Submission to the Forum on Information and Democracy's Working Group on

AI and its Implications for the Information and Communication Space

IT for Change

November 2023

Table of Contents

IT for Change's Submission to the

Forum on Information and Democracy's Working Group on

AI and its Implications for the Information and Communication Space[1]

Introduction

Artificial intelligence (AI) has ushered in a transformative era in the information and communication space, redefining the way we process, disseminate, and interact with information. Algorithms power search engines and recommendation and content curation systems, enhancing the efficiency of information retrieval and customization. Natural language processing capabilities enable AI to understand and generate human-like text, contributing to the development of advanced chatbots and virtual assistants that facilitate seamless communication. Moreover, AI plays a pivotal role in analyzing vast datasets, enabling more accurate predictive analytics and data-driven decision-making. While these capabilities have benefits in many use cases, their deployment without sufficient guardrails has resulted in the worrying increase in disinformation, misinformation, and deep fakes, amplifying the dangers of information chaos and putting increasing strain on our public sphere, democratic institutions, and social life. The UN Secretary-General in his recent policy brief on Information Integrity on Digital Platforms flagged advances in AI – including image generators and video deepfakes – as a major threat to information integrity today.[2]

In this submission, we highlight certain issues and provide recommendations addressing the implications of AI on the information and communication space in relation to (i) the development and deployment of AI systems; (ii) accountability and liability in AI systems; and (iii) global cooperation and international governance of AI.

[1]This submission builds on the work of IT for Change and is authored by Merrin Muhammed Ashraf, with direction and review from Anita Gurumurthy.
[2]UN Secretary-General. (2023). *Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms.* United Nations. https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf

1. Development and Deployment of AI Systems

AI systems and recent advancements in the form of generative AI (Gen AI) and Large Language Models (LLMs) have had a disruptive effect on the information and communication space. The democratization of synthetic content creation at scale[3] has lowered trust in the information ecosystem, thereby jeopardizing public health,[4] public safety and order,[5] the integrity of electoral process,[6] and efforts to mitigate climate change.[7] AI models can be potent in terms of the risks they pose for information integrity[8] not only because they increase the possibilities for generating disinformation, misinformation, and hate speech (passed off convincingly as facts), but also since they facilitate the rapid and targeted dissemination of such content, that too, at scale, by malicious actors.[9] By influencing the creation, dissemination, reach, and consumption of news and information, AI systems have implications not only for individuals' access to information and freedom of expression, but also for human dignity,[10] personal autonomy, and the capacity of individuals to make free and informed decisions, especially in their role as citizens.[11]

This is not all. In the overall context of platform capitalism, the viral dissemination at scale of disinformation and hate speech is a negative externality of the prevalent business model of most digital platforms and it affects trust in public institutions, generates political and economic instability, and erodes democracy. The unaccountable power of platform companies and their privatized algorithmic apparatus thus have individual and structural consequences for democracy.

While AI holds immense potential to improve access to information and knowledge, detect disinformation, and regulate online content, the actual realization of this potential is limited by the black-box logic that governs the working of today's AI systems, the unreliability and biases in datasets, and the different

[3]Diresta, R., & Willner, D. (2023). *White House AI Executive Order Takes On Complexity of Content Integrity Issues.* Tech Policy Press. *https://techpolicy.press/white-house-ai-executive-order-takes-on-complexity-of-content-integrity-issues/*
[4]Davey, M. (2023). *'Alarming': Convincing AI vaccine and vaping disinformation generated by Australian researchers.* The Guardian. https://www.theguardian.com/australia-news/2023/nov/14/alarming-convincing-ai-vaccine-and-vaping-disinformation-generated-by-australian-researchers
[5] Center for Countering Digital Hate. (2023). *Misinformation on Bard, Google's new AI chat.* https://counterhate.com/research/misinformation-on-bard-google-ai-chat/#about*;* Hsu, T., & Thompson, S.A. (2023). *Disinformation researchers raise alarms about A.I. chatbots.* The New York Times. https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html
[6]Panditharatne, M., & Giansiracusa, N. (2023). *How AI Puts Elections at Risk — And the Needed Safeguards.* Brennan Centre for Justice. https://www.brennancenter.org/our-work/analysis-opinion/how-ai-puts-elections-risk-and-needed-safeguards
[7]Galaz, V.,et al. (2023). *Climate misinformation in a climate of misinformation*. Research brief. Stockholm Resilience Centre (Stockholm University) and the Beijer Institute of Ecological Economics (Royal Swedish Academy of Sciences). http://arxiv.org/abs/2306.12807
[8]UN Secretary-General's Policy Brief on Information Integrity on Digital Platforms (supra 1) defines information integrity as "accuracy, consistency, and reliability of information".
[9]UN Secretary-General. (2023). *Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms.* United Nations. https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf
[10]European Data Protection Supervisor. (2018). *Ethics Advisory Group Report 2018.* https://edps.europa.eu/data-protection/our-work/publications/ethical-framework/ethics-advisory-group-report-2018_en
[11]Ibid.

