

# / If you want to help humanity, don't fall for “AI” hype

AlgorithmWatch’s response to the Interim Report *Governing AI for Humanity* by the UN Secretary-General's AI Advisory Body

**Berlin/Zurich, March 31, 2024**

We welcome the UN's initiative to advance recommendations for the international governance of “Artificial Intelligence”. Overall, we also welcome that the report argues that technological advancement should serve the public interest. However, we are concerned about major issues; In particular, the report construes a false balance between risks and opportunities related to “AI”.

*Following this introduction, we document the responses we submitted under the section headings the online form for submission specified.*

The benefits of “AI”<sup>1</sup> are visible to us when we look at navigation systems, machine translation or drug development. But “AI” currently mainly helps to foster the interests of advertisers (based on the largest collection of data about humans in history), to encourage us to spend as much time as possible on social media platforms, to track workers’ performance, or to fabricate disinformation and deep fakes.

Alas, the enormous achievements in combating the climate crisis and world hunger that have been heralded for decades have not materialized. As a case in point, the interim report argues under the headline “early promises of AI helping to address climate change” that one of the promises is “using advanced climate modeling tied to information about urban mobility and behavior patterns to create new early warning systems, allowing for more effective delivery of post conflict/disaster relief and recovery.” Hailing “AI” to manage harm and damage as a contribution to combating the climate crisis can at best be seen as sarcasm, at worst as cynicism, given the fact that “AI” – and above all “generative AI” – not only consumes huge

---

<sup>1</sup> We use quotation marks throughout the text to highlight the fact that what constitutes “AI” is underdefined and the term is mainly used strategically to achieve specific goals.



amounts of energy, water and rare earth minerals, but also enables business models that contribute significantly to exacerbating the climate crisis.

Instead of the misguided dichotomy of “opportunities” vs. “risks”, we therefore encourage an alternative narrative – for three reasons.

1. We should not focus on (often speculative) opportunities or risks in the future, but on benefits or negative impacts that today affect real people.
2. Some of the negative impacts of "AI" (e.g., infringements on human rights) cannot simply be outweighed by some of the potential benefits (e.g., efficiency gains). We should thus not look at the chances/risks or benefits/harms as mere lists of potential trade-offs.
3. The actual potential of “AI” could also be greater than the mere net sum of [benefits minus negative impacts] – but it is a potential we can only harness if we ensure that "AI" itself and the industry behind it became sustainable, just, and genuinely driven by the interests of humanity rather than those of a few companies. To do so, we need to address the political economy, the power asymmetries, the resource consumption and the exploitative working conditions behind these technologies.

If we do not resist "AI" hype and the techno-solutionist trap, the outcome will be novel problems and wasted resources.

## Opportunities and Enablers

This entire section reeks of techno-solutionism. Paragraphs such as "similarly, possibilities exist with respect to environmental problems, making education more accessible, helping ease poverty and hunger, and making cities safer" without any concrete examples are empty "AI" hype marketing. In many, if not most cases poverty and hunger are caused by a large number of factors that interact in a complex web of dependencies, from colonialist exploitation (in past and present) to power-asymmetries and wars. The use of some kind of magical "AI" systems will not make a difference as long as the tools themselves (and the industry behind it) are not sustainable and oriented towards the common good – and it will not exempt society from taking the responsibility of tackling these problems. To create this impression will result in a dangerous distraction from necessary steps to address the causes of these problems.



Similarly, “AI” systems can be useful tools for a number of tasks which may lead to energy efficiency. But claiming that “AI” can be a game changer in the climate crisis sounds absurd at best. The reason we are failing to address the climate crisis is not because we are lacking informational and technological know-how – but because we do not take the necessary measures.

Furthermore, “AI” systems, especially generative “AI” including large language models (LLMs), cause a huge environmental concern. Researchers warn us about the carbon impact of “AI”<sup>2</sup> or draw our attention to “AI”’s intense water- and mineral-usage<sup>3</sup>. Civil society calls on decision-makers to focus on the ecological impact of “AI” and to address these in necessary policy measurements<sup>4</sup>. Even OpenAI’s CEO acknowledges how resource-intensive certain “AI” models are and says we need even more power for future models<sup>5</sup> – but suggests addressing this by building nuclear fusion reactors, in which he has a massive business interest. In light of the fact that such technology is nowhere near practical application, we must assume that the enormous energy requirements he sees need to be met by the use of fossil fuels and atomic energy, both of which are sources we should work hard to ramp down.

