European Law Open (2022), page 1 of 27
doi:10.1017/elo.2022.47

CAMBRIDGE UNIVERSITY PRESS

CORE ANALYSIS

Reclaiming transparency: contesting the logics of secrecy within the AI Act

Madalina Busuioc[1], Deirdre Curtin[2]* and Marco Almada[3]

[1]Professor of Public Governance, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, [2]Professor of European Union Law, European University Institute, Florence, Italy and [3]Doctoral researcher, European University Institute, Florence, Italy
*Corresponding author. E-mail: deirdre.curtin@eui.eu

(Received 16 June 2022; revised 19 October 2022; accepted 1 November 2022)

Abstract

Transparency is widely acknowledged as a core value in the governance of artificial intelligence (AI) technologies. However, scholarship on AI technologies and their regulation often casts this need for transparency in terms of requirements for the explanation of algorithmic outputs and/or decisions produced with the involvement of opaque black-box AI systems. Our article argues that this discourse has re-interpreted and reshaped transparency in fundamental ways away from its original meaning. The target of transparency – in most cases, the provider of AI software – determines and shapes what is made visible to the outside world, and there is no external check on the validity and accuracy of such mediated accounts and explanations, opening transparency up for manipulation. Through a theoretically informed and critical analysis of the transparency provisions in the European Union's AI Act proposal, the article shows that the substitution of transparency with mediated explanations faces important technical constraints, creates opportunities and incentives for both providers and public-sector users of AI systems to adopt opaque practices, and reinforces secrecy requirements that gag accountability in practice. An approach to transparency as disclosure thus becomes necessary, even if not sufficient in and of itself, to ensure the accountable development and use of AI technologies in the European Union. Transparency needs to be reclaimed as a core concept, accountability tailored and reinforced and the necessity for secrecy re-examined and cordoned off.

Keywords: (algorithmic) transparency; artificial intelligence; public accountability; explanation; AI Act

1. Introduction

Transparency has been critical to artificial intelligence (AI) debates, not least because artificial intelligence algorithms come with significant transparency challenges. Due to their so-called black-box nature,[1] AI algorithms raise unprecedented opacity challenges[2] by virtue of their *technical complexity* (powerful AI algorithms such as neural networks, or deep learning, are highly opaque in their functioning) as well as due to the *proprietary nature* of many AI algorithmic systems deployed both in public and private domains. At the same time, transparency is crucial in view of the core areas where AI algorithms are increasingly relied upon and the high-stakes

[1]F Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Reprint Edition, Harvard University Press 2016).

[2]J Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' 3 (2016) Big Data & Society 1.

implications of their deployment.[3] For instance, the reliance by tax authorities in the Netherlands on a 'learning algorithm' as decisional aid in childcare benefits assessments resulted in state-sanctioned discrimination.[4] The AI algorithm disproportionately flagged citizens with a migration background, causing widespread harm. Among the repercussions faced by wrongly accused families were dire financial circumstances, bankruptcies, lasting mental health damage, broken families and thousands of children taken into foster care.[5] Victims that attempted to challenge the system were told that officials could not access the algorithmic inputs, with public officials reportedly justifying decisions 'because the algorithm said so'.[6] The scandal speaks acutely to the high-stakes implications of the reliance on AI systems in the absence of systemic transparency as to their functioning and adequate (institutional) guardrails and safeguards on their deployment, in line with administrative law imperatives.[7]

Calls for more transparency are all around us, in the public sphere and for government actors, across almost all policy areas, but also increasingly in the private sphere too. Transparency is what can be called a 'floating signifier',[8] a malleable concept that is empty of specific content but rather refers to form. That form, at its core, is a medium that is seen through, rather than looked at directly.[9] Koivisto has described its original meaning as 'the promise of unmediated visibility'.[10] As a normative metaphor, 'it promises legitimacy by making an object or behaviour visible'.[11] Birchall describes transparency as the invisible medium 'through which content is brought to our attention, into the visible realm'.[12]

Conceptually, transparency is closely linked to accountability, yet differs from it on important counts.[13] Its ambitions are more modest as it does not proclaim – in and of itself – to justify, to

[3]See also M Busuioc, 'Accountable Artificial Intelligence: Holding Algorithms to Account' 81 (2021) Public Administration Review 825; N Diakopoulos, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes* (Tow Center for Digital Journalism 2014).

[4]Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers kinderopvangtoeslag (Autoriteit Persoonsgegevens 2020) Research Report z2018-22445 <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf> accessed 1 May 2022; Ongekend onrecht. Parlementaire ondervragingscommissie Kinderopvangtoeslag (*Tweede Kamer*, 2020) <https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf> accessed 1 May 2022.

[5]M Heikkilä, 'A Dutch Algorithm Scandal Serves a Warning to Europe – The AI Act Won't Save Us' *POLITICO* (30 March 2022) <https://www.politico.eu/newsletter/ai-decoded/a-dutch-algorithm-scandal-serves-a-warning-to-europe-the-ai-act-wont-save-us-2/> accessed 1 May 2022; G Geiger, 'How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud' (*Vice*, 1 March 2021) <https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud> accessed 1 May 2022.

[6]D Hadwick and S Lan, 'Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany' 13 (2021) World Tax Journal 6.

[7]On the broader debates about the impact of automation in administrative law, see, inter alia, DK Citron, 'Technological Due Process' 85 (2008) Washington University Law Review 1249; R Calo and DK Citron, 'The Automated Administrative State: A Crisis of Legitimacy' 70 (2021) Emory Law Journal 797; C Harlow and R Rawlings, 'Proceduralism and Automation: Challenges to the Values of Administrative Law' in E Fisher, J King and A Young (eds), *The Foundations and Future of Public Law: Essays in Honour of Paul Craig* (Oxford University Press 2020) 275–98; J Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making' 39 (2019) Legal Studies 636; D Curtin, 'The EU Automated State Disassembled' in E Fisher, J King and A Young (eds), *The Foundations and Future of Public Law: Essays in Honour of Paul Craig* (Oxford University Press 2020) 233–56; S Ranchordás, 'Empathy in the Digital Administrative State' 71 (2022) Duke Law Journal 1341.

[8]C Birchall, *Radical Secrecy: The Ends of Transparency in Datafied America* (University of Minnesota Press 2021) 4.

[9]Deirdre Curtin, '"Accountable Independence" of the European Central Bank: Seeing the Logics of Transparency' 23 (2017) European Law Journal 28.

[10]I Koivisto, *Thinking Inside the Box: The Promise and Boundaries of Transparency in Automated Decision-Making* (European University Institute 2020) Working Paper AEL 2020/01 3. On this very point, see also Curtin (n 9); G Michener and K Bersch, 'Identifying Transparency' 18 (2013) Information Polity 233.

[11]Koivisto (n 10) 3.

[12]C Birchall, 'Radical Transparency?' 14 (2014) Cultural Studies ↔ Critical Methodologies 77.

[13]M Bovens, 'Analysing and Assessing Accountability: A Conceptual Framework' 13 (2007) European Law Journal 447.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

explain, to control nor to hold power to account but rather to *render visible* that which is hidden, and in doing so, to open up possibilities for oversight that otherwise would not be there. As such, transparency is an indispensable first step (albeit not sufficient on its own) towards the realisation of other goals – a *necessary condition* for the functioning of accountability is the acquisition of accurate and reliable information by relevant forums. While transparency often gets unjustly 'maligned' for failing to realise these related (proximate) goals,[14] in fact transparency or information disclosure is a vital first phase of accountability, but this information then needs to be next taken up, questioned, scrutinized, and prodded by forums, explained, debated, and justified as part of accountability processes, with the possibility for consequences to arise, should actor explanations and justifications fall short of expectations.[15] If this fails to materialize and transparency is not taken up further to have meaningful effects, it is a failure of accountability as a check on power, and speaks to the need to bolster our accountability mechanisms and processes rather than to the limitations of transparency.

The value of transparency is thus instrumental to other goals and follows rights. If new rights are attributed, transparency may be obliged to ensure their enjoyment. If rules expand, such as free movement or the internal market, then so too does the scope of transparency.[16]

In the new digital context of automated decision-making and the use of AI there is a vigorous debate on the meaning and reach of transparency,[17] which has, in turn, impacted how transparency is being legislated in this area. A *key argument* advanced by this article is that transparency in this context (and relatedly, in recent legislative efforts) has shifted from its original meaning of visibility to the adoption of a completely different logic, namely that of *communication,* where the *target* of transparency determines, shapes, and influences the content of what is made visible to the outside world. Many authors – and policy-makers – now equate transparency as meaning (only) communication in the sense of explanation. Explainability[18] (and related concepts such as explicability[19] or understandability[20]) dominate AI transparency and governance debates, with explainability advanced as a way to address transparency problems raised by opaque black-box models. With its emphasis on explainability, the discourse on transparency in relation to AI has reinterpreted and reshaped transparency in fundamental ways, away from this original and literal meaning. In these understandings, transparency is no longer about immediate visibility but rather about a significant re-framing occurring towards *explanation*. In the name of facilitating understanding, a form of heavily mediated, pre-digested information provision is being advanced, often to a limited group of users and notably, *in the absence of any external check on the validity*

[14]See for instance, M Ananny and K Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' 20 (2018) New Media & Society 973. Ananny and Crawford claim transparency fails to 'control' and 'govern' algorithmic systems, while this is not within transparency's remit in and of itself.

[15]Bovens (n 13).

[16]See further A Buijze, 'The Six Faces of Transparency' 9 (2013) Utrecht Law Review 3.

[17]For a snapshot of the various accounts of AI transparency in the EU legal order, see M Finck, 'Automated Decision-Making and Administrative Law' in P Cane and Others (eds), *The Oxford Handbook of Comparative Administrative Law* (Oxford University Press 2020); L Edwards and M Veale, 'Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking For' 16 (2017) Duke Law & Technology Review 18; T Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in T Wischmeyer and T Rademacher (eds), *Regulating Artificial Intelligence* (Springer International Publishing 2020); M Brkan and G Bonnet, 'Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas' 11 (2020) European Journal of Risk Regulation 18; M Fink and M Finck, 'Reasoned A(I)Administration: Explanation Requirements in EU Law and the Automation of Public Administration' 47 (2022) European Law Review 376.

[18]S Wachter, B Mittelstadt and C Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' 31 (2018) Harvard Journal of Law & Technology 842.

[19]AI HLEG, *Ethics Guidelines for Trustworthy AI* (European Commission 2019); L Floridi and Others, 'AI4People–An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' 28 (2018) Minds and Machines 689.

[20]Ananny and Crawford (n 14) 985.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

and accuracy of these mediated accounts and explanations, opening transparency up for manipulation.

This directly connects with the 'dark side' of transparency. The dark side of transparency is secrecy and secrecy has expanded. We argue that the conceptual substitution of transparency for explanation risks reinforcing secrecy in this context. In an institutional world where transparency is substituted with explainability, the AI 'black box' no longer needs to be opened and disclosure becomes seemingly redundant. The corollary is that secrecy and proprietary protections can then be kept, expanded, and legitimized, with transparency in its core meaning discarded as 'inadequate'.[21] Black boxes (produced by technical and/or legal opacity), become the accepted norm. And this reframed logic, as we will see below, feeds directly into how AI transparency is to be legislated in practice in this context, weakening accountability of providers and use by public authorities with transparency becoming a rudimentary facsimile of its former self.

