Forum on Information & Democracy

Working Group on Artificial Intelligence and its implications for the Information and Communication Space

Call for Contributions

1/ Introduction

The rapid development and advancements in artificial intelligence (AI), generative artificial intelligence and even artificial general intelligence (AGI), and its application in all spheres of life has important implications for the global information and communication space. AI systems are becoming easily accessible to the general public and can be used with little or no programming knowledge. Generative AI tools such as ChatGPT, as well as image or video creation tools, enable anyone to become a creator of content at a speed and efficiency previously unseen. Predictive AI systems are playing a growing decision-making role in the curation, moderation and verification of information. Thus, AI systems are dramatically changing our information and communication space.

Launched by the Forum on Information and Democracy, the implementing organization of the International Partnership for Information and Democracy, and composed of 14 international experts, the Working Group will develop urgent policy recommendations. In its endeavor **the Working Group is calling upon experts worldwide to contribute in sharing their ideas for regulatory, self-regulatory and technological solutions to mitigate the risks of AI systems for the information space and incentivise positive impact**.

2/ Submit a contribution

2.1 Contributions

To address the implications of AI on the global information and communication space, we invite contributors to answer the questions below. Contributions do not need to cover all aspects listed below, but can be focused on specific ones. This list is not exhaustive and contributions from other areas of thought are welcome. Contributions can also take the form of already existing papers or opinion pieces on the topics listed below.

The working group's **preliminary definition of an AI system** is "*a machine-based system that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.*"

1. **AI Systems' Development and Deployment**

Forum on Information & Democracy

How can we effectively establish and implement *guardrails for the responsible development and deployment of AI systems impacting the creation, dissemination, and consumption of information? Are there any proposals or best practices we should look at?*

In addressing this, you might consider:

- How can we incentivize AI developers and deployers to adhere to **agreed norms or principles** during the design and deployment process and that the objectives of their systems are aligned with democratic values?
    - What role can Code of Conducts play in this, and how can we ensure their effectiveness?
    - How can we establish **incentives**, both regulatory and market-driven, to foster the responsible and ethical development of AI systems specifically designed for creating, distributing, and consuming information, with the aim of supporting the realization of a free, pluralistic, and democratic information space?

- What framework should be used to define different classes of risks in the information and communication space when **assessing AI systems**? Which criteria should be used (impact severity, reach, duration, probability, etc.)?
    - How should AI systems be categorized based on their risk profile and what specific obligations and restrictions should apply to each category?
    - What should fundamental rights impact assessments focus on and which risk mitigation strategies could be put in place?
    - Who should be involved in third-party conformity prior assessments and regular audits (independent certifiers, civil society, public, etc)? Are they needed for all AI systems?

- Should specific AI systems be treated as **basic infrastructure or public utilities**? Why or why not?
    - Is there merit in having a public dataset available for training of AI systems? If so, how could it be created and managed? How could we incentivize potential contributors?

- How should **intellectual property laws** adapt to keep pace with AI developments, and how can we enforce them?
    - Who should hold the copyright for AI-generated content, especially when it builds upon pre-existing intellectual property?
    - How can AI models be designed to ensure transparency and provide proper attribution to copyrighted work?

- How can AI's **data privacy and data protection** challenges be tackled?
    - Should users be afforded specific privacy and data protections when interacting with AI systems?
    - How can we uphold the principle of data minimization and safeguard rights such as the right to erasure and rectification?

- How can we ensure that **AI training datasets** are comprehensive and reflect the diversity of the intended user base?
    - Do you see a role for third-party compliance services in verifying the integrity and governance of AI training data sets?

- How can we enhance the **transparency, explainability, and contestability of AI systems**?
    - What disclosure practices should be put in place to ensure stakeholders access to AI system documentation, methodologies and data, which is necessary to assess their societal impacts? What access levels are appropriate for various actors (public, AI Authority, researchers, auditor, etc.)?
    - What mechanisms can be established to guarantee corrections or modifications of AI systems and their models when risks are detected?
    - Do you believe that every publicly-released generative-AI system should have an associated detection mechanism to identify content generated by it? What alternative technological solutions could address the issue of authenticity and verification?

- How should **open-source AI systems** be treated concerning these guidelines?
    - Would uniform rules inadvertently discourage open-source development or research-oriented systems?

2. Accountability and Liability in AI Systems

How should we design and implement accountability and liability regimes for AI systems used to create, disseminate, and consume information? Are there any proposals or best practices we should look at?