priorities of the actors controlling the development and deployment of these systems,[12] often private actors with undue media dominance or concentration, all of which are at odds with an enabling environment for media diversity and independent voices.

In addition to the issues related to reliability and fair representation in datasets, the en-masse appropriation of text, visual, and audio material available in the public domain for the training of Gen AI models also raises concerns about creating knowledge enclosures. Open datasets are often used to build proprietary and closed systems, with little to no benefit, economic or otherwise, to the people or community whose data is used.[13] Further, "it is likely that Gen AI will undercut efforts to preserve and maintain traditional knowledge practices, as it continues to be used as a knowledge production and management technology."[14]

Recommendations

a. Classification, Assessment, and Mitigation of Risks from AI Systems

Understanding and mitigating the risks of AI systems deployed in the information and communication space is crucial to safeguard information integrity, protect civil liberties and rights, and enhance public trust in the information ecosystem. The interest in balancing risk mitigation with the need to foster the innovation potential of AI has given rise to risk-based classification of AI systems and proportionate regulatory requirements.

When it comes to assessing the risks of AI systems in the information and communication space, there needs to be a life-cycle approach, meaning that the risks of AI systems should be assessed not just at the development stage, but also at the stage of deployment and the subsequent life of the system.[15] This is important in view of the fact that many AI models – like LLMs and General Purpose AIs (GPAIs) used in the information and communication space – are generic in nature and their impact will depend on the specific context in which downstream deployers use or adapt it, and the individuals and groups that interact with or are affected by the system. As Prof. Lilian Edwards notes, "AI products are dynamic, not static products – their behavior (and successful implementation) will change with new data, new uses, and new

[12]UNESCO. (2020). *Artificial Intelligence: Media and Information Literacy, Human Rights and Freedom of Expression*. https://unesdoc.unesco.org/ark:/48223/pf0000375983/PDF/375983eng.pdf.multi
[13]UNCTAD Secretariat. (2023). *Issues Paper on Data and Development.* https://unctad.org/system/files/information-document/CSTD2023-2024_Issues01_data_en.pdf
[14]IT for Change. (2023). *IT for Change's Submission to the U.S. Copyright Office on Artificial Intelligence Study.* https://www.regulations.gov/comment/COLC-2023-0006-9159 .
[15]OECD. (2022). *OECD Framework for the Classification of AI Systems.* https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1700046307&id=id&accname=guest&checksum=A316B731523E849F0A1680C7B146F50D

integrations, which in turn changes their risk profiles and requires continuous evaluation."[16] Therefore, a risk-based approach must assess the risk of AI technology in the context of specific uses and applications, rather than the risk of the technology in the abstract.

