

Lifting the Veil of the Future

Arendt and Justice-Centred Artificial Intelligence

Esmeralda Colombo

17 July 2026

Contents

Epigraph	3
I. Introduction	4
II. Background, assumptions, and limitations	7
III. The constitutive gap for justice-centred AI in energy matters	12
A. Introduction	12
B. The conceptual grounding of justice-centred AI	13
C. Constitutive principles: towards principled regulation	16
IV. Lifting the veil of the world	19
A. Regulating technology from an earthbound perspective	19
B. AI regulation for the AI–energy power couple	21
V. Three justice-centred principles for AI in energy systems	25
A. Introduction	25
B. First principle: the categorical imperative reversed	25
C. Second principle: <i>diligentia diligentis, or the diligence of a reasonably prudent person</i>	30
D. Third principle: the right to have rights	33
VI. Institutionalising justice-centred AI principles	37
A. Introduction	37
B. Institutionalising the reversed categorical imperative	37
C. Institutionalising <i>diligentia diligentis</i>	39
D. Institutionalising the right to have rights	41
E. Fiduciary duties or rights?	44
VII. Conclusion	46

Abstract: This article develops the concept of justice-centred artificial intelligence (AI) through Hannah Arendt’s thinking on the danger of a machine world replacing the real world and the freedom achieved in communities of equals. It explores how regulation might advance localised forms of justice concerning AI development and deployment reframing corporate accountability from transparency to enforceable duties. It focuses on energy systems and engages with European data law *acquis*, proposing three principles for justice-centred AI: a reversed categorical principle, fiduciary obligations grounded in the *diligentia diligentis* standard, and the right to have rights. It presents justice-centred AI as a constitutional project for the EU, aligning technology with rights, ecology, and political freedom.

Keywords: European digital constitutionalism, Hannah Arendt, justice-centred AI, energy justice

The laws hedge in these new beginnings and guarantee the preexistence of a common world, the permanence of a continuity that transcends the individual life span of each generation, and in which each single man in his mortality can hope to leave a trace of permanence behind him.

H. Arendt, *Thinking without a Banister: Essays in Understanding 1953–1975*, p 46.

I. Introduction

This article proposes the novel concept of justice-centred artificial intelligence (AI) through a lens, Hannah Arendt’s political thought, and a use case, the widespread use of AI in energy systems. Inspired by existing research and Arendt’s humanistic perspective, this article reclaims the principles by which we act and the criteria by which we judge and conduct our lives¹ to establish a digital constitutionalism integrative of ecological elements.² Such an interdisciplinary perspective aims to investigate existing regulatory gaps at the intersection of AI and energy, which the International Energy Agency (IEA) has dubbed the new ‘power couple’.³

This article proceeds in four main sections. Section II presents the methodology. Section III defines the gap between the democratic ideal of collective deliberation and the reality of individuals behaving as energy consumers. Section IV explores the role of international and regional law in filling the gap by offering a baseline of three principles, as explored in Section V, for justice-centred AI, particularly in energy matters. Finding these principles in need of institutionalisation, Section VI proposes three methods to embed the constitutive principles of justice-centred AI through both centralised and decentralised means.

This article adopts a European digital constitutionalism perspective, situating its analysis within the framework of EU constitutional law, including the Charter of Fundamental Rights of the European Union and the evolving body of secondary legislation governing AI, most notably the AI Act. While informed by international human rights and environmental law, the primary geographic and jurisdictional scope of this contribution is the European Union.

¹ H Arendt, *The Life of the Mind* (ed M McCarthy) (Harcourt, 1981), Thinking, p 71. The author is deeply grateful to Professor Catherine Barnard, Editor-in-Chief, and to the two anonymous reviewers for their exceptionally insightful, constructive, and complementary comments, which significantly strengthened the manuscript. Heartfelt thanks go also to the brilliant copyeditor generally provided by the Cambridge Yearbook of European Legal Studies. The author also wishes to thank Professor Konrad Lachmayer for the invitation to discuss an earlier version of this work at Sigmund Freud University in 2024. The ensuing discussion contributed substantially to sharpening this paper’s arguments, thanks to the discussion also with the researchers there present. This research benefited from the funding by and presentations within the Catholic University Centre (2023–2026). The research leading to these results has further received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 101028505.

² See *infra* B.

³ V Rozite, J Miller, and S Oh, ‘Why AI and Energy Are the New Power Couple’, *IEA* (2 November 2023), <https://www.iea.org/commentaries/why-ai-and-energy-are-the-new-power-couple>.

From a normative standpoint, the article operates at the intersection of *de lege lata* and *de lege ferenda*. On the one hand, it interprets existing EU legal materials, including fundamental rights, general principles of EU law, and regulatory instruments. On the other hand, it allows for a normative reconstruction of these materials through the articulation of justice-oriented principles intended to guide the development and deployment of AI systems beyond the current regulatory framework. These principles therefore should be read not as purely descriptive of the law as it stands or as abstract philosophical ideals but as constitutional proposals embedded within the EU legal order, aimed at addressing identified normative gaps.

This focus responds to the absence of a coherent set of guiding principles in the EU Artificial Intelligence Act (AI Act), a structural shortcoming that undermines also the possibility to shape a regulatory model centred on human rights.⁴ The European Parliament’s introduction of such principles, largely by replicating those developed by the High-Level Expert Group on AI, did not substantially improve the AI Act on this issue.⁵ The resulting framework risks functioning as a vademecum of pre-existing values, often redundant with established EU law (eg data protection, non-discrimination) or too vague to effectively guide AI development and deployment.⁶ In this sense, the absence of clear guiding principles in the AI Act is not merely a legislative gap but a constitutional one, also vis-à-vis more principled pieces of EU digital legislation such as the General Data Protection Regulation (GDPR).⁷

This approach aligns with the literature on European digital constitutionalism, which explores in particular the protection of fundamental rights and the regulation of private digital actors.⁸ In particular, the EU is currently transitioning from digital liberalism, prioritising economic innovation, to digital constitutionalism, which explores (i) the horizontal application of human rights and (ii) the alignment of digital law with existing regulatory frameworks.⁹

A necessary limitation of the proposed principles concerns their jurisdictional and enforcement reach. The proposal is inspired by the EU AI Act’s extraterritorial logic (Art 2), which extends its scope beyond formal market placement or deployment. In particular, the Regulation anticipates scenarios in which it governs AI systems that are operated outside the Union and yet the resulting outputs are used within the EU, thereby preventing regulatory circumvention and ensuring effective protection of

⁴ A Mantelero, ‘The AI Act: A Realpolitik Compromise and the Need to Look Forward’ in I Spiecker, L Schertel Mendes, and R Campos (eds), *Digital Constitutionalism* (Nomos, 2025), p 332.

⁵ Ibid.

⁶ Ibid, 332–33.

⁷ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance) OJ L 119, 4.5.2016, pp 1–88.

⁸ O Pollicino and F Paolucci, ‘Digital Constitutionalism’ in L Floridi and M Taddeo (eds), *A Companion to Digital Ethics* (Wiley, 2025).

⁹ Ibid.

fundamental rights. The practical enforceability of such extraterritorial obligations, however, remains largely untested. This is particularly relevant for the present proposal's reliance on duties of *diligentia diligentis* (the diligence of a reasonably prudent person) imposed on AI providers and deployers, which presupposes a degree of regulatory reach that may not fully materialise, especially in transatlantic contexts, given widely different regulatory approaches and risks of regulatory arbitrage (see Section IV.B).

Within this context, the reference to Hannah Arendt serves a methodological function. Arendt's idea of a 'right to have rights' and her call for a legal order capable of addressing humanity as a whole, as later explored, are invoked to interpret the constitutional fabric of current AI regulatory efforts in EU law. In this sense, Arendt's thought is mobilised to illuminate the tension between universality and situated legal orders, as well as to support the articulation of principles that, while grounded in EU law, aspire to broader normative relevance. Accordingly, this article understands the EU as more than a regulatory actor. At the present juncture, the EU rather seems a constitutional laboratory in which the principles governing AI can be developed, tested, and potentially projected beyond its borders.

II. Background, assumptions, and limitations

Offering an Arendtian take on AI and its place in the world of international law is not as baffling as it may initially appear. Notably, Arendt was a thinker endowed with a profound historical consciousness and a pioneer of many of the key themes of the 21st century: progress, the crisis of modern science, the rise of the bureaucrat and mass societies, as well as the opposition between culture and technology. Moreover, AI was first conceptualised during Arendt's life, becoming as much a product of the 20th century as it is of the 21st century.

The term 'artificial intelligence' was coined in 1956 by information theorist Claude Shannon during a conference at Dartmouth College that was sponsored by the US Defense Advanced Research Projects Agency.¹ However, Alan Turing's pivotal paper in 1950, published in *Mind*, already suggested new avenues of inquiry: can a machine think? Can a machine be linguistically indistinguishable from a human? Turing also introduced new methods of inquiry, such as the Turing Test, a tool to evaluate a machine's capability to demonstrate intelligent behaviour akin to that of a human in an 'imitation game' of sorts.² The Turing Test posed a philosophical inquiry into how minds work, a theme also dear to Arendt.

Arendt's engagement with questions that later resonated with debates on AI must be situated within her broader philosophical project. While AI was still in its infancy in 1958, Arendt identified the change in the constellation and mutual relationship of human capabilities from *animal laborans* to *homo faber* up to individuals capable of action (*homo politicus*), with all their self-created risks in an increasingly technological world.³

Building on this insight, Arendt articulated a framework of human capacities that structures the way we relate to nature, the world, and one another, which can be classified into three capacities: labour—dictated by natural needs; work—poised to create a safe space for existence; and action—the coming together as equals in a community.

¹ S Bringsjord and NS Govindarajulu, 'Artificial Intelligence' in EN Zalta and U Nodelman (eds), *The Stanford Encyclopedia of Philosophy (Summer 2026 Edition)*, forthcoming URL <https://plato.stanford.edu/archives/sum2026/entries/artificial-intelligence/> (currently available pre-archive at <https://plato.stanford.edu/entries/artificial-intelligence/#HistAI>).

² HJ Levesque, *Common Sense, the Turing Test, and the Quest for Real AI* (MIT Press, 2019), p 59.

³ H Arendt, *Between Past and Future* (Penguin Classics, 2006), p 62.

Further, the ability to act *and* speak differentiates human beings from other animals or the operations of machines.⁴ Even more importantly, Arendt’s phenomenological deconstruction reveals how the machine world has become a substitute for the real world, even though such a ‘pseudo-world’ cannot fulfil the most important task of human action, which is to provide mortals with a more permanent and stable dwelling than themselves.⁵ To enable such a dwelling, Arendt complemented her theory of action among individuals (*vita activa*) with a theory of activity within individuals (*vita contemplativa*), namely the life of the mind.⁶ In Arendt, the law’s place is within *action*: it serves as the walls for the political life of citizens, endowing power with institutional stability and increasing the scope of legitimate action.⁷

Arendt emphasised that, for the first time in history, the human capacity for action had begun to dominate all other capacities—the capacities of *homo faber* and *human animal laborans*, and the capacity for contemplation. Indeed, since the 20th century, the human capacity for action has been epitomised by technology, which has arisen as the meeting ground of history and nature.⁸ To be sure, technology has proven able to act not only over but *into* nature as we used to act in human affairs.⁹ The technological world we have created differs from the mechanised, post-Industrial Revolution era of *homo faber* and presents new conundrums. In this context, human action no longer fabricates objects but instead generates natural processes and channels them ‘into the human artifice and the realm of human affairs’.¹⁰ It unleashes a possibly ‘endless new chain of happenings whose eventual outcome the actor is utterly incapable of knowing or controlling beforehand’.¹¹

In this article, AI encompasses a rapidly evolving family of technologies comprising machine-based systems that—for explicit or implicit objectives—infer from the input they receive ‘how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments’ with which they interact.¹² Machine learning models have made it possible for AI to simulate human intelligence by swiftly evaluating data and inputs, generating new data, and even

⁴ Arendt, *The Life of the Mind* (n 1), Thinking, p 70: humans are ‘subject to labor in order to live, motivated to work in order to make themselves at home in the world, and roused to action in order to find their place in the society of their fellow-men’, see here, p. 70: <https://maytemunoz.net/wp-content/uploads/2016/10/hannah-arendt-the-life-of-the-mind.pdf>.

⁵ H Arendt, *The Human Condition*, 2nd ed (University of Chicago Press, 1998).

⁶ M Borren, ‘Hannah Arendt: Plural Agency, Political Power and Spontaneity’ in T Keiling and C Erhard (eds), *Routledge Handbook of Phenomenology of Agency* (Routledge, 2020), p 160.

⁷ H Arendt, ‘The Great Tradition I: Law and Power’ (2007) 74 *Social Research* 713, 717. See H Arendt, *Men in Dark Times* (Mariner Books, 1970), pp 81–82.

⁸ Arendt, *Between Past and Future* (n 12), p 61.

⁹ Ibid, pp 58–59.

¹⁰ Ibid, pp 59–60.

¹¹ Ibid.

¹² See Recommendation of the Council on Artificial Intelligence (2017, revised in 2024), <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>, I.

possessing the ability to self-program and alter their own codes.¹³ Natural language processing (NLP) is used to understand, extract, and use key information from text. Building on NLP, large language models (LLMs) are computing systems loosely inspired by the neurons in the brain, creating artificial neural networks (ANNs).¹⁴ In the form of language models, most recently, AI has been able to generate human-like text based on the input it receives, enabling natural-sounding conversations and providing responses.¹⁵

The clean energy sector is taking early steps to incorporate AI,¹⁶ which has the potential to decarbonise as well as democratise energy systems¹⁷ while contributing to the UN-developed Sustainable Development Goals (SDGs).¹⁸ On the other hand, AI's dark sides are increasingly emerging as a result of its energy consumption and carbon footprint,¹⁹ its built-in biases,²⁰ and its inability to truly understand meaning, operating more as a stochastic 'parrot'.²¹ Because LLMs can reorganise without any grounding and awareness of reality, they are not intelligent in the sense that they possess the capacity to understand. Rather, LLMs are endowed with a semblance of

¹³ Ibid.

¹⁴ D Valdenegro, 'A Large Language Models Digest for Social Scientists' (Leverhulme Centre for Demographic Science, University of Oxford), <https://osf.io/m74vs/download>, p 1.