These dynamics are not new and closely map onto familiar debates from more traditional areas, outside and beyond the scope of AI. For example, part of the ecosystem of transparency within the European Union (EU) has always been the rules on secrecy, featuring alongside mentions of transparency in many and varied legislative instruments and policy areas. Now, more than two decades after the adoption of the transparency regulation on access to documents,[22] a draft regulation is proposed that will regulate professional secrecy within the EU in a legislative instrument that also impacts on the requirements of professional secrecy and trade secrecy within the AI Act.[23] Such rules and practices throw up opaque filters that block visibility or only give it to specific audiences (and subject to confidentiality requirements) or in highly secluded ways, also for accountability forums such as courts or parliaments, with the risk of cordoning such areas from oversight.[24]

With the adoption of regulation on AI, the EU is seemingly the first political and legal system to define in legislation what transparency must be taken to mean in this context. It turns out that information governance under the AI Act is mainly about secrecy not about visibility, about concealment rather than about disclosure and if there is disclosure it is to be a limited and secluded one, with no participation or accountability envisaged. Our article explores the provisions of the (draft) AI Act on transparency, secrecy, and related aspects of governance to understand the implications in the context of the wider conceptualization of transparency and accountability, in the broader context of European governance. We do so as European lawyers, accountability scholars and a computer scientist/lawyer at a crucial moment in the debate on the regulation of AI in Europe and ultimately globally in a context where the language of transparency is misappropriated. It is borrowed and used to imply a promise of visibility, of public accountability and with a vista of participation, none of which are or can in fact be realized in practice as currently framed. We critically contest existing conceptualizations and call for some recalibration of the core concepts as applied in the AI context to ground the normative promise implied in the use of the term transparency. We argue that transparency needs to be reclaimed as a core concept, accountability tailored and reinforced and the necessity for secrecy re-examined and cordoned off.

[21]Ananny and Crawford (n 14), 982–84.

[22]See in general terms the blog with some overview, HCH Hofmann and P Leino-Sandberg, 'An Agenda for Transparency in the EU' (*European Law Blog*, 23 October 2019) <https://europeanlawblog.eu/2019/10/23/an-agenda-for-transparency-in-the-eu/> accessed 13 June 2022.

[23]Proposal for a Regulation of the European Parliament and of the Council on information security in the institutions, bodies, offices and agencies of the Union 2022 (COM(2022) 122 final, 22 March 2022).

[24]D Curtin, 'Second Order Secrecy and Europe's Legality Mosaics' 41 (2018) West European Politics 846. See, in general, V Abazi, *Secrecy and Oversight in the EU: Law and Practices of Classified Information* (Oxford University Press 2019).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

2. Hiding in plain sight

A. Entangled logics of secrecy and transparency

The concepts of transparency and of secrecy are to be considered in relation to one another and as each other's complement. As we have seen, the very word transparency connotes the ability to see through. Visibility is literally in the name transparency: it implies a subject seeing as well as the object being seen. In this sense, it is a two-way street: a beholder and an object form a complex set of affairs.[25] The beholder may force the object to react to its 'gaze' and in this sense may be controlling. In the European Union, transparency has as a floating signifier carried an almost impossible burden over the years. It would allegedly, according to the Commission, make the Union closer to the people, and stimulate a more informed and involved debate on EU policy.[26] The evolution of the debate and practice of transparency has been closely connected with and in part seen as a remedy for the so-called 'democratic deficit' of the EU. Democracy requires institutional accountability and the preliminary *sine qua non* for that is the provision of reliable information.

Secrecy, on the other hand, can be defined as all those behaviours whereby one party intentionally conceals information from another.[27] Secrecy creates a form of opacity, which makes it hard to discover who takes the decisions, what they are, and who gains and who loses – precisely the opposite of what transparency is meant to achieve.

There are arguably two main competing forms or logics within transparency which will be detailed below also as a framing to move beyond the approach taken in the AI Act. The first logic concerns a *logic of disclosure* or *public access* and comes closest to the core meaning of transparency – transparency in its unadulterated self, amounting to public release of information through the direct disclosure of documents (such as reports, meeting transcripts, etc). The system is one of passive disclosure by distinct actors once disclosure is requested but the idea is that over time more active disclosure through public registers of documents and special public databases will also occur. At the same time, this logic includes whole areas being removed from the rules on disclosure for various legal and political reasons (including rules on secrecy). The very essence of rules on professional secrecy is to define a system of insiders and outsiders, those who are secluded and those who are excluded.

The second logic is substantively mediated and at no point involves actual visibility or seeing through. This is the *logic of communication* or explanation and in its very essence it does not employ the medium of transparency but rather of 'describing', communicating, or explaining aspects of what is kept hidden or not revealed. The disclosure involves a triad between the subject listening, the object not being seen and the secret keeper revealing the rationale or explanation.

This exposes a fundamental qualitative difference between the two logics: the logic of disclosure is one of little or no mediation, whereas the logic of communication often relies on active curation behind closed doors and subsequent narration around essentially undisclosed information. The limited mediation involved in the logic of disclosure is the act of formally granting access (through a public register or subject to an access to documents request). Information that is hidden or not public necessarily requires technically granting access to the document or information in question. This stands in stark contrast to the substantive mediation involved in the logic of communication where the specific information is not revealed in its granular and authentic original detail ('raw') but rather the narrated reasoning behind it in the words only of the mediator.

[25] I Koivisto, *The Anatomy of Transparency: The Concept and Its Multifarious Implications* Working Paper EUI/MWP 2016/09.

[26] D Curtin and AJ Meijer, 'Does Transparency Strengthen Legitimacy?' 11 (2006) Information Polity 109.

[27] See S Bok, *Secrets: On the Ethics of Concealment and Revelation* (Oxford University Press 1984); A Gutmann and DF Thompson, *Democracy and Disagreement* (Belknap Press 1998) ch 3; O Pfersmann, 'Norme de secret, normes secrètes et État moderne' 26 (2006) *Cités* 115; DE Pozen, 'Deep Secrecy' 62 (2010) Stanford Law Review 257; KL Scheppele, *Legal Secrets: Equality and Efficiency in the Common Law* (University of Chicago Press 1988). See also Western European Politics 41 (2018).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

B. Logic of disclosure: a public magnifying glass

The logic of disclosure thus pertains to transparency as visibility, reflecting its semantic pedigree of 'light' and 'sight'.[28] It speaks to 'the ability to *look clearly* through the windows of an institution',[29] with visibility described as 'a necessary condition for there to be transparency': 'Transparency is about information, and if information is not *visible* then the first and primary meaning of the parent-word, "transparent" – having to do with light and visual properties – loses its relevance'.[30]

The notion of directly (but not necessarily fully) seeing amounts to a form of immediate verification that speaks to transparency's enduring appeal: 'So long as we can witness the reality with our own eyes, we do not need verbal explanations, which are, by virtue of transparency, indirectly considered less reliable than direct visual observation'.[31] This notion of 'immediate observability' stands in 'a tense relationship' with mediation: 'the more the object of transparency becomes mediated, the less immediacy there can be'.[32] Relatedly, the emphasis on 'granularity' – the disclosure of raw, detailed, or close-to-the-source data (such as minutes, transcripts, datasets etc) – regarded as an important dimension of transparency, is a means to reduce the risk of such information being strategically processed or gamed.[33] Of course, transparency as disclosure is not without trade-offs.[34] For instance, disclosure of (granular or raw) information can come at the expense of simplification, which can render such information less accessible or user friendly.[35] Nevertheless, while there are always trade-offs involved, transparency as disclosure serves a crucial function pertaining to informational reliability: 'Because raw data is usually less mediated, it typically reflects fewer opportunities for officials to "cook" or "game" it out of professional or political motivations'.[36]

There have been some significant legislative changes in practice that have sought to create an ecosystem around transparency as disclosure for the EU. The earliest and by now most developed were the efforts to copper-fasten in law and in practice informational transparency. This dates from 2001, more than two decades ago, and has been incrementally expanded and fought over both by individuals of one type or another – lawyers, experts, academics, journalists, NGOs, etc – as well as by institutions and institutional representatives (eg members of parliament). Yet transparency in this informational sense has remained troubled and troubling, just as is the case in other parts of the world where there is an even longer tradition of freedom of information (for example the United States).[37] It is inevitably dependent on the practice of the institutions, their willingness to pro-actively provide information as well as passively on request and also to a large extent on the institutions meant to police it.

As we have seen, this logic of public access comes closest to the understanding of transparency as visibility. It does not patronise the citizen but rather values the role that the public and the informed citizen can play in a wider democratic perspective. In this perspective, transparency

[28]Michener and Bersch (n 10).

[29]MGW den Boer, 'Steamy Windows: Transparency and Openness in Justice and Home Affairs' in V Deckmyn and I Thomson (eds), *Openness and Transparency in the European Union* (European Institute of Public Administration 1998) 105, emphasis added.

[30]Michener and Bersch (n 10) 238.

[31]Koivisto (n 10) 10.

[32]I Koivisto, 'Transparency in the Digital Environment' 8 (2021) Critical Analysis of Law 1, 2.

[33]G Porumbescu, A Meijer and S Grimmelikhuijsen, *Government Transparency: State of the Art and New Perspectives*, 1st edn (Cambridge University Press 2022) 11; Michener and Bersch (n 10).

[34]For a discussion of such trade-offs, see Porumbescu, Meijer and Grimmelikhuijsen (n 33) 10–13. See also C Hood and D Heald (eds), *Transparency: The Key to Better Governance?* (Oxford University Press 2006); DE Pozen, 'Transparency's Ideological Drift' 128 (2018) Yale Law Journal 100.

[35]Porumbescu, Meijer and Grimmelikhuijsen (n 33) 11–12.

[36]Michener and Bersch (n 10) 239.

[37]On this topic, see DE Pozen and M Schudson (eds), *Troubling Transparency: The History and Future of Freedom of Information* (Columbia University Press 2018).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

is seen as a fundamental citizens' right and as a means towards securing public accountability.[38] It implies that all arms of government – the executive, the entire public administration as well as parliaments – should be subject to the requirement of openness or public access.[39] The deeper democratic meaning of why openness and transparency are important is based on the logic that 'increased openness (...) enables citizens to participate more closely in the decision-making process and guarantees (...) greater legitimacy and is more effective and more accountable to the citizen in a democratic system'.[40] This classic way of viewing transparency recognizes a relationship between a beholder (seeing) and an object (being seen) and mirrors an accountability relationship between an actor (any actor) and an accountability forum (judicial, parliamentary, audit etc). This is the case even if reality is such that the beholder cannot see straight through and there are significant shutters and blinds in place. In other words, transparency in this logic is not only an end value (full sight) but also plays a critical instrumental role with respect to its potential to further legitimacy, enhance participation rights, reason-giving in the administration and public accountability. At the same time, transparency is different from these concepts – it is often a key pre-requisite, a facilitator, towards the realisation of these other values. And it performs these functions instrumentally even if full sight is not possible.

Following Bovens,[41] transparency is a *necessary* but insufficient condition in and of itself for accountability. Direct citizen accountability through transparency[42] may often be an illusion. On the other hand, the media and other stakeholders can and do use transparency to force change. Similarly, elected representatives can use transparency to demand and enact accountability of public bodies and authorities on behalf of citizens.[43] Transparency, in turn, also enables the media to work as fire alarms, triggering vertical (parliamentary) accountability. Importantly, even if the citizen does *not* directly 'understand by seeing' alone, third-party stakeholders (lawyers, academics, experts, NGOs, advocacy networks, etc) can *publicly* contribute to understanding as well as verify and act as *checks on understandings that are being advanced*. These other forums (with more capacity, resources, expertise) can and do act as fire alarms and demand accountability on behalf of adversely affected citizens. This becomes much harder however, if not next to impossible, under conditions of secrecy.