For a model of risk-based assessment and classification of AI systems, the OECD Framework for the Classification of AI systems is recommended. As per this framework, AI systems can be classified into different risk categories based on four dimensions:

1. Context: The context in which an AI system is developed and deployed, including stakeholders that deploy an AI system, the stakeholders impacted by its use, and the sector in which an AI system is deployed.
2. Data: This includes data classifiers used, the source of the data, its structure, scale, and how it was collected.
3. Type of algorithm: The degree of transparency and/or explainability, robustness, and implications for human rights, privacy, and fairness depend on the type of model as well as the model-building and inferencing processes. For example, systems using neural networks are often seen as having the potential to provide comparatively higher accuracy but less explainability than other types.
4. Task: The kind of task to be performed and the type of output expected vary across AI systems, from forecasting and content personalization to detection and recognition of voice or images.[17]

These dimensions are applied to different use cases of AI to understand the risks under each dimension and thereby classify the relevant systems as 'unacceptable risk, high-risk, medium-risk, and low-risk'. Further, while some of these dimensions are more relevant to AI 'in the lab', i.e., AI system's conception and development, others are more relevant to AI 'in the field', i.e., use and evolution of an AI system after deployment.[18] In this way, the Organisation for Economic Co-operation and Development's (OECD) framework of risk classification allows a life-cycle approach to AI risk assessment and enables the institution of appropriate ex-ante as well as ex-post regulation of AI systems.

[16]Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions.* Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf
[17]OECD. (2022). *OECD Framework for the Classification of AI Systems.* https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1700046307&id=id&accname=guest&checksum=A316B731523E849F0A1680C7B146F50D; Also see, Singh, N. (2021). *Technology Regulation: Risk-based approaches to Artificial Intelligence governance, Part II.* CCG Blog. https://ccgnludelhi.wordpress.com/2021/06/23/technology-regulation-risk-based-approaches-to-artificial-intelligence-governance-part-ii/
[18]OECD. (2022). *OECD Framework for the Classification of AI Systems.* https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1700046307&id=id&accname=guest&checksum=A316B731523E849F0A1680C7B146F50D

There should also be clear reviewable criteria to assess risks under each of these dimensions in specific use cases or scenarios.[19] These criteria could mirror those listed under Article 7 of the EU's Artificial Intelligence Act (AI Act) which include: the extent to which the use of an AI system has caused or is likely to cause harm to the health and safety or adverse impact on the fundamental rights; the extent to which potentially harmed or adversely impacted persons are dependent on the outcome produced with an AI system or are in a vulnerable position in relation to the user of an AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances, or age; and extent to which outcome produced is easily reversible.[20]

When it comes to the creation, dissemination, and consumption of information, the most serious risks in this sector relate to the behavioral effects of AI on personal autonomy (e.g., addictive design and exploitative targeting) and the resulting effects on public discourse. These include the impact of AI-enabled political targeting of fair democratic debate and the spread of mis/disinformation and violent and extreme content.[21] Evaluations for high-risk AI use by large search engines and social media companies should also include their potential impacts on journalism and information-sharing, including the spread of harmful content or the burying of legitimate news online.[22]

Apart from ex-ante assessment, there should also be ex-post assessment of impact of such AI systems on fundamental rights, particularly, freedom of expression, access to information, right to human dignity, and privacy. Further, the impact on fundamental rights should be assessed not just at an individual level, but at a societal level. Research shows that AI development teams tend to be non-diverse and rarely include representation from the marginalized groups most impacted by biased or unfair systems.[23] A report by the Future of Life Institute argues that recommendation algorithms "may cause societal-level harms, even when they cause only negligible harms to individuals" (by, for instance, tipping the balance in an election by discouraging wavering voters from turning up).[24] Further, algorithmic systems construct new groups whose commonalities do not easily fit into existing rules for discrimination and protected characteristics.[25]

---

[19]Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions.* Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf
[20]European Union. (2021). Article 7, *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*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[21]Centre for Data Ethics and Innovation. (2020). *AI Barometer Report.* https://assets.publishing.service.gov.uk/media/5ef0c4c1e90e0741f91db12d/CDEI_AI_Barometer.pdf
[22]Ibid.
[23]Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions.* Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf
[24]Future of Life Institute. (2022). *FLI Position Paper on the EU AI Act.* https://futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x76795
[25]Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions.* Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf

Hence, risk assessment and fundamental rights impact assessment of AI systems should take into account "risks to groups and communities; individual and structural discrimination caused by contexts of deployment; environmental impacts; effectiveness; transparency; contestability; and the views and wishes of end users and affected communities."[26]

Further, as UNESCO recommends, the process of risk assessment and monitoring should involve "broad participation of all stakeholders, including, but not limited to, vulnerable people or people in vulnerable situations. Social, cultural and gender diversity should be ensured, with a view to improving learning processes and strengthening the connections between findings, decision-making, transparency, and accountability for results."[27] To realize this, appropriate processes should be instituted to ensure the involvement of civil society and the general public in risk assessment and monitoring in an effective manner.

