

'Multi-stakeholder model as a viable option for a pro-social AI governance'

Author: Tamar Navdarashvili, Ph.D. Student in International and Public Law, Ethics and Economics for Sustainable Development (LEES)

Affiliation: University of Milan

Abstract

As newly fueled Artificial intelligence (AI) systems gain ground in various sectors of social life and promise to bring societal and economic benefits, dangers of power concentration, algorithmic surveillance, data extractive practices, and inequality challenge global society in unusual ways (Vipra & Korinek, 2023; O'Neill 2016; D'Ignazio & Klein 2020; Crawford, 2021; Zuboff, 2019; Pistor 2020). Against this backdrop, public and private actors are called upon to provide a proper governance framework to manage AI's downsides and leverage its benefits. In this regard, many scholars advocated for a proactive, collaborative private-public governance structure with corporations and governments co-regulating and overseeing AI ethics, while promoting a safe environment for AI innovation (Lobel, 2012, 2022; Ulicane et al., 2021; Selbst, 2021). The EU's AI governance path somewhat mirrors this idea with a hybrid approach (Francés-Gómez, 2023), taking the lead by the adoption of ethical guidelines as a moral compass, suggesting the list of essential ethical principles for trustworthy AI (Floridi et al., 2018; Jobin et al., 2019; Vesnic-Alujevic et al., 2020), and prescribing a common governance system rules with the newly arrived flagship AI Act, which promises to protect fundamental rights, address AI-related risks, and transform principles into tangible protection.

Yet, AI governance is not a solved issue, and concerns about who (and how) controls means of prediction, especially data and computational infrastructure, become a serious challenge. In this vein, one might notice that AI is mainly a private venture and leading big tech companies arranged as corporations, with hierarchical governance structures, are those who steer the AI trajectories, advantaged with significant control over substantial AI infrastructure (OECD, 2024), privileged with data ownership and the power of monetizing the control over that (Pistor, 2020). So far, corporations made a noteworthy (at least formal) effort to implement AI ethics programs, taking social responsibility to tackle AI perils and placing governance/ethical schemes in ethical business policies. Albeit voluntary activities of self-studies and self-audits are important endeavors, private AI ethical statements, and efforts might be suspected of 'ethics washing' and 'machinewashing' (Bietti, 2020; Seele & Schultz, 2022; Schultz et al., 2024), carrying the dangers of ethics remaining merely cosmetic and turning out to be self-serving. Doubts persist since profit-motivated and market-driven firms might not be incentivized to welcome new costs related to ethical deliberations, risk assessment, and compliance with governance requirements (Francés-Gómez, 2023; Auld et al., 2022). Additionally, the uncertainty around the uniform definition of AI, and the opaqueness and proprietary nature of the systems, can be a strong shield for the 'privatized' and non-transparent ethics to circumvent 'AI responsibility'. Those trends, in turn, can seriously undermine internal stakeholders, for example, employees' well-being (rights, interests, inclusivity), and cause negative externalities for broader groups of stakeholders as well. In this context, former Google's top AI ethics researcher Timnit Gebru's story—ousted from Google over a dispute with the company relating to

the publication of an article on the potential risks and harms of large language models can be a clear example (Vincent, 2021).

Given the scenario—giant AI corporations tend to privatize ethics and essential infrastructure, preferring hierarchy over markets since it can decrease costs inside the firm, expanding control rights and benefiting asymmetrically (Pistor, 2020), all these make corporate self-governance legitimacy doubtful and needs sophisticated inspection of how institutionalized forms of stakeholder inclusion and alternative viable organizational model might be fundamental for AI governance. In this respect, recalling the transaction cost theory approach (Hansmann, 1996), the predominant economic perspective on governance tends to support the viewpoint that pursuing multiple, likely heterogeneous interests of stakeholders renders the firm less efficient. And, for the sake of reducing governance costs, it is reasonable to delegate control power to the sole stakeholder while protecting and coordinating others' interests through contracts. However, it has been scrutinized that such a hierarchical ownership structure and an exclusion of pertinent stakeholders from governance may also generate costs and trigger abuses of authority (Borzaga & Sacchetti, 2015; Sacconi, 1999). Alternatively, the need for a multi-stakeholder approach (as an extended governance model of CSR) to AI governance turns out to be relevant, which implies the extension of fiduciary duties to all stakeholders—empowering each stakeholder and balancing their interests and rights (Fia & Sacconi 2019). Indeed, multi-stakeholder cooperatives also proved to be efficient in many contexts, characterized by mechanisms of inclusion and cooperation and the capacity to reduce costs of exclusion of relevant stakeholders from the organization (Borzaga & Sacchetti, 2015).

Considering these tendencies, this contribution advocates for a shift of institutional mindset toward an alternative organizational multi-stakeholder model as a viable option for pro-social AI governance through ingeniously integrated ethics and economics vision. Ideally, the model has the potential to improve transaction cost efficiency, promote transparency and social dialogue among relevant stakeholders; and guard against the corporate AI ethical statements being simply rubber stamps, while preventing extractive logic of business and power abuses, with tangible prospects of being complementary to public policies and regulations.