The Public Good Obligation: Updating the UGAI for AI's environmental and informational risks

Jens Meijen (Ulysses AI)
Niharika Gujela (Hertie School of Governance)

Abstract

This paper argues that the 2018 Universal Guidelines for AI should be updated with one new principle that encompasses several recent global developments: the Public Good Obligation, or the idea that AI should not harm publicly accessible commodities that benefit all of society. The principle aligns with the UGAI's fundamental logic of "do no harm", leaves room to incorporate future developments, and covers AI's multi-faceted risks that have proliferated over the past five years, especially due to its widespread adoption. In the paper, we devote particular attention to AI's impact on the environment and on the internet, both of which can be approached as public goods.

AI and governance

Global, regional, and national governance efforts have attempted to address AI's risks and create guidelines to responsibly harness its many opportunities. Examples include the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles, the Global Partnership on AI (GPAI), the forthcoming EU AI Act, and national AI strategies. One example of a governance framework issued by a diverse coalition of actors was the Universal Guidelines for AI (UGAI), issued in 2018. While any direct causal relation between the UGAI principles and AI policies that have subsequently been developed is methodologically difficult to establish, it is nonetheless clear that the UGAI principles dovetail well with the general direction of responsible AI governance efforts. With five years of technological and political development behind us as well as binding European and US AI legislation on the horizon, an update of the UGAI seems timely. While an update could go in many different directions and should include a diversity of perspectives, we believe that any new Universal Guidelines should at least include a thorough consideration of AI's changed impact on and place in society from 2018 up to now.

Between 2018 and 2023, the world has seen a series of transformative events - from a global pandemic and the proliferation of online disinformation to widespread wildfires and other extreme weather events. Over the same period, artificial intelligence may well have had its breakthrough moment. We've seen the meteoric rise of LLMs and chatbots, a budding AI arms race between major powers, AI driving skyrocketing share prices of large companies (e.g. Nvidia) and drawing unprecedented amounts of funding (e.g. Anthropic, Databricks, Hugging Face…), but also rising concerns about AI and its potential dangers (Tyson and Kikuchi, 2023). Either way, AI has evolved from an arcane technology that was once the sole domain of technical experts to a highly visible concept that continuously receives mainstream media attention and is a top priority in several fields.

AI's environmental and informational risk

However, widespread adoption implies a greater impact on society and therefore greater and broader risks. We find two issues particularly pressing. First, AI systems have an understudied but nonetheless considerable impact on the environment due to their exorbitant resource requirements. Training and operating AI systems consumes large amounts of energy and water while other materials (like tungsten and aluminium) must be mined to create GPUs. These issues have been covered in several reputable media outlets, so they have moved from niche issues to the heart of public debates on AI (see e.g. Saul and Bass, 2023; Guerrini, 2023; Jariwala and Lee, 2023; Sing, 2023). It should be clear, then, that the widespread adoption of resource-intensive AI systems risks contributing disproportionately (and for many perhaps unexpectedly) to climate change, environmental destruction, and biodiversity loss. This may be particularly worrying if those systems are essentially used for entertainment purposes (as is the case for many users of image and text generators) without users being aware of the resources they consume.

Second, recently developed LLMs can create convincing natural language in the blink of an eye, which means AI has the potential to generate massive quantities of online (dis)information that is fast and cheap to produce and disseminate (Pan et al., 2023; also see Knight, 2023). The effects of LLMs have yet to be fully discovered, but some observers have argued that the internet is already suffering the consequences of AI, and particularly LLMs, with the proliferation of AI-generated text (Kniaz, 2023). If left unchecked, the internet may eventually become a desolate information wasteland of AI-generated text that simply regurgitates and rehashes what has already been written many times over. The internet would then be little more than bots talking to each other instead of humans interacting to exchange information. This is what we here propose to call the "informational risk" of AI - its potential negative consequences for how we collectively store, access, and evaluate information via the internet. While these environmental and informational risks may not be entirely new, AI's widespread adoption may exacerbate them. Essentially, AI now affects everyone on the planet and everyone on the internet.

Both the environment and the internet can be seen as (partial) public goods (see Chin, 2021). The environment is often considered a public good because it exhibits the key characteristics of non-excludability and non-rivalrous consumption. Non-excludability implies that it is challenging or prohibitively expensive to exclude individuals or groups from benefiting from environmental resources or services, such as clean air, clean water, or a stable climate. Non-rivalrous consumption means that one person's use or enjoyment of these environmental resources does not necessarily diminish their availability to others. The environment is a shared resource that can benefit society as a whole, and it often requires collective efforts and policies to manage and protect it for the common good. The internet, too, can at least in part be approached as a public good (Haugen, 2020; Ros-Galves and Rosa-Garcia, 2015). While not all information that can be found on the internet is available free of charge, a good portion of it is (such as Wikipedia, social media, or news websites). Likewise, internet access requires at least buying a device to browse, but it is nonetheless difficult to exclude individuals from access. Also, consuming information on the internet is non-rivalrous, as reading a news article or Wikipedia page does not make it unavailable for others.

