SHEDDING LIGHT ON OPACITY: WHY JUSTIFICATION MATTERS FOR EXPLAINABILITY

http://dx.doi.org/10.4314/ezumezu.v2i1.3

Submission: March 22, 2025 **Accepted: February 06, 2026**

Jacopo MORELLI
Department of Cultural and Linguistic Studies,
University of Modena Reggio-Emilia, Italy
Email: jacopomorellifilosofia@gmail.com
ORCID No.: https://orcid.org/0000-0001-5397-7884

Abstract

This paper contends that while technical advancements in transparency are essential, they alone cannot fully resolve the challenges posed by Artificial Intelligence (AI) opacity. I propose three distinct ways of meaningfully conceptualizing opacity, each highlighting different aspects of how and why certain AI systems resist transparency. These can be described as (1) technical opacity, which stems from the complexity of machine learning models; (2) commercial opacity, which arises from proprietary restrictions and intellectual property concerns; and (3) societal opacity, which reflects broader issues of accessibility, literacy, and power asymmetry in the use of AI systems. By distinguishing these domains, I argue that a meaningful notion of explainability must address all three. To this end, I suggest a conceptual framework that integrates Rainer Forst's critical theory of justification with Kate Vredenburgh's practical approach to AI ethics. This synthesis aims to deepen our understanding of explainability, ensuring that AI systems operate not only transparently but also justifiably.

Keywords: AI opacity; explainable artificial intelligence; transparency; accountability; justification.

Introduction[1]

The problem of opacity arises when decision-making processes, especially those involving algorithms and artificial intelligence (hereafter AI), become difficult to comprehend or scrutinize. This lack of understanding can lead to issues of fairness, accountability, and trust, as people may not know how or why certain decisions are made, and this is why concerns about opacity have since influenced debates on AI ethics and regulation. To address these challenges, one of the first widely recognized frameworks to emerge was the domain of "explainable AI" (XAI). Michael van Lent, William Charles Fisher and Michael Mancuso (2004) were among the first to popularize the concept of XAI to describe their system's ability to clarify the actions of AI-controlled units in simulation gaming.[2] In this early phase, explainability primarily referred to the *functional transparency* of models, making the internal operations of a system understandable to developers or end-users. More recently, however, scholars have expanded and refined this notion to encompass broader epistemic, ethical, and societal dimensions of AI decision-making (ALI n.d.n. 2023; KALASAMPATH n.d.n. 2025). These contributions

[1]This paper was first presented in an early form at the *Logics of Artificial Intelligence* colloquium, hosted by the Conversational School of Philosophy at the University of Calabar, Nigeria, in celebration of UNESCO World Logic Day. I am grateful for the valuable feedback received during the colloquium, as well as for the insightful comments provided by the two anonymous reviewers of the initial manuscript submitted to this journal.

[2]They cite as forerunners for XAI the works done on MYCIN (SHORTLIFFE 1976) and on Explainable Expert System (NECHES, SWARTOUT & MOORE 1985).

underscore that explainability is not a purely technical problem but one that intersects with governance, trust, and moral responsibility.

This paper situates itself within this broader and more recent movement, extending the discussion by arguing that although technical advancements in transparency are crucial, they are insufficient on their own to resolve the multifaceted challenges posed by AI opacity. I propose to analyze the problem of opacity through three interconnected dimensions: technical, commercial, and societal. This tripartition reveals that explainability, when reduced to mere interpretability and transparency, falls short of addressing opacity in all its nuances. To effectively encompass these three dimensions, I contend that we need a concept of explanation deeply intertwined with that of justification. To specify my favored conception of justification, I follow Rainer Forst's critical framework but also integrate it with Kate Vredenburgh's practical proposal to show how this can be implemented to design AI systems that can provide justifying explanations. By combining these perspectives, I aim to demonstrate how a critical theory of justification can be operationalized to create AI systems that not only function transparently but also align with broader moral, political, and legal norms.

Three Levels of Opacity

The multidimensional nature of AI opacity has been increasingly recognized in recent scholarship. Florian J. Boge (2022) distinguishes between *epistemic* and *functional* opacity, Alessandro Facchini and Alberto Termine (2022) propose a taxonomy of *opacity types* emphasizing how opacity can arise at several stages of an AI system's lifecycle, and Kashyap Haresamudram, Stefan Larsson, and Fredrik Heintz (2023) in turn, outline three levels of AI transparency — *traceability*, *explainability*, and *justifiability* — each corresponding to a different way of addressing opacity. Together, these accounts suggest that opacity is not a singular problem but a complex, layered phenomenon encompassing epistemic, procedural, and social dimensions.