C. Logic of communication: mediation and explanation

Yet, some authors – and policy-makers – approach *transparency as (only) communication.*[44] As such, they accept that transparency is essentially mediated, and 'excessively simplified and thus are blind to the complexities of the contemporary state, government information and the public'.[45]

In a governance context, communication amounts to the intentional release of *redacted information* on the substance of decisions and of (some of) the facts and reasons on which they are based. Such release of information previously intentionally concealed, is (normally) to an outside audience, which may be affected by the decision but is not involved in the decision-making. It thus makes the outside actor (market, citizen, or parliament) aware of the existence of what is *not*

[38]Curtin (n 9) 39.

[39]C Grønbech-Jensen, 'The Scandinavian Tradition of Open Government and the European Union: Problems of Compatibility?' 5 (1998) Journal of European Public Policy 185.

[40]Joined Cases C-39/05 P and C-52/05 P *Turco v Council of Ministers* (2008) ECLI:EU:C:2008:374, para 45, reiterating the preamble of Regulation 1049/2001.

[41]Bovens (n 13).

[42]V Mabillard and R Zumofen, 'The Complex Relationship between Transparency and Accountability: A Synthesis and Contribution to Existing Frameworks' 32 (2017) Public Policy and Administration 110.

[43]A Meijer, 'Transparency' in M Bovens, RE Goodin and T Schillemans (eds), *The Oxford Handbook of Public Accountability* (Oxford University Press 2014) 507–24.

[44]On this point, see Curtin (n 9) 32.

[45]M Fenster, 'Transparency in Search of a Theory' 18 (2015) European Journal of Social Theory 150, 150.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

known, for example that deliberations, negotiations and other elements of constituent decisions have taken place – but does not give transparency of the actual process and content of that decision-making – at least not in unredacted form.

The logic of communication reveals almost unlimited discretion to the object or target of transparency as to what is revealed and what is kept secret. It is not the beholder controlling the object but rather the opposite. In effect, *the logic of communication is a logic of secrecy*, not disclosing the actual secret but narrating a story around it either openly or (more commonly) to a specified number of stakeholders, respecting certain rules on professional secrecy. Information is pre-digested, recounted and 'spun' into an account at the intermediary's control. By contrast, in a disclosure logic of transparency, the object is less in control and is forced to assume a more reactive and defensive position by virtue of interactions with others (parliaments, the public, ombudsmen or courts).

The turn towards explanation in algorithmic transparency debates signals, as we will see below, a similar move towards a logic of communication through the provision of pre-digested, simplified information. While such an approach is meant to facilitate understanding on the part of the beholder, it simultaneously affords the object (rather than the beholder) control over openness. We become 'passive recipients of simplified information' 'increasingly dependent on translating intermediaries',[46] which simultaneously opens opportunities for undue influence and distortion: 'Explanation includes more human influence than sheer transparency'.[47]

It is important to recognize that efforts to reframe transparency in this manner – away from its core meaning of disclosure – are not unprecedented, nor unique to AI algorithmic debates. Such arguments have been recurrently and strategically deployed, for instance, by public actors to advance preferred (and often self-serving) versions of transparency. For instance, under pressure to increase its transparency, the European Central Bank (ECB) advanced notions of transparency as mediated information provision, in an effort to control the parameters of its own transparency.[48] It resisted providing access to raw information in favour of the object explaining and translating the rationale. For instance, it argued explicitly in its annual reports: 'Transparency means not only releasing information, *but also structuring that information in such a way that the public can understand it* . . . Transparency requires central banks to clearly *explain how they interpret* and implement their mandates'.[49] In doing so, it strategically followed a deliberate and autonomous strategy to assert its discretion and ensure a version of transparency to its own liking, affording itself control over what it *chooses* to reveal.

Understanding communication as a form of transparency in rationale is heavily contested. Scholars rightfully dispute that it is in fact a form of transparency that can really lead to accountability with the secret keeper able to retain absolute control over what is released, when and how.[50] The very word communication (or explanation) implies linearity of the process and passivity on the part of the beholder or recipient. Transparency in rationale enables the secret keeper to enjoy almost unlimited discretion to autonomously decide what to intentionally reveal and what not to reveal and with what slant to 'communicate' it. It may generate some legitimacy for the secret keepers, but it is also *more vulnerable to manipulation*.

3. AI debates: transparency appropriated

A. Hand in glove: explanation and secrecy protections

With its emphasis on explainability, the discourse on transparency in relation to AI has re-interpreted and reshaped transparency in ways similar to what has been described above,

[46]Koivisto (n 10) 20.
[47]Ibid.
[48]Curtin (n 9).
[49]ECB, *Annual Report* (2003), at 142, emphasis added.
[50]See Curtin (n 24); B Rittberger and KH Goetz, 'Secrecy in Europe' 41 (2018) West European Politics 825.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

towards a logic of communication. In these understandings, transparency is no longer about seeing or visibility but rather a significant re-framing occurs towards explanation, which is to facilitate understanding – essentially, a form of heavily mediated information provision is being advanced. This shift is visible not only recurrently in scholarly debates[51] but also in policy documents. For example, the High-level Expert Group on Artificial Intelligence in its ethical guidelines for trustworthy AI describes transparency in this context as 'closely linked with the principle of explicability'[52] and defines transparency as '[i]ncluding traceability, explainability and communication'.[53]

This recasting of transparency is motivated by the fact that while in its original meaning transparency 'assumes that when we see by ourselves, we can understand what is happening',[54] this assumption, to a significant extent, no longer holds true when it comes to complex AI algorithmic systems' functioning. In a digitalised word, the promise of seeing by knowing seemingly starts to fray, with the potential to render transparency's claim to knowability more tenuous.[55] Machine learning models such as neural networks (or 'deep learning'), the most important part of ML these days, can be extremely complex in their operation: hundreds of layers deep, thousands of features and millions (and even billions) of weights that contribute to one predictive outcome.[56] The sheer size of the parameter space and the complexity of feature interactions give rise to unprecedented opacity. By contrast to the traditional contexts we are accustomed to demand transparency of, in the context of AI, visibility or *disclosure* does not automatically lead to understanding: 'Seeing inside a system does not necessarily mean understanding its behavior and origins'.[57]

This, in turn, has led some authors to altogether disavow the importance of transparency as seeing inside the black box (ie disclosure of underlying training data, source code, and model) as a means to govern algorithmic models: 'transparency is an inadequate way to govern algorithmic systems'.[58] Various authors have highlighted the limitations of traditional understandings of transparency in this context, advocating for a move away from its core understanding. For instance, de Fine Licht and de Fine Licht speak of the 'harms' of 'full transparency' arguing that disclosing how systems goals are defined, coded, and implemented would make '[f]inding the relevant information … as difficult as finding the proverbial needle in a haystack',[59] while Wischmeyer claims that information providing measures fail to impart a sense of agency upon those affected by algorithmic systems.[60] Transparency becomes relegated to: 'unnecessary and always insufficient to simply look inside structures' in the context of AI.[61]

At the same time, such conclusions can be contested in important ways. While lay citizens might not understand through disclosure alone, this does not remove its relevance when it comes to important fire-alarm forums, such as the media, regulators, or scholars. We have seen for

[51]See for instance, Wachter, Mittelstadt and Russell (n 18); M E Kaminski, 'The Right to Explanation, Explained' 34 (2019) Berkeley Technology Law Journal 189, s V; Brkan and Bonnet (n 17) s II; P Hacker and J-H Passoth, 'Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond' in A Holzinger and Others (eds), *xxAI – Beyond Explainable AI* (Springer International Publishing 2022) 343–73; L Naudts, P Dewitte and J Ausloos, 'Meaningful Transparency through Data Rights: A Multidimensional Analysis' in E Kosta and R Leenes (eds), *Research Handbook on EU Data Protection Law* (Edward Elgar Publishing 2022) s 3.4.

[52]AI HLEG (n 19) 18.

[53]Ibid., 14.

[54]Koivisto (n 10) 9.

[55]I Koivisto, *The Transparency Paradox* (Oxford University Press 2022).

[56]Busuioc (n 3) 829.

[57]Ananny and Crawford (n 14) 980.

[58]Ibid., 982.

[59]K de Fine Licht and J de Fine Licht, 'Artificial Intelligence, Transparency, and Public Decision-Making' 35 (2020) AI & Society 917, 922.

[60]Wischmeyer (n 17) 87.

[61]Ananny and Crawford (n 14) 985.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

instance, computational journalists[62] and scientists[63] play critical roles in this respect, with the added obstacle of having to additionally reverse-engineer secret models to obtain some semblance of model disclosure to be able to investigate model behaviour and exercise crucial fire alarm roles. Such forums now must try to second-guess and imperfectly approximate features of secret models, with much higher thresholds to become activated (if at all) in their fire alarm roles, with proprietary protections acting as 'a key constraint . . . making it difficult for independent researchers to dissect them', left to 'only guess as to the actual mechanisms' behind disparities.[64]

In addition to being qualified as unnecessary in dominant debates, transparency is simultaneously presented as an impossible ideal. Disclosure, the argument goes, would raise unreasonable demands of private actors to forgo proprietary protections, potentially harming innovation and/or creating opportunities of gaming by those targeted though the deployment of AI systems. For instance, Joshua Kroll and co-authors argue that:

> [d]isclosure of source code is often neither necessary (because of alternative techniques from computer science) nor sufficient (because of the issues analyzing code) (. . .). Furthermore, transparency may be undesirable, such as when it permits tax cheats or terrorists to game the systems determining audits or security screening or discloses private information[65]

advocating for a move away from transparency as visibility (or disclosure). Yet, the effectiveness of secrecy as an antidote for gaming is far from uncontested in the technical literature.[66] Cynthia Rudin points out that well-designed systems ensure attempts to game the system align with overall goal improvement – for instance, attempting to 'game' one's credit rating by going out of red aligns with the ultimate policy goal of improving one's credit worthiness.[67] What is more, many characteristics – such as past record and family history – cannot actually be gamed but are immutable. It is rather sub-optimal systems based on poor proxies – such as the use of water usage measurements as an indicator of benefits fraud[68] – that are likely to lend themselves to gaming. Such setups raise important questions whether public systems should not rather be aimed towards overall system improvement (through well-designed systems) rather than trying to 'catch out' citizens with poorly designed models protected by secrecy.

These critiques notwithstanding, the risk of misuse of disclosed information has had an enduring appeal and is often invoked as part of arguments for the inadequacy of transparency as disclosure. In these contexts, where it is argued opening the black box could prove harmful to private interests and/or the achievement of end goals, explanation techniques of black boxes have been advanced as an alternative to disclosure, as a way to have our transparency cake and eat

[62]J Angwin and Others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 1 May 2022.

[63]See eg a study that diagnosed racial bias in a widely used health care algorithm, deployed on more than 200 million people in the US context: Z Obermeyer and Others, 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' 366 (2019) Science 447.

[64]Obermeyer and others (n 63) 447.

[65]J Kroll and Others, 'Accountable Algorithms' 165 (2017) University of Pennsylvania Law Review 633, 633–4.

[66]See, eg D Pavlovic, 'Gaming Security by Obscurity', in S Peisert and Others (eds), *NSPW'11: Proceedings of the 2011 New Security Paradigms Workshop* (Association for Computing Machinery 2011) 125–40; B Biggio and F Roli, 'Wild Patterns: Ten Years after the Rise of Adversarial Machine Learning' 84 (2018) Pattern Recognition 317; A Venturi and C Zanasi, 'On the Feasibility of Adversarial Machine Learning in Malware and Network Intrusion Detection' in *2021 IEEE 20th International Symposium on Network Computing and Applications (NCA)* (2021).

[67]C Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' 1 (2019) Nature Machine Intelligence 206.

