I, EU AI Act (2021) uniformly share the risks highlighted in the Tabassi et al. (2023) AIRMF Appendix B. Annex I (a) can broadly cover ChatGPT, LLMs, Generative AI, and machine learning (ML). These approaches heavily rely on data and pose challenges in applying traditional verification and risk management strategies.

AI scientists have made progress in providing frameworks for acknowledging and identifying risks of AI research. This is enforced through the open dissemination of key attributes of ML — AI model cards proposed by Mitchell et al (2019) and dataset datasheets proposed by Gebru et al. (2021). AI scientists are also actively working on mitigating the challenges highlighted in Tabassi et al. (2023) AIRMF Appendix B, for instance, making AI models less of a black box by enhancing transparency, explainability, and trust.

ANNEX B – Examples and Precedents that Provide Regulatory Insights and Lessons

Surgical Robot Usage

Surgical robot usage dates back to the 1980s. They are representative of critical applications with heightened risks. Surgical robots can possess varying degrees of autonomy and "intelligence" as described in Attanasio et al (2021). In the EU, the popular Da Vinci surgical robot is classified as a Class IIb medical device. Barattini et al (2019) describe the process of certification by a "notified body" that audits the quality system of the manufacturer, reviews the product and technical documents to assess the safety of the product. In the United States, the Food and Drug Administration requires manufacturers to demonstrate the medical benefits and safety of the product, including the possibility of clinical trials. In Australia, the Da Vinci robot is put in Class IIb by the Therapeutic Goods Administration and requires an audit of the product with clinical evidence, risk management reports, and performance data.

Autonomous Vehicles

Autonomous vehicles are an example of risk-prone applications of so-called 'intelligent' driving systems. This area has been influenced by big corporations and received significant media attention. In California, USA, the Department of Motor Vehicles controls the deployment of autonomous vehicles. Manufacturers need to disclose operational details including design domains, limitations, and publicly disclosed assessment of safety. In Australia, the National Transport Commission (2021) has presented a policy paper on regulatory frameworks for automated vehicles. They suggest regulations for controlling the entry of a new product into the market, compliance, enforcement, and liability.

Safety-Critical Software Systems

Software systems have been used in rigorously verified applications like aviation. Safety-critical systems like flight control software must comply with the DO-178C standard developed by Radio Technical Commission for Aeronautics and used by regulators in the US and EU. The applicable formal verification and model checking closely related to Annex I (b), EU AI Act (2021). These AI technologies rely primarily on rules and logic.

Guidance on Generative AI

In July 2023 the Digital Transformation Agency (DTA) and the Department of Industry, Science and Resources (DISR) released interim guidance on government use of publicly available Generative AI platforms. China has also recently released interim measures for the management of Generative AI.

Labour Demands to Regulate AI

The Writers Guild of America Strike (2023) conveys clear proposals around regulating AI, including: "Regulate use of artificial intelligence on MBA-covered projects: AI can't write or rewrite literary material; can't be used as source material; and MBA-covered material can't be used to train AI." In mid-July 2023, the SAG-AFTRA have joined in striking, demanding, "protections for members against misuse of artificial intelligence, as well as a definition of acceptable use of the technology." Outcomes from these actions will have implications for the intersection of AI and creative professions, as well as for labour rights. These applications present distinct risk characteristics that are more cross-cutting with the emergence of broader, often unforeseen network effects.

Limiting Data Access for Generative AI

In July 2023, Twitter temporarily imposed a 'rate limit' on access to combat data scraping as reported by Q.ai, Forbes (2023). This is described as a response to Generative AI technologies using data obtained from the internet. Driven by the effect of AI companies, a similar change of restricting erstwhile 'public' data access on the internet played out on Reddit, as reported by Mike Isaac, New York Times (2023). While these actions are from commercial players rather than governments, they make clear how data access and data ownership are imminent regulatory targets that require systematic attention.

REFERENCES

Annex I, Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts (EU AI Act Annex I), (2021). https://artificialintelligenceact.eu/annexes/

Antoniak, M., Lucy L., Sap, M., & Soldaini, L. (2023). Using large language models with care. https://blog.allenai.org/using-large-language-models-with-care-eeb17b0aed27

Attanasio, A., Scaglioni, B., De Momi, E., Fiorini, P., & Valdastri, P. (2021). Autonomy in surgical robotics. Annual Review of Control, Robotics, and Autonomous Systems, 4, 651-679.

Autonomous vehicle deployment program. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-deployment-program/

Barattini, P., Vicentini, F., Virk, G. S., & Haidegger, T. (Eds.). (2019). Human-robot interaction: Safety, standardization, and benchmarking. CRC Press.