Earlier discussions, however, tended to focus more narrowly on what can be termed technical opacity (CASTELVECCHI 2016; CREEL 2020), which stems from the inherent complexity of machine-learning models, especially those utilizing advanced architectures such as deep neural networks. To situate this form of opacity within a broader systematic perspective, I propose a tripartite framework that distinguishes among *technical*, *commercial*, and *societal* opacity. This framework complements existing classifications by extending the analysis beyond epistemic and procedural concerns to include the institutional and relational barriers that determine who can access, interpret, or contest AI-driven decisions. In doing so, it expands the analyses of explainability, normativity, and accountability to incorporate dimensions of *epistemic justice* that are often overlooked. To clarify how these forms of opacity operate in practice, I examine each dimension in turn.

Technical opacity remains, indeed, the most familiar and widely discussed dimension, popularized through the metaphor of the "black box". This metaphor captures the epistemic inaccessibility of a model's inner workings, the difficulty of tracing how specific outputs result from given inputs, even when the underlying algorithms are understood in principle. The root of this challenge lies in the way these models operate: they analyze vast amounts of data, identify intricate patterns, and generate predictions through highly complex, multi-layered computations. While the underlying algorithms can be understood at a mathematical or theoretical level, the sheer scale of parameters, often numbering in the millions or even billions, and their intricate interactions make it extraordinarily difficult to trace how specific decisions are reached. This lack of interpretability has profound practical implications, particularly in high-stakes domains such as healthcare, criminal justice, and autonomous vehicles. In these fields, understanding the reasoning behind an AI-driven decision is not just a matter of curiosity; it is essential for ensuring safety and fairness (CASTELVECCHI 2016).

The second dimension of opacity I identified as commercial opacity arises from the proprietary nature of many AI systems. Companies and organizations frequently treat their algorithms, models, and datasets as closely guarded trade secrets, justified by the need to maintain a competitive edge and protect intellectual property. While such practices are understandable from a business perspective, they create significant barriers to external scrutiny, independent validation, and accountability. Regulatory bodies, researchers, and even individuals directly affected by these systems are often left in the dark, unable to fully examine how decisions are made, what biases may be embedded within the algorithms, or whether the systems adhere to ethical and legal standards. This lack of transparency not only hinders the ability to detect and address potential harms but also erodes public trust in AI technologies. Concerns about misuse, abuse, or unintended consequences are amplified when the inner workings of these systems remain inaccessible, fostering skepticism and resistance toward their adoption (LU 2020). In this context, commercial opacity becomes a critical challenge, as it raises urgent questions about how to balance innovation and private interests with the need for public transparency and oversight.

The third dimension of opacity in the tripartition is that of societal opacity, which highlights a profound disconnect between the creators and deployers of AI systems and those who are subject to their outcomes. On the one hand, developers and engineers, often operating within highly specialized technical domains, may lack a deep understanding of the broader social, cultural, and ethical implications of their work. This limited perspective can lead to the design of systems that inadvertently perpetuate biases, reinforce inequalities, or fail to account for the diverse contexts in which they are deployed. On the other side, users and stakeholders (particularly those from non-technical backgrounds) often lack the technical literacy or resources needed to critically engage with, question, or challenge these systems. This knowledge gap exacerbates existing power imbalances, leaving individuals and communities vulnerable to decisions that they neither fully understand nor have the capacity to influence. For instance, marginalized groups, who are frequently disproportionately affected by biased or discriminatory AI systems, may find themselves at a great disadvantage when advocating for fairness or accountability. Without access to the technical expertise, tools, or platforms necessary to interrogate how these systems operate, they are often left powerless to contest decisions that affect their lives (DIAKOPOULOS 2015).

These forms of opacity cut across all three levels of traceability, explainability, and justifiability (HARESAMUDRAM, LARSSON & HEINTZ 2023), revealing that the challenge lies not merely in a lack of technical transparency or information disclosure, but in a broader normative asymmetry between those who design and deploy AI systems and those who are affected by their outcomes. In this sense, the tripartite framework proposed here underscores that resolving opacity ultimately requires more than enhancing explainability or strengthening regulation; it calls for a more equitable distribution of epistemic responsibility among all stakeholders in the AI lifecycle.