¹⁵ Ibid, p 8.

¹⁶ Ibid; NM Kumar et al, 'Distributed Energy Resources and the Application of AI, IoT, and Blockchain in Smart Grids' (2020) 13 *Energies* 5739; L Yu et al, 'A Review of Deep Reinforcement Learning for Smart Building Energy Management' (2021) 8 *IEEE Internet Things Journal* 12046; R Pierdicca et al, 'Automatic Faults Detection of Photovoltaic Farms: SolAIr, a Deep Learning-Based System for Thermal Images' (2020) 13 *Energies* 6496; A Khanna et al, 'Biodiesel Production from Jatropha: A Computational Approach by Means of Artificial Intelligence and Genetic Algorithm' (2023) 15 *Sustainability* 9785.

¹⁷ E Colombo, 'Storage for Energy Justice: The Role of Energy Communities in EU Law' (2024) 35 *Yearbook of International Environmental Law* 1.

¹⁸ Z Fan et al, 'Deep Learning and Artificial Intelligence in Sustainability: A Review of SDGs' (2023) 15 *Renewable Energy, and Environmental Health* 1; UN General Assembly (UNGA), 'Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development', Res A/78/L.49 (2024). See also A Knauer, 'The First United Nations General Assembly Resolution on Artificial Intelligence', *EJIL:Talk!* (2 April 2024).

¹⁹ R Verdecchia, J Salloou, and L Cruz, 'A Systematic Review of Green AI' (2023) 7 *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1.

²⁰ On the 'alignment problem', see R Berkowitz, 'Humanity in the Nuclear and AI Ages', *Hannah Arendt Center for Politics and Humanities* (16 July 2023), <https://hac.bard.edu/amor-mundi/humanity-in-the-nuclear-and-ai-ages-2023-07-16>.

²¹ EM Bender et al, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' in *FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2021), pp 610–23.

agency²² and process immense volumes of data, which remain only shadows of reality, much like in Plato’s allegory of the cave.²³

This article proposes a baseline of three principles for justice-centred AI, particularly in energy matters, from an Arendtian perspective. The turn to Arendt for conceptualising norms rests on two reasons. First, in Arendt the law is constitutive of a political community as it ‘creates first of all a space in which it is valid, and this space is the world in which we can move with freedom’.²⁴ Harking back to the etymology of the Greek term for law, *nomos*, Arendt calls it a space within which defined powers may be legitimately exercised.²⁵ The law should be constitutive—centring on constitutional and empowering rules, thereby opening new avenues for the plurality of human conduct—rather than solely regulative, hinging on the experience and content of prescriptions.²⁶ Second, Arendt’s perspective on humanity is ‘earthbound’ rather than human-centred, opening a more ecological set of principles to integrate AI technology within energy systems.²⁷

In this article, energy justice is the ‘goal of achieving equity in both the social and economic participation in the energy system, while also remediating social, economic, and health burdens on those historically harmed by the energy system’.²⁸ Ultimately, the goal is to trigger some further thinking, willing, and judging, in Arendt’s sense, around digital constitutionalism ‘as the embodiment of the limits to the exercise of powers in a networked society’.²⁹ Digital constitutionalism is poised to emancipate the debate on law and technology from the shackles of a technocratic, intellectual property, or solely privacy-bounded perspective,³⁰ particularly in the European Union. At the same time, no magical solution to harness AI for energy justice can be offered. More realistically, as for all types of justice, energy justice is a receding horizon,³¹ slipping

²² L Floridi, *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities* (Oxford University Press, 2023), p 48.

²³ State of the Union 2024 (23 May | Refettorio—Badia Fiesolana | EUI), www.youtube.com/watch?v=tS4Er1-S11Q&t=8507s, at 5:51–52.

²⁴ H Arendt, ‘Was ist Politik? Fragmente aus dem Nachlaß’ (ed U Ludz) (Piper, 2003), p 122, as translated by H Lindahl, ‘Give and Take: Arendt and the Nomos of Political Community’ (2006) 32 *Philosophy and Social Criticism* 881, 882.

²⁵ FM Cornford, ‘From Religion to Philosophy: A Study in the Origins of Western Speculation’ (Harper Torchbooks, 1957), p 30, as cited by H Arendt, *On Revolution* (Penguin Books, 1963), 186–87.

²⁶ M La Torre, ‘Hannah Arendt and the Concept of Law: Against the Tradition’ (2013) 99 *Archiv für Rechts- und Sozialphilosophie* 400.

²⁷ For an ecological approach to Arendt, see O Belcher and JJ Schmidt, ‘Being Earthbound: Arendt, Process and Alienation in the Anthropocene’ (2021) 39 *Environment and Planning D: Society and Space* 103.

²⁸ S Baker, S DeVar, and S Prakash, *The Energy Justice Workbook* (2019), p 5, <https://iejusa.org/wp-content/uploads/2019/12/the-energy-justice-workbook-2019-web.pdf>.

²⁹ O Pollicino, ‘Introduction’ in G De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (Oxford University Press, 2022), p xiii.

³⁰ *Ibid.*, p xiv.

³¹ K Tester, *Conversations with Zygmunt Bauman: Zygmunt Bauman and Keith Tester* (Polity Press, 2001), p 49, with reference to utopia and socialism. See similarly DJ Hill, ‘International Justice’

away as soon as it appears close. The quest for justice is not likely to receive an answer even in ‘just societies’, as societies are ‘just’ only insofar as they continue to question their level of justice.³²

While this article emphasises the growing power of corporate actors in AI-enabled energy systems, historical experience cautions that the concentration of power within the state has often posed equal, if not greater, risks to democratic governance. This dual concern reflects Arendt’s own diagnosis of totalitarianism, where the entwinement of bureaucratic rationality and technological systems enabled unprecedented forms of domination.³³ From a republican perspective, the relevant concern is not the identity of the actor, state, or corporate but the existence of arbitrary power capable of interfering with freedom.³⁴

Overall, this article is structurally limited by the realisation that AI governance is an emerging field of study. In fact, the regulation of AI unfolds under conditions of epistemic and normative uncertainty: the technologies under the AI umbrella keep evolving, knowledge of risks remains fragmented,³⁵ many harms are emergent and difficult to anticipate, and regulatory frameworks often institutionalise rather than resolve uncertainty.³⁶ At the same time, efforts to embed human values into AI systems may raise further challenges, as values are plural, context-dependent, and subject to contestation.³⁷ Accordingly, regulatory proposals should be understood not as definitive solutions but as iterative and revisable interventions within complex socio-technical systems.

(1896) 6 *Yale Law Journal* 1, 2–3; A Dershowitz, *Rights from Wrongs: A Secular Theory of the Origin of Rights* (Basic Books, 2004), pp 8–9; A Sen, *The Idea of Justice* (Allen Lane, 2009), pp viiff and 8–9; F Stella, *La giustizia e le ingiustizie* (Il Mulino, 2006), esp p 13ff; and M Galanter, ‘Access to Justice in a World of Expanding Social Capability’ (2010) 37 *Fordham Urban Law Journal* 115, 125. See the metaphor of justice as the ocean floor that the eye cannot see when one is out at sea but is nonetheless there, concealed by the depths too deep, Alighieri (1321), *Paradiso*, Canto XIX, 58–63.

³² See Z Bauman, *Society under Siege* (Polity Press, 2002), p 54. See the ungraspable ‘justice’ component of international law in T Ahmed and D French, ‘Situating Climate Change in (International) Law: A Triptych of Competing Narratives’ in S Farral, T Ahmed, and D French (eds), *Criminological and Legal Consequences of Climate Change* (Hart, 2012), p 252.

³³ H Arendt, *The Origins of Totalitarianism* (Meridian Books, 1962), p 157ff.

³⁴ On the ideal of freedom as non-domination, see P Pettit, *Republicanism: A Theory of Freedom and Government* (Clarendon Press, 1997), p 130ff.

³⁵ BW Wirtz, JC Weyerer, and I Kehl, ‘Governance of Artificial Intelligence: A Risk and Guideline-Based Integrative Framework’ (2022) 39 *Government Information Quarterly* 1.

³⁶ S Carey, ‘Regulating Uncertainty: Governing General-Purpose AI Models and Systemic Risk’ (2026) 17 *European Journal of Risk Regulation* 123.

³⁷ C Orwat et al, ‘Normative Challenges of Risk Regulation of Artificial Intelligence’ (2024) 18 *Nanoethics* 10.

III. The constitutive gap for justice-centred AI in energy matters

A. Introduction

A central challenge for justice-centred AI in energy contexts is what I term the ‘constitutive gap’. This gap reflects the distance between the democratic ideal of collective deliberation and the lived reality of individuals no longer ‘acting politically’ but ‘merely behaving’ as economic producers, consumers, and dwellers.¹

In the energy context, AI is a double-edged sword, poised to either perpetuate or possibly improve the current centralised and insufficiently democratic decision-making processes for energy systems.² Artificial neural networks and expert systems have been used for over 30 years in the energy sector, optimising the efficiency of several tasks and becoming the most utilised digital technology in electricity systems.³ On the other hand, companies in the energy sectors, depending on their development and deployment of AI systems, are often deemed ‘safety-critical systems’, namely high-risk AI systems whose failure would put the health and well-being of citizens at risk.⁴

Even the most comprehensive regulatory framework on AI, the EU AI Act, tackled the role of citizens through the means of the administrative state while missing the opportunity to fully recognise individuals’ rights.⁵ On the one hand, the EU AI Act has set the highest bar in AI regulatory requirements through proactive gover-

¹ S Benhabib, ‘The Embattled Public Sphere: Hannah Arendt, Jürgen Habermas, and Beyond’ in E Ullmann-Margalit, *Reasoning Practically* (Oxford University Press, 2000), pp 165–66.

² JC Stephens, ‘Energy Democracy: Redistributing Power to the People through Renewable Transformation’ (2019) 61 *Environment: Science and Policy for Sustainable Development* 4, 9.

³ F Heymann et al, ‘Operating AI systems in the Electricity Sector under European’s AI Act—Insights on Compliance Costs, Profitability Frontiers and Extraterritorial Effects’ (2023) 10 *Energy Reports* 4538, 4538–39.

⁴ Ibid, 4540.

⁵ Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828 (Artificial Intelligence Act), Brussels, 14 May 2024 (OR en) 2021/0106(COD) PE-CONS 24/24 (hereinafter, EU AI Act).

nance,⁶ both content and institution-wise. Content-wise, it has set forth a ‘risk-based’ approach, whereby the higher the risk of causing societal harm, the stricter the rules. Institution-wise, pursuant to the AI Act, an AI Office has been established within the EU Commission to ensure enforcement, as supported by an AI Board of member states’ representatives, a scientific panel of independent experts, and an advisory forum for stakeholders to provide technical expertise to the AI Board and the Commission, which adds to national enforcement authorities. On the other hand, the Achilles’ heel of the AI Act, beyond its level of generality regarding AI technologies and domains of application, is its effective enforcement, which is based on enforcement by public authorities alone.⁷ Member states will decide on penalties, and the main guidance point from the AI Act is the imposition of a range of administrative fines.⁸ Only high-risk AI systems are required to carry out a fundamental rights assessment, which *may* be judicially interpreted as conferring enforceable rights onto natural persons or groups.⁹

The wide use of AI by energy actors makes the lack of basic statistics on the quantity and structure of electricity market agents even more striking.¹⁰ Meanwhile, the literature is patchy regarding citizen participation in centralised systems and representational forms of democratic governance.¹¹ Such gaps raise questions on how to achieve more transparency across the actors tasked with adopting and enforcing AI systems in the energy space. To bridge the constitutive gap between current energy systems and the democratic ideal of collective deliberation, at least two re-conceptualisations are needed: the meaning of AI and its principled regulation.

B. The conceptual grounding of justice-centred AI

The vast array of qualifications of AI—trustworthy, human-centred, ethical, explainable, responsible, and fair—fail to include a justice component. Justice-centred AI is a relatively overlooked concept, whereby some scholars and leaders in AI ethics have

⁶ On proactive AI governance, see OJ Gstrein, N Haleem, and A Zwitter, ‘General-Purpose AI Regulation and the European Union AI Act’ (2024) 13 *Internet Policy Review* 1, 8ff and 16ff.

⁷ EU AI Act, Art 3(45)–(47). See similarly Gstrein et al (n 52), p 17ff.

⁸ Upon the infringement of *prohibitions*, fines are up to 35,000,000 EUR or, if the offender is an undertaking, up to 7% of its total worldwide annual turnover for the preceding financial year, whichever is higher (EU AI Act Arts 5 and 99(3)). Upon infringing *lower-level* obligations, administrative fines of up to 15,000,000 EUR or, if the offender is an undertaking, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher (EU AI Act Art 99(4)). Upon infringement of incorrect, incomplete, or misleading information to public bodies in reply to a *request*, administrative fines can amount to up to 7,500,000 EUR or, if the offender is an undertaking, up to 1% of its total worldwide annual turnover for the preceding financial year, whichever is higher (EU AI Act Art 99(5)).

⁹ Ibid, Art 27.

¹⁰ Heymann et al (n 49), 4551.

¹¹ M Wahlund and J Palm, ‘The Role of Energy Democracy and Energy Citizenship for Participatory Energy Transitions: A Comprehensive Review’ (2022) 87 *Energy Research and Social Science* 1.

focused on ensuring that AI systems promote equity, inclusion, transparency, accountability, the ethical use of data, representation, participation, the correction of historical biases, redress, and remedies.¹² A growing critique suggests that embedding ethical values into AI systems risks introducing distortions into their mathematical structure, potentially generating opacity and unpredictability.¹³ However, this view presupposes a neutrality of computational systems that has been extensively challenged in the philosophy of science,¹⁴ as well as in mathematics¹⁵ and law.¹⁶ Rather than introducing values *ex post*, AI systems are already shaped by implicit normative choices of data, models, thresholds, and objectives. The new form of agency introduced by AI¹⁷ is intrinsically shaped by goals and thus carries an inherent axiological dimension. Artificial agency, in fact, emerges from the interplay between programmed objectives and learned behaviours, constituting ‘a computational, goal-driven form of agency defined by human purposes’.¹⁸ In this sense, AI agency marks a shift away from biologically grounded purposiveness towards engineered goal-directedness.¹⁹ The question, therefore, is not whether values should be embedded but how they can be made explicit, contestable, and subject to democratic judgement. In the specific domain of energy law, the need for justice-centred AI becomes particularly acute. Energy systems are deeply intertwined with questions of access, affordability, sustainability, and governance, all of which are central to energy justice.