In addressing this, you might consider:

- How can we clearly define and distribute **responsibility for AI-generated content and decision** among data providers, AI system developers, system deployers and users?
    - Who should bear the responsibility when AI outputs violate existing intellectual property rights or data privacy? Who needs to prove evidence of fault?
    - Who should bear responsibility for AI systems' moderation, verification, curation, recommendation, and ad-targeting and delivery decisions? How much autonomy should AI systems have, especially in high-stake scenarios? What role should human intervention play?
    - Should certain usages be forbidden or restricted, and if so, how can we implement and enforce these restrictions?
    - What mechanisms should be available to users to report potential issues or biases in AI-driven decisions and outputs?

    - How can we ensure that AI developers and deployers prioritize safety and continually monitor and refine AI systems to address emerging challenges?

- What rights and responsibilities should **platforms hosting AI-generated content** (e.g., digital platforms, search engines, media) have?
    - To what extent should they be responsible? What kind of liability approach should apply?
    - Should they have an obligation to verify and/or label AI-generated content before distribution?
    - What mechanisms should be available to users to report or challenge AI-generated content, and what obligations should platforms have in responding to these challenges?

- What **rights and duties should users and subjects of AI systems** have?
    - Should they have the right to know when they are interacting with an AI system, and if so, how should this be disclosed?
    - What kind of complaint and redress mechanisms should users and AI subjects have, and who should enforce them? What standard of proof should be required ? What remedies should be available for successful claims (e.g., compensation, corrective action, system overhauls/deletion)?
    - What ethical obligations should users, including journalists, researchers and highly influential actors, have when creating or distributing AI-generated content?
    - Should there be mandatory disclosure when users utilize AI tools to create content? How can disclosure about the origin of content and creator be reconciled with freedom of expression and protection of sources?

- What role should AI developers and deployers play in providing **AI literacy** about the systems they offer, thus enabling users and subjects, including journalists, to better understand the tools they use?

3. Governance of AI Systems

How can we design and implement an effective AI governance framework for the deployment and monitoring of AI systems impacting how we create, disseminate, and consume information? Are there any best practices we should look at?

In addressing this, you might consider:

- Is there a need for an **independent AI Authority**?
    - Would a singular, centralized oversight institution be more effective, or should there be specialized decentralized entities for distinct AI sectors ?
    - What should be its core roles and responsibilities ?
    - Which enforcement tools should be available to the AI Authority?
    - How should the AI Authority be funded?
    - How should national courts interface with the AI governance framework?

- Should we implement **licensing processes** to uphold digital safety standards of AI systems, and if so based on which criteria and implemented by which institution?
    - Should licensed AI systems undergo periodic reviews to ensure continued compliance?

- How can we **monitor and ensure compliance** with agreed principles and regulations?
    - How can we employ external audits for AI system accountability?
    - What feedback mechanisms should be established for users and affected parties to report potential harm or concerns arising from AI system interactions?
    - How can we ensure that stakeholders, from developers to end-users, are adequately educated about AI regulations, potential risks, and protective measures?

- How can the **auditing process** be designed to evaluate the technical components of an AI system, including algorithms' robustness, bias mitigation, and transparency, as well as their impact on pluralism and freedom of expression?
    - Should continuous cybersecurity audits be in place to detect any potential vulnerability appearing each time the code of an AI system and/or infrastructure is altered ? Should a Vulnerability Disclosure Policy be required?
    - What sanctions should be available for non-compliance (e.g., recommendations, financial penalties, revocation of licenses)?
    - Should there be an appeal process for challenging audit results or sanctions, and how should they be structured?
    - Should audit results be disclosed to the public? How can we balance transparency with concerns about proprietary information?

- How can we ensure **democratic oversight and civil society involvement**?
    - Should the AI Authority be able to rely on an Advisory Board?
    - In what capacities can civil society organizations and the general public be integrated into AI governance ?
    - How can we establish sustainable funding models and mechanisms to support the involvement of Civil Society Organizations?

- Given many AI systems' global reach and impact, how can we develop and implement **international AI governance** mechanisms to address cross-border and cross-jurisdictional issues?
    - What measures or frameworks are needed to ensure ethical and legal consistency across countries?
    - To what extent can compatibility regimes facilitate a common ground and help us overcome the inherent difficulties in harmonizing AI regulations, standards, and ethical considerations across different regions and cultures?

4. Final Thoughts

Is there any other relevant information or perspective you wish to share that can assist the working group in its mandate?