[68]This specific indicator appeared in the precursor to the SyRI system used for welfare fraud detection in the Netherlands: C van Veem, 'Profiling the Poor in the Dutch Welfare State. Report on Court Hearing in Litigation in the Netherlands about Digital Welfare Fraud Detection System ("SyRI")' (*Center for Human Rights and Global Justice | NYU School of Law*, 1 November 2019) <https://chrgj.org/2019/11/01/profiling-the-poor-in-the-dutch-welfare-state/> accessed 1 May 2022.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

it. This amounts to understanding without seeing (without disclosure), and with proprietary protections safely in place.

B. XAI to the rescue? Having your transparency cake and eating it?

The move (or swerve, rather) towards transparency-as-explanation is epitomised in technical solutions in the form of the explainable AI (or XAI). This umbrella term refers to a variety of techniques attempting to explain black-box model behaviour.[69] Traditionally developed by engineers to investigate model behaviour and troubleshoot models, such techniques are being advanced as a solution to black-box transparency problems, this time in institutional and governance contexts. These techniques generally involve interfacing a black-box model with a second post-hoc explanation model.[70] Essentially, using a simpler algorithm to explain the black box, to effectively re-digest the black-box model into something we can understand. The post-hoc explanation model is often a different model altogether to the underlying black box, with different input features than the original black-box model.[71] Rather than seeing with our own eyes, with explainable AI, providers of AI systems develop explanation models of their black-box models, playing a mediating role between the AI system and those affected by it. At first sight, in light of their justificatory connotations, such approaches seem to rhyme with the demands of reasoned decision-making (and reasoned administration, when public authority is concerned). What is more, the appeal of such approaches simultaneously stems from the fact that they also bypass opening the black box. With such techniques, the underlying model can remain secret or undisclosed safeguarding intellectual property protections – with the explanation model facilitating understanding without actually seeing inside the black box.

While seemingly summoning the promise of transparent and reasoned decision-making, we argue below that the transparency-as-explanation approach ultimately fails to deliver on this promise. What we potentially stand to lose in the process is precisely what is at stake with the logic of communication more broadly: simply put, the emphasis on *explanation opens up transparency for influence*. Mediated explanations open possibilities for abuse and raise questions about the reliability and truthfulness of what is revealed. There is now an intermediary (often the provider of the AI system) controlling, through technical means, that what is revealed, in a context where the intermediary has high stakes in what is being revealed or kept secret. Effectively, the quality, fidelity and truthfulness of the account provided hangs on the trustworthiness of an intermediary with heavy vested interests in the content of the information presented.

The issue is not a theoretical one: while such techniques can be valuable for designers and system engineers to understand and investigate their models in the design process, we are seeing growing critique of these approaches when it comes to their ability to allow us to really understand specific outcomes[72] and afford meaningful transparency of system functioning. This raises serious

[69]For different explanation techniques see eg MT Ribeiro, S Singh and C Guestrin, '"Why Should I Trust You?": Explaining the Predictions of Any Classifier' in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery 2016); SM Lundberg and S-I Lee, 'A Unified Approach to Interpreting Model Predictions' in *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems December 2017*, 4768–77; Wachter, Mittelstadt and Russell (n 18); P Jonathon Phillips and Others, *Four Principles of Explainable Artificial Intelligence* (National Institute of Standards and Technology 2021) s 6.

[70]Ribeiro, Singh, and Guestrin (n 69).

[71]Ibid.

[72]See, *inter alia*, Rudin (n 67); S Barocas, AD Selbst and M Raghavan, 'The Hidden Assumptions behind Counterfactual Explanations and Principal Reasons' in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (ACM 2020); H Lakkaraju and O Bastani, '"How Do I Fool You?": Manipulating User Trust via Misleading Black Box Explanations' in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery 2020); S Bordt and Others, 'Post-Hoc Explanations Fail to Achieve Their Purpose in Adversarial Contexts' in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2022).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

questions about their suitability as tools of transparency *in a governance and institutional context*, which we discuss in Section 5 below. But first we critically explore the provisions of the draft EU AI Act in terms of access, disclosure, communication, and secret-keeping and demonstrate how the re-framing of transparency has bled into legislative efforts in this regard.

4. Entangled logics within the EU AI Act: what, who, how?

An earlier form of the debate on explanation as *ersatz* transparency has appeared in the context of data protection law. Under the GDPR,[73] the use of automated decision-making systems attracts the application of specific transparency provisions,[74] which prompted intense scholarly debate on the existence of a right to an explanation of automated decisions.[75] As a result of these debates, transparency has already become a central theme in the accountability regimes for algorithmic decision-making.[76] Its role, however, is likely to be significantly amplified by the new legislative instrument under construction in the EU: the AI Act.[77]

The AI Act establishes a uniform framework for the development, marketing, and use of AI systems throughout the EU.[78] The term 'AI system' refers to a software developed with certain techniques[79] that can 'for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'.[80] This definition captures a broad range of systems currently in use or that have been planned in an European context, as the set of techniques encompasses relatively simple approaches such as decision trees or linear regression models as well as complex techniques such as deep neural networks. The AI Act's definition of AI also covers a broad range of potential applications of AI technologies, such as autonomous vehicles navigating roads with little or no human input, recommender systems personalizing the items shown to consumers, and risk scoring systems used to detect potential cases of tax and benefits fraud. Each of these technologies imposes a different set of risks upon its users, third parties, or society, which the AI Act addresses through a broad range of technical and governance measures.[81]

[73]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) OJ L 119/1.

[74]Under Articles 13(2)(f), 14(2)(g), and 15(1)(h) GDPR, a natural person subject to an automated decision based on their personal data has the right to obtain information about the existence of such a decision, its significance and envisaged consequences, as well as meaningful information about the logic involved in decision-making. While these provisions do not speak of 'explanation', the term is used in Recital 71, which mentions safeguards that should be adopted in automated decision-making.

[75]For an overview of the scholarly debate on whether the GDPR establishes a right to an explanation, as well as its potential contours, see Naudts, Dewitte and Ausloos (n 51) s 3.4. This question has been recently referred to the CJEU in *Dun & Bradstreet Austria* (C-203/22, application lodged in 16/03/2022).

[76]Busuioc (n 3); M Wieringa, 'What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability' in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2020) 4–5.

[77]Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021) 206 final, 21.4.2021) (AI Act). Any mentions to Articles of the AI Act refer to the Commission text, unless specified otherwise. Citations to page numbers direct the reader towards the AI Act proposal's explanatory memorandum.

[78]Recital 1 AI Act.

[79]Annex I AI Act lists three broad groups of techniques defined as AI: machine learning approaches, logic- and knowledge-based approaches, and "[s]tatistical approaches, Bayesian estimation, search and optimization methods".

[80]Article 3(1) AI Act.

[81]On the AI Act as risk-based regulation, see G De Gregorio and P Dunn, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' 59 (2022) Common Market Law Review 473.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

While the Act is justified as a means to pursue various reasons of public interest, such as the protection of fundamental rights,[82] its primary legal foundation is Article 114 of the TFEU, which empowers the EU to adopt unifying measures concerning the functioning of the internal market.[83] As a result of this market focus, essentially on the 'product' and the rules governing it in the interests of free movement, the main regulatory object is the 'AI system'.[84] While a discussion of legal basis is beyond the scope of this article,[85] it is important to note that this choice has important implications for the scope of transparency, its targets and beneficiaries. Despite this reliance on a market harmonization instrument, the AI Act also governs the use of AI systems in the public sector, as seen, *inter alia*, in the list of high-risk applications in Annex III AI Act. At the same time, because of this product safely regulation focus, the Act insufficiently regulates the use of AI systems by public authorities, directing instead most of its regulatory attention to the behaviour of providers.

Thus, consistent with this market focus, most obligations in the AI Act are directed towards the provider of the AI system, that is, the actor 'that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark'.[86] Providers are responsible for specifying the technical properties of any given AI system, which is why the AI Act sets up various requirements that the providers of high-risk AI systems must observe.[87] When it comes to transparency, too, most requirements, are again *focused on the provider* rather than the user for instance – even though such users can be public authorities deploying AI systems in the exercise of salient public tasks. These requirements oblige the provider to supply information about the system to various other actors: the users of the provided AI systems, the general population, and the authorities designated to ensure the application and enforcement of the AI Act.[88] Understanding the transparency regime established by the AI Act requires understanding how information flows between those actors and the secrecy rules that set limits to disclosure and communication requirements.

A. Transparency towards users

The term 'transparency' is used sparingly in the main body of the Act, mostly in the context of the relationship between providers and users of an AI system. High-risk AI systems are subject to specific transparency requirements under Article 13 of the AI Act. Under this article, any high-risk AI system must be transparent enough to allow *users* – the ones putting an AI system into use under their own authority,[89] which are not necessarily the end users of a system – to interpret its outputs and use them 'appropriately'. Transparency, so construed, must be ensured

[82]Recital 1 AI Act.

[83]Recital 2 AI Act.

[84]Article 1 AI Act defines the subject matter of the Act largely in terms of AI systems, which are defined in Article 3(1) AI Act.

[85]But see, inter alia, N Smuha and Others, *A Response to the European Commission's Proposal for an Artificial Intelligent Act* (LEADS Lab @University of Birmingham 2021); L Edwards, *Regulating AI in Europe: Four Problems and Four Solutions* (Ada Lovelace Institute 2022) Expert Opinion <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf> accessed 31 March 2022.

[86]Article 5(1) AI Act.

[87]See Recitals 42, 45, 51 AI Act.

[88]Article 59 AI Act.

[89]Article 5(4) AI Act. The transparency requirements between users and providers presuppose that these roles are played by different actors, but users may also be providers in some circumstances. The most straightforward case is when an actor uses a system that has been developed under its own specifications. In addition, Article 28 AI Act specifies that users are considered providers if they place on the market or put into service a high-risk AI system under their own name or trademark, modify the purpose of a high-risk AI system already in the market or in service, or make substantial modifications to a high-risk AI system.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

throughout the design and development of the AI system,[90] a formulation that suggests providers are required to adopt technical approaches and measures that foster transparency.[91] Under a minimalist reading of this transparency requirement, it would be sufficient to communicate to the user instructions on how to interpret the output, such as potential biases or underlying assumptions of the decisional model. A more expansive approach to communication would require providers to explain how the system arrives at its outcomes to users. But, under a logic of disclosure, proper interpretation and use of the outputs requires information about the system itself and the decision logic embedded in it. The current formulation of transparency in the AI Act does not exclude any of these possible interpretations, giving providers *ample leeway* in the choice of technical means for complying with Article 13(1) AI Act. This risks perpetuating sub-optimal transparency practices and opaque systems accompanied by poor or incomplete transparency measures that provide little insight into actual system functioning.

The AI Act includes a second source of transparency for users: mandatory use instructions. Article 13(2) AI Act requires that any high-risk AI system be accompanied by instructions to users, which must include at least:[92] the contact information of the provider; the characteristics, capabilities, and limitations of the system, such as its level of accuracy;[93] any changes to the system and its performance which have been pre-determined by the controller at the moment of the initial conformity assessment; the human oversight measures adopted in the system; and the expected life cycle of the system, including necessary updates and maintenance practices.

However, compliance with these requirements requires the provider to *give up little information about the inner workings of the system*. Instead, all the required information can be provided in an entirely black-box fashion, in which the provider specifies how users should interact with the system and supplies statistics purporting to represent the properties of system outputs.[94] As a result, even a fully compliant provider still retains non-transparent control over information about the technical aspects of the system they originate. A logic of communication paired with secrecy is what is at stake.

Providers are also required to adopt technical measures meant to ensure users have access to information about the everyday operation of AI systems. Any high-risk AI system must include tools that enable automated logging of what the AI system does,[95] thus enabling the traceability of its operations throughout its life cycle.[96] Providers are also required to identify the circumstances in which human oversight is needed to prevent or minimize risks to health, safety, and fundamental rights,[97] and adopt technical measures that provide access to information necessary for overseeing the operation of the AI system.[98] These formulations are not incompatible with a logic

[90]Article 13(1) AI Act.