Explanations provided by AI systems, even when technically accurate, can still provoke frustration, skepticism, and harm if they fail to address the deeper moral and relational concerns of those affected (ZEDNIK 2021). For example, an AI system used in criminal sentencing might explain its recommendations through statistical correlations; yet it seems essential for judges to fully explain and justify to the victim how they rely on such correlations to avoid potential injustices. [3] This is due to the fact that agents expect other agents not only to be transparent but also to be accountable. Accountability requires going beyond descriptions of how a decision was made to articulate reasons that address the moral and political implications of their actions.

Only by integrating these moral dimensions into our understanding of opacity can we advance toward forms of AI governance that are genuinely justifiable, trustworthy, and accountable. Yet this raises a further question: what does it mean for an explanation or decision to be truly justifiable? To pursue this question, we must first clarify what *justification* itself entails and how it differs from merely describing the reasons behind an action.

Reasons, Motivations, Justifications: Forst's Account

What does articulating reasons for actions truly entail? At its core, this process seems to address at least two distinct types of questions: one concerning how one *ought* to act (justificatory or normative reasons) and another concerning what *actually* motivates someone to act (motivating reasons); these are two distinct, although related, matters: while justificatory reasons provide a rational or moral justification for an action (offering grounds that make it the right or reasonable thing to do) motivating reasons refer to the psychological or practical considerations that explain why an agent actually performs that action. In other words, justificatory reasons answer the question "Should one do this?" whereas motivating reasons answer "Why did one, in fact, do this?". Therefore, when we discuss explanations in the context of AI, what exactly are we referring to? Are we merely seeking a descriptive account of what was done and why it was done? Or are we also probing whether the action aligns with normative principles that could justify it? To meaningfully engage with debates surrounding the problem of opacity, it is crucial to confront this ambiguity.

My proposed answer is, I believe, simple and intuitive: just as in everyday human decision-making, we are not satisfied with explanations based solely on motivational reasons, the same applies to AI. Motivational reasons are not enough because we need normative reasons that justify actions in a way that can withstand critical scrutiny. This broader conception of explanation is indispensable for addressing the problem of opacity in all its dimensions, from technical transparency to social accountability. Two key questions must therefore be further clarified: how can we distinguish justifying reasons from non-justifying ones, and how can we design AI systems to provide such explanations, ensuring they operate not only transparently but also justifiably?

To address the first question, we need clear criteria for what it means to justify an action. I suggest following Rainer Forst, who defines justification through two appealing criteria: reciprocity and generality (FORST 2012, 6).[4] The criterion of reciprocity emphasizes that, in a situation of justification, no one can claim rights or privileges for themselves while denying them to others. Reciprocity requires mutual respect among the people involved, preventing one party from positioning itself as superior or projecting its own interests, values, or needs onto others. Reciprocity, therefore, demands that the reasons provided to justify a norm or action must be such that those proposing them would

[3]This topic remains nonetheless the subject of vigorous debate, with some scholars advocating for the full automation of human judges, arguing that they are already incapable of providing such explanations as efficiently as AI systems (DANZIGER, LEVAV & AVNAIM-PESSO 2011).

[4]While Forst develops his account of the right to justification primarily within a political-philosophical framework, this paper extends his approach to the applied domain of artificial intelligence. The aim is to show how Forst's theory, though not originally conceived with AI in mind, offers valuable conceptual tools for addressing questions of opacity, accountability, and normative legitimacy in AI systems.

be willing to accept them for themselves, overcoming logics of unilateral power or privilege. This ensures that justification is not based on arbitrary or partial premises (2012, 189).

The criterion of generality, instead, establishes that the reasons supporting a norm or action must be shareable by all those affected, without excluding or marginalizing anyone. Generality demands inclusivity in the justification process, ensuring that the objections and perspectives of every person involved are taken into account. No individual or group can be excluded from the process of justification. The reasons must be such that every reasonable person, regardless of their position or particular interests, can accept them. "Generality, then, means that the reasons for such norms need to be shareable among all persons affected, not just dominant parties" (FORST 2012, 146).