Once AI systems are categorized based on risks, high-risk systems could be subject to more intensive standards like ensuring the quality of datasets used, technical documentation and record-keeping, transparency and provision of information to users, human oversight, and robustness, accuracy, and cybersecurity. Additionally, providers of AI systems[28] should document the development process, intended applications, and limitations of their models, and provide this valuable information to deployers or downstream users.

Periodic regulatory impact assessments should also be conducted to enable the government and regulators to understand whether the regulatory framework has been effective to achieve the desired objective.[29] The OECD defines Regulatory Impact Analysis as a "systemic approach to critically assessing the positive and negative effects of proposed and existing regulations and non-regulatory alternatives".[30]

While a risk-based approach to the classification and regulation of AI systems is useful, we should be cautious about legitimizing a check-box approach to risk assessment and compliance, which can drive AI models to the lowest common denominator. To guard against this tendency, it is essential that AI regulation be rooted in the principle that "rights are non-negotiable and they must be respected regardless of a risk level associated with external factors".[31] A risk-based approach should not be seen as an

[26]Ibid. Also see, Friedrich-Ebert-Stiftung. (2023). *China's Regulation on Algorithms.* https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf
[27]UNESCO. (2021). *Recommendations on the Ethics of Artificial Intelligence.* https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi
[28]By the term, 'providers of AI systems', we mean those who develop an AI system with a view to place it in the market or put it into service under their own name or trademark.
[29]Singh, N. (2021). *Technology Regulation: Risk-based approaches to Artificial Intelligence governance, Part II.* CCG Blog. https://ccgnludelhi.wordpress.com/2021/06/23/technology-regulation-risk-based-approaches-to-artificial-intelligence-governance-part-ii/
[30]OECD. (n.d.). *Regulatory Impact Assessment.* https://www.oecd.org/regreform/regulatory-policy/ria.htm
[31]Fanny, et al. (2021). *The EU should regulate AI on the basis of rights, not risks.* Access Now. https://www.accessnow.org/eu-regulation-ai-risk-

alternative to a rights-based and rights-respecting approach, but rather as a scalable and proportionate approach to compliance.[32] Adopting a precautionary principle, the burden to prove that the AI system in question does not violate fundamental human rights should be on the provider or deployer of the system, as the case may be.[33]

b. Intellectual Property

Training and data mining- (TDM) related activities serve as a critical step towards developing Gen AI models. These activities include the use of complete and exact replicas of original work, not just work protected by copyrights but also public domain knowledge that lack intellectual property protection, including the traditional knowledge commons.[34] As IT for Change submitted in its comments to the US Copyright Office: "Gen AI systems are *stochastic or agere sine intelligence* in nature, i.e., they act without understanding exactly what they are generating as a result. In effect, the Gen AI system simply creates a layer of noise in order to construct art or text. This, however, is a reconstruction of the art or text that is part of its training data. A grant of copyright on Gen AI outputs, would, therefore, render indigenous artists without the means to exercise their intellectual property rights, making them vulnerable to knowledge theft and misappropriation."[35]

To safeguard the plurality and diversity of knowledge and information in society, it is important to safeguard indigenous and traditional art and knowledge from being reduced to mere training data for AI models and appropriated by private actors that control them. Firstly, in line with the predominant anthropocentric approach of the copyright law, AI-generated work should not be classified as copyrightable as there is no involvement of personhood. Even if the approach of copyright law changes with the broadening of the notion of personhood, AI-produced works should not be granted copyright without examining the implications of fair use for the intellectual and knowledge commons.[36]

Secondly, the social and economic value generated from work produced by AI trained on datasets that include indigenous art and traditional knowledge should be fairly distributed to the relevant community. A collective licensing regime could be instituted to limit the reuse of work in violation of the cultural commons. The services of a collective management organization (CMO) can be utilized to negotiate

based-approach/

[32]Article 29 Data Protection Working Party. (2014). *Statement on the role of a risk-based approach in data protection legal frameworks.* https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf

[33]World Commission on the Ethics of Scientific Knowledge and Technology. (2005). The Precautionary Principle. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000139578

[34]IT for Change. (2023). *IT for Change's submission to the U.S. Copyright Office on Artificial Intelligence Study.* https://www.regulations.gov/comment/COLC-2023-0006-9159 .