[91]Such an interpretation is consonant with the reliance on technical design measures elsewhere in the EU technology regulation framework, notably in the GDPR: L Jasmontaite and Others, 'Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR' 4 (2018) European Data Protection Law Review 168.

[92]Article 13(3) AI Act.

[93]Ibid., 13(3)(b).

[94]On the absence of a clear-cut disclosure obligation in the AI Act, see B Kuzniacki and Others, 'Towards EXplainable Artificial Intelligence (XAI) in Tax Law: The Need for a Minimum Legal Standard' 14 (2022) World Tax Journal s 4.1.2.

[95]Article 12(1) AI Act. On the role of automated logs for the governance of AI systems, see eg JJ Bryson and A Theodorou, 'How Society Can Maintain Human-Centric Artificial Intelligence' in M Toivonen and E Saari (eds), *Human-Centered Digitalization and Services* (Springer Singapore 2019) 317; D Curtin and M de Goede, 'Bits, Bytes, Searches and Hits: Logging-in Accountability for EU Data-Led Security' in D Curtin and M Catanzariti (eds), *Data at the Boundaries of (European) Law* (Oxford University Press 2022).

[96]Article 12(2) AI Act. On traceability as a tool for human-centric AI, see Bryson and Theodorou (n 95).

[97]Article 14(2) AI Act. Appointing these overseers is a task left to users under Article 29 AI Act.

[98]Ibid., Article 14(3)(a) and 14(4). On the cognitive challenges for oversight and potential biases in human-machine interactions, see eg S Alon-Barkat and M Busuioc, 'Human-AI Interactions in Public Sector Decision-Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice' (2022) Journal of Public Administration, Research and Theory, <https://doi.org/10.1093/jopart/muac007>.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

of disclosure; however, they, too, again, can also be addressed by purely communicative means such as explanation techniques.[99] Once again, the AI Act leaves providers with considerable latitude to provide 'appropriate' transparency without any meaningful disclosure of the system's inner workings.

Finally, the AI Act also includes provisions regarding transparency of certain classes of AI systems towards users. One such use concerns AI systems designed for interacting with natural persons, such as the chatbots often used to automate customer service. Providers of such systems must design them in such a way that informs the natural persons in question that they are interacting with an AI system and not with a human.[100] By mandating disclosure of the artificial character of the system, the AI Act seeks to close opportunities for impersonation and deception,[101] which can be harmful even if the system itself is not used for a high-risk purpose.[102] Nevertheless, disclosure remains limited to the artificial character of the system, leaving providers with free rein regarding what information they will communicate about the system itself.

B. Transparency to the general public

Providers are subject to various, if somewhat shallow, transparency duties to users of certain AI systems. However, the users covered by the AI Act are seldom the only ones put at risk by the operation of an AI system. In fact, in most public sector applications, one could argue there is little risk for users themselves, as the risk resides rather with the citizen(s) affected by the user's deployment of the AI system. User-facing transparency measures such as those described above might be of little use for many people adversely impacted by AI systems.

In fact, one of the main lines of AI Act criticism among scholars and civil society is that individuals are afforded no right to obtain information about systems.[103] Instead, the AI Act requires communication to the public of certain kinds of information about AI systems. Users of emotion recognition and biometric categorization systems are required to disclose the existence of these systems to the people exposed to them,[104] and users must indicate when they use an AI system to generate a deep fake.[105] These communication requirements seek to ensure people will not be fooled into thinking they are dealing with a human rather than a machine. Nevertheless, they do not equip the general population with any mechanisms for obtaining information about the AI systems used for such purposes.

High-risk AI systems are subject to more extensive disclosure requirements. Before a high-risk AI system is placed in the market or put into service, its provider or authorized representative must register it in an EU database of high-risk AI systems.[106] During the registration process, providers or their representatives must provide various types of information,[107] including the

[99]For a pre-AI Act proposal in this direction, see A Bibal and Others, 'Legal Requirements on Explainability in Machine Learning' 29 (2021) Artificial Intelligence and Law 149, s 3.3.

[100]See Article 52(1) AI Act.

[101]Recital 70 AI Act.

[102]For example, one might be more inclined to trust a customer service channel if they believe they are interacting with a human and not with a machine.

[103]See, inter alia, H van Kolfschooten, 'EU Regulation of Artificial Intelligence: Challenges for Patients' Rights' 59 (2022) Common Market Law Review 106; Smuha and others (n 85) s 4.3.1; EDRi and Others, 'An EU Artificial Intelligence Act for Fundamental Rights. A Civil Society Statement' (30 November 2021) s 5 <https://www.accessnow.org/cms/assets/uploads/2021/11/joint-statement-EU-AIA.pdf> accessed 3 December 2021.

[104]Article 52(2) AI Act. This obligation does not apply to systems authorized by law for the detection, prevention, and investigation of criminal offences.

[105]Article 52(3) AI Act. Any use of AI for generating deep fakes must be communicated to the recipients of said fake content, except for lawful uses of deepfakes for law enforcements purposes and situations in which the unlabelled use of deep fakes is necessary to the exercise of the rights to freedom of expression and freedom of the arts and sciences.

[106]Article 51 AI Act. This database is to be maintained by the Commission, in collaboration with the Member States, as specified by Article 60(1) AI Act.

[107]Annex VIII AI Act provides a comprehensive list of the information needed.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

Member States in which the system is or has been placed on the market, put into use, or otherwise made available; the intended purpose of the AI system; information about conformity and certification markings; and an electronic version of the instruction for use mentioned above.[108] Since the registered information is required to be accessible to the public,[109] the database discloses to the public information about the operations of high-risk AI systems.[110]

However, the reach of this disclosure is limited considerably by various aspects of the AI Act. Only high-risk systems that are placed on the market or put into service in the Union market must be registered,[111] and only the provider – or their representative – is obliged to register AI systems into the database.[112] Since many deployments of AI systems are made by users who are not providers of the AI systems they use,[113] the database is likely to be silent about many of the applications of high-risk AI systems in real-world contexts.[114] Even in the cases in which an application is registered into the database, the actual information present in the database offers little disclosure about the AI system itself.[115] Such an approach, in fact, may end up legitimizing high-risk AI applications by offering formal compliance with communicative affairs as an alternative to scrutiny over the system and the purposes for its use.[116] Transparency in the AI database is, therefore – and yet again – a largely communicative affair, in which the general public has access to bits of information selected and curated by the providers of AI systems.

C. Transparency in the governance of AI systems

We are not arguing that there are no deeper requirements for disclosure in the AI Act. Such requirements, however, are formulated in the language of product safety legislation that evolved in a completely different time frame and for very different regulatory objects. Drawing from the New Legislative Framework for harmonized product safety legislation,[117] the AI Act requires providers of high-risk AI systems to demonstrate *ex ante* compliance with the technical requirements laid down for such systems[118] and adopt a post-market monitoring system to follow the risks posed by the system placed on the market.[119] Each of these processes is grounded on information about the AI system and its operation, which means their operation depends on various transparency requirements laid down in the corresponding provisions of the AI Act.

Before placing a high-risk AI system on the market, providers must draw up a written EU declaration of conformity for each system[120] and affix the CE marking to the system or its

[108]Under Point 11 of Annex VIII AI Act, the instructions for use must not be provided for high-risk AI systems in the areas of law enforcement and migration, asylum, and border control management.

[109]Article 60(3) AI Act.

[110]In doing so, the AI Act builds upon previous experiences of public disclosure of AI applications, such as the municipal AI registries adopted by Amsterdam and Helsinki: L Floridi, 'Artificial Intelligence as a Public Service: Learning from Amsterdam and Helsinki' 33 (2020) Philosophy & Technology 541.

[111]Article 51(1) AI Act.

[112]Users would only be obliged to register AI systems if they used the system for a new purpose or substantially modified the system: Article 28 AI Act.

[113]J Cobbe and J Singh, 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges' 42 (2021) Computer Law & Security Review 105573.

[114]Amendment 172 of the Draft Report of the European Parliament proposes that users must register high-risk AI systems used by public authorities or EU institutions, bodies, and agencies – or on behalf of such entities. So far, no such requirement has been proposed for uses of high-risk AI in the private sector.

[115]See the discussion on instructions of use above.

[116]See, eg C Cath and F Jansen, 'Dutch Comfort: The Limits of AI Governance through Municipal Registers' [forthcoming] *Techné: Research in Philosophy and Technology*.

[117]AI Act, Proposal (n 77) 14.

[118]Article 19 AI Act. On the conformity assessment procedure, see Article 43 AI Act.

[119]Article 61 AI Act.

[120]Article 48 AI Act.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

documentation.[121] Such conformity statements can usually be obtained without disclosing information to entities beyond the controller, as the AI Act determines that most systems can be subject to internal controls for conformity.[122] Even if the provider must rely on a third-party assessment for conformity of a given AI system,[123] this third party is bound by secrecy requirements, which restrict further disclosure of the information obtained as part of a conformity assessment.[124] As a result, the *ex ante* mechanisms of the AI Act provide a clear example of transparency seen through a logic of communication.

Communication is also an important element in post-market monitoring. Under Article 61 AI Act, providers are required to 'actively and systematically' collect, document, and analyse data about the performance of the AI systems they provide, in order to evaluate compliance with the requirements imposed upon high-risk AI systems.[125] In particular, providers are obliged to report serious incidents and malfunctions of a high-risk AI system,[126] as well as situations in which the use of the system may present risks[127]; for example, because of its application to a new purpose or safety failures discovered after the system was placed on the market. Providers thus act as an information nexus for AI systems placed on the market or put into service, collecting information from users and other sources,[128] analysing it and reporting potential issues to the relevant authorities.

Post-market enforcement of the AI Act rests upon market surveillance authorities (MSAs).[129] For AI systems used by EU institutions, agencies, and bodies, this role is to be played by the European Data Protection Supervisor.[130] The scenario is somewhat more fragmented at the national level: as a rule, national supervisory authorities also play market surveillance roles,[131] but systems subject to other instruments of EU product harmonization legislation or used by financial institutions regulated by EU financial services legislation are subject to the corresponding supervisory authorities.[132] In addition, market surveillance for certain categories of high-risk AI systems used for law enforcement or immigration, asylum, and border control management purposes falls either to the data protection authority or the authority supervising the sector.[133] MSAs are expected to cooperate to pursue their shared goal of carrying out activities and taking

[121]Article 49 AI Act.

[122]Under Article 43(2) AI Act, most high-risk systems are subject to the conformity assessment procedure defined in Annex VI AI Act, which is based solely on internal controls.

[123]External controls may be required for some of the applications listed in Annex III AI Act (Article 43(1) AI Act) or if the system is required to undergo third-party conformity assessments as result of other applicable legislation on product safety.

[124]Article 33(5) AI Act establishes confidentiality requirements regarding any information disclosed during certification procedures.

[125]Article 61(2) AI Act. Information may be obtained from users, who are obliged under Article 29(4) AI Act to disclose information to providers about serious malfunctions or incidents in system operation, as well as about any situations in which the systems present risks even if used in accordance with the instructions of use.

[126]Article 62 AI Act. Under Article 29(4) AI Act, users of high-risk AI systems are required to communicate any serious incident or malfunction to providers, and Amendment 227 of the Draft Report from the Parliament seeks to extend this reporting duty to include communication to the relevant authorities.

[127]Article 22 AI Act. Under Article 29(4) AI Act, users of high-risk AI systems must inform the provider whenever they find out that using the system according to instructions may present risks in the sense of Article 65(1) AI Act.