Forst argues that these two criteria are both necessary and sufficient conditions for achieving a genuinely normative justification. Only by adhering to both these criteria are we able to efficiently protect every individual's most fundamental right: the right to justification. This right guarantees that no person is subjected to norms or decisions without being provided with reasons that they, as autonomous agents, could not reasonably reject (FORST 2012, 19). This means that the justification must be reciprocal, ensuring that no one is subjected to norms or rules they could not also impose on others, and general, meaning it must apply universally to all parties involved. Forst contends that this fundamental right extends beyond being a right owed to individuals in particular situations; instead, it is a universal principle that forms the cornerstone of a just and equitable social order, ensuring that all individuals are treated with respect and accountability within the framework of shared norms and decisions, safeguarding individuals as autonomous and equal agents.

Toward a Truly Normative Right to Explanation for AI: Vredenburgh's Account

Similar to Forst, Kate Vredenburgh argues that individuals possess a fundamental moral interest in informed self-advocacy and that realizing this interest necessitates the protection of a specific right: the right to explanation. Vredenburgh, however, unlike Forst, directly tackles this issue in the field of AI and is therefore more equipped to give us an answer to our second main question, namely, how can we design AI systems to provide explanations that protect our self-advocacy.

Vredenburgh starts his analysis by identifying three different types of explanations that individuals might have a right to. 1) *Intentional Explanation* explains actions based on an agent's motivating reasons. Reasons the agent considers as justifying their actions. 2) *Causal Explanation* explains actions based on causes, which might include motivating reasons but can also involve non-reason-based factors, such as physical causes and impositions. 3) *Normative Explanation* explains actions in terms of normative reasons. Reasons that objectively count in favor of an action, irrespective of whether the agent acted on them (VREDENBURGH 2022, 211–13).

Vredenburgh here uses explanation as a broad, general term that precedes reason and encompasses the distinctions we previously introduced, such as motivating and justificatory reasons. In this broader sense, explanations are processes that provide information (and/or reasons) about "why" a decision or action was undertaken. It is crucial, however, to avoid misinterpreting Vredenburgh's focus on explanations as being limited solely to the intentional-motivational aspects or the procedural "how" of decision-making, as some problematic trends within the XAI movement seem to do (RETZLAFF n.d.n. 2024); Vredenburgh explicitly underlines that normative reasons are central to her conception of explanation and self-advocacy (VREDENBURGH 2022, 217).

Vredenburgh, however, is also aware of a potential dilemma between the effectiveness and feasibility of the right to explanation regarding AI, and the justificatory burden it imposes on both rights holders and duty bearers (VREDENBURGH 2022, 219-220). According to this first part of the dilemma, more complex and accurate explanations are too costly for rights holders to use, but simpler, more intelligible explanations are also ineffective if they are superficial and do not provide adequate clarity (they fail to capture all relevant factors). The second aspect of the dilemma focuses on the costs borne by duty bearers. If explanations were provided in a personalized manner for each individual, the cost for decision-makers would be intolerable; conversely, if explanations were too abstract, the right to explanation would become ineffective for rights holders.

The usual proposed solution to the first dilemma (having personalized explanations) seems to create problems for the second one. We arrive thus at an impasse: "either the explanations are abstract enough to be tolerably costly to decision- makers, but the right is undermined because it becomes ineffective for rights bearers; or, to be effective, explanations ought to be personalized to the rights bearer, but the right is undermined because such personalization is intolerably costly for decision-makers" (VREDENBURGH 2022, 220). To resolve this dilemma, Vredenburgh proposes that explanations should consist of high-level descriptions of the relevant rules provided by decision-makers or free experts; the fact that some legal systems already incorporate them demonstrates that their implementation is reasonably cost-effective. These descriptions aim for functional transparency, understanding how inputs connect to outputs, without requiring detailed knowledge of the code or the process that generates the output (VREDENBURGH 2022).[5] This solution seeks to balance the effectiveness of the right to explanation with feasibility for duty bearers, making it tolerably costly for both parties.

Vredenburgh's proposal thus directly addresses the challenge of societal opacity by promoting forms of explanation that are both accessible and normatively meaningful to those affected by AI decisions. By emphasizing high-level, rule-based explanations provided by accountable human agents, her model redistributes epistemic responsibility between experts and laypersons, thereby mitigating the moral and cognitive asymmetries that characterize societal opacity. In this sense, her solution complements the tripartite framework advanced here, demonstrating how normative justification can guide the design of explanation practices that are not only efficient but also socially and ethically responsive.