Following Thomas Franck, fairness is constituted of a procedural aspect—proper process—and a substantive aspect—distributive justice.²⁰ The procedural and substantive aspects are intrinsically related. In fact, legal systems are perceived as fair so long as the rules satisfy ‘the participants’ expectations of a justifiable distribution of costs and benefits’ and ‘the rules are made and applied in accordance with what the

¹² Eg V Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor* (St Martin’s Press, 2018); C O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, 2016).

¹³ On the ethicist versus engineering framing, see J Harris and V Dubljević, ‘Navigating the Ethics of Artificial Intelligence’ (2025) 5 *Encyclopaedia* 2025 1, 4.

¹⁴ B Xavier, ‘Biases within AI: Challenging the Illusion of Neutrality’ (2025) 40 *AI and Society* 1545. While acknowledging that no part of science can be considered value-free, what is crucial is to examine the constraints governing how such values are incorporated and applied, HE Douglas, *Science, Policy, and the Value-Free Ideal* (University of Pittsburgh Press, 2009), p 56.

¹⁵ DM Monte-Serrat and C Cattani, ‘Towards Ethical AI: Mathematics Influences Human Behavior’ (2023) 13 *Journal of Humanistic Mathematics* 469.

¹⁶ H Weerts et al, ‘The Neutrality Fallacy: When Algorithmic Fairness Interventions Are (Not) Positive Action’ in FAccT’24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery, 2024), p 2064.

¹⁷ L Floridi, ‘AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis’ (2025) 38 *Philosophy and Technology* 1.

¹⁸ *Ibid.*, 30.

¹⁹ *Ibid.*

²⁰ TM Franck, *Fairness in International Law and Institutions* (Clarendon Press 1998), p 7.

participants perceive as the right process’.²¹ In particular, distributive justice favours *change*.²² In the contested arena of values and their trade-off in a pluralistic society, distributive justice is also the foreground for *deliberation*, where ‘[h]eterogeneity and interpretative conflict’ are a resource, rather than a barrier, to deliberative problem-solving.²³

Distributive justice can be particularly challenging and difficult to incorporate into legal regimes, including integrating AI into energy matters. Initially, the environmental movement sidelined distributive justice as a distraction,²⁴ operating under the assumption that environmental policies would automatically yield a better environment for all. Since the 1980s, one of the most notable legal developments has been the rise of the environmental justice movement. At that time, less powerful communities realised that they were disproportionately impacted by environmental degradation but had limited possibilities to influence relevant decision-making processes. They thus resorted to legal action. Around the same time, the World Charter for Nature, adopted by the UN General Assembly in 1982, emphasised ecological justice by ensuring that all individuals have the right to seek legal redress for environmental damage or degradation. It acknowledged state sovereignty over natural resources while promoting public participation in environmental decision-making, thus allowing individuals to be recognised through actual involvement in and impact on decisions affecting their environment.²⁵ In the mid-2010s, the concept of climate justice was incorporated as a subset of ecological justice, addressing the disproportionate burden of climate change impacts on poor and marginalised communities.²⁶

Energy justice emerged in an academic context in 2010.²⁷ In 2015, one of the first comprehensive works on energy justice defined it as a mechanism poised to achieve procedural and distributive justice.²⁸ Rather than socio-technical fixes, energy justice

²¹ Ibid.

²² Ibid.

²³ P Nanz, ‘Democratic Legitimacy and Constitutionalisation of Transnational Trade Governance: A View from Political Theory’ in C Joerges and E Petersmann (eds), *Constitutionalism, Multilevel Trade Governance and Social Regulation* (Hart, 2006), p 59ff.

²⁴ J Purdy, ‘The Long Environmental Justice Movement’ (2018) 44 *Ecology Law Quarterly* 809, 814–17.

²⁵ See UNGA, World Charter for Nature, Res 37/7 (28 October 1982), points 22–24.

²⁶ Both citations are from RS Abate, ‘Introduction’ in RS Abate (ed), *Climate Justice: Case Studies in Global and Regional Governance Challenges* (Environmental Law Institute (ELI), 2016), p xxxiii.

²⁷ L Guruswamy, ‘Energy Justice and Sustainable Development’ (2010) 21 *Colorado Journal of International Environmental Law and Policy* 231. See similarly A McHarg, ‘Energy Justice: Understanding the “Ethical Turn” in Energy Law and Policy’ in I del Guayo et al (eds), *Energy Justice and Energy Law* (Oxford University Press, 2020), 15 fn 1.

²⁸ BK Sovacool and MH Dworkin, ‘Energy Justice: Conceptual Insights and Practical Applications’ (2015) 142 *Applied Energy* 435, 435–36.

requires transformative politics.²⁹ Instead of simply decarbonising energy systems,³⁰ energy justice should facilitate the participation of interested communities in the design, function, and ownership of energy systems,³¹ particularly those devoid of the opportunities and the right to act.³² Ultimately, like energy justice and other conceptions of justice, justice-centred AI in energy matters would need to encompass procedural, substantive, and recognition aspects, ensuring that algorithmic decision-making in the energy sector advances, rather than undermines, democratic ideals and equitable outcomes.

C. Constitutive principles: towards principled regulation

As noted earlier, constitutive principles, rather than haphazard regulation,³³ are highly needed to bridge the constitutive gap between current energy systems and the democratic ideal of collective deliberation. Pursuant to the modern constitutional tradition, such principles are constitutive in the sense of a constituent power reclaiming agency and self-government in a landscape shaped mainly by others.³⁴ Artificial intelligence makes what Nico Krisch calls the post-national sphere even more vivid, where ‘the structure of governance is the result of multiple interacting moves, organic growth and choices of powerful actors’.³⁵ It is challenging to determine the ‘who’ and the ‘what’ of political outcomes in the resulting networks of actors where the category of fate, not of constitution, seems to rule.³⁶ In the digital age, power has not simply shifted from states to corporations. It has been reconfigured through the rise of digital platforms and data-driven business models, which enable private actors to exercise forms of economic, epistemic, and infrastructural power traditionally associated with public authority.³⁷ Rather than being displaced, states have been entangled with the rise of corporate digital power, as they ‘created the conditions for extraction, legit-

²⁹ Baker et al (n 37), p 25.

³⁰ See similarly D Fairchild and A Weinrub, *Energy Democracy: Advancing Equity in Clean Energy Solutions* (Island Press, 2017), p 11.

³¹ Baker et al (n 37), p 26.

³² E Colombo, ‘The Politics of Silence: Hannah Arendt and Future Generations’ Fight for the Climate’ (2023) 17 *Vienna Journal on International Constitutional Law (ICL Journal)* 43, at 56.

³³ H Roberts et al, ‘Global AI Governance: Barriers and Pathways Forward’ (2024) 100 *International Affairs* 1275–86; K Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial* (Yale University Press, 2021).

³⁴ N Krisch, ‘*Pouvoir constituant* and *pouvoir irritant* in the Postnational Order’ (2016) 14 *International Journal of Constitutional Law* 657.

³⁵ Ibid, 667.

³⁶ Ibid, 668.

³⁷ B Schneier and NE Sanders, *Rewiring Democracy: How AI Will Transform Our Politics, Government, and Citizenship* (MIT Press, 2025), p 12.

imized its underlying logic, and often shielded it from democratic oversight’.³⁸ This development coexisted with a long-term expansion of state regulatory and administrative capacity since the post-World War II period. Ultimately, the commodification of human behaviour has given rise to surveillance capitalism, a new economic order reclaiming human experience as free raw material for its transformation for analysis and sales.³⁹ This model is distinct from, yet potentially mutually reinforcing with, government surveillance, and its expansion may contribute to more democratic disorder and de-institutionalisation.⁴⁰

If a power of new beginnings can nonetheless be rescued with regard to the ‘power couple’ of AI and energy (see Section I), it would be a *pouvoir irritant*—meaning, several critical stings irritating what can be deemed established constituted power.⁴¹ From an Arendtian perspective, this type of constitutional power, albeit fragmented, allows for the ‘constitution of political freedom’, which creates new centres of power and new productive capacities among citizens.⁴² Combining the ideas of Arendt with those of Alexander Hamilton, constitutive principles provide an opening—a ‘placeholder’—for societal aspirations to establish themselves out of reflection and choice.⁴³ The recent drive towards regulating the use of data and data-powered technologies⁴⁴ is here described as one of the possible forms of digital constitutionalism, particularly ‘as a

³⁸ B Kilic, ‘The Geopolitics of Surveillance Capitalism’ (Carr-Ryan Center for Human Rights, Harvard Kennedy School, Harvard University, 27 October 2025, issue 2025–07), p 4.

³⁹ S Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs, 2019), p 112ff.

⁴⁰ S Zuboff, ‘Surveillance Capitalism or Democracy? The Death Match of Institutional Orders and the Politics of Knowledge in Our Information Civilization’ (2022) 3 *Organization Theory* 1, 5.

⁴¹ In general, Krisch (n 80), pp 673, 678; G Teubner, *Constitutional Fragments: Societal Constitutionalism and Globalization* (Oxford University Press 2012), p 62.

⁴² Arendt (n 33), p 147; J Muldoon, ‘Arendt’s Revolutionary Constitutionalism: Between Constituent Power and Constitutional Form’ (2016) 23 *Constellations* 596, 599.

⁴³ The Federalist No. 1 (Alexander Hamilton). See also Krisch (n 80), p 677.

⁴⁴ OECD.AI (2021), powered by EC/OECD (2021), Database of National AI Policies, <https://oecd.ai>. At the EU level, see, eg, the EU AI Act, the AI Liability Directive: Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (Text with EEA Relevance) SEC(2022) 344 final, SWD(2022) 318 final, SWD(2022) 319 final, SWD(2022) 320 final, and the Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L 2024/2853. See also Regulation (EU) 2019/881 on the European Union Agency for Cybersecurity and on Information and Communications Technology Cybersecurity Certification (Cybersecurity Act) and Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 (Text with EEA Relevance) PE/86/2018/REV/1 OJ L 151, 7.6.2019, pp 15–69 and Directive (EU) 2020/1828 (Data Act) (Text with EEA Relevance) PE/49/2023/REV/1 OJ L, 2023/2854, 22.12.2023.

reaction to private norms and external interferences from other standards of protection'.⁴⁵

⁴⁵ G De Gregorio, 'How Does Digital Constitutionalism Reframe the Discourse on Rights and Powers? A Theoretical Lens to Understand Digital Policy Developments', *Ada Lovelace Institute* (7 December 2022).

IV. Lifting the veil of the world

A. Regulating technology from an earthbound perspective

In Arendt’s terms, humankind has always been tempted to lift the veil of the future with the aid of technology, which can be broadly defined as the application of scientific knowledge for practical purposes.¹ Such will has accelerated with the idea of progress that emerged in the modern age and its drive to subject the world to its rule.² With the rise of science in the modern age, progress emerged as a new notion in the 17th century, becoming ‘the most cherished dogma of *all* men living in a scientifically oriented world’.³ In the 18th and 19th centuries, with the transposition of progress from a concept of natural science to one concerning history and human affairs, exemplified by Hegel and Marx, a theory of ethics was ‘treated in the perspective of History and on the assumption that there is such a thing as Progress of the human race’.⁴

Early on, Arendt underscored the first boomerang effects of science’s great triumphs, namely that the ‘truths’ of the modern scientific worldview can be demonstrated through our know-how—mathematical formulas and technological proofs—but cannot be fully understood in thought and speech.⁵ Owing to this incapability of thought, we risk becoming enslaved not so much by our machines as by our know-how, becoming ‘thoughtless creatures at the mercy of every gadget which is technically possible’.⁶ Presciently, Arendt underlined the difficulties of communicating the ‘truth’ of the modern scientific worldview, which is a sheer inability ‘to understand, that is, to think and speak about the things which nevertheless we are able to do’ through our action *into* nature.⁷ In our technological world’s communicative gap, machines would overtake our brain, Arendt states, ‘so that from now on we would indeed need artificial machines to do our thinking and speaking’.⁸

¹ Arendt, *The Life of the Mind* (n 1), Will, p 159.

² Ibid, p 158, referring to Nietzsche and Heidegger.

³ Arendt, *The Life of the Mind* (n 1), Thinking, p 26.

⁴ Ibid, p 216.

⁵ N Hassenfeld, ‘Even the Scientists Who Build AI Can’t Tell You How It Works’, *Vox* (15 July 2023), www.vox.com/unexplainable/2023/7/15/23793840/chat-gpt-ai-science-mystery-unexplainable-podcast.

⁶ Arendt, *The Human Condition* (n 14), p 3.

⁷ Ibid.

⁸ Ibid.

If technology cannot be fully thought of or spoken about, it strikes us in its apoliticality. The loneliness of human beings before technology emerges from the silence of traditional political and philosophical accounts of the world, which are themselves inadequate to interpret scientific and technological advances. Before the precipice of progress, we are also left without a banister towards meaningfulness. Insofar as human beings live and move and act in this world, they experience meaningfulness only because they can talk with and make sense of each other and themselves, which makes them political beings.⁹

If the fear that human beings could be enslaved by technology has also characterised other authors,¹⁰ Arendt's view shifts the focus from humans to the Earth, proving particularly heuristic for the purposes of this article.¹¹ In Arendt, the Earth is the quintessence of the human condition, turning 'earthly nature' into a unique set of habitats where human beings can move and breathe without artifice. Through life thus intended, human beings remain earthbound, namely related to 'all other living organisms'. Conversely, the human artifice separates us from this 'mere animal environment'.¹² In this context, humankind seems to exchange existence—a gift given from nowhere, secularly speaking—for something that humankind has made *itself*, technology and the artifice more broadly.¹³ Towards the same rebellious end of exchanging nature for artifice, this future human being has proved capable of destroying 'all organic life on earth', prompting questions on the following: which is the direction towards which we are using our new and scientific-technical knowledge? Far from being nostalgic for a pre-scientific era, Arendt tackled the ethics of technology because it is 'a political question of the first order', which cannot be left to professional scientists or politicians.¹⁴

Accordingly, Arendt approached the technology question from an intrinsically 'earthbound' perspective. In this way, Arendt seems to foreground some of the tenets of the Anthropocene, whereby human and geological times intersect¹⁵ in the 'progress' towards self-destruction, as mediated by technocratic dreams of world domination.¹⁶ In

⁹ Ibid.