[128]Such as publicly available information about the systems, as well as system logs produced in compliance with Article 12 AI Act, to the extent that they remain under control of the provider.

[129]Article 63(1) AI Act.

[130]Article 63(6) AI Act.

[131]As specified in Article 59(1) AI Act, national supervisory authorities are authorities designated or established for ensuring the application and implementation of the Act. Member States are required under Article 59(3) AI Act to disclose the authority – or potentially authorities – designed for that purpose. As of this moment, no such authorities have been formally designated, but the use of the European Data Protection Supervisor as the supervisory authority for EU systems suggests data protection authorities are likely to play similar roles at the national level.

[132]Article 63(3–4) AI Act.

[133]Member States must specify which one: Article 63(5) AI Act.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

measures to ensure AI systems comply with the harmonization requirements, as many AI systems are likely to be used in more than one Member State.[134]

MSAs need considerable amounts of information to adequately carry out their duties. In part, information is supplied by providers as they discharge their reporting duties outlined above. However, MSAs also have access to mechanisms for proactively obtaining information. These authorities can request access to technical documentation about high-risk AI systems,[135] which providers must draw up and keep updated with the information needed to demonstrate compliance with the requirements of the AI Act.[136] In addition, MSAs have the power to organize technical tests of the high-risk AI system to detect potential breaches of obligations under Union law intended to protect fundamental rights.[137] These provisions require providers – and users, where applicable – to supply substantial amounts of information to MSAs, either spontaneously or upon demand by the authority. Nevertheless, these provisions still follow largely a logic of communication, as the access of the MSA to the system is mediated by the content of the reports and technical documentation or by the reported results of software tests.

Some mechanisms available to MSAs offer the potential of unmediated disclosure of AI systems. Upon a reasoned request, these authorities can obtain access to the source code of a high-risk AI system to the extent such access is needed to assess the conformity of the high-risk AI system with AI Act requirements.[138] MSAs can also obtain access to the data used in the training process of a high-risk AI system,[139] a provision that is particularly relevant for the analysis of machine learning systems. These mechanisms force providers – and users, where applicable – to disclose information about the inner workings of an AI system.

Based on these provisions, MSAs can scrutinize high-risk AI systems from angles unavailable to the general population or even to most users of high-risk AI systems, introducing elements of disclosure into a communication-driven approach to transparency. Yet, the actual impact of these disclosure instruments might not be enough to ensure the widespread transparency expected of AI systems. In the absence of independent fire alarms, such as those provided by civil society activists and complaints to authorities from concerned individuals, the MSAs' attention is likely to be directed only to these issues brought to their attention by provider reports.

Transparency is further limited by the secrecy requirements surrounding the activities of MSAs. This is an important limitation. Under Article 70 AI Act, these authorities must respect the confidentiality of information and data they obtain as they perform their duties. This provision affords special protection to certain interests, such as national and public security[140] and intellectual property rights and trade secrets,[141] which often clash with the public interests promoted by transparency in AI systems and elsewhere.[142] The confidentiality requirement from Article 70 AI Act does not create formal limits to the information-gathering powers conferred to the MSAs.

[134]Article 63(7) AI Act. The Draft Report from the European Parliament (prepared by the IMCO and LIBE committees) includes a large number of amendments that ascribe enforcement powers to the European Commission in cases of widespread infringements of the Act, covering at least three Member States.

[135]Article 50(a) AI Act. Article 64(3) AI Act extends the right to request access to technical documentation to national public authorities or bodies supervising or enforcing 'the respect of obligations under Union law protecting fundamental rights in relation to the use of high-risk AI systems'. These authorities or bodies must inform the relevant MSA in case of any such request.

[136]Article 19 AI Act obliges providers to draw up technical documentation, which must include the information listed in Annex IV AI Act.

[137]Article 64(5) AI Act.

[138]Article 64(2) AI Act.

[139]Article 64(1) AI Act.

[140]Article 70(1)(c) AI Act.

[141]Article 70(1)(a) AI Act.

[142]On public interest as a limitation to trade secret privileges, see K Foss-Solbrekk and AK Glenster, 'The Intersection of Data Protection Rights and Trade Secret Privileges in "Algorithmic Transparency"' in E Kosta and R Leenes (eds), *Research Handbook on EU Data Protection Law* (Edward Elgar Publishing 2022) 163–183.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

But it nonetheless creates obstacles to these authorities as they seek to obtain and disclose information about AI.

Part of this comes from the introduction of upstream opacity, as providers and users might be subject to confidentiality requirements that restrict the kinds of information they can communicate to third parties.[143] But confidentiality requirements also reduce downstream flows of information, as the MSAs are constrained in their ability to share information with third parties that might otherwise contribute to the governance framework,[144] such as civil society watchdogs. In particular, access to information on high-risk AI systems used in law enforcement and asylum, immigration, and border control management is subject to various constraints: direct access to their documentation is only possible to staff members of the MSAs holding the appropriate level of security clearance.[145] Furthermore, any mediated disclosure of information by the MSA in such domains of application requires prior consultation of the user of the AI system – that is, the public body deploying the system.[146] While such a requirement is in line with the 'originator control' principle that is prevalent in these sectors,[147] it nevertheless introduces friction in the cooperation between MSAs and considerably narrows down the possibilities for subjecting this information to the scrutiny of actors other than the MSAs directly involved with the system. These upstream and downstream effects of opacity ensure that providers and certain types of users – particularly public sector users of high-risk AI systems for law enforcement and asylum, migration, and border control management – remain in the driver's seat of transparency under the AI Act. Disclosure, once again, can only go where communication allows it to.

5. Reclaiming transparency

In what follows, and building on the above, we zoom in on what we regard as three major deficits undermining the promise of transparency and its potential to facilitate accountability and open opportunities for oversight.

A. The limits of explanation

Overall, the information ecosystem foreseen in the AI Act is tilted towards protecting the interests of providers and users at the expense of the individuals and the overall accountability system. As we saw above, key provisions on transparency target *obligations of providers towards users* – enshrined in Article 13, under the heading 'transparency and provision of information to users' – rather than to individuals affected by these systems.[148] This is likely a by-product of the existing models of product liability that informed the approach taken by the Act as well as its legal basis in the internal market. The result is a curious and unsatisfactory halfway house: while the AI Act is meant, *inter alia*, to govern the use of AI in the public sector, and while it touts protecting public

[143]See, eg the provisions in the German Tax Code preventing the publication of information about the risk management system used by tax authorities (Section 88(5) of the *Abgabenordnung*). While such a prohibition might not be directly opposable to the relevant MSA, it introduces friction in the flows of information between the users of the system and the providers of its components.

[144]Article 70(3) AI Act ensures such constraints do not affect the exchange of information and warnings between Member States, the Commission, and certification bodies, or obligations of these parties to provide information under Member State criminal law. Article 70(4) AI Act allows the Commission and Member States to share confidential information with regulatory agencies of third countries, so long as such exchange is necessary and covered by arrangements providing an adequate level of confidentiality.

[145]Article 70(2) AI Act.

[146]Article 70(2) AI Act, which also establishes the need to consult the originating national competent authority, that is, the authority that obtained information from the user.

[147]Curtin (n 24) 853. See, eg Article 22 Europol Regulation (Regulation (EU) 2016/794 of the European Parliament and of the Council on the European Union Agency for Law Enforcement Cooperation (OJ L 135/53)).

[148]See also, Smuha and Others (n 85).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

values such as fundamental rights, its transparency measures, geared towards product safety risks, offer little remedy for the unique risks the use of AI by public authorities imposes upon individuals affected by such deployments: Those adversely affected by the system, and the citizenry at large, are mostly forgotten when it comes to information and disclosure rights.[149] The shift from a logic of disclosure to a logic of communication thus erodes the role of transparency in the oversight of public sector decision-making, as the public is deprived of access to any sources of information about AI systems other than those heavily mediated by the very actors against which transparency should provide a safeguard.

What is more, the disclosure *obligations of providers (to users*), too, while the more elaborate set of information obligations in the draft Act, remain *overly underspecified*. Requirements to ensure that the operation of high-risk AI systems they develop are 'sufficiently transparent' to allow for user interpretation of output and appropriate use, ensuring an 'appropriate type and degree of transparency'[150] leave discretion to providers to make key transparency choices. With conformity a matter of provider self-assessment, what exactly 'sufficiently transparent' for interpretation or an 'appropriate type and degree of transparency' entails, becomes a matter of provider assessment and ultimately, provider choice. Moreover, while providers are mandated to build high-risk AI systems for human oversight,[151] the draft Act leaves the measures to be put in place to facilitate interpretation at provider discretion, and 'as appropriate to the circumstances'.

This will likely perpetuate below par transparency practices opening the door to providers to stack the deck in their favour and provide little insight as to actual system functioning. For instance, popular explanation techniques of black boxes that have received a lot of press as potential transparency 'fixes' – such as counterfactual explanation models[152] – allow for multiple explanations: when applied to the same model, under the same circumstances, different techniques for XAI (or in fact, the same explanation technique deployed by a different developer) may highlight different attributes as being crucial for a particular decision. Since any mathematical model is built upon procedures of abstraction, idealization, and approximation, choices on which elements to abstract, idealize, or approximate afford considerable discretion to the builder of the explanation model: '[. . .] left to their own devices, decision makers are afforded a remarkable degree of power to pursue their own welfare through these choices'.[153] As a result, the provider of an explanation – which is usually conflated with the provider of the black-box model being explained in the first place – has considerable leeway to decide which features to disclose in the explanation, opening up opportunities for gaming by providers to self-serving ends: 'while designed to restore power to decision-subjects, partial explanations grant a new kind of power to the decision maker, to use for good or abuse as desired'.[154]

Relatedly, an often overlooked aspect with respect to explanation models of black boxes more generally is that they represent 'approximations' of underlying models rather than faithful renditions of their logic.[155] An explanation model is not merely a simplified version of the original model obtained by removing irrelevant factors; instead, it is a different model that approximates the behaviour of the black box under study, and, in doing so, might ascribe different weights to the features used in the black-box decision, or even use different features or attributes altogether.[156]

[149]Indeed, it has been argued that the AI Act introduces a filter that dismisses public interests that cannot be easily described in product safety terms: M Almada and N Petit, *The Primrose Path to AI Regulation: Combinatorial Troubles and Path Dependence in the AI Act* (Governing Artificial Intelligence: Designing Legal and Regulatory Responses, Dublin, 3 June 2022).

[150]See Article 13 AI Act.

[151]See Ibid., Article 14.

[152]Wachter, Mittelstadt and Russell (n 18).

[153]Barocas, Selbst and Raghavan (n 72) 87. On the various degrees of freedom available to the designer of an explanation model, see Bordt and Others (n 72) s 4.

[154]Barocas, Selbst, and Raghavan (n 72).

[155]Rudin (n 67).

[156]Ribeiro, Singh and Guestrin (n 69).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

In the words of leading computer scientist Cynthia Rudin in her influential *Nature* article warning against the use of such techniques in high-stakes decision-making:

> Explainable ML methods provide explanations that are not faithful to what the original model computes. (. . .) They cannot have perfect fidelity with respect to the original model. If the explanation was completely faithful to what the original model computes, the explanation would equal the original model, and one would not need the original model in the first place, only the explanation.[157]

While post hoc explanation techniques can be valuable for developers and system engineers 'as part of the knowledge discovery process'[158] to investigate model behaviour, explanations are often partial or incomplete, failing to provide sufficient information to fully understand what the underlying black box is doing.[159] There is a real risk, therefore, that explanations – even when produced in good faith – amount to partial and as such misleading descriptions of model functioning,[160] rendering them of limited use as tools of oversight from a legal perspective. They cannot provide decision subjects the insights needed to subsequently take up and contest specific decisions and demand accountability.