2.2 Format

We welcome your contributions to the questions above by submitting a comment through the following questionnaire. You may also submit a comment directly to our email: contributions@informationdemocracy.org in English, French, or Spanish by November 19, 2023.

The call for contributions is open to any individual or institution working in any capacity on these issues. We will not attribute any individual comments to you by name. However, with your permission we will thank you for your contribution in the Acknowledgements section of the working group's final report. If you agree, please include your biography or description of the institution in your contribution.

3/ About the Working Group

3.1 The effects of AI systems on the information and communication space

This prevalence of AI is raising important challenges:

- Generative AI produces content that, in addition to copyright and privacy concerns, raises challenges as to the authenticity and trustworthiness of that content. AI can invent sources, create misinformation and deep fakes. That information can be used unintentionally or intentionally to spread mis- and disinformation, sow discord and perpetrate deception. It is already nearly impossible to distinguish AI and human-created content, diminishing trust in the global information space. While existing legislation on data privacy and copyright does generally apply to AI training data sets, the means to effectively implement the legislation is lacking. The existing legislation needs to be developed further to tackle the challenges arising from AI.
- Beyond content creation, generative AI can be employed to create fake social media accounts, amplify information and give it credibility, thus contributing to an information chaos.
- AI systems are taking crucial decisions in the information and communication sphere, as the mere amount of information available and content created is exceeding by several orders of magnitude human capacities to consume, sort, moderate and verify this information. AI recommender systems decide which information we consume everyday, giving them tremendous power over the information space. Yet, these tools are not without flaws. They often suffer from biases, and their decision-making processes are not transparent but rather arbitrary. Furthermore, these algorithms frequently prioritize content that aligns

with users' pre-existing beliefs, leading to the creation of filter bubbles and echo-chambers. As AI systems increasingly shape our information ecosystem, the influence of private enterprises on democracy is growing. Currently, mainly private enterprises are deciding on the rules, terms of use of their systems, moderation policies, training data, etc. It is thus the objectives of private companies that determine the outcomes for the information space and not democratic principles that are used as the basis for building and deploying AI systems.

- Overall, AI creates a heightened manipulability of the information and communication space. It makes it easier to create difficult-to-recognise false information. It also enables tracking citizens' preferences and behavior more closely to target content and messages. This can have detrimental effects on information pluralism and media diversity.

Regulating the creation, deployment and use of artificial intelligence is high on the political agenda, with a global debate not only about AI-specific laws but also about leveraging existing regulation for responsible AI development, deployment, and use. The OECD Principles on Artificial Intelligence adopted in 2019, the UNESCO Recommendation on the Ethics of Artificial Intelligence adopted in 2021 and the ongoing negotiations for an EU Artificial Intelligence Act are just some examples.

Yet, the most pressing challenges as to AI's influence on the information space remain. They concern:

- The design and deployment processes and rules of AI systems to minimize risks to a pluralistic, free and trustworthy global information space supportive of democracy.
- Who is accountable and liable for the AI-created and disseminated content as well as the decisions taken by AI.
- The governance and enforcement of rules regarding AI.

3.2 Mandate

The working group will develop a comprehensive set of principles and recommendations to ensure that AI-powered systems—especially AI-based algorithms used for content moderation, curation, verification, and ad-targeting and delivering and generative AI technology that can produce text, audio, imagery, and synthetic data—strengthen the integrity of the information and communication space rather than diminish it. Recognizing the varied legislative and regulatory landscapes across countries, the working group will propose policy recommendations that can be applied in different contexts.

In particular, the working group will look at the following issues:

Objective 1: Define guardrails for the development and deployment of AI systems impacting the information and communication space.

Most products released on markets, from pharmaceuticals to toys, must adhere to strict rules during their development and production before becoming available to consumers. However, AI systems do not currently face such stringent standards. This lack of oversight poses a risk to a pluralistic, free, and trustworthy global information environment essential for democracy**.**

To fill this regulatory gap, the working group will engage with experts from various industries and countries to formulate robust guardrails and regulations for designing, developing, and deploying AI systems. Their primary goal is to safeguard the information space and ensure that AI systems are built in a way that they have a positive impact on the information space.

Specifically, the working group will examine ways to protect intellectual property and privacy rights. It will also provide guidance on establishing risk assessments, audits, and human rights due diligence to scrutinize the potential impact of new AI systems before their launch. Furthermore, the group will identify best practices for ensuring the integrity of training data sets to address issues of bias and discrimination and explore solutions to enhance the transparency, explainability, and contestability of AI systems for all stakeholders.