¹⁰ Eg U Galimberti, *Psiche e techne: l'uomo nell'età della tecnica* (Feltrinelli, 2008); E Severino, *Essenza del nichilismo* (Adelphi, 1982), pp 196–97; M Heidegger, 'The Question Concerning Technology and Other Essays' (trans W Lovitt) (Garland, 1977); M Horkheimer, *Eclipse of Reason* (Oxford University Press, 1947).

¹¹ In a similar strand, see H Jonas, 'Technology and Responsibility: Reflections on the New Tasks of Ethics' (1973) 40 *Social Research* 31.

¹² On both points, Arendt, *The Human Condition* (n 14), p 2. On another earthbound perspective, less concerned with the life of the mind, see H Jonas, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age* (University of Chicago Press, 1984).

¹³ Arendt, *The Human Condition* (n 14), pp 2–3.

¹⁴ Ibid.

¹⁵ D Chakrabarty, 'The Climate of History: Four Theses' (2009) 35 *Critical Inquiry* 197.

¹⁶ D Danowski and E Viveiros de Castro, *The Ends of the World* (trans R Nunes) (Polity Press, 2017), p 113ff; Belcher and Schmidt (n 36), 111.

response to the ensuing alienation from earthbound concerns, Arendt demands that science be reincorporated into the arenas of discussion and deliberation that underpin the political sphere.

B. AI regulation for the AI–energy power couple

Previously, I argued for the need to thoroughly regulate technology from an earth-bound perspective, which counters the idea that technology develops independently and that law cannot steer innovation in technology.¹⁷ This regulatory need is even more overt in AI regulation because AI is inherently difficult to regulate owing to its innovation ratio and the character of general-purpose technology, opening the flood-gates to use and misuse. Artificial intelligence regulation breaks down into at least three macro areas concerning (1) spending (eg private investments and public funding); (2) competition, innovation, and ownership (eg intellectual property and reduction of rent extraction); and (3) ethics and safety (eg human rights, consumer protection, bias and discrimination, transparency, self-replicability, and explainability). Such macro areas are not orthogonal and create trade-offs, for instance, between private investments and fair competition or consumer protection and innovation. Because of insufficient regulation, AI as a new power source has yet to be socially or politically legitimised through effective governance.¹⁸

The first national AI strategies appeared in 2017 and now number over 50, comprising over 930 policy initiatives across 71 jurisdictions.¹⁹ Several national strategies align with the 2019 Organisation for Economic Co-operation and Development (OECD) AI Principles, updated in May 2024²⁰—the first intergovernmental standard on AI. The OECD AI Principles identify five complementary value-based principles for the responsible stewardship of trustworthy AI: inclusive growth, sustainable development, and well-being; human-centred values and fairness; transparency and explainability; robustness, security, and safety; and accountability. The OECD AI Principles recognise that AI has the potential to contribute positively to sustainable global economic activity while potentially having disparate effects within and among societies.²¹ Accordingly, they emphasise the need for a well-informed, whole-of-society public debate to capture the beneficial potential of the technology while limiting the risks associated with it.²² While the principles stop short of considering AI in energy systems and seem

¹⁷ ‘Democracy’s Discontent: Michael Sandel with David Brooks’ (1 December 2022), www.youtube.com/watch?v=wVUVQNFjt5M&t=4s, minute 40ff.

¹⁸ S Vallor, Senate Homeland Security and Governmental Affairs Committee, Issues Testimony from University of Edinburgh Director, *Targeted News Service* (2 December 2023), p 2.

¹⁹ See OECD AI, <https://oecd.ai/en/wonk/national-policies-2>.

²⁰ Ibid.

²¹ OECD Recommendation of the Council on Artificial Intelligence, preamble paras 6 and 7.

²² Ibid, preamble para 8.

to conflate energy with climate and other AI sustainability issues, they importantly emphasise the need to empower stakeholder engagement.²³

The Council of Europe’s (COE) Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law was agreed upon by 57 states in May 2024 and is the first binding international treaty on AI. From the start, the aspiration has been for the convention to be adopted worldwide by non-COE states and to remain future-proof by *not* regulating technology,²⁴ which is a remarkable, laissez-faire policy statement. Notwithstanding, the convention reaffirms a commitment to a series of foundational human rights documents: the 1948 Universal Declaration of Human Rights, the 1950 Convention for the Protection of Human Rights and Fundamental Freedoms, the 1966 International Covenant on Civil and Political Rights, the 1966 International Covenant on Economic, Social and Cultural Rights, and the 1961 European Social Charter, as revised in 1996, as well as their respective protocols; the 1981 Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data and its protocols; the 1989 United Nations Convention on the Rights of the Child; and the 2006 United Nations Convention on the Rights of Persons with Disabilities.²⁵

Importantly, pursuant to Article 7 of COE’s the framework convention, each state party to the convention shall ‘adopt or maintain measures to respect human dignity and individual autonomy in relation to activities within the lifecycle of artificial intelligence systems’. As the other side of dignity, pursuant to Article 9 of the convention, each state party shall also adopt or maintain measures to ensure accountability and responsibility for adverse impacts on human rights. To fulfil the protection of dignity and human rights through accountability, state parties shall also impose accessible and effective remedies and procedural safeguards.²⁶

Another regional instrument, the EU AI Act, was the first comprehensive regional law on AI. It is meant to protect human rights and goes one step further by underscoring the need to reduce the consumption of energy and other resources by AI systems.²⁷ It also enshrines the possibility to create AI regulatory sandboxes that the European Commission would facilitate through delegated acts to avoid fragmentation across the Union.²⁸ Such regulatory sandboxes would enable a liberal, albeit controlled, use of personal data to develop certain AI systems in the public interest, such as for ‘en-

²³ Ibid, preamble para 1.1. See similarly European Declaration on Digital Rights and Principles for the Digital Decade (2023/C 23/01).

²⁴ <https://www.coe.int/en/web/artificial-intelligence/cai> on the Committee on Artificial Intelligence (CAI), serving until 2025, and the succeeding Steering Committee for New and Emerging Digital Technologies (CDNET).

²⁵ Council of Europe (COE) Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law [CETS 225]—Artificial Intelligence, [5.IX.2024] (hereinafter COE AI Convention), preamble paras 15 and 16.

²⁶ Ibid, Arts 14–15.

²⁷ EU AI Act Art 40(2).

²⁸ Ibid, Art 58(1).

ergy sustainability’.²⁹ Notably, the AI Act acknowledges the specificities of the latter compared to environmental, climate, or biodiversity matters.³⁰

Beyond such regional instruments, national attempts have mushroomed without attaining a sufficient level of specificity regarding the power couple of AI and energy. What stands out is regulatory divergence, in particular, among the EU’s preventative approach,³¹ the US digital laissez-faire,³² and China’s ‘strong government’ approach whereby, in case of conflict, national interests are meant to prevail over individual citizen rights.³³ Ultimately, the risk of regulatory arbitrage is real.

At the global level, G7 or country-initiated summits have yet to tackle the risks and opportunities of AI in energy systems,³⁴ while UN efforts to coordinate AI policy seem more promising. After a Mexican-led UN initiative received little interest in 2017,³⁵ in March 2024, the first UN resolution on AI, proposed by the United States, was adopted by the General Assembly in a consensus vote. In particular, the resolution calls upon member states and other stakeholders to refrain from or to cease using AI systems that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights, especially of those who are in vulnerable situations. At the same time, the resolution draws an equivalence test between online and offline rights protection throughout the life cycle of AI systems.³⁶

Although the March resolution stops short of considering AI in energy systems, it reaffirms a commitment not only to the UN Charter and the Universal Declaration of Human Rights but also to Agenda 2030, thus underscoring the role of AI in enabling the SDGs.³⁷ The resolution thus entrusts international law with providing the appropriate basis for devising AI governance systems that are ‘interoperable, agile,

²⁹ Ibid, Art 59(1)(a)(iii).

³⁰ Ibid, Art 59(1)(a)(ii).

³¹ EU AI Act, paras 47, 75, 115; Art 7(2)(k)(ii), although a precautionary approach to digital law has been presented as more heuristic, J Mazur, ‘Automated Decision-Making and the Precautionary Principle in EU Law’ (2019) 9 *Baltic Journal of European Studies* 3.

³² The White House, *A National Policy Framework for Artificial Intelligence: Legislative Recommendations* (20 March 2026), www.whitehouse.gov/wp-content/uploads/2026/03/03.20.26-National-Policy-Framework-for-Artificial-Intelligence-Legislative-Recommendations.pdf; The White House, *Promoting the Export of the American AI Technology Stack* (Executive Order, 23 July 2025), www.whitehouse.gov/presidential-actions/2025/07/promoting-the-export-of-the-american-ai-technology-stack/.

³³ J Cheng and J Zeng, ‘Shaping AI’s Future? China in Global AI Governance’ (2023) 32 (143) *Journal of Contemporary China* 794–810, 803.

³⁴ See the Bletchley Declaration by Countries Attending the AI Safety Summit (1–2 November 2023), as well as the agreement by G7 leaders on International Guiding Principles on Artificial Intelligence (AI) and a voluntary Code of Conduct for AI Developers under the Hiroshima AI Process, <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems>.

³⁵ E Zapata and C Gomez-Mont, ‘Mexico: The Story and Lessons behind Latin America’s First AI Strategy’, *Apolitical* (18 June 2020).

³⁶ UNGA (n27), preambular para 5.

³⁷ Ibid, preambular paras 1, 2, 6, 7, and 8.

adaptable, inclusive, responsive to the different needs and capacities of developed and developing countries alike and for the benefit of all'.³⁸

Last, it handed the baton of further international AI lawmaking to the Summit of the Future, where the Global Digital Compact was annexed to the Pact for the Future in September 2024.³⁹ In the Digital Pact, references to AI sustainability are few and far between, mainly revolving around Objective 1(e), which addresses sustainability across the life cycle of digital technologies with explicit reference also to SDG 7 on clean, affordable energy and SDG 13 on climate.⁴⁰

Overall, the regulatory frameworks discussed here provide some specificity regarding the sustainability requirements of AI systems. However, likely owing to gaps in international law on AI in energy systems, they tend to overlook the implications of AI from an energy justice perspective, as well as the role of communities and their deliberations in AI-enabled energy systems. In this respect, the reviewed frameworks seem to epitomize a trend in international law where transparency has increasingly supplanted accountability, participation, and even distributive justice.⁴¹

³⁸ Ibid, preambular para 9.

³⁹ Ibid, preambular para 11. See Resolution 76/307 of 8 September 2022 and Decision 77/568 of 1 September 2023 and its first revision: 'Pact for the Future: Rev.1' (14 May 2024), www.un.org/sites/un2.un.org/files/sotf-pact-for-the-future-rev.1.pdf.

⁴⁰ Pact for the Future, Global Digital Compact and Declaration on Future Generations, A/79/L.2, https://unsdg.un.org/sites/default/files/2025-03/sotf-pact_for_the_future_adopted.pdf.

⁴¹ A Peters, 'The Transparency Turn of International Law' (2015) 1 *Chinese Journal of Global Governance* 3, 15.

V. Three justice-centred principles for AI in energy systems

A. Introduction

I have argued that a justice-centred approach to AI is necessary, as technology evolves far more rapidly than regulation, while regulatory purposes remain so broad that they risk diverting efforts towards overly abstract initiatives. Existing legal and policy frameworks therefore provide little traction for addressing the concrete practices through which AI shapes society. What is missing is not more regulation at a higher level of generality but a principled orientation capable of guiding both governance and evaluation. In short, we lack a normative framework of justice with which to evaluate the practices that constitute AI.¹

The following three constitutive principles can offer a starting point for prospective discussions on the communitarian and individual role of AI in energy systems. In particular, these constitutive principles aim to illuminate how regulatory frameworks can harness technology for localised forms of political activity to ultimately enhance energy justice in AI contexts. After introducing the first principle, by which AI must be treated strictly as a means, never as an autonomous end, and the second principle, hinging on fiduciary duties to strengthen corporate accountability, the third principle approaches the role of AI in energy systems through the right to have rights, which would be relevant in at least three ways: as an ‘irritant’ to centralised powers at both public and private levels; as a reinforcer of human dignity against the over-exploitation of nature; and as an enabler of democratic deliberation in the increasing opposition between culture and technology.

B. First principle: the categorical imperative reversed

The first principle of justice-centred AI consists of a reversed categorical imperative. Informed by Kant’s second formulation of the categorical imperative, or principle of humanity, we ought to ‘treat humanity, whether in (our) own person or in that of any

¹ On the absence of a moral framework, see I Gabriel, ‘Toward a Theory of Justice for Artificial Intelligence’ (2022) 151 *Daedalus* 218.

other ... as an end withal, never as means only'.² Pursuant to this principle, providers and deployers of AI should not only treat humanity as an end but also treat AI as a *means* rather than an end—an approach that is not always observed. The relevance of the categorical imperative in AI is owing to its foundational role in the emergence of human rights in the 20th century and its permeation of the *corpus iuris* of international law into the present day,³ as well as to the existence of strikingly similar principles across cultural traditions and in secular humanism.⁴ Further, in its Renaissance tradition, dignity is the power to opt for different paths of development—"thou art confined by no bounds"—rewarding those who develop their intellect.⁵ At the same time, human dignity entails the capacity inherent in human nature to assume obligations vis-à-vis others, including nature.⁶ Importantly, human dignity and individual autonomy are also enshrined in the COE's AI Convention.⁷ Overall, the practical significance of the categorical imperative in AI matters is self-knowledge—knowledge of ourselves as rationally efficacious and responsible agents.⁸

Conversely, a distinct narrative seems presently at work in AI matters. According to this, non-human systems are analogous to human minds, and with sufficient training human-like intelligence can flourish,⁹ meaning that our humane capabilities are outmoded and non-essential for superhuman intelligence.¹⁰ According to this narrative, in the inexorable course of progress, AI is a superhuman power beyond our control

² I Kant, *Grundlegung zur Metaphysik des Sitten* (Akademie Ausgabe, vol. IV, 1911), p 429, as translated by Arendt, *The Life of the Mind* (n 1), Willing, p 125.