[35]Ibid.

[36]Ibid.

licenses for a specific class of work (such as images) for a specified category of workers, regardless of whether they are members of the CMO.[37]

Thirdly, the new ethical standard of fair learning that is proposed to evaluate AI models that use copyrighted material should also account for traditional and indigenous knowledge as well. In other words, the risk of cultural appropriation and enclosure must also be a significant factor in the assessment of 'fair learning'.[38]

c. Integrity, Fairness and Public Nature of AI Datasets

To mitigate the problem of bias, discrimination, and exclusion in the datasets used to train AI models, it is necessary to institute mechanisms to ensure audit and oversight by an independent regulatory authority to verify the accuracy and integrity of datasets. For instance, China's Interim Measures for the Management of Generative Artificial Intelligence Services also provides for content moderation to ensure the training dataset does not contain any illegal material.[39] There should be effective oversight and monitoring over this content moderation activity as otherwise, it will give the system owner significant power to decide what material is harmful and what is permitted.[40] While synthetic data holds the potential to correct gender and race-based data gaps and bias in training datasets, "Regulators must also note that synthetic data isn't free from bias. In fact, mimicked data creates a feedback loop of biases, as the flaws in the original dataset quickly replicate during both generation and discrimination."[41]

As much as it is important to ensure the integrity and accuracy of datasets, it is also important to ensure that the development and deployment of AI tools do not create data enclaves or enclosures controlled by a few, large digital technology firms.[42] As Kean Birch notes, "the value of digital data for a firm do not reflect ownership and property rights per se, but rather diverse modes of access and use restrictions created through economies of scale, network effects, intellectual property, limited interoperability, contractual arrangements, etc." As a result, data access is not available as a widely available resource, and this is not conducive to the development of small-scale and independent AI tools by diverse actors. To overcome this, a commons approach to management of publicly available datasets, as proposed by Tomasso Fia, is

[37]Ibid.
[38]Ibid.
[39]China Law Translate. (2023). *Interim Measures for the Management of Generative Artificial Intelligence Services.* https://www.chinalawtranslate.com/en/generative-ai-interim/
[40]Norwegian Consumer Council. (2023). *Ghost in the Machine: Addressing Consumer Harms of Generative AI.* https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf
[41]Sahai, A. (2023). *Artificial Data for Artificial Intelligence: Could This Be a Game Changer?* Bot Populi. https://botpopuli.net/artificial-data-for-artificial-intelligence-could-this-be-the-game-changer/; Competition & Markets Authority. (2023). *AI Foundation Models: Short Version.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1185507/Short_version_.pdf
[42]Birch, K. (2023). *Data Enclaves: Valuing Google's Data Assets.* IARIW – CIGI. https://iariw.org/wp-content/uploads/2023/10/Birch.pdf

needed.[43] According to Fia, grasping raw non-personal data as a commons means making data available to a wide number of actors for the fulfilment of fundamental rights and enhancement human flourishing.[44]

In pursuance of this commons approach to the management of public datasets used by AI models, strong institutional safeguards need to be put in place to protect social sector datasets, especially where there is a risk of proprietization of core development functions through AI models (such as in health, education, and welfare).[45] Further, access to public domain and open government data should be allowed only on a conditional basis, subject to purpose limitations and clear sunset clauses on use.[46] There should also be reciprocity guarantees in common data pools whereby private model developers who build on public data layers are mandated to share back and enrich the commons.[47]

2. Accountability and Liability in AI Systems

Fixing accountability and liability for the development and deployment of AI systems is fraught due to the risk of unexpected behaviors or outcomes from the AI system and due to the involvement of multiple actors controlling it at different stages of the life cycle. This is of particular concern in the information and communication space, which has witnessed the rising use of LLMs. General Purpose AI models that are technologically complex and opaque evolve as they are fed with more data, and deployed and adopted in different use cases, and involve often-complicated actor chains behind a consumer-facing Gen AI system.[48] The responsibility of news media and journalists using AI tools to generate content also raises concerns about harm to information integrity and the public's access to information.