Objective 2: Define accountability and liability regimes for the developers, deployers, users and subjects of AI systems with regards to the information produced, curated and recommended, moderated and verified through these AI systems.

As AI systems become increasingly integrated into our information and communication space, there's a growing ambiguity about who should be responsible for their output and decisions. These systems rely on complex supply chains, with each phase involving various actors, each overseeing a distinct component of the system's training, development and deployment.

Given AI's potential to yield erroneous, inaccurate, biased, or discriminatory outcomes and its vulnerability to misuse—from manipulating public opinion to spreading mis- and disinformation or generating deep fakes and other doctored content—it is essential to determine who bears responsibility for any resulting harm or loss.

The working group will explore possible accountability regimes for developers, deployers, and users concerning decisions made by AI systems used for content moderation, verification, curation, recommendation, and ad-targeting and delivery. In this context, the group will assess to what extent decision-making can be entrusted to AI systems.

In addition, the working group will scrutinize the respective responsibilities and duties of developers, deployers, and users as they pertain to AI-generated content. This includes thoroughly investigating the challenges and implications of generative AI for intellectual property rights, data privacy, and editorial oversight. The group also intends to examine how liability should be structured for the use of generative AI to create social media accounts and amplify information.

The working group will also discuss the rights and duties of distributors of AI-generated content, like digital platforms and media, and closely examine the rights and duties associated with the use of AI systems. Central to this, the group will explore initiatives to promote AI literacy to enable users and AI subjects, including journalists, to better understand the tools they use.

Objective 3: Recommend governance options for the deployment and monitoring of AI systems impacting the information and communication space.

With AI systems becoming increasingly integrated into our lives, it is crucial to establish robust governance mechanisms that ensure their ethical and responsible use, mitigate potential misuse, and prevent oversight lapses. While several initiatives to govern AI are being launched, no clear governance frameworks currently exist. These gaps in oversight and accountability can result in inconsistent practices, unreliable or problematic system outputs, potential harm to users and the broader society, and ultimately erode trust in the information and communication space.

To address this, the working group will look at governance options to monitor the deployment of AI systems, ensure compliance with development regulations, and enforce strict adherence to established accountability regimes.

The working group will specifically explore how to establish licensing processes that adhere to digital safety standards. It will also delineate the powers and institutional setup of oversight mechanisms and institutions tasked with overseeing development regulations and the accountability regime. Integral to this is a reflection on how to establish rigorous auditing mechanisms of AI systems' behavior and their development processes, complemented by appropriate enforcement mechanisms and sanctions.

Furthermore, the working group will consider potential enforcement mechanisms in case of regulatory non-compliance. A particular emphasis will be placed on the need for establishing democratic oversight and enforcement processes, underscoring the critical role civil society can play in AI governance.

3.3 Deliverables and objectives

The working group aims to provide governments, policymakers, and relevant stakeholders with concrete recommendations for regulation, co-regulation, and self-regulation. These recommendations will be based on consultation with experts and written contributions received through the present call. The working group's report will be addressed to the 51 signatory States of the International Partnership for Information and Democracy, civil society, digital platforms, and tech companies.

3.4 Organization

The Forum on Information and Democracy

Forum on Information & Democracy

The Forum on Information and Democracy is the implementing organization of the International Partnership for Information and Democracy endorsed by 51 countries. Founded by 11 NGOs and research centers around the world, its mission is to implement democratic safeguards in the global information and communication space.

The Forum gathers international working groups composed of experts whose mandate is to publish regulatory and self-regulatory recommendations addressed to various stakeholders (States, platforms, civil society, etc.).

The Working Group

The working group is co-chaired by Laura Schertel Mendes, Lawyer and Professor of Civil Law at the University of Brasilia and Senior Researcher at Goethe University, Frankfurt am Main and Jonathan Stray, Senior Scientist at the Berkeley Center for Human-Compatible AI. It is composed of 14 international experts and its main mission is to provide expertise to the team of rapporteurs in charge of drafting the report and advise the Forum's Board of Directors on the final recommendations. The members meet periodically to help develop the recommendations.

The team of rapporteurs

The team of rapporteurs ensures the integration of contributions and drafts the recommendations of the regulatory and self-regulatory framework. Viviana Padelli, lead-rapporteur, is an economist and public policy professional working at the intersection of tech and democracy. Lia Chkhetiani is a political scientist and lawyer and Kaye Celine Palisoc is a computational and public policy analyst.