³ AA Cançado Trindade, 'Some Reflections on the Principle of Humanity in Its Wide Dimension' in R Kolb and G Gaggioli (eds), *Research Handbook on Human Rights and Humanitarian Law* (Edward Elgar, 2013), pp 188–97; AA Cançado Trindade and C Barros Leal, *The Principle of Humanity and the Safeguard of the Human Person* (Capes, 2016), p 9.

⁴ A Fisher et al, *Religious Ethical Systems*, <https://open.library.okstate.edu/introphilosophy/chapter/religious-ethical-systems/>.

⁵ G Le Moli, *Human Dignity in International Law* (Cambridge University Press, 2022), p 56. See also Pope Francis, Address of His Holiness Pope Francis to the G7 (Borgo Egnazia, 14 June 2024): 'We would condemn humanity to a future without hope if we took away people's ability to make decisions about themselves and their lives, by dooming them to depend on the choices of machines. We need to ensure and safeguard a space for proper human control over the choices made by artificial intelligence programs: human dignity itself depends on it'.

⁶ Dicastery for the Doctrine of the Faith, Declaration *Dignitas Infinita* on Human Dignity (2 April 2024), para 27.

⁷ COE AI Convention, Art 7.

⁸ C Bagnoli, 'The Practical Significance of Kant's Categorical Imperative' in M Timmons (ed), *Oxford Studies in Normative Ethics Volume 11* (Oxford Academic, 23 December 2021), without considering AI matters.

⁹ Vallor's testimony (n 109), p 4.

¹⁰ Ibid, p 7. On the definition of AI as a purely economical system: 'a machine that can think and reason just like a human', as a system that can 'surpass human capabilities in a majority of economically valuable tasks'. See M Sullivan, 'Why Everyone Seems to Disagree on How to Define AGI', *AI Decoded* (18 October 2023), www.linkedin.com/pulse/why-everyone-seems-disagree-how-define-artificial-general-ai/.

and, thus, ungovernable, which greatly exaggerates its risks and distracts from other existential threats, notably climate change and the energy transition.¹¹ The implication is regulatory laissez-faire insofar as new AI inventions, ideally subsidised by public investment, can ensure safety from the AI bogeyman threatening human eradication.¹²

Differently, from an Arendtian perspective, people's dignity demands that they are seen, every single one in their particularity, which contradicts the very idea of progress as the law of the human species.¹³ Along these lines, the principle of humanity strengthens self-governance—namely the capability to think, will, and judge for oneself how best to live—which underpins the civil and political liberties guaranteed by constitutions, international law, and democratic life more generally.¹⁴ As Shannon Vallor expounds, such faculties of self-determination allow for decisions about how best to live to be 'made in political cooperation with those whose fates are intermingled with ours'¹⁵—not by AI. Technology can respond to the *how* but not to the *why* and the *for what*: why we make one choice and not another.

In the 'power couple' of AI and energy, one of the practical consequences of the reversed categorical principle is a set of constitutive counter-powers to citizens. Notably, when AI is deployed in energy systems, individuals should be the masters of the data that are extracted from them. This first categorical principle can, for instance, entail introducing contract clauses on data sovereignty and security design and ensuring data encryption to preserve citizens' privacy.¹⁶ This counter-power also includes applications to community energy, a global phenomenon by which energy initiatives are owned, developed, decided upon, and often managed by local communities rather

¹¹ Vallor's testimony (n 109), p 8.

¹² Ibid. 'Thus, at a time when artificial intelligence programs are examining human beings and their actions, it is precisely the *ethos* concerning the understanding of the value and dignity of the human person that is most at risk in the implementation and development of these systems. Indeed, we must remember that no innovation is neutral. Technology is born for a purpose and, in its impact on human society, always represents a form of order in social relations and an arrangement of power, thus enabling certain people to perform specific actions while preventing others from performing different ones. In a more or less explicit way, this constitutive power dimension of technology always includes the worldview of those who invented and developed it.'

¹³ Arendt, *The Life of the Mind* (n 1), Willing, p 272. See also on the idea of progress J Ortega y Gasset, *Meditación de la Técnica y otros ensayos sobre ciencia y filosofía* (Revista de Occidente en Alianza Editorial, 1998).

¹⁴ See also Vallor's testimony (n 109), p 1.

¹⁵ Ibid.

¹⁶ Bria, 'A Digital Green Deal for Europe's Technological Sovereignty' (Inaugural lecture 2021/2022 Universidad Oberta de Catalunya), www.uoc.edu/portal/_resources/CA/documents/la_universitat/llicons-inaugurals/Llico-Inaugural-2021-2022-EN.pdf; F Bria, 'Building Digital Cities from the Ground Up Based Around Data Sovereignty and Participatory Democracy: The Case of Barcelona' (2019) 73 *CIDOB* 83, 87. The General Data Protection Regulation (GDPR) mandates the use of encryption to protect personal data and end-to-end encryption in communication services, A Atadoga, 'A Comparative Review of Data Encryption Methods in the USA and Europe' (2024) 5 *Computer Science and IT Research Journal* 447.

than corporate actors.¹⁷ Presently, however, even some of the most ambitious regulatory frameworks on energy communities at the EU level fail to ensure an effective governance model warranting local ownership and deliberation.¹⁸

The reform proposed would be a rights-based model of informational self-determination, grounded in fundamental rights, rather than in property rights, and enforced through regulatory obligations. This approach is clearly reflected in the landmark *Google Spain v AEPD*, where the Court of Justice affirmed that control over personal data derives from fundamental rights rather than proprietary claims.¹⁹ Differently from the EU AI Act, the GDPR already establishes a comprehensive set of data subject rights, including access (Art 15), erasure (Art 17), and portability (Art 20), which together operationalise a form of individual control over personal data.²⁰ The Data Act extends this logic beyond personal data by introducing rights of access and sharing with respect to data generated by connected devices, including in the energy sector, thereby addressing asymmetries between users and data holders in increasingly data-driven infrastructures.²¹ Broadly stated, in EU law, digital rights amount to the opportunity to be seen and heard in the public realm, in an Arendtian sense,²² or to withdraw from it, expounding one's positive and negative freedom as digital sovereignty.

Although the concept of European digital sovereignty was foreshadowed in the Digital Services Act and furthered in the Digital Markets Act and the Data Governance Act,²³ its embodiment appears legally unfinished and unsatisfactory in the EU AI

¹⁷ See, eg, R Leonhardt et al, 'Advancing Local Energy Transitions: A Global Review of Government Instruments Supporting Community Energy' (2022) 83 *Energy Research and Social Science* 1.

¹⁸ REScoop, Recommendations on Directive (EU) 2023/2413, pp 2–3.

¹⁹ Case C-131/12 *Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* EU:C:2014:317, Judgment of 13 May 2014, para 99, as now reflected in Art 17 of the GDPR, which reinforces the principles established by the CJEU and establishes the 'right to be forgotten'. See also D Messina, 'The Right to Be Forgotten' in C Schepisi, V Capuano, and S Lattanzi, *An Introduction to Digital Data Law in the EU Regulatory Framework and Future Perspectives* (Giappichelli, 2025), p 20ff.

²⁰ On the limited legal requirements in the EU AI Act, compared to the GDPR, see Mantelero (n 4), p 315ff.

²¹ Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA Relevance) PE/49/2023/REV/1 OJ L, 2023/2854, 22.12.2023, Arts 5–11.

²² Arendt, *The Human Condition* (n 14), p 50.

²³ On European digital sovereignty and the need to improve the Digital Services Act, see A Turillazzi et al, 'The Digital Services Act: An Analysis of Its Ethical, Legal, and Social Implications' (2023) 15 *Law, Innovation and Technology* 83. On the Digital Markets Act and the difficulty to define the meaning of digital sovereignty, see C Hoeffler and F Mérand, 'Digital Sovereignty, Economic Ideas, and the Struggle over the Digital Markets Act: A Political-Cultural Approach' (2024) 31 (8) *Journal of European Public Policy* 2121–46. On the EU's limits in operationalising data sovereignty through the Data Act and the Data Governance Act, see H Carrapico and B Farrand, 'EU Data Sovereignty: An Autonomy–Interdependence Governance Gap?' (2025) 13 *Politics and Governance* 1.

Act, which reflects a shift from a value-based debate to a risk-centred and mainly safety-oriented regulation,²⁴ resulting in a framework that lacks the normative depth required to guide long-term technological development. In this sense, the proposed principle seeks to complement the existing legal framework by addressing a residual gap: resolving the collective and systemic dimensions of data governance, particularly in contexts such as AI-driven energy systems where value is generated through aggregation, inference, and cross-sectoral data flows.

The need for stricter governance rules, notably on community energy, would be an ‘irritant’ to what presently appears to be the corporate capture of energy communities whenever private investors (eg investment funds) secure majority equity and capture the public incentives supporting such initiatives without any guarantee that crucial AI-harvested data, decision-making power, and sufficient ownership remain under the control of the actual prosumers, namely those who produce, store, or consume energy.²⁵

Approaching the role of AI in energy systems through a reversed categorical imperative would be relevant in at least three ways. First, the ‘circle of dignity’ underlying fundamental rights aligns with the foundation of contemporary international law, and, as argued by Ginevra Le Moli, Kantian dignity relates to respect and is relational rather than self-referential.²⁶ On this point, the reversed categorical imperative also aligns with energy justice metrics, in particular with respect to participation and to better sharing of economic and social benefits with communities that previous energy systems have neglected, thanks to technology, instead of despite it.²⁷ Second, stricter governance rules for AI in energy systems are also required to avoid the captive phenomenon of socialising bailouts and privatising gains,²⁸ which has long been a feature of late capitalism and appears to be ramping up in AI financing.²⁹ By allowing corporations to profit—accumulating public money without strict guarantees—from AI-extracted data, any regulation would treat AI and its experimentations as an end rather than a means. Third, from an Arendtian perspective, ‘human dignity needs a new guarantee which can be found only in a new political principle, in a new law on earth, whose validity this time must comprehend the whole of humanity while its power must remain strictly limited’.³⁰ To take this insight seriously means recognising that dignity cannot rest on abstract declarations alone. Rather, it requires embodiment in a framework of rules that are not only strict (see Section V.C) but also enforceable (see Section V.D).

²⁴ Mantelero (n 4), p 315.

²⁵ See REScoop and Ecopower, *Financing Guide for Energy Communities*, www.sccale203050.eu/wp-content/uploads/2023/02/SCCALE203050_financingguide_energycommunities.pdf, p 27, on equity investors.

²⁶ Le Moli (n 137), p 21.

²⁷ Baker et al (n 37).

²⁸ M Mazzucato, ‘We Socialize Bailouts. We Should Socialize Successes, Too’, *New York Times* (1 July 2020).

²⁹ Vallor’s testimony (n 109).

³⁰ Arendt, *The Origins of Totalitarianism* (n 42), p ix.

C. Second principle: *diligentia diligentis*, or the diligence of a reasonably prudent person

The second principle of a justice-centred AI consists of a clear set of fiduciary duties to be imposed on AI providers and deployers. Fiduciary duties generally provide a legal infrastructure for relationships spanning the care, custody, and administration of persons, property, and organisations.³¹ Fiduciary duties arise where one party is entrusted with discretionary authority over another's interests and is legally bound to exercise that power solely for the beneficiary's benefit.³² These duties appear across legal domains, from classic private law relationships—such as trustee–beneficiary or agent–principal—to corporate and public governance contexts.³³

The first conceptualisation dates to Roman law: this duty, *diligentia diligentis patris familiae* (the diligence of the diligent good father), in the *Corpus Juris Civilis* later evolved into canon law and Medieval English case law.³⁴ Although fiduciary duties are widely debated,³⁵ at least three can be identified: the duties of loyalty, prudence, and impartiality.

The duty of loyalty requires acting with undivided fidelity, untainted by self-interest. For AI providers and deployers, duties of loyalty would appear, such as the obligation to devise safety systems of precautionary action and disclosing in plain language the potential and risks of using AI-extracted data in energy systems. In this respect, regulators are starting to govern information markets to ensure the sufficient quality and volume of information.³⁶ Such regulation can counterbalance structural asymmetries owing to an unregulated information market, avoiding distortions.³⁷ As a result, the transparency layered through disclosure duties may initially substitute for public participation,³⁸ especially when the general public has few means to enforce entities' disclosure duties.

³¹ EJ Criddle, PB Miller, and RH Sitkoff, 'Introduction' in EJ Criddle, PB Miller, and RH Sitkoff (eds), *The Oxford Handbook of Fiduciary Law* (Oxford University Press, 2019), pp xix and xxiii.

³² R Valsan, 'Fiduciary Duties' in A Marciano and GB Ramello (eds), *Encyclopaedia of Law and Economics* (Springer, 2025).

³³ Ibid.

³⁴ J Getzler, 'Fiduciary Principles in English Common Law' in EJ Criddle, PB Miller, and RH Sitkoff (eds), *The Oxford Handbook of Fiduciary Law* (Oxford University Press, 2019), p 471.

³⁵ Eg FJ Cavaliere et al, 'The Battle over Fiduciary Responsibility of Pension Fund Managers: The Trump Administration Takes On the ESG Movement' (2021) 30 *Southern Law Journal* 1, 22.

³⁶ On the supply and demand of information in environmental matters, see JS Applegate, 'Bridging the Data Gap: Balancing the Supply and Demand for Chemical Information' (2008) 86 *Texas Law Review* 1365.

³⁷ See in general JE Salzman and DA Kysar, 'Harnessing the Power of Information for the Next Generation of Environmental Law' (2008) Duke Science, Technology and Innovation Paper No. 26, pp 5 and 14.

³⁸ L Catá Backer, 'Transparency between Norm, Technique and Property in International Law and Governance: The Example of Corporate Disclosure Regimes and Environmental Impacts' (2013) 22 *Minnesota Journal of International Law* 1, 6.