The impact of AI on information integrity is further exacerbated by digital platforms like social media, search engines, media, etc., that host harmful and illegal AI-generated content and do not take measures to remove them and instead profit from its virality. Therefore, apart from providers and deployers of AI systems, there should also be accountability and liability regimes targeted at digital platforms that host AI-generated content and facilitate its amplification through platform affordances for monetary reasons.

Finally, for any accountability and liability regime for AI systems to be effective, it is important to empower users and subjects of AI systems with certain rights in relation to their direct and indirect interactions with

[43]Fia, T. (2021). *An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons.* Global Jurist 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3698914
[44]Ibid.
[45]Gurumurthy, A., & Bharthur, D. (2023). *Reframing AI Governance through a Political Economy Lens.* IT for Change. https://itforchange.net/reframing-ai-governance-through-a-political-economy-lens
[46]Ibid.
[47]Ibid.
[48]Norwegian Consumer Council. (2023). *Ghost in the Machine: Addressing Consumer Harms of Generative AI.* https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf

AI systems. This is necessary to level the information and capability asymmetry between them and the AI providers/deployers, and thereby seek redress for the rights violations and harms suffered due to the operation of the AI system.

Recommendations

a. Responsibility of Providers and Deployers for AI-generated Content

Providers of AI models must be required to foresee and prevent the harms caused by their deployment to provide goods and services, even when those products evolve in ways not specifically desired or foreseeable by their manufacturers. AI providers, particularly providers of high-risk AI, should be required to comply with some duty of care requirements such as creating a quality management system, including a risk management system, meeting data quality criteria, ensuring human oversight, accuracy, robustness, and cybersecurity measures, as well as creating technical documentation.[49] If a harm arises from an AI system and it can be shown prima facie that the provider failed to comply with the duty of care requirements, then there should be a presumption of causality in favor of the claimant. This reversal of the burden of proof is essential in light of the specific characteristics of AI, including complexity, autonomy, and opacity, that makes it prohibitively difficult and expensive for claimants to meet the requirements for a successful liability claim.[50]

To ensure the safe and fair use of AI systems in the information and communication space, the accountability and liability regime should target not just the providers of AI systems but also the deployers of such systems.[51] A deployer is a person or entity who puts into service or uses an AI system. Many of the duty of care requirements such as ensuring human oversight can only be effectively implemented in the deployment stage.[52] Therefore, both providers and deployers of general-purpose AI should share responsibility for assessing its conformity with fundamental rights and with the safety standards and essential requirements as legislatively prescribed.[53]

[49]European Union. (2021). *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*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[50]European Commission. (2022). *Proposal for a directive of the European parliament and of the council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0496
[51]Demircan, M. (2023). *Deployers of High-Risk AI Systems: What Will Be Your Obligations Under the EU AI Act?* Kluwer Competition Law Blog. https://competitionlawblog.kluwercompetitionlaw.com/2023/06/02/deployers-of-high-risk-ai-systems-what-will-be-your-obligations-under-the-eu-ai-act/
[52]Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions.* Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf
[53]Ibid.

Under the AI Act, some obligations are imposed on deployers of high-risk AI systems which includes using the high-risk AI system in accordance with the instructions of use, issued by the providers; complying with sectoral legislations, ensuring input data is relevant, monitoring the high-risk AI system's compliance with its own terms of use, keeping logs, and conducting data protection impact assessment.[54] Similarly, under Article 52, deployers are required to comply with extra transparency measures for certain AI systems such as emotion recognition systems, biometric categorization systems, and Gen AI models that produce image, audio, or video content that appreciably resembles existing persons.[55] However, critics of the AI Act have pointed out that these constitute only minimum obligations, and urged legislators to impose more obligations, especially to analyze the potential impact on fundamental rights, equality, accessibility, public interest, and the environment to consult with affected groups and to take active steps to mitigate potential harms.[56] Deployers of Gen AI in consumer-facing interfaces and services must be obligated to disclose how the generated content is influenced by commercial interests of providers, deployers, or third parties. This is particularly relevant when the content generated serves to inform consumer choices, such as content generated in the context of search queries.[57] Further, deployers of Gen AI systems must be obligated to disclose whenever consumers are interacting with a Gen AI system, and whether consumer-facing systems are using AI to affect the outcome of a decision.[58] Public and private organizations must be obligated to disclose whenever content has been generated by Gen AI, when that content may have an effect on decisions affecting consumers, consumer rights more broadly, or democratic processes.[59]