Biddle, N. (2023). Views of Australians towards science and AI ANU centre for social research and methods. The ANU Centre for Social Research and Methods. https://csrm.cass.anu.edu.au/sites/default/files/docs/2023/7/Views_of_Australians_towards_science_and_AI_-_For_web.pdf

Blueprint for an AI bill of rights. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Braithwaite, V. (1995). Games of engagement: Postures within the regulatory community. Law & Policy, 17(3), 225-255.

Braithwaite, V. (2014). Defiance and motivational postures. In Encyclopedia of criminology and criminal justice. Springer.

Da Vinci by Intuitive. https://www.intuitive.com/en-gb/products-and-services/da-vinci

Davis, J. L., Williams, A., & Yang, M. W. (2021). Algorithmic reparation. Big Data & Society, 8(2). https://doi.org/10.1177/20539517211044808

Device technologies Australia pty ltd - Robot, surgical, operation unit (97348). https://www.tga.gov.au/resources/artg/97348

DO-178() https://www.rtca.org/do-178/

Drahos, P. (2017). Regulatory theory: Foundations and applications. ANU Press.

Ensuring American leadership in automated vehicle technologies: Automated Vehicles 4.0 https://www.transportation.gov/av/4

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A. (2023). Trust in Artificial Intelligence: A global study. The University of Queensland and KPMG Australia. 10.14264/00d3c94

Hendrycks, D., Mazeika, M., & Woodside, T. (2023). An overview of catastrophic AI risks. arXiv preprint arXiv:2306.12001.

Hendrycks, D., Carlini, N., Schulman, J., & Steinhardt, J. (2021). Unsolved problems in ML safety. arXiv preprint arXiv:2109.13916.

Interim guidance for agencies on government use of generative AI platforms. https://architecture.digital.gov.au/guidance-generative-ai

Isaac, M. (2023). Reddit wants to get paid for helping to teach big A.I. systems. New York Times. https://www.nytimes.com/2023/04/18/technology/reddit-ai-openai-google.html

Mądry, A. (2023). Testimony for US Senate subcommittee on human rights and the law - Artificial Intelligence and human rights. https://www.judiciary.senate.gov/committee-activity/hearings/artificial-intelligence-and-human-rights

MEPs ready to negotiate first-ever rules for safe and transparent AI. https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229).

Montgomery, C. (2023). Testimony in US Senate hearing on oversight of AI: Rules for Artificial Intelligence. https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence

Pan, A., Chan, J. S., Zou, A., Li, N., Basart, S., Woodside, T., ... & Hendrycks, D. (2023, July). Do the rewards justify the means? measuring trade-offs between rewards and ethical behavior in the machiavelli benchmark. In International Conference on Machine Learning (pp. 26837-26867). PMLR.
Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts (EU AI Act), (2021), COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
Q.ai. (2023). Has Elon Musk killed Twitter? New rate limit leaves users fuming. Forbes. https://www.forbes.com/sites/qai/2023/07/04/has-elon-musk-killed-twitter-new-rate-limit-leaves-users-fuming/?sh=525d5aba4655
The regulatory framework for automated vehicles in Australia: Policy paper. https://www.ntc.gov.au/sites/default/files/assets/files/NTC%20Policy%20Paper%20-%20regulatory%20framework%20for%20automated%20vehicles%20in%20Australia.pdf
Rijke, J., Brown, R., Zevenbergen, C., Ashley, R., Farrelly, M., Morison, P., & van Herk, S. (2012). Fit-for-purpose governance: A framework to make adaptive governance operational. Environmental Science & Policy, 22, 73-84.
SAG-AFTRA television, theatrical and streaming contracts expire without a deal. https://www.sagaftra.org/sag-aftra-television-theatrical-and-streaming-contracts-expire-without-deal
Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., ... & VIRK, G. (2023). Identifying sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. arXiv preprint arXiv:2210.05791.
Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., ... & Dafoe, A. (2023). Model evaluation for extreme risks. arXiv preprint arXiv:2305.15324.
So, W., Lohia, P., Pimplikar, R., Hosoi, A. E., & D'Ignazio, C. (2022, June). Beyond fairness: Reparative algorithms to address historical injustices of housing discrimination in the US. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 988-1004).
Tabassi, E. (2023). AI risk management framework. National Institute of Standards and Technology. https://doi.org/10.6028/nist.ai.100-1
WGA negotiations---Status as of May 1, 2023. https://www.wgacontract2023.org/the-campaign/wga-negotiations-status-as-of-5-1-2023