As we discussed above, AI may harm the environment through its enormous resource requirements and "pollute" the internet with automatically generated content. AI therefore risks leading to a tragedy of the commons, in which the developers and distributors of AI systems (usually large or Big Tech companies) continue to exploit the environment and the internet for AI-driven profit by "overgrazing" the online commons (Angwin, 2023), spoiling both for everyone by straining them so much that their value as shared public resources diminishes.

Relevance for the UGAI - the Public Good Obligation
So what does this mean for the UGAI? The new risks associated with more widespread use of AI imply a potential deterioration of the environment and of the quality of the internet as a source of information. We expect these issues to become more pressing as AI adoption continues to increase. The UGAI principles, issued in 2018, do not really have room to include these public good problems in existing principles. One option would be to expand the public safety principle to include public goods. However, this would imply that the original meaning of the principle could be lost: it primarily concerns the obligation for institutions to assess the public safety risks that arise from deploying AI systems *that direct or control physical devices* (emphasis ours) - while the environment and the internet could arguably involve public safety risks, they are not physical devices *per se*. The problem we identify above therefore does not fit into the existing structure of the UGAI.

This is why we propose to add "The Public Good Obligation" to the UGAI.

> The **Public Good Obligation** principle ensures that the development, training, and deployment of AI systems should not harm publicly accessible commodities that benefit (almost) all of society. AI should not harm things that we collectively benefit from, such as the Internet or the environment.

This aligns with the UGAI's basic "do no harm" philosophy and covers several modern problems, including (but not limited to) environmental pollution, climate change, and disinformation. To be clear, this is not to say that all AI systems should by definition positively impact one or more public goods. Many systems have private benefits as their primary purpose, such as to contribute to business efficiency through financial analysis, which we do not see as inherently bad. Such systems used for private gains should, however, not destroy the environment (e.g. if an AI decides to frack for oil in a vulnerable part of an ocean) or "pollute" the internet (e.g. an AI designed to generate fake news articles at scale). Furthermore, we believe that providers of AI systems should better inform users of the environmental or other societal harms (the use of) an AI system could cause. This would allow individuals to make better decisions on whether or not to use an AI system at all, especially considering AI's environmental footprint. We believe this would be a timely and elegant update that would cover new risks while leaving room for future developments.

References

Chin, M. (2021). 'What are global public goods?' *IMF.com*
https://www.imf.org/en/Publications/fandd/issues/2021/12/Global-Public-Goods-Chin-basics

European Parliament (2023). 'EU AI Act: first regulation on artificial intelligence.'
europarl.europa.eu.
https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Global Partnership on Artificial Intelligence (2020). *The Global Partnership on Artificial Intelligence*. https://gpai.ai/about/

Guerrini, F. (2023). 'AI's Unsustainable Water Use: How Tech Giants Contribute To Global Water Shortages.' *Forbes.com.*
https://www.forbes.com/sites/federicoguerrini/2023/04/14/ais-unsustainable-water-use-how-tech-giants-contribute-to-global-water-shortages/

Haugen, H.M. (2020). 'The crucial and contested global public good: principles and goals in global internet governance.' *Internet Policy Review* 9(1): 1-23.
https://policyreview.info/pdf/policyreview-2020-1-1447.pdf

Kniaz, R. (2023). 'The Incoming Tidal Wave Of Data Pollution In AI.' *Forbes.com*
https://www.forbes.com/sites/robkniaz/2023/05/09/the-incoming-tidal-wave-of-data-pollution-in-ai/

Knight, W. (2023). 'It Costs Just $400 to Build an AI Disinformation Machine.' *Wired.com*
https://www.wired.com/story/400-dollars-to-build-an-ai-disinformation-machine/

Pan, Y., Pan, L., Chen, W., Nakov, P., Kan, M., Wang, W. (2023). 'On the Risk of Misinformation Pollution with Large Language Models.' *Computation and Language,* 2305.
https://arxiv.org/abs/2305.13661

Ros-Galves, A. and Rosa-Garcia, A. (2015). 'Private provision of a public good: cooperation and altruism of Internet forum users.' *International Journal of the Commons* 9(2): 720-743.
https://www.jstor.org/stable/26522846

Saul, J. and Bass, D. (2023). 'Artificial Intelligence Is Booming—So Is Its Carbon Footprint.' *Bloomberg.com.*
https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure

Singh, M. (2023). 'As the AI industry booms, what toll will it take on the environment?' *The Guardian.*
https://www.theguardian.com/technology/2023/jun/08/artificial-intelligence-industry-boom-environment-toll

The Organization for Economic Cooperation and Development (2019). *OECD AI Principles.* https://oecd.ai/en/ai-principles#:~:text=The%20OECD%20AI%20Principles%20promote,human%20rights%20and%20democratic%20values.

The Public Voice (2018). *Universal Guidelines for Artificial Intelligence.* https://thepublicvoice.org/ai-universal-guidelines/

Tyson, A. and Kikuchi, E. (2023). 'Growing public concern about the role of artificial intelligence in daily life.' Pewresearch.org. https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/

United Nations Educational, Scientific and Cultural Organization (2022). *Recommendation on the Ethics of Artificial Intelligence.* https://unesdoc.unesco.org/ark:/48223/pf0000381137