While the Council of Europe proposes the removal of liability for the upstream provider of general-purpose AI,[60] critics of this position point out that "responsibility cannot and should not be allocated to the deployer alone, since the power to control and modify such infrastructure, alongside technical resources, largely lies with the upstream provider."[61] Responsibility should be joint with the provider and, "a much more nuanced appraisal must be made of what duties should lie where at what point in time, and who is

[54]European Union. (2021). Article 29, *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*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[55]European Union. (2021). Article 52, *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*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[56]European Digital Rights. (2021). *Introduce obligations on users of high-risk AI systems.* https://edri.org/wp-content/uploads/2022/05/Obligations-on-users-AIA-Amendments-17022022.pdf
[57]Norwegian Consumer Council. (2023). *Ghost in the Machine: Addressing Consumer Harms of Generative AI*. https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf
[58]Ibid.
[59]Ibid.
[60]Council of the European Union. (2021). *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 – Progress report.* https://data.consilium.europa.eu/doc/document/ST-13802-2021-REV-1/en/pdf
[61]Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions*. Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf

empowered either legally or by practical control, power or access to data and models, to make changes".[62] Such a nuanced approach is particularly important to account for instances when deployers make substantial modifications to the high-risk AI system in a way that they effectively become producers of the AI system.

b. Responsibility of Platforms Hosting AI-generated Content

Digital platforms such as social media, search engines, media hosting platforms, etc., should institute mechanisms to proactively detect illegal and harmful content that is generated using AI tools and remove it or label it as appropriate. Platforms should also adopt measures such as triggering an internal viral circuit breaker to prevent the algorithmic amplification of AI-generated unlawful or harmful content. In addition to adopting suitable technology measures, the platforms should also engage human reviewers and collaborate with fact-checkers to determine the authenticity of a piece of content. Swift, accessible, and effective grievance redressal mechanisms should also be made available for users to report and demand the offensive piece of content to be removed. Platforms should not benefit from safe harbor protection if there is a systematic or deliberate failure and gross negligence on their part by continuing to host AI-generated unlawful or harmful content.

Digital platforms should strengthen the safety and ethics team by allotting them with necessary resources and investment, and should have in place long-term roadmaps to systematically address these issues.[63] Further, as Nadah Feteih writes, "While underrepresented employees at the company provide valuable perspective and context which helps identify problems affecting their communities, their reports and personal escalations should not replace sufficient testing and red teaming to find and fix issues preemptively."[64]

c. Rights and Duties of Users and Subjects of AI Systems

All fundamental human rights of users and subjects of AI systems should be safeguarded against the risks posed by AI systems, and in the context of information and communication space, the right of users to access diverse and plural information, right to personal autonomy, and right to privacy and dignity need to be specially safeguarded. Users and subjects of AI systems should also have the following rights: the right to be protected from unsafe and effective systems, the right to not be subjected to discrimination by algorithms, the right to be free from abusive data practices and to have agency over how data about them

[62]Ibid.
[63]Feteih, N. (2023). *When AI Systems Fail: The Toll on the Vulnerable Amidst Global Crisis.* Tech Policy Press. https://techpolicy.press/when-ai-systems-fail-the-toll-on-the-vulnerable-amidst-global-crisis/
[64]Ibid.

is used, the right to know that an automated system is being used and to understand how and why it contributes to outcomes that impact them, and the right to opt out of AI systems, where appropriate and to have easily accessible grievance redressal mechanisms.[65] While individual rights of users are important, it is also important to balance them with the right of the public to access plural and diverse information and knowledge, the right to democratic participation, and the interest in safeguarding information integrity and public trust in the information ecosystem.