Such concerns are further echoed by work on 'misleading explanations' showing how user trust can be manipulated through unreliable explanations that do not accurately reflect the issues and biases in the underlying black box.[161] What is more, recent computer science work also illustrates how explanation models can be purposefully manipulated by their owners to hide biases from external auditors, further speaking to the vulnerabilities and shortcomings of such methods.[162] These insights perfectly mirror concerns with transparency as communication more broadly: its potential for influence, placing the object rather than the beholder in control to shape what is made visible, and in doing so, to reshape the parameters of its own transparency and accountability.

Given these concerns, leaving the technical measures to be put in place to afford interpretation of system's output and transparency up to the providers of AI systems, as the draft regulation does, is highly problematic. If model functioning is not actually transparent, diagnosing failure and enacting accountability becomes next to impossible. The absence of robust and unambiguous legal standards as to what measures are required to afford meaningful interpretation of system outputs is a serious shortcoming that will detract from effective transparency and accountability in this area.

B. User disclosure duties as public authorities: on the need to incentivise responsible user behaviour

Users of high-risk AI systems, while recipients of information from providers as detailed in Section 4, are themselves in turn subject to perfunctory, token obligations – primarily amounting to complying with a set of instructions of use provided by providers with no meaningful transparency duties of their own under the Act. In the context of public sector applications, these duties are particularly insufficient, as they fail to attend to the specific needs of transparency in the activities of public authorities, again likely the result of the product safety and market focus reflected in

[157]Rudin (n 67) 207.

[158]Ibid.

[159]Rudin (n 67) 208.

[160]C Rudin, C Wang and B Coker, 'The Age of Secrecy and Unfairness in Recidivism Prediction' 2 (2020) Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/7z10o269/release/5> accessed 1 May 2022.

[161]Lakkaraju and Bastani (n 72) 79.

[162]D Slack and Others, 'Counterfactual Explanations Can Be Manipulated', in *Advances in Neural Information Processing Systems* 34 (2021).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

the Act's legal basis. As we saw in the Dutch context and the *toeslagenaffaire*, uses of AI systems by public authorities can bring about fundamental, life-changing implications for individuals. The lack of disclosure duties *of users* towards those adversely affected by the systems they deploy is disconcerting given the growing trend visible across public sectors to rely on algorithmic systems managed by external suppliers to implement consequential public tasks, often subject to secrecy provisions as to these systems' set-up and functioning. As they do so, public bodies are effectively abdicating consequential public tasks to private providers with the prospect of little or no oversight of system functioning.[163] In this context, the failure to include meaningful information duties for public authorities as a particular class of users within the scope of the legal act that is to regulate the EU-wide use and deployment of AI (also in the public sector) for years – if not decades – to come is a glaring omission.

The absence of safeguards on this will lead to the continued adoption of AI models that preclude scrutiny of their functioning. It lets users of high-risk AI systems 'off the hook' and encourages them – given that they lack disclosure duties of their own towards those affected by their systems – to continue to purchase secret and proprietary models. This is disconcerting in a context where public authorities are found to rely on overly-complex, proprietary models to implement public tasks *even when* non-proprietary alternatives of equivalent performance exist or can be developed.[164] Reliance on proprietary models entails that algorithm design, methodology and functioning will remain hidden to the public bodies that deploy them, to individuals adversely affected by decisions informed by them and/or to citizens, exacerbating information asymmetries vis-à-vis private providers and removing opportunities for meaningful oversight. For instance, a report by the Netherlands Court of Auditors noted that public bodies have little insight into the quality of algorithms they rely on when these are managed by external providers:

> Ministries that have outsourced the development and management of algorithms have only a limited knowledge of these algorithms. (. . .) the responsible minister does not have any information on the quality of the algorithm in question nor on the documents underlying compliance with the relevant standards and refers to the supplier instead.[165]

By contrast, requiring strict transparency duties of users such as public authorities will encourage them to be demanding users and consumers of AI and to in turn push high standards upstream, demanding openness and high transparency standards of providers. For instance, rather than purchasing proprietary AI models, public authorities could contractually require providers to forgo proprietary protections to be able to sell high-risk AI systems into the public sector.[166] Where transparency cannot be secured in this context, serious consideration should be given to whether the use of such systems in high-stakes public sector is justified. In this vein, some MEPs have already proposed that the deployment of AI systems should be subject to fundamental rights impact assessments.[167]

C. Gagging accountability: absent or silenced 'fire alarms'

In terms of enforcement, while the AI Act affords in principle considerable information disclosure to national competent authorities (MSAs) in their enforcement roles, in the absence of independent fire alarms, providers hold the 'chokehold' on monitoring and whether public regulators become 'activated' in their enforcement roles to begin with. This is because providers themselves

[163]See Finck (n 17); Busuioc (n 3).

[164]For multiple examples of this in a public sector context, see Rudin (n 67).

[165]Algemene Rekenkamer, *Understanding Algorithms* (Netherlands Court of Audit 2021) 42 <https://english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms> accessed 3 May 2022.

[166]Rudin (n 67).

[167]Heikkilä (n 5).

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

are largely responsible for carrying out post-market monitoring. As discussed above, market supervisory authorities rely on *provider notification* of incidents and malfunctioning and there is no mechanism provided for individuals adversely affected to lodge a complaint with market surveillance authorities.[168] This is a significant omission from an accountability perspective, as the enforcement system lacks much-needed 'fire alarms' providing independent feedback on system (mal)functioning. Users of AI systems, too, while they are to monitor and play a role in flagging malfunctioning systems, they do so vis-à-vis *the provider* (rather than directly to regulators). Such a setup, by concentrating post-market monitoring responsibilities with providers, simultaneously creates a closed information circuit with providers as *central nodes* controlling access to information on system functioning. This exacerbates a heavy dependence on providers doing their due diligence in monitoring and indeed, on them truthfully reporting on their models' malfunctioning. To know even where to begin to look, regulators need to rely on providers – a dependency which stands to aggravate already significant informational deficits. Such deficits are likely to be further compounded by anticipated resources and expertise shortages of MSAs themselves, which are likely to be significant,[169] especially so in light of the complexity, technical nature and the sheer size of the regulatory domain. These constraints are particularly salient in cases in which supervisory authority is ascribed to under-specialized regulators (remember that the AI Act does not foresee the creation of purpose-specific regulators), which might lack the resources to cultivate extensive in-house expertise in AI, already struggling to keep up with existing responsibilities (as plainly seen with some national data protection authorities).[170]

Even in the cases when an MSA has already obtained access to aspects of the AI system – for example, access to training data is not conditioned upon a reasonable request – rendering the system transparent actually requires considerable technical work,[171] which resource-constrained authorities are unlikely to perform in the absence of reasons that warrant attention to that particular system. Consequently, even the unmediated disclosure tools available to MSAs depend upon communication practices by providers and users of high-risk AI systems.

Given this informational dependence, MSAs' enforcement will likely be geared towards issues flagged by the provider, especially given envisaged capacity constraints to independently monitor and police patrol the crowded domains under their purview. This regulator informational dependence is particularly problematic given stakes shaping provider disclosure incentives – dominant private actors in this area have been found to go to great lengths to quash internal dissent and prevent the publication of internal research evidencing harm.[172]

Importantly, it is also unclear to what extent providers will even be able to exercise such post-marketing monitoring roles meaningfully vis-à-vis some categories of users such as public authorities in secrecy-dominated areas (such as law enforcement or asylum, as noted above). Adequate monitoring and enforcement will likely be absent precisely in the highly sensitive domains where

[168]See the proposal in the IMCO–LIBE draft report to provide natural persons with a right to lodge a complaint with the national supervisory authority aiming to speak precisely to such concerns.

[169]Initial estimates from the European Commission estimate each Member State would need from 1 to 25 full-time staff for enforcing the AI Act: European Commission (n 77) 12. Such estimates have been criticized as overly optimistic: M Veale and FZ Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' 22 (2021) Computer Law Review International 97; Smuha and Others (n 85). The proposed increase in Commission responsibilities in the IMCO–LIBE Draft Report also raises questions regarding the availability of resources for AI Act enforcement at the EU level, as the Commission also plays a central role in the enforcement of the Digital Markets Act and Digital Services Act.

[170]See, *inter alia*, J Ryan and A Toner, *Europe's Enforcement Paralysis. ICCL's 2021 Report on the Enforcement Capabilities of Data Protection Authorities* (Irish Council for Civil Liberties 2021).

[171]See, eg B Allen, 'Source Code Isn't' in TS Mullaney et al. (eds), *Your Computer Is on Fire* (The MIT Press 2021) 273–95.

[172]Google, for instance, fired the co-leads of its own AI ethics group over a critical paper on the risks of large language models, key to the company's business model, which Google tried to have retracted and prevent from getting published: T Simonite, 'What Really Happened When Google Ousted Timnit Gebru' (*Wired*, 8 June 2021) <https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/> accessed 3 May 2022.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

they are most needed, given the sensitive nature of the data and state secrecy protections characteristic of these fields.[173]

What is more, while market supervisory authorities (once activated) have the possibility for extensive access to information, they are simultaneously bound, as we saw above, by confidentiality clauses. As a result, there is little room for this information to feed back into much needed cross-border regulatory coordination in this area (deepening fragmentation), or in fact, into public discussion or debate on this information, preventing MSAs, in turn, from fulfilling their own fire alarm roles towards other public forums (such as parliaments, other national regulators, courts, civil society watchdogs, or the media). As a result, there is little room for public discussion or debate on this information or opportunities for other forums to take it up further. This in turn removes the raw ingredients for accountability – information and deliberation – from public space. Thus, due to both upstream and downstream opacity, accountability is depleted of the very lifeblood that sustains it: transparency.

6. Conclusions: seeing accountability

There is no accountability in the AI Act – only governance, and that governance model is inspired largely by a very different context (product liability) and era (largely physical goods). But when the users (not the end users!) of high-risk AI systems are to a significant extent public authorities such as the Dutch tax authorities in our opening example (or law enforcement and border guards) then this narrow product-based approach is clearly insufficient and not tailored to the reality of use by public authorities of all kinds at the EU level itself but also very much at the national level too.

To sustain accountability, there is a need to first and foremost *reclaim transparency*. The reframing of transparency as explanation in AI debates has moved the goalposts away from the core meaning of transparency, opening it up for influence. In our contribution, we challenge accounts that deny the value of disclosure and reduce transparency to predigested explanations in legal and regulatory contexts, leading to an exponential growth in secrecy. These reworked logics have now also bled into the draft AI Act's approach to transparency, which to a large extent restricts transparency to a communicative affair. This is problematic as the currency of accountability is information – the acquisition of reliable information by accountability forums is critical to 'jumpstart' accountability processes. Despite their close pedigree, communication is not the same as transparency. While effective communication is important for public bodies for instance, to get their message out, to convey accessible public information to citizens about their work and services, to build and rally support to sustain their authority, it ultimately amounts to the provision of curated and redacted information, mediated from a specific institutional or organizational perspective, and shaped and advanced in view of such interests. This only becomes reinforced when the narrators are private actors in nature with strong vested interests in the AI products they sell. To the extent that the narratives they advance are partial, incomplete, or simply inaccurate, they can only be of questionable value for meaningful algorithmic transparency and accountability.