## Risks and Challenges

Regarding what “AI” regulation should look like, UN High Commissioner for Human Rights, Volker Türk, recommends that: *“The starting point should be the harms that people experience and will likely experience.”*

We agree that harms – or rather impact – need to be identified first. We would, however, recommend focusing on current, existing benefits, impact and harms “AI” causes, as opposed to future theoretical scenarios. Focusing great efforts on *“yet unknowable harms”*, as the interim report suggests, should happen after

---

<sup>2</sup> The carbon impact of artificial intelligence: <https://www.nature.com/articles/s42256-020-0219-9>

<sup>3</sup> An Elemental Ethics for Artificial Intelligence: [An Elemental Ethics for Artificial Intelligence: Water as Resistance Within AI’s Value Chain by Sebastián Lehuedé :: SSRN](#)

<sup>4</sup> Dr. Mollen, Vieth-Ditlmann, SustAI Magazine 3<sup>rd</sup> Edition, Just Measure it: [Just Measure It - sustAI \(algorithmwatch.org\)](#)

<sup>5</sup> [Sam Altman Says AI Using Too Much Energy, Will Require Breakthrough Energy Source \(futurism.com\)](#)



we tackle currently existing problems. We are concerned about the “existential threat” rhetoric in the interim report, which follows ideologically driven narratives of so-called ‘longtermist’ perspectives, fueled by enormous financial resources provided by tech firms and investors<sup>6</sup>. This is dangerous because: (i) it distracts from real, existing problems and (ii) from this perspective, one could legitimize present harms in order to realize a potential greater good in the future. This is diametrically opposed to a human rights-based perspective. We would instead advise focusing attention and resources on existing problems.

We welcome that the interim report stresses that certain “risks are more a product of humans than AI” and that certain people are more exposed to risks of “AI”.

In general, if we continue to advance “AI” as we do currently, i.e., primarily focusing on its “opportunities” while half-heartedly trying to mitigate some of its “risks”, then this technology will continue to benefit the interests of those in a privileged position – and therefore remain a powerful tool to cement and exacerbate existing injustices. This real threat already manifests itself in various ways. The supply chains of large “AI” models largely rely on exploitative working conditions, “AI” tools are built on training data of the past that is not sufficiently diverse and contains existing societal patterns of injustices – hence their outputs adopt and cement these and can discriminate against people that are already overly at risk of discrimination. “AI” tools are also used to manage, track, and surveil people at their workplace or online.

We welcome that the interim report highlights some of the risks of biometric surveillance. We suggest the report also stresses the risks of post remote biometric identification, emotion recognition systems, and biometric categorization that can lead to mass surveillance in publicly accessible spaces and pose a severe threat to human autonomy. Especially with regard to emotion recognition, there needs to be a focus on thorough assessments of the models used, as their validity is questionable. This relates to a general point: There is a lot of “AI” snake oil, i.e., tools that are generally advertised with great marketing

---

<sup>6</sup> AI doomsayers funded by billionaires ramp up lobbying (politico.com)



efforts, promising to solve tasks magically, but that are not based on scientifically solid and ethically responsible research.

## Guiding Principles to guide the formation of new global governance institutions for AI

We welcome that the interim report highlights the importance of an “AI” governance framework which respects human rights and the rule of law, and we largely agree with many of the guiding principles on international “AI” governance. However, the guiding principles could start with emphasizing the important objective that “AI” governance means that society governs the development and use of “AI” rather than the opposite. Moreover, we believe the interim report should strengthen its human rights-based approach. As former UN High Commissioner for Human Rights Michelle Bachelet writes:

*“The need for a human rights-based approach to new technologies in general, and artificial intelligence in particular, has been recognized by a growing number of experts, stakeholders and the international community.[1] A human rights-based approach offers a toolbox to help societies to identify ways to prevent and limit harm while maximizing the benefits of technological progress.”*