Further, as Prof. Lilian Edwards argues, "Users must be given a chance to have their views considered both before the product is certified as valid to enter the market, as well as rights to challenge the legality of a system after it is placed on the market. Civil society, as the representatives of users, must be empowered and resourced to enter the standard-setting process on their behalf."[66]

In addition to safeguarding the rights of users and subjects of AI systems, to safeguard the integrity of the information and communication, it is also necessary for particular classes of users like journalists, researchers, and highly influential actors to have certain duties and high degree of responsibility when creating or distributing AI-generated content and follow full disclosure. This duty could be enforced through sector regulators or industry bodies.

3. International Governance of AI Systems

Enhanced global cooperation in the governance of AI is essential to achieve harmony in regulatory approaches, particularly in assessing risk and determining liability in order to ensure equal protection of human rights for citizens across jurisdictions. Competing visions of development also characterize national visions of AI and how the balance between individual rights and social good is calibrated.[67] However, it is imperative to prevent the fragmentation of the digital policy space and build global policy consensus for norm-building, and ensure equitable development and access to technology for all countries.[68]

Recommendations

Binding common standards and approaches through the development of a common global architecture should be evolved to guide future trajectories of AI design, development, and use.[69] There should be global

[65]The White House. (2022). *Blueprint for an AI Bill of Rights.* https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf
[66] Edward, L. (2022). *Regulating AI in Europe: four problems and four solutions.* Ada LoveLace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf
[67]Pomares, J., & Abdala, M. (2020). The future of AI governance. *Global Solutions Journal, 5.* https://www.global-solutions-initiative.org/wp-content/uploads/2020/04/GSJ5_Pomares_Abdala.pdf
[68]IT for Change. (2023). *Submission of Inputs for the Global Digital Compact by the Global Digital Justice Forum.* https://itforchange.net/submission-of-inputs-for-global-digital-compact-by-global-digital-justice-forum
[69]Ibid.

benchmarks and standards on due diligence and risk assessment, as well as the margin of error that is acceptable for different AI models and full disclosure of the same.[70] International standards from standards development organizations like the ISO/IEC and IEEE can help ensure that global AI systems are ethically sound, robust, and trustworthy, that opportunities from AI are widely distributed, and that standards are technically sound and research-driven, regardless of sector or application.[71] Further, joint platforms for regulatory sandboxes could also be explored.[72]

To ensure that citizens have access to equal rights and opportunities across territories, countries must ensure that AI providers based in their jurisdiction whose systems impact people outside of their jurisdiction are subject to the same requirements as those within their jurisdiction, or higher standards of requirements in case the other jurisdiction guarantees better protection and advanced safeguards. Compliance with algorithmic accountability and transparency regulation should be mandated as a precondition for transnational AI service providers to have market access, especially in developing countries.[73] Further, digital trade rules that prevent nation-states from enforcing transparency and accountability regulations on AI services and application providers and deployers (such as the prohibition of source code transfer) should be rejected.[74]

While a multistakeholder approach is essential for the international governance of AI, such an approach should create and uphold a democratic global digital governance regime through a multilateral, people-centered, public policy mechanism for internet, platform, data, and AI technologies, that are free from corporate capture.[75]

[70]These thresholds can be appropriately tiered to ensure that models with higher impact have greater requirements to fulfill. See, Gurumurthy, A., & Bharthur, D. (2023). *Reframing AI Governance through a Political Economy Lens.* IT for Change. https://itforchange.net/reframing-ai-governance-through-a-political-economy-lens
[71]Kerry, C.F., et al. (2021). *Strengthening international cooperation on AI.* Brookings, https://www.brookings.edu/articles/strengthening-international-cooperation-on-ai/
[72] Ibid.
[73]IT for Change. (2023). *Submission of Inputs for the Global Digital Compact by the Global Digital Justice Forum.* https://itforchange.net/submission-of-inputs-for-global-digital-compact-by-global-digital-justice-forum
[74]Ibid.
[75]Ibid.