The duty of prudence requires fiduciaries to act with caution, care, and diligence in managing the systems for their beneficiaries.³⁹ In particular, for AI providers and deployers, duties of care under the duty of prudence would include, for example, the obligation to examine the factual basis and ecological impact of AI applications in energy systems, including the carbon budget associated with the energy intensity managed and optimised through such systems. This rule would also preclude providers and deployers from merely replicating market investments and applications that are ultimately financed by energy consumers, thereby privileging short-term profits while mispricing sustainability risks.⁴⁰

The duty of impartiality refers to good-faith efforts to identify, respect, and balance diverse interests among beneficiaries when carrying out fiduciary responsibilities.⁴¹ In particular, for AI providers and deployers, duties of impartiality would flesh out, for instance, in the obligation to consider not only the stakes of present generations but also the intergenerational duty that, in principle, should underpin impartiality.⁴² In this sense, impartiality entails ‘taking the viewpoints of others into account’.⁴³ Equity would provide ‘standards for allocating and sharing resources and for distributing the burdens of caring for the resources and the environment in which they are found’,⁴⁴ including energy resources, with the principle of equity *between* generations underlying the very notion of sustainable development.⁴⁵ A notable example would be the requirement that AI providers and deployers work on a more ecological principle of reducing LLMs’ carbon footprint while maintaining the same level of effectiveness.

Approaching the role of AI in energy systems through fiduciary principles would be relevant in at least three ways. First, it incorporates several legal traditions and cultures. In fact, comparable fiduciary duties can be found in both common and civil law traditions, usually complemented by duties arising from other legal sources, such

³⁹ C Woods and R Urwin, ‘Putting Sustainable Investing into Practice: A Governance Framework for Pension Funds’ (2010) 92 *Journal of Business Ethics* 1, 13.

⁴⁰ In parallel to fiduciary duties held by pension funds, see J Singh Bachher, AD Dixon, and AHB Monk, *The New Frontier Investors: How Pension Funds, Sovereign Funds, and Endowments Are Changing the Business of Investment Management and Long-Term Investing* (Palgrave Macmillan, 2016), pp 39–40; KL Johnson and FJ de Graaf, ‘Modernizing Pension Fund Legal Standards for the 21st Century’ (2009) 2 (1) *Rotman International Journal of Pension Management*, <https://ssrn.com/abstract=1408691>.

⁴¹ Restatement of Trusts (Third), comments to §79(1); Johnson and de Graaf (n 172), 44, 46.

⁴² Woods and Urwin (n 171), 13.

⁴³ H Arendt, *Lectures on Kant’s Political Philosophy* (ed R Beiner) (University of Chicago Press, 1992), p 42.

⁴⁴ E Brown Weiss, ‘Intergenerational Equity’ in R Wolfrum (ed), *Max Planck Encyclopedia of Public International Law (MPEPIL)* (2013).

⁴⁵ IPCC WG III, *Climate Change 2014*, ch 4, s 4.2.2, 14, as referenced by C Redgwell and L Rajamani, ‘And Justice for All? Energy Justice in International Law’ in I del Guayo et al., *Energy Justice and Energy Law* (Oxford University Press, 2020), p 60.

as statutes or contracts.⁴⁶ At the same time, fiduciary principles would offer a more concrete policy response to the present call to make AI providers and deployers accountable. A notable framework is the OECD AI Principles, by which all AI actors should apply a systematic risk management approach to each phase of the AI system life cycle, including across their value chain, based on their roles, context, and ability to act.⁴⁷

Second, fiduciary duties, as described earlier, require a type of due diligence in connection to sustainability that has long existed in environmental and energy matters but has been consistently overlooked in AI regulatory attempts. Sustainability due diligence should cover limits on the use of planetary resources, social issues—notably human rights protection—and economic and governance issues, with recurring assessments at an interval of at least three years—potentially more in case of business changes—across the global value chain.⁴⁸

Third, fiduciary duties would offer concrete procedural steps to fulfil the nexus of human rights with environmental principles in energy matters,⁴⁹ notably prevention, sustainability as an umbrella principle, and the intra- and intergenerational equity principles, which feature among the core principles of energy justice in international law. From an Arendtian perspective, the role of procedures as irritants that countervail the power of corporate entities and bureaucratic societies can humanise the ability to handle things that are not yet properly encompassed by the faculty of the will.⁵⁰ Within the myth of progress, when driven solely by profit, businesses have revealed ‘the utter amorality of the profit motive and its indifference to consequences’, as Beth Stephens recalled more than 20 years ago in reference to companies that sold revolutionary data management systems to the Nazis.⁵¹

Overall, the power couple of AI and energy can transform corporate accountability from current conversations on transparency to more concrete, outright fiduciary duties arising from human rights and environmental principles. Nonetheless, fiduciary duties should be painstakingly designed as enforceable. An Arendtian perspective would deny the factual validity of human rights corresponding to fiduciary duties not per se but to the extent that human rights do not entail any enforceable praxis on their own.⁵² In this

⁴⁶ See JT Walsh, ‘The Fiduciary Foundation of Corporate Law’ (2002) 27 *Journal of Corporation Law* 333.

⁴⁷ OECD AI Recommendation, IV, 1.5.

⁴⁸ B Sjøfjell, ‘The Financial Risks of Unsustainability: A Research Agenda’, University of Oslo Faculty of Law Legal Studies Research Paper Series No. 2020–18 (2020), p 197.

⁴⁹ On this nexus, see Redgwell and Rajamani (n 177).

⁵⁰ Arendt, *The Life of the Mind* (n 1), Thinking, p 76. See also W Scobie, ‘Questions of Corporate Responsibility: Can Hannah Arendt’s “Thoughtlessness” Apply to Companies and Their Actors?’ (2017) 42 *Alternative Law Journal* 55.

⁵¹ R Bernstein, ‘IBM’s Sales to the Nazis: Assessing the Culpability’, *New York Times* (7 March 2011), E8, as cited by B Stephens, ‘The Amorality of Profit’ (2002) 20 *Berkeley Journal of International Law* 45.

⁵² Y Sari, ‘Arendt, Truth, and Epistemic Responsibility’ (2018) 2 *Arendt Studies* 149.

sense, human rights are, in principle, inalienable if a prerequisite is secured—namely the right to possess rights and to enforce them as a citizen of a political community—as we will see in the following description of the third principle for justice-centred AI (Section V.D).

D. Third principle: the right to have rights

The third principle of justice-centred AI is comprehensive. It is the right to have rights, an Arendtian syntagm meant to secure human dignity that consists of the right to enforce one's rights while engaging in meaningful political action.⁵³ Arendt's argumentation for this right springs from the historical experiences of the 20th century, when human rights, 'supposedly inalienable, proved to be unenforceable even in countries whose constitutions were based upon them—whenever people appeared who were no longer citizens of any sovereign state'.⁵⁴ Although Arendt directed her main critique to the calamity of the rightless and stateless, her proposal of the right to have rights remains striking in its normative content. According to Arendt, political action in a situated community is the means to exercise freedom, while the right to have rights is denied whenever individuals no longer belong to a community.⁵⁵

Along these lines, the right to have rights springs from participation as members of a group whereby we 'guarantee ourselves mutually equal rights'.⁵⁶ Consequently, community is never a given. It is a construct that citizens should actively design and maintain through opinions, contracts, and promises.⁵⁷ In Arendt, the civic bond is provided by neither common interests nor the common good. Instead, it is the common world where we can disagree and dissent—that is, debate about what is between us, the *inter-esse*.⁵⁸

In this world outside and between us, public interest is the measure for public life, securing the permanence that lies in the public sphere and materialises in institutions.⁵⁹ If any social contract can be envisaged in Arendt, it would be a horizontal contract lifting each individual out of isolation while also limiting their individual power in view of relational power, or action-in-concert, in a shared world.⁶⁰ This essentially horizontal and political understanding of the separation of powers led Arendt to understand the

⁵³ On the right to have rights as the new law on Earth, a new principle that will secure human dignity, see JD Ingram, 'What Is a "Right to Have Rights"? Three Images of the Politics of Human Rights' (2008) 102 *American Political Science Review* 411.

⁵⁴ Arendt, *The Origins of Totalitarianism* (n 42), p 293.

⁵⁵ Ibid, p 295.

⁵⁶ Ibid, p 301.

⁵⁷ M Borren, *Amor mundi: Hannah Arendt's Political Phenomenology of World*, PhD thesis, Universiteit van Amsterdam, (2010), p 117.

⁵⁸ Ibid.

⁵⁹ Ibid, pp 113–14.

⁶⁰ Ibid, p 117.

separation of power as implying more power—in the sense of more *loci* of power—to buttress a robust political community.⁶¹ The nation-state’s notion of sovereignty, a dangerous megalomania originating from absolutism,⁶² cannot ensure the separation of powers as a counterbalance of powers through the multiplication of its *loci*. In Arendt, ‘there can only be democracy ... where the centralisation of power in the nation-state has been broken, and replaced with a diffusion of power into the many power centres of a federal system’ of sorts.⁶³

In Arendt, human freedom is a matter not of metaphysics but of fact.⁶⁴ It is experienced primarily not in will and thought but in action, requiring a political sphere for such action-in-concert.⁶⁵ To operationalise freedom as an essentially political phenomenon, Arendt turns to two forms of direct political participation: Thomas Jefferson’s ‘ward system’, as inspired by the town hall meetings of the American Revolution, and the citizens’ councils mushrooming in 19th and 20th-century Europe.⁶⁶ Rather than the demise of modern representative democracies, Arendt emphasises the need to embody effective democratic practices through localised forms of political engagement. The latter, as ‘little republics’, would institutionally cement the *pouvoir constituant* of the people and prevent centralised powers from acting on any tyrannical tendencies.⁶⁷ The right to have rights brings the law to the centre by requiring spaces that preserve the pre-political condition of the existence of a common world of *inter-esse*.⁶⁸ The need for ‘little republics’ also shows that our society often aims to leverage AI to increase our control over an uncertain future, while AI can—paradoxically—reduce our agency over the future, thus exposing the myth of progress on which modernity has so largely relied.⁶⁹

Transposing Arendt’s reflections on the ‘power couple’ of AI and energy, energy consumers face the same paradox of stateless individuals whenever the nation-state does not guarantee their right to have rights. In the alienating world of energy consumers, the right to have rights would require not only an opportunity to impact the lawmaking process but also a ward system whereby, through speech, individuals can make choices

⁶¹ Ibid, p 198. See similarly to Montesquieu and differently from Rousseau and the tenets of the French Revolution, Arendt (n 33), p 149ff.

⁶² H Arendt, *Thinking without a Banister: Essays in Understanding 1953–1975* (Schocken Books, 2018), p 261. On sovereignty as manifestative of another type of dignity, *dignitas*, see Le Moli (n 137), ch 3.

⁶³ Arendt (n 194), p 261.

⁶⁴ Ibid, p 241.

⁶⁵ Ibid, p 229.

⁶⁶ Ibid, p 209.

⁶⁷ Arendt (n 33), pp 248–49. See also J Muldoon, ‘The Lost Treasure of Arendt’s Council System’ (2011) 12 *Critical Horizons* 396, 401.

⁶⁸ Arendt (n 194), p 717, saying it with Heraclitus, ‘a people must fight for their laws as they fight for the wall (*teichos*) of their city’—walls offer the pre-political condition of the existence of the city.

⁶⁹ H Nowotny, *In AI We Trust: Power, Illusion and Control of Predictive Algorithms* (Polity Press, 2021).

about their freedom. In this ward system, humans would also appropriate and, as it were, dis-alienate ‘the world into which, after all, each of us is born as a newcomer and a stranger’.⁷⁰ Decentralised forms of decision-making and self-government are required in energy matters, all the more so when AI systems are engaged, thus demanding the painstaking work, yet to happen, of institutional translation from democracy in paper to democracy in practice. Such work would reboot political freedom not only within energy communities but also through less formalised ways of debating AI and energy matters.

In the power couple of AI and energy, one of the practical consequences of the right to have rights is the possibility for AI to accelerate discussions and the implementation of ecological citizenship beyond the value of transparency alone. By supplanting the universalism customarily associated with citizenship, ecological citizenship can be understood as providing spaces of ‘insurgent’ citizenship.⁷¹ In this way, laws would offer practical pathways to thematically cope with the unfulfilled promises of representative democracy through the creation of new, specialised spaces of deliberation, such as energy communities where the institutional translation of the right of rights ought to be taken more seriously.

Truly, meaningful political participation cannot be mandated by law. Overly prescriptive forms of participation risk formalism, reinforcing passivity and encouraging citizens to externalise responsibility to the state. The function of law, therefore, is not to compel participation but to enable and structure it, by lowering barriers to engagement, creating accessible fora, and ensuring that participation has tangible effects on decision-making. In this context, approaching the role of AI in energy systems through the right to have rights would be relevant in at least three ways. First, the right to have rights would be an ‘irritant’ to centralised powers at both public and private levels, which constitute the large majority of energy systems and which, in Arendt, always risk stagnating in the absence of any counterpower.⁷² Second, it aligns with a possible constitutive stage of human dignity, as foreshadowed by Ginevra Le Moli. In fact, it limns a more ecocentric understanding of the knots between humans and nature against the over-exploitation of nature.⁷³ In this further circle of dignity, dignity rises to a more political form where ‘[p]lurality is the law of the earth’⁷⁴ because ‘[m]an is not merely conditioned by his environment; he conditions the environment, and the environment then conditions him in turn’.⁷⁵

⁷⁰ Arendt, *The Life of the Mind* (n 1), Thinking, p 98.

⁷¹ On ecological citizenship from an Arendtian perspective, with no AI consideration, see PA Latta, ‘Reading Environmental Justice as Citizenship: An Arendtian Perspective’ (CPSA, 2006), www.cpsa-acsp.ca/papers-2006/Latta.pdf.

⁷² Arendt (n 33): ‘Power can be divided without decreasing it, and the interplay of powers with their checks and balances is even liable to generate more power, so long, at least, as the interplay is alive and has not resulted in a stalemate’.

⁷³ Le Moli (n 137), p 327.

⁷⁴ Arendt, *The Life of the Mind* (n 1), Thinking, p 19.

⁷⁵ Arendt, *The Human Condition* (n 14), p 326.

Third, among the suggested three principles, the right to have rights opens up the conditions for the more legitimate design and use of AI in energy systems because it is the most political among the three principles. In fact, it rests on the assumption that the direction of technology should be the object of democratic deliberation. In the increasing opposition between culture and technology,⁷⁶ a common world of deliberation, promises, and contracts can shape a type of culture that better conciliates technology with nature.