“AI” systems can pose a threat to or violate our fundamental rights, and undermine our rule of law and democratic institutions. It is essential that “AI” governance be guided by the necessity and proportionality principle. Are there alternatives to “AI” systems, i.e. systems that can be more effectively tested and more easily explained than black-box “Machine Learning” models, and that are more energy-efficient and cost-efficient, to perform a specific task? Does the benefit of developing or using an “AI” system outweigh the negative impact to our fundamental rights and planetary boundaries? These are some of the guiding questions which should precede the uptake of “AI”. It must be ensured that the benefits outweigh the negative impacts, and that those benefits are serving the public, i.e., everyone on the globe – and not just a handful of powerful corporations.

We agree that “AI should be governed inclusively, by and for the benefit of all”. We would add in this section that the field of “AI” is highly politicized: there are great power imbalances among the various stakeholders – industry, Big Tech,



security agencies, civil society, watchdog organizations, academics – which need to be addressed by empowering those with less resources and to put a limit on the hard and unfair lobbying of the powerful. We also recommend addressing the issue of market concentration as opposed to just noting it descriptively.

The report could also highlight – next to rule-setting, which is at its core – that governance must also include measures such as promoting interdisciplinary research on AI as well as on its societal implications, including journalistic research, and raising public awareness and evidence-based public debate.

### **Institutional Functions that an international governance regime for AI should carry out**

The Institutional Functions recommendations are shockingly under-ambitious and need to be revised. 6 of the 7 of the “recommended institutional functions for international AI governance” address observational, soft law, and self-regulatory measures only, and the section “Compliance and accountability based on norms” is factually flawed and remains underdeveloped. To state that “a regional effort for an AI treaty is already underway” is an understatement, as both the EU’s AI Act and the Council of Europe’s Convention on Artificial Intelligence were in the final stages of being agreed when the report was published. Furthermore, even though it is clear that both small and major gaps exist in current regimes to tackle the negative impact “AI” can have, its development and use does not happen in a legal vacuum. Existing laws apply – and should be interpreted accordingly, from national and international legal norms protecting human rights, anti-discrimination, labor law, administrative law, criminal law, to competition law and other fields. Also, a number of laws specifically addressing “AI”-based aspects have been in effect for years (e.g., Japan’s Act on Improving Transparency and Fairness of Digital Platforms, or the EU’s Digital Services Act on the use of – often “AI”-based – recommender systems by platforms).

Current, we witness an increase in conflicts between countries instead of cooperation, so it is perfectly understandable that the UN does not consider it feasible to develop a global binding treaty on “AI” governance. Still, the Advisory Body must not turn a blind eye towards the fact that there should be a clear



focus on *hard law* regimes to address the development and use of “AI”. One of the Body’s most important contributions to the global discussion should be the analysis of existing hard law instruments and to develop best-practice guidelines on how they could be mirrored in other parts of the world. We recommend reworking section 62 toward a clear focus on binding rules.

Along the same lines, we welcome that the report highlights that civil society should play a key role in building evidence for policy, assessing impact, and holding key actors to account during implementation.

But first of all, it needs to be clear that it is first and foremost the role of governments, and in some part international organizations (like the Council of Europe, the African Union, the UN, and others), to ensure compliance with existing rules, and therefore accountability of both governments’ and private entities’ use of “AI”-based systems.

Secondly, the Advisory Body should develop meaningful and practical suggestions for how to ensure that civil society is empowered and resourced to effectively assume these tasks. Lastly, while exchanges with private actors might be helpful, it must be made explicit that private actors are not responsible for introducing governance measures, given the societal division of labor in states based on the rule of law.

## About AlgorithmWatch

AlgorithmWatch is a human rights organization based in Berlin and Zurich. We fight for a world where algorithms and “Artificial Intelligence” (“AI”) do not weaken justice, democracy, and sustainability but strengthen them. Such systems should make societies more just, democratic, inclusive and sustainable, be it with regard to alleged or claimed race and gender, sexual orientation, abilities, age, wealth, class, or resource consumption.

<https://algorithmwatch.org/>