Several solutions are paramount in our salvage effort. First, we argue that unambiguous disclosure responsibilities must be placed on users of high-risk AI systems, especially public authorities, of whom as citizens we are entitled to expect and demand high standards. This will encourage them to give adequate consideration as to whether reliance on high-risk AI models is truly justified in specific circumstances – pushing them to become discerning users of AI systems – as well as to make strong transparency demands, in turn, of providers and of the models they purchase. It will encourage public sector users of high-risk AI systems to purchase tools that are compatible with the public nature of their roles and that allow public administrators to discharge their continued responsibilities to citizens as well as to afford scrutiny. Recurrent scandals involving AI algorithm deployment for public tasks, including models deployed in highly sensitive areas without being

[173]Smuha and Others (n 85) 39.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

checked for bias,[174] failure to meet basic requirements,[175] discrimination and widespread harm through the deployment of sub-optimal algorithmic systems on already vulnerable citizens[176] or the deployment of non-transparent models over transparent alternatives of equivalent performance,[177] demonstrate that this duty of care is often not undertaken.

Second, intellectual property and secrecy considerations cannot trump crucial public values considerations in critical domains: administrative duties (duties to give reason and disclosure obligations), our ability to exercise public accountability, or individual rights considerations. Reliance on proprietary models entails that algorithm set-up, methodology, and functioning will remain undisclosed to public bodies, to those adversely affected by decisions informed by it and/or to citizens, removing any opportunities for scrutiny and meaningful oversight. This is unacceptable in a context where AI informed decision-making is making inroads into highly consequential public domains. To be able to comply with their transparency and reason-giving duties, public authorities can – and should be required to – contractually require private providers to forgo proprietary protections to sell into the public sector. This could be secured through outright purchasing models from providers (rather than leasing them) to address proprietary concerns, or, at the very least, using procurement processes to set up contractual transparency and open access standards.[178] Alternatively, models can be developed in-house by public authorities.[179] This will become critical to public administrators' ability not only to exercise meaningful control of third-party systems they rely on in their decision-making but also to them being able to comply with their duty to give reasons for administrative decision-making when algorithms are used in this context. However, in-house development must be accompanied by a re-assessment and relinquishment of various legal forms of secrecy used to restrict access to public sector algorithms,[180] lest one end up exchanging private secrecy for public secrecy with no gains in visibility for society as a whole.

Third, it becomes necessary to critically assess upfront the use of opaque techniques for achieving the purposes for which AI is deployed. While it has become commonplace in the literature to argue there is a trade-off between transparency and model efficiency,[181] such trade-offs are not necessarily present, as inherently interpretable models can achieve the outcomes needed

[174]For instance, London Metropolitan Police trialled live facial recognition in 2017 without checking it for bias: H Margetts and C Dorobantu, 'Rethink Government with AI' 568 (2019) Nature 163. See also K Peachey, 'Post Office Scandal: What the Horizon Saga Is All About' (*BBC News*, 23 April 2021) <https://www.bbc.com/news/business-56718036> accessed 13 June 2021; Algemene Rekenkamer (n 165).

[175]The Netherlands Court of Audit recently found that 6 out of 9 audited algorithms used in the Dutch public sector do not meet basic requirements with inadequate checks on performance, bias, unauthorized access, data leaks. Algemene Rekenkamer, *An Audit of Algorithms* <https://english.rekenkamer.nl/publications/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government> accessed 18 October 2022.

[176]See for instance, Heikkilä (n 5); Geiger (n 5); L Amoore, 'Why "Ditch the Algorithm" Is the Future of Political Protest' (*The Guardian*, 19 August 2020) <http://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics> accessed 20 December 2020; V Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (First edition, St Martin's Press 2018); R Xenidis, 'Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience' 27 (2020) Maastricht Journal of European and Comparative Law 736.

[177]Rudin (n 67).

[178]LM Ben Dor and C Coglianese, 'Procurement as AI Governance' 2 (2021) IEEE Transactions on Technology and Society 192; J Raso, 'AI and Administrative Law' in F Martin-Bariteanu and T Scassa (eds), *Artificial Intelligence and the Law in Canada* (LexisNexis 2021) 181.

[179]In jurisdictions such as the US and Brazil, in-house development is reportedly more common than outsourcing as an approach for developing public-sector AI systems: RC de Fassio and C Langevin, *Unpacking AI Procurement in a Box: Insights from Implementation* (World Economic Forum 2022) 6. No such numbers are readily available for the EU or its Member States, but the possibility of software development by administrative bodies should not be dismissed out of hand.

[180]See, eg the provisions in the German Tax Code preventing disclosure of information on the system used for fiscal risk assessment (n 143).

[181]See, eg P Hacker and Others, 'Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges' 28 (2020) Artificial Intelligence and Law 415, 431.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

for many given public sector applications.[182] Multiple scholars have shown how inherently transparent models of equivalent performance to black-box ones, used for instance, for consequential and problematic applications like criminal justice risk assessment, can be developed – demonstrating that the reliance on black-box models in these domains by public authorities is unwarranted[183] (even from a performance standpoint alone[184]): 'in criminal justice, there is no evidence of a loss in predictive accuracy for using a transparent model'.[185] In fact, while inherently interpretable models require strong expertise to develop, they are easier to assess and troubleshoot for false assumptions, inaccuracies and bias as well as transparent from the outset[186] – yet another powerful argument advocating for their reliance in public sector contexts over black box alternatives. While there are, indeed, specific domains in which opaque models clearly outperform transparent models available so far,[187] in many cases public sector applications can be met by interpretable models of suitable performance while remaining transparent and preserving other relevant safeguards, giving us hope that 'seeing' and 'understanding' need not be a zero-sum game in the age of automation.

Finally, even if the use of an opaque model turns out to be unavoidable, it should be accompanied by suitable safeguards along the lines discussed above, such as the disclosure of their source code for public scrutiny through an open-source model, a practice that is now taken up even by some industry actors on large black-box models.[188] Disclosure matters: the source code, choices made during the design and training – all these are part and parcel of a model's validity and reliability, and whether such a model can and should be deployed. In contrast, opaque models allow errors to remain undetected and proliferate. Take, for instance, COMPAS, the proprietary risk recidivism algorithm widely used in US criminal justice and flagged for racial bias by ProPublica. Despite being recurrently prodded, imperfectly reversed engineered and partially reconstructed by computational journalists and computer scientists alike, the inescapable conclusion remains in the absence of actual model disclosure: 'COMPAS may still be biased, but we can't tell'.[189] Ultimately, as some computer scientists are increasingly warning, 'this lack of transparency is precisely what allows errors to propagate and results in damage to society (. . .) Merely being able to explain black box models is not sufficient to resolve this – the models need to be fully transparent'.[190]

Our aim is by no means to disavow the value of accessible information on model functioning. But rather to point out that disclosure – seeing *inside* the box – may in fact assist understanding and serves a critical verification and validation function that the logic of communication alone

[182]See, very emphatically on this point, Rudin (n 67).

[183]See, *inter alia*, N Tollenaar and PGM van der Heijden, 'Which Method Predicts Recidivism Best? A Comparison of Statistical, Machine Learning and Data Mining Predictive Models' 176 (2013) Journal of the Royal Statistical Society. Series A (Statistics in Society) 565; C Rudin and B Ustun, 'Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice' 48 (2018) Interfaces 449; J Dressel and H Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' 4 (2018) Science Advances eaao5580.

[184]That being the case, it is questionable that suitability should in fact be assessed from a performance standpoint alone. Optimizing exclusively for performance, especially if small performance gains from an opaque model come at the price of fundamental rights or compromise the ability of domain experts to understand the models that they rely on in critical public sector domains, would be unsatisfactory. See, eg P Ohm, 'Throttling Machine Learning', in M Hildebrandt and K O'Hara (eds), *Life and the Law in the Era of Data-Driven Agency* (Edward Elgar 2020) 214–29.

[185]Rudin, Wang and Coker (n 160) 35.

[186]Rudin (n 67) 207, 208.

[187]For example, current advances in natural language processing draw heavily from large language models of substantial complexity, which cannot be reduced without considerable distortion of model operation.

[188]For an example in the context of the disclosure of a large language model, see: 'Meta AI Is Sharing OPT-175B, the First 175-Billion-Parameter Language Model to Be Made Available to the Broader AI Research Community' <https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/> accessed 16 May 2022.

[189]S Corbett-Davies and Others, 'A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear.' (*Washington Post*, 17 October 2016) <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> accessed 6 June 2022.

[190]Rudin, Wang and Coker (n 160) 35.

https://doi.org/10.1017/elo.2022.47 Published online by Cambridge University Press

cannot realise. Reasoned decision-making is, and remains, a core pillar of administrative law, which bears directly on the permissible uses of technology by governments. Yet, exchanging transparency for provider-mediated explanations in the absence of any external check on the validity of these provider-produced accounts is a transparency sleight-of-hand: 'enhanced' understanding of potentially fabricated (or partial) accounts does not transparency make.

We repeat that our aim as lawyers, accountability scholars and a computer scientist is, at a crucial moment in the development of AI regulation (it happens to be in Europe as the first attempt), to critically focus on an aspect that could still be overlooked in the final versions[191]: the rescue of disclosure within transparency as something distinct from explanation and other forms of communication as well as the corresponding relevance and primacy of secrecy obligations of various kinds gagging the future practice of accountability. The subject matter is no less than the accountability of contemporary and future public administration – in Europe, nationally and maybe at some point globally. Yet, the way this is being dealt with is modelled on product liability and governance from some decades back. It is not the purpose of our article to ask for a different legal basis or a more fundamental approach (although arguments can and have certainly been made in this regard[192]) but to zoom in on a core issue that is very much debated in the AI context: transparency not as access to raw ingredients but transparency as a heavily mediated phenomenon orchestrated and controlled by provider communication with really no supervision at all on this point. It is the mirage of transparent and reasoned decision-making rather than its actual manifestation. Without spelling it out in so many words in a consistent manner this is what the draft AI Act effectively currently does, and it may well slip in under the radar as representing what seems to be a premature and precipitated consensus on what transparency should (and even must) mean in the AI context.

We seek to reclaim transparency back to an original core meaning, in any event to some independent public supervisory authorities (bolstered by substantial powers, expertise and resources) and in some instances also publicly through public accountability forums or otherwise. We thus seek to prevent the idea and reality of transparency from being appropriated by those who give it a narrow, discretionary, and exclusively mediated substance, reinforcing self-serving industry narratives that effectively hollow out transparency. In so doing, our interest is neither dogma nor doctrinal but simply to reconsider the point of it, the means, and the audiences it speaks to. At the same time, we ask to see transparency as nonetheless the essential foundation stone for deeper reflections on the meaning of public accountability. When the users are in fact public authorities operating and using high-risk systems in areas hugely sensitive for citizens and their lives, the urgency and saliency is acute. The citizen is after all not a dataset.[193]

Acknowledgements. The authors wish to thank the participants of the Faculty Seminar held at the European University Institute (EUI) in Florence on 11 May 2022, in particular Nicolas Petit, Martijn Hesselink, Vigjilenca Abazi and Sarah Tas.

Competing interests. The authors have no conflicts of interest to declare.

[191]Not only the JURI opinion includes different proposals on what transparency is and what it is meant to achieve, but civil society proposals also push back against the construction of transparency in the AI Act proposal. But there is very little discussion on this in the various Council Presidency drafts that are being leaked rather than being shared. For an overview not provided by the EU, see K Zenner, 'Documents and Timelines: The Artificial Intelligence Act (Part 3)' (*Digitizing Europe*, 14 May 2022) <https://www.kaizenner.eu/post/aiact-part3> accessed 13 June 2022.

[192]See, inter alia, Edwards (n 85).

[193]This is the title of a report by the Dutch national ombudsman: E Govers and Others, *The Citizen Is Not a Dataset. Vision on the Appropriate Use of Data and Algorithms by Public Sector Authorities* (Nationale Ombudsman 2021).

Cite this article: Busuioc M, Curtin D, and Almada M. Reclaiming transparency: contesting the logics of secrecy within the AI Act. *European Law Open*. https://doi.org/10.1017/elo.2022.47