In the words of Romano Guardini, under whom Arendt studied in her university education, this premise of deliberative spaces is not anti-technological, but it would ensure that ‘the process of technology worldwide will really achieve the great things that it can and should’.⁷⁷ That technology brings a new measure of freedom is, in principle, a gain; the ‘value of freedom, however, is not fixed solely by the question “Freedom from what?” but decisively by the further question “Freedom for what?”’. In fact, in absolute modernity, conflicts unfold between ‘the embodiment of ideas in social practice (*ethos*) and the rationale behind practice (*logos*)’.⁷⁸

⁷⁶ Eg R Guardini, *Letters from Lake Como: Explorations in Technology and the Human Race* (Eerdmans, 1994), www.thettedkarchive.com/library/romano-guardini-letters-from-lake-como.

⁷⁷ Ibid, p 19.

⁷⁸ E Beltramini, ‘The Government of Evil Machines: An Application of Romano Guardini’s Thought on Technology’ (2021) 9 *Scientia et Fides* 257, 259.

VI. Institutionalising justice-centred AI principles

A. Introduction

Previously, I explored the constitutive gap between the democratic ideal of collective deliberation and the reality of individuals behaving as energy consumers. To help bridge this gap, I further argued that international law could be a gap-filler by offering a baseline of three principles for justice-centred AI in energy matters: the categorical principle reversed, a set of fiduciary principles through the *diligentia diligentis* (the diligence of a reasonably prudent person) standard, and the provision of the right to have rights. Finding these principles in need of institutionalisation, I here propose three methods to embed them through both centralised and decentralised means. What I mean by this is that institutionalisation initiatives are needed across levels of government, spanning local and global governance fora, the latter embodied in the OECD AI Policy Observatory and the UN, given the newly inaugurated season of AI resolutions and the 2024 Pact for the Future.

B. Institutionalising the reversed categorical imperative

To review, the first principle on the categorical imperative reversed holds that providers and deployers ought to treat humanity as an *end* and AI as a *means*. To uphold this humanity principle and its underlying tenet of human dignity, one can envisage at least three institutional arrangements. First, knowledge hubs would foster AI literacy, often recalled in policy documents, albeit at a high level of generality and without factualisation.¹ Because AI affects society at large and is rapidly evolving in unpredictable ways, AI literacy hubs would nurture reflection and nuanced scepticism in such a way that the emergence of machine intelligence is steered through critical intelligence towards planetary sapience in the service of a viable long-term future. At the same time, AI literacy hubs can counteract the frightening feeling of ‘solipsistic

¹ Eg EU AI Act, Art 4. OECD AI Recommendation, para 2.2.

freedom’—‘the “feeling” that my standing apart, isolated to everyone else, is due to free will, that nothing and nobody can be held responsible for it but myself’.²

Artificial intelligence knowledge hubs have already been proposed, albeit mainly in the Nordic region in Europe,³ and would align with the United Nations Educational, Scientific and Cultural Organization’s (UNESCO) guidelines and frameworks for ethical AI.⁴ Artificial intelligence literacy hubs would also help reconceptualise the public sphere for the ‘electro-iconographic societies’ of late capitalism⁵ and help the youth confront the iconographic medium of the electronic means of representation,⁶ including through online activities.

Second, AI companies would be required to embed safety institutions in their organisational structure. Safety institutions are presently rare and under-incentivised, or used to ‘seek government and philanthropic funding to invest in technical study of the unknown, longer-term risks of future “frontier” models that could be more dangerous than those we have today’.⁷

Three, AI bias should be the object of stricter discussions and regulations concerning training materials and output control. Artificial intelligence bias counters the very principle of dignity by lumping individuals under discriminatory labels, practices, assumptions, and assertions, thus worsening the algorithmic echo chambers, which thoroughly diminishes our political sense of a shared reality and enables new forms of political manipulation.⁸ The COE’s AI Convention also encourages equality and non-discrimination protections.⁹

Overall, from such institutionalisations of the principle, a more solid understanding would follow about how AI systems function and how they can be used, which would in turn increase the acceptance of AI.¹⁰ Understanding is needed not only of the known risks and uncertainties posed by AI but also of the unknowns attached to

² Arendt, *The Life of the Mind* (n 1), Willing, p 196. ‘Democracy’s Discontent: Michael Sandel with David Brooks’ (n 108), minutes 38–40.

³ Eg www.nordforsk.org/node/1374.

⁴ United Nations Educational, Scientific and Cultural Organization (UNESCO), Recommendation on the Ethics of Artificial Intelligence (SHS/BIO/REC-AIETHICS/2021), <https://unesdoc.unesco.org/ark:/48223/pf0000380455>, para 44. UNESCO, *Records of the General Conference, Forty-First Session, Paris, 9–24 November 2021*, vol. 1, *Resolutions*, annex VII.

⁵ Benhabib (n 47), p 166.

⁶ Ibid, p 177.

⁷ Vallor’s testimony (n 109).

⁸ Ibid, p 5. The EU AI Act, preambular para 27, recalls the AI High-Level Expert Group on Artificial Intelligence’s seven non-binding ethical principles, while failing to justify why such principles are non-binding or should not be binding throughout the AI Act.

⁹ COE AI Convention, Art 10.

¹⁰ G Schiavo, S Businaro, and M Zancanaro, ‘Comprehension, Apprehension, and Acceptance: Understanding the Influence of Literacy and Anxiety on Acceptance of Artificial Intelligence’ (2024) 77 *Technology in Society* 1.

the technology, even for AI developers and researchers.¹¹ Further, these reform suggestions would widen the chances for humans to engage in reflection in a non-cognitive, non-specialised sense as a natural need of human life, thus becoming a prerogative of not only a few so-called scholars and experts but all humans, regardless of cultural, educational, and social conditions.¹²

C. Institutionalising *diligentia diligetis*

To review, the second principle rests on the diligence required of AI providers and deployers through fiduciary duties. It has the potential to tackle the risks associated with the use of AI in energy systems, notably ‘lack of transparency, decline of human autonomy, cybersecurity, market dominance, and price manipulation on the electricity market’.¹³

To uphold fiduciary duties and the underlying tenet of *diligentia diligetis*, one can envisage at least three institutional arrangements. First, fiduciary duties should be discussed and made mandatory in order to limit managerial autonomy for the self-preservation of energy system functions.¹⁴ This approach will replace non-mandatory and individually defined standards of care with measurable metrics for which AI providers and deployers are accountable,¹⁵ with an enhanced role for AI safety institutes and third-party controls. Accordingly, fiduciary duties can become enforceable *functional* duties of disclosure, auditing, and control processes.¹⁶

Second, beyond this governance level, a set of due diligence processes should be detailed to further tackle the growing horizontal relationships connecting individuals with private digital companies capable of competing with public authorities.¹⁷ Arti-

¹¹ JM White and R Lidskog, ‘Ignorance and the Regulation of Artificial Intelligence’ (2021) 25 *Journal of Risk Research* 488.

¹² Arendt, *The Life of the Mind* (n 1), Thinking, p 191. This first constitutive principle may be deemed to pertain to thinking, Arendt, *The Life of the Mind* (n 1), Thinking, p 100: ‘The sheer naming of things, the creation of words, is the human way of *appropriating* and, as it were, disalienating the world into which, after all, each of us is born as a newcomer and a stranger’ (italics in the original).

¹³ I Niet, R van Est, and F Veraart, ‘Governing AI in Electricity Systems: Reflections on the EU Artificial Intelligence Bill’ (2021) 4 *Frontiers in Artificial Intelligence* 1. These authors find that the ‘AI Act addresses well the issue of transparency and clarifying responsibilities, but pays too little attention to risks related to human autonomy, cybersecurity, market dominance and price manipulation’.

¹⁴ G Teubner, ‘Corporate Fiduciary Duties and Their Beneficiaries: A Functional Approach to the Legal Institutionalization of Corporate Responsibility’ in KJ Hopt and G Teubner (eds), *Corporate Governance and Directors’ Liabilities* (De Gruyter, 1985), p 166. This second constitutive principle may be deemed to pertain to willing, Arendt, *The Life of the Mind* (n 1), Willing, p 178: following Nietzsche’s interpretation, ‘technology’s very nature is the will to will, namely, to subject the whole world to its domination and rulership’.

¹⁵ Ibid, pp 167 and 172.

¹⁶ Ibid, p 167.

¹⁷ Pollicino (n 38), pp xiii–xiv. G De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (Oxford University Press, 2022), p 9ff.

ficial intelligence due diligence would require AI developers to conduct mandatory pre-deployment sustainability assessments, including the impact of AI on both ecological and social values, notably fundamental rights.¹⁸ For high-risk AI systems, such as energy matters, mandatory algorithmic impact assessments would be coupled with independent third-party audits throughout their life cycle. Algorithmic impacts would also be needed in order to ensure responsible deployment decisions and a safety profile across the AI value chain over time.¹⁹ Given the scope for discretion in risk assessments and relatively limited experience in AI matters, when compared to privacy,²⁰ the EU AI Act requires guidance and disambiguation in this respect.²¹

Third, and connected to the second point, the other side of the principle *diligentia diligentis* is enforceability, namely the institution of ‘stronger mechanisms for contestability, liability, and redress for avoidable and significant AI harms, to reinternalise the costs of preventable harm and developer negligence currently being imposed upon vulnerable publics’.²² Current victims of AI have few paths of remedy and redress at their disposal, while such mechanisms have been key in building and sustaining the innovation and safety cultures of other industries, notably civil aviation, civil engineering, and pharmaceuticals.²³

Overall, the proposed systems of legal accountability, contestability, and redress would better incentivise AI developers and deployers to meet a high standard of care²⁴ in the face of current political failures to offer proper compliance incentives or effective, flexible, and iterative regulation.²⁵ Moreover, this approach would counter the mushrooming yet slight effectiveness of voluntary codes of conduct, which are upheld even by what has been deemed the stricter AI regulatory framework, namely the EU AI Act.²⁶ Further, stopping short of institutionalising limits to AI design and deployment, concerns would only increase with regard to the fate of the rule of law in an algorithmic society.²⁷

¹⁸ On pre-deployment assessment including the impact of AI on fundamental rights only, see Vallor’s testimony (n 109), pp 9–10.

¹⁹ See similarly *ibid.* See also AD Selbst, ‘An Institutional View of Algorithmic Impact Assessments’ (2021) 35 *Harvard Journal of Law and Technology* 1; A Cosentini et al, ‘Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights’ (6 March 2025), <https://ssrn.com/abstract=5168579>; A Mantelero, ‘The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Operational Methodologies’ (2024) 54 *Computer Law and Security Review* 54. See also United Nations Development Programme (UNDP), *Human Rights Impact of AI Assessment Toolkit* (2025).

²⁰ Mantelero (n 4), p 317.

²¹ C Orwat et al, ‘Normative Challenges of Risk Regulation of Artificial Intelligence’ (2024) 18 *Nanoethics* 1, 11.

²² Vallor’s testimony (n 109), p 10.

²³ *Ibid.*

²⁴ *Ibid.*

²⁵ *Ibid.*, p 4.

²⁶ Art 95 of the EU AI Act.

²⁷ G De Gregorio, ‘The Normative Power of Artificial Intelligence’ (2023) 30 *Indiana Journal of Global Legal Studies* 55.

D. Institutionalising the right to have rights

To review, the third principle posits that individuals ought to be secured the right to enforce their rights while engaging in meaningful political action. Rather than warranting participation, the right to have rights guarantees the conditions for effective democratic practices through localised forms of political engagement. To uphold the right to have rights and the underlying fulfilment of human dignity in its more political form, one can envisage at least three institutional arrangements. First, public and private actors ought to recognise that the ‘decentred’ public sphere, including electronic media, is where most value debates are waged and ‘subaltern’ politics is carried out by previously excluded and marginalised groups.²⁸ As for more traditional forms of the public sphere, the task there would be ‘to challenge the secret logic of power, hierarchy, and domination’ in narratives and means.²⁹ At the same time, offline and online public spheres should be taken seriously and moderated to foster an ethics of the limit while new technologies induce a feeling of unlimitedness.

Second, it should be acknowledged that the demand for the right to have rights in the public sphere is particularly challenging because the public sphere is a nostalgic trope.³⁰ Democracy rests on the idea of an autonomous public sphere for self-governance through collective deliberation. There exists, however, a hiatus between this regulative ideal and ‘the increasingly desubstantialized carriers of the anonymous public conversation of mass societies’.³¹ This hiatus turns the ideal of democracy into a constitutive fiction, causing constant anxiety.³²

Third, because the power of encoding thoughts into language is still the hallmark of our ‘unique cognitive liberty and the heart of our political capacity’,³³ the institutionalisation of the right to have rights ought to emphasise thought and speech as ‘*the greatest interest and most distinctive achievement of man*’.³⁴ As notably encouraged at the OECD, EU, and Nordic levels,³⁵ regulators can support safe ‘sandbox’ testing environments for thought and speech with regard to new AI solutions. Instead of including only expert knowledge (ie, researcher and innovator participation),³⁶ AI regulatory sandboxes should be informed by a ward system of multi-stakeholder debate and cooperative meetings built in federal units under a municipality, for example town council meetings built into federal units. A type of council meeting would also happen online

²⁸ Benhabib (n 47), p 176.

²⁹ Ibid, p 175.

³⁰ Ibid, p 164.

³¹ Ibid, p 165.

³² Ibid.

³³ Vallor’s testimony (n 109), p 11, referring to the cybernetics pioneer Norbert Wiener.

³⁴ N Wiener, *The Human Use of Human Beings: Cybernetics and Society* (Houghton Mifflin, 1950), p 95 (italics in the original).

³⁵ See eg www.oecd.org/publications/regulatory-sandboxes-in-artificial-intelligence-8f80a0e6-en.htm and www.nordforsk.org/node/1374. See EU AI Act, Arts 57–62.

