
PRELUDE

Dear Reader,

As both the volume and range of the studies appearing in this fine bouquet will clearly demonstrate, in the last few years AI Ethics became a burgeoning field that increasingly permeates business and policy, beyond its academic roots. But does the field of AI ethics and governance as well as its sizeable global community of practitioners pull their weight? Does this domain and its fervent advocates fulfill their purpose as embodying adequate social control over the spreading of a technology of unprecedented power, exponentiality and fickleness: AI?

We belong to the school of thought that treats AI Ethics very holistically - and equally pragmatically. Our compeers also believe that in its broadest definition, ownership of Ethical AI belongs to an informed society, composed of responsible digital citizens who drive related social movements - not just to professional practitioners.

But as this question is being often approached in the present series of studies, is control of AI technology possible? Can regulation solely aim at that? We are of a belief that even if *control* is and will stay elusive, *steering* or *stewardship* should be set as the prime civilizational objective - and that with carefully selected and crafted combination of methods.

However, the farther we find ourselves from the comfort of our studies and enter the labyrinth of fieldwork, the more we encounter distortions, misrepresentations and reductionisms that jeopardize the success of the ethical and responsible AI mission. The novelty and cross-disciplinary complexity of this realm vividly showcases the shortcomings of trying to solve new problems with old tools and mindsets, as well as exposes the unsolved burdens we still carry from previous technological waves. For many stakeholders who by and large use “digital”, “data” and “AI” interchangeably (especially in policy), the specifics of AI still need to sink in: namely its distinctive capabilities of autonomous decision-making, learning capability and the high level of potential opacity (Héder 2020b). On the other end of the spectrum, many technology leaders in charge of AI governance who understand its fundamentals would prefer to reduce the intricacies of the AI Ethics problem set to just data and model bias - and solve it with a dedicated tool that merely checks boxes.

Waking up in an AI world caused us to try to wrap our heads around this set of novel phenomena. These first attempts led to the creation and proliferation of AI Ethics guidelines, numbered in the hundreds of manifestoes by now. While they have been and will be criticized, - sometimes reasonably so -, for being too numerous and obscure, too much overlapping but not too useful, and too self-important, we believe being principled is an unavoidable, highly necessary step - but not the destination. Mostly originating from organizations trying to fill a trust gap in the present state of global governance, one can explain many of their shortcomings to be mainly derived from the factors of a global international order being constantly battered by dissenting incumbent and aspiring hegemon - and consequentially losing significance.

Since many of the studies in this collection come from the CEE region that should be better known for its own, ingrained version of critical thinking, - probably rooted in its grim history overloaded with enforced dogmatism -, one cannot overlook the pattern of wake-up calls that define themselves as “realist”-see for instance Gyulai and Ujlaki (2021) in this issue. In this world becoming multipolar and with authoritarian instincts on the rise, one can even argue a temporary advantage in the realm of AI that benefit those who primarily grasp AI from the perspective of power and dismiss the ethical side as a nice-to-have or just noise. A ruthless AI race mindset permeates both the realms of geopolitics and that of Big Tech: never before was “the winner takes it all” taken so literally. Moreover, as realists are painfully aware, reining in one of the most concerning AI application fields, Autonomous Weapons Systems (AWS) seems distant. Most the UN Security Council Members are opposed to a binding global ban based on New Cold War reflexes, while many smaller states would certainly be in favor as well as the general public. (In one study of this issue, Aron Dombrovski (2021) is offering a nuanced perspective as the advocate against oversimplified AWS bias.)

China is set out to become a hyperpower built on an AI engine and no stake is too high for them. New Zealand treats data as a tribal heirloom that many generations curate for AI to solely serve citizen well-being. Dubai leaps ahead with its own version of techno-absolutism and deploys vast swaths of government AI services that are measured against the goal of raising citizen happiness levels. The Pope calls for multi-stakeholder global work on human-centric, ethical and responsible AI to preserve Creation. These examples clearly raise the questions: how to be values-based and human-centered in AI with global applicability, while also being mindful of cultural differences, sectoral interests and societal priorities around the globe? How to rethink and redesign our global institutional frameworks and fill them with new meaning to successfully bridge the trustworthiness gap, the most ominous social plague of our civilizations(s)? How to convince Big Tech (equals Big AI) to internalize ethical and responsible AI as a strategic imperative that is key to attract and retain 21st century conscious customers - and not a nice-to-have, borderline ethics-washing parlor trick, a dark possibility Vică, Voinea and Uszkai (2021) elucidate in the current issue?

The emerging AI world – especially in the West – currently has an AI-ready vision and societal model deficit, and that needs remedying first. Our biggest bet is on an informed and responsible society: the emerging class of digital citizens and consumers, professionals and thought leaders who increasingly demand being in charge (Heder 2020a) of their privacy and choices, judge the ethical decision of their employers (and move on if need be), and are ready to stand up for an AI Future that prioritizes human well-being as the *ultima ratio*. Our world could benefit from *AI as augmented intelligence*, a machine-assisted extension of what makes us human -and not the path of artificiality which inevitably dehumanizes. The task of researchers and the AI-savvy is to speak the truth and do their best to make AI understood for conscious citizens worldwide *sans* sensationalism and obfuscation, so that societies could understand what is at stake, what are the new rules, and convert technospeak to challenges and solutions that impact their very lives. This

issue of studies makes a great contribution to this mission of great significance and therefore deserves your kind reading - which you will hopefully find equally profound and enjoyable.

Author Information

Dr. George A. Tilesch, Founding president of the PHI Institute for Augmented Intelligence
Co-Author of *BetweenBrains: Taking Back Our AI Future*
Member of the Board of Advisors, Experfy – Harvard Innovation Labs
AI Ambassador, John von Neumann Computer Society

References

- Dombrovski Áron. "The Unfounded Bias Against Autonomous Weapons Systems." *Információs Társadalom* 21, no. 2 (2021).
<https://doi.org/10.22503/inftars.XXI.2021.2.2>.
- Héder Mihály. "A Criticism of AI Ethics Guidelines." *Információs Társadalom* 20, no. 4 (2020a): 57.
<https://doi.org/10.22503/inftars.XX.2020.4.5>.
- Héder Mihály. "The Epistemic Opacity of Autonomous Systems and the Ethical Consequences." *AI & SOCIETY*, (July 30, 2020b).
<https://doi.org/10.1007/s00146-020-01024-9>.
- Gyulai Attila, and Ujlaki Anna. "The political AI: a realist account of AI regulation." *Információs Társadalom* 21, no. 2 (2021).
<https://doi.org/10.22503/inftars.XXI.2021.2.3>.
- Constantin Vică, Cristina Voinea, and Radu Uszkai. "The emperor is naked: moral diplomacies and the ethics of AI." *Információs Társadalom* 21, no. 2 (2021).
<https://doi.org/10.22503/inftars.XXI.2021.2.6>.