³⁶ www.nordforsk.org/node/1374.

and target the youth in order to inspire their self-empowerment and self-governance, for instance with culture—which alone is able to arrest and move us.³⁷

In Arendt’s constitutional scheme, such councils would play a deliberative and decision-making function ‘in opposition to the party system as an antidote to their electoral manipulation and depoliticising effect’.³⁸ Councils can discuss and attempt to solve problems through ‘incompletely theorised agreements’, which Cass Sunstein foreshadows as key to legal and political reasoning for their social stability function. They facilitate convergence, reduce the costs of disagreement, and help demonstrate humanity and mutual respect in a liberal democracy.³⁹ Agreements are incompletely theorised because participants agree on the results but not on the underpinning rationale.⁴⁰ Incompletely theorised agreements that reduce the cost of perpetual disagreement⁴¹ are the best option for people with limited time and capacities,⁴² making them widespread. These agreements serve as the foundations for both rules and analogies⁴³ and are well suited to a world containing social dissensus on large-scale issues. At the same time, such agreements remain provisional and contestable, requiring iterative testing and refinement, which is an approach that aligns with experimental regulatory tools such as AI sandboxes in the EU AI Act.

Nonetheless, AI citizens’ councils are no guarantee of justice⁴⁴ and would require a range of cases in which the principles are tested against others and refined,⁴⁵ which would align with the requisite for AI regulatory sandboxes. This type of agreement can also align with what has been called ‘algor-ethics’, namely the moderation of algorithms and AI programs pursuant to discrete values. Notwithstanding value pluralism, where ‘a single set of global values’ seems impossible to find, we can still retrieve shared principles governing the role of AI in our common world,⁴⁶ and incompletely theorised agreements can be a tool to such an end. While citizens’ councils may rely on digital infrastructures to enable coordination, their legitimacy cannot rest on purely virtual interaction. In Arendtian terms, meaningful political action presupposes a shared space of appearance rooted in the world. Accordingly, councils should be conceived of as

³⁷ Arendt, *Between Past and Future* (n 12), p 201.

³⁸ Muldoon (n 88), 604 and Muldoon (n 199). See also Arendt (n 33), pp 215–16. See also JF Sitton, ‘Hannah Arendt’s Argument for Council Democracy’ (1987) 20 *Polity* 80.

³⁹ C Sunstein, ‘Practical Reason and Incompletely Theorized Agreements’ in E Ullmann-Margalit, *Reasoning Practically* (Oxford University Press, 2000), pp 98, 105, 113.

⁴⁰ Ibid, p 99.

⁴¹ Ibid, p 98.

⁴² Ibid, p 106.

⁴³ Ibid, pp 99 and 100.

⁴⁴ Ibid, p 112.

⁴⁵ Ibid, p 113.

⁴⁶ Pope Francis (n 137). On algorethics, see P Benanti, *Oracoli: Tra algoretica e algocrazia* (Luca Sossella Editore, 2018). The third proposed constitutive principle may be deemed to pertain to judgement, Arendt, *The Life of the Mind* (n 1): Judging, p 266: The enlarged mentality required of judgement takes others’ possible judgements into account ‘because I am human and cannot live outside the company of men’.

‘onlife’ institutions: digitally extended, yet anchored in situated deliberation, where participants engage with the material consequences of AI systems within their local environments. Citizens’ councils would function as standing deliberative bodies that scrutinise and inform in real time the design and deployment of AI technologies, especially of high impact, in line with the emerging field of participatory AI.⁴⁷

Within this architecture, councils would also operate as distributed sensing nodes, capturing localised environmental externalities and societal impacts of AI systems, thereby contributing to a polycentric model of governance grounded in context-sensitive knowledge.⁴⁸ In fact, AI governance in energy systems can be fruitfully understood through the lens of common-pool resource (CPR) dilemmas, albeit in a non-classical sense. While AI as such does not constitute a CPR, its underlying infrastructures and impacts, particularly compute-intensive processes, energy consumption, and associated environmental externalities, display features of rivalry and collective vulnerability that generate coordination problems analogous to those identified by Elinor Ostrom in CPRs.⁴⁹ In parallel, the epistemic layer of AI, including models and data, exhibits characteristics of a knowledge commons. This hybrid configuration calls for governance approaches that move beyond centralised, one-size-fits-all regulation, and that focus on value communication (even ‘cheap talk’) and trust to avoid resource overuse.⁵⁰

Within this framework, AI councils can be conceptualised as key institutional mechanisms for addressing CPR-like dilemmas in AI-driven energy systems, notably as trust accelerators, fostering repeated interaction, and early-warning systems for over-consumption, identifying emerging risks related to energy use, data extraction, and environmental impact before they crystallise into systemic failures. Crucially, councils would institutionalise deliberation on the long-term and intergenerational impacts of AI, both positive and negative,⁵¹ embedding anticipatory governance capable of addressing the temporal asymmetries inherent in AI development.

A central mechanism would be the introduction of mandatory AI life-cycle disclosures, including energy consumption, carbon intensity, material dependencies, and data practices, ensuring accountability by rendering the material infrastructures of AI visible and contestable. In this way, the ward system provides the institutional infrastructure through which AI may be enacted as a form of distributed, reflexive, and

⁴⁷ P Ahrweiler et al, ‘Inclusive Technology Co-design for Participatory AI’ in P Ahrweiler (ed), *Participatory Artificial Intelligence in Public Social Services: From Bias to Fairness in Assessing Beneficiaries* (Springer, 2025). On onlife, see L Floridi, *The Onlife Manifesto: Being Human in a Hyperconnected Era* (Springer, 2015).

⁴⁸ E Ostrom, ‘Beyond Markets and States: Polycentric Governance of Complex Economic Systems’ (2010) 100 *American Economic Review* 641.

⁴⁹ Ibid.

⁵⁰ Ibid, 641–42.

⁵¹ See similarly O Hauser, ‘Climate Change, Intergenerational Fairness, and the Promises and Pitfalls of Artificial Intelligence’ (2025) 88 *Environmental and Resource Economics* 2689.

ecologically bounded governance, aligning technological development with planetary limits while preserving democratic agency.

Beyond oversight, the ward system also enables the deployment of AI as a sustainability enabler, positioning citizens' councils as co-designers of AI-enabled ecological transformation. In this capacity, councils would deliberate on and prioritise the use of AI in domains such as energy system optimisation, biodiversity monitoring, and climate risk management, thereby shaping socio-technical pathways aligned with sustainability goals. This participatory orientation reflects broader calls for mission-oriented and democratically guided innovation for planetary-scale computation,⁵² moving beyond technocratic approaches.

Crucially, for such participation to be meaningful rather than symbolic, councils would need to be complemented by AI literacy hubs, designed to enable informed judgement. In Arendtian terms, the capacity for political action presupposes not expertise as such but the ability to form and express opinions in a shared world, thus engaging in political activity not as a means to an end but as an end in itself.⁵³ Literacy hubs would therefore function not as technocratic training centres but as enabling conditions for participation, bridging the gap between highly technical AI systems and the public reasoning required for their legitimate governance.

At the same time, the councils would be featured by power moving 'neither from above nor from below' but being 'horizontally directed so that the federated units mutually check and control their powers'.⁵⁴ They could be imagined as a linchpin between public perceptions and debates and the institutions of liberal representative democracies. In Arendt, this federal network of councils would eventually produce parliamentary representatives,⁵⁵ for instance through bodies whose preventive advice should be strictly accounted for to pass AI regulation.

E. Fiduciary duties or rights?

A clarification is necessary regarding the relationship between the institutionalisation of fiduciary duties (second principle) and the 'right to have rights' (third principle), both grounded in the instrumentality of AI (first principle). The argument advanced in this article does not presuppose a direct horizontal application of human rights law to private AI providers or deployers. Such a claim would sit uneasily within the classical vertical structure of both EU fundamental rights law, anchored in Article 51(1) of the Charter of Fundamental Rights of the European Union, and the European Con-

⁵² BH Bratton, *The Stack: On Software and Sovereignty* (MIT Press, 2016).

⁵³ T Tömmel and M Passerin d'Entreves, 'Hannah Arendt' in EN Zalta and U Nodelman (eds), *The Stanford Encyclopedia of Philosophy (Spring 2025 Edition)*, <https://plato.stanford.edu/archives/spr2025/entries/arendt/>.

⁵⁴ Muldoon (n 88), 604.

⁵⁵ Ibid.

vention on Human Rights system, in which protection against private interference is mediated through positive obligations of the state rather than direct claims against private actors.⁵⁶

Instead, the proposed principles would operate through a mediated and institutionalised form of horizontality, characteristic of contemporary EU constitutionalism. The proposed *diligentia diligentis* obligations, as described earlier, are conceived of as legally structured duties of care embedded in contractual, regulatory, and compliance architectures governing AI development and deployment. Similarly, the Arendtian and more deliberative ‘right to have rights’ is understood not as a directly enforceable subjective entitlement against private actors but as a principle of institutionalised contestability that requires the legal order, primarily through the EU legislator, to construct and maintain spaces of meaningful deliberation, objection, and revision. This is particularly relevant in AI governance, where AI systems in energy governance are usually developed and deployed without meaningful *ex ante* public deliberation. This generates a structural asymmetry: individuals and affected communities are subjected to algorithmic forms of decision-making before having had any genuine opportunity to shape their design, objectives, or distributive consequences. The ‘right to have rights’, in this context, must therefore operate not only *ex ante*, through participation in design and regulatory processes, but also retrospectively, by reopening spaces for contestation, revision, auditability, and collective oversight of systems that are already operational.

On this reading, the proposed principles are rendered effective through their proceduralisation within EU law. Without such institutional mediation, rights risk remaining formally recognised yet practically unenforceable, a concern already implicit in Arendt’s critique of rights as dependent on membership in a political-legal order capable of guaranteeing their effectiveness.⁵⁷

⁵⁶ P Schumacher, ‘Horizontal Effects’ in A Marciano and GB Ramello (eds), *Encyclopedia of Law and Economics* (Springer, 2025). Cf GC Moniz, ‘A natureza híbrida da horizontalidade no direito da proteção de dados da União Europeia: a Carta de Direitos Fundamentais, o RGPD e o Tribunal de Justiça’ (2025) 2 *Revista de Investigações Constitucionais*, <https://doi.org/10.5380/rinc.v12i3.100151>.

⁵⁷ Arendt, *The Origins of Totalitarianism* (n 42), p 230ff.

VII. Conclusion

In this article, I have attempted to untether technological considerations from much of contemporary political philosophy.¹ Building on Arendt’s earthbound and constitutive legal perspective, I have argued that current AI regulatory frameworks are highly inadequate and destabilising for energy justice and political life. My central proposition lies in three constitutive principles for enabling a ‘justice-centred AI’ and recreating the common world that has been lost in the ever-growing alienation and mass society of energy consumers, starting from the European Union. To institutionalise such principles, I offer a framework where bonds and ties for the future are enshrined through institutional practice ‘without, however, destroying the fundamental contingency of action’.²

More generally, as for other technological inflection points of the past (eg the printing press or the railways),³ AI presents us with both a technical and a cultural shift. To overcome the outright commodification of data and other grave crises (eg climate change and energy transition consensus), current regulatory frameworks need to understand their intrinsically cultural, as opposed to normative, value.⁴ In particular, the law can contribute to the much-needed cultural transformation of our relationship with AI. It can counter the doom-and-gloom narratives on AI takeovers of humanity, especially in the high-risk field of energy systems, by imposing non-voluntary obligations on AI providers and deployers and places of deliberation for non-expert stakeholders. Such findings open new avenues for future research on fiduciary obligations to be extended beyond corporate actors to include public authorities deploying AI systems, particularly where such systems affect access to essential services such as energy.⁵

Overall, AI laws can contribute to displaying AI ‘as our own creative power’: belonging to humans by being human-made and human-chosen.⁶ As a matter of fact,

¹ Gabriel (n 133).

² Borren (n 189), p 109.

³ K Crawford and R Calo, ‘There Is a Blind Spot in AI Research’ (2016) 538 *Nature* 311.

⁴ J Øyrehagen Sunde, *1000 år med norsk rettshistorie: ei annleis noregshistorie om rett, kommunikasjonsteknologi, historisk endring og rettstat* (Dreyers Forlag, 2023); C Volk, ‘Hannah Arendt and the Law’ (2013) 11 *International Journal of Constitutional Law* 261.

⁵ On states’ fiduciary duties, see E Fox-Decent and EJ Criddle, ‘The Fiduciary Constitution of Human Rights’ (2009) 15 *Legal Theory* 301.

⁶ S Vallor, *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking* (Oxford University Press, 2024), pp 11–12.

AI is thus neither artificial nor intelligent.⁷ In contrast, AI systems mirror our own intelligence back to us.⁸ Such backward knowledge has great potential but can also ‘magnify, occlude and distort what is captured in their frame’.⁹ More dangerously, it can ‘induce a type of self-forgetting ... that looses our grip on our own human agency and clouds our self-knowledge’,¹⁰ preventing us from solving existential crises.

Conclusively, Arendt helps us reimagine our common world by entrusting technology with a mediating, rather than a purpose-led, function. Technology introduces scientific findings ‘into the everyday world of appearances and renders them accessible to common-sense experience’.¹¹ Conversely, questions raised by thinking—questions of meaning—are unanswerable by common sense¹² and, thus, by technology. By preserving the recognition of human thinking, willing, and judging through justice-regulated AI, with healthy optimism, we can consider that humans will continue to pose such unanswerable questions of meaning, remaining capable of new beginnings and a common world protected by the solid walls of laws.¹³

⁷ Ibid, p 7. On AI as an artefact able to act without being intelligence, see L Floridi, ‘AI as *Agency without Intelligence*: On ChatGPT, Large Language Models, and Other Generative Models’ (2023) 36 *Philosophy and Technology* 15.

⁸ Vallor (n 109), p 2.

⁹ Ibid, p 13.

¹⁰ Ibid, p 2.

¹¹ Arendt, *The Life of the Mind* (n 1), Thinking, p 57.

¹² Ibid, p 58.

¹³ Ibid, p 62.

The Ted K Archive

Esmeralda Colombo
Lifting the Veil of the Future
Arendt and Justice-Centred Artificial Intelligence
17 July 2026

Cambridge Yearbook of European Legal Studies, pp. 1–30. DOI:
<https://doi.org/10.1017/cel.2026.10033>
The Author(s), 2026. Published by Cambridge University Press on behalf of Centre
for European Legal Studies, Faculty of Law, University of Cambridge. Creative
Commons Attribution-NonCommercial-NoDerivatives licence
(<http://creativecommons.org/licenses/by-nc-nd/4.0>)

www.thetedkarchive.com