Justification Across Technical, Commercial, and Societal Opacity

The analyses developed by Forst and Vredenburgh converge on the centrality of normative justification, albeit articulated through partially different terminologies and theoretical frameworks. This shared emphasis, I suggest, provides a crucial lens for clarifying what is at stake in the threefold conception of opacity introduced above. In this final section, I make explicit how a justification-centered framework helps to illuminate the distinct failures of justification associated with technical, commercial, and societal opacity, and how it offers guidance for addressing these issues in ways that go beyond informational transparency alone.

Technical opacity represents a failure at the level of justificatory intelligibility in the sense that the reasoning behind a system's decisions is inaccessible to those affected. Even if the algorithm produces outputs reliably, the absence of a human or responsible

[5]Vredenburgh mentions a distinction between different types of algorithmic transparency, including functional transparency, structural transparency, and execution transparency. Functional transparency, which focuses on high-level rules, is considered sufficient for the right to explanation, unlike the other forms of transparency, which might be too complex or costly.

agent capable of explaining and justifying those decisions renders the system normatively opaque. From a Forstian perspective, justification requires that reasons be intelligible and assessable by those subject to the decision. When no one, neither the system itself nor a responsible developer, can provide such an explanation, the justificatory relation breaks down, and affected individuals are left unable to determine whether the decision aligns with principles they could reasonably accept.[6]

Commercial opacity, by contrast, represents a failure at the level of justificatory access. Proprietary protections, trade secrecy, and intellectual property regimes may restrict the disclosure of information necessary to evaluate whether AI-driven decisions conform to shared norms and legal standards. While such restrictions may be economically justified, they pose a normative problem when they prevent affected parties from accessing the reasons that govern decisions about them. In Forst's terms, commercial opacity allows certain actors to exempt themselves from reciprocal justification by invoking interests that cannot themselves be justified to all affected parties. Vredenburgh's emphasis on high-level, rule-based explanations is particularly relevant here: even when full technical disclosure is infeasible or unjustified, decision-makers remain obligated to provide normative explanations that articulate the principles, constraints, and objectives guiding the system's use. These explanations need not reveal proprietary details, but they must be sufficient to enable informed self-advocacy and contestation.

Finally, societal opacity reflects the most pervasive and structurally entrenched failure of justification. It arises when disparities in technical literacy, institutional power, and social standing prevent certain groups from effectively participating in justificatory practices at all. Here, opacity is not reducible to missing information; rather, it is produced by unequal capacities to demand, interpret, and challenge reasons. Marginalized individuals may formally receive explanations, yet remain unable to use them to advocate for themselves or to question the legitimacy of decisions. Vredenburgh's account directly addresses this problem by grounding the right to explanation in the moral interest of informed self-advocacy. When explanations are designed with this interest in mind, they function not merely as informational tools but as enabling conditions for justificatory agency. In this sense, societal opacity is mitigated not by increasing technical detail but by designing explanation practices that are responsive to the normative position of those affected.

This framework, I believe, shows why opacity is best understood not as a single epistemic deficit but as a family of justificatory failures that manifest differently across technical, commercial, and societal contexts. From this perspective, efforts to improve explainability that focus solely on interpretability or transparency, addressing only the first levels, are insufficient. What is ultimately required is the institutionalization of justificatory practices that ensure reciprocity and generality in the reasons governing AI-assisted decisions. It also clarifies what it means to operationalize a critical theory of justification in the design and governance of AI systems. Operationalization does not require attributing moral agency to AI systems themselves. Rather, it involves embedding justificatory responsibilities within the sociotechnical arrangements through which AI systems are developed, deployed, and regulated. These responsibilities include: specifying the normative principles that guide system design and use; ensuring that explanations articulate these principles at an appropriate level of abstraction; and

[6] If only a narrow group of experts can understand and assess the grounds of a decision, then the justificatory relation becomes asymmetrical, violating again the right to justification but on a different level, that of societal opacity that we will discuss in the next page.

assigning clear accountability to human agents and institutions for the reasons provided. In this way, justification becomes a practical standard against which different forms of opacity can be diagnosed and addressed, rather than a purely abstract ideal.

Conclusion

I have argued that addressing the problem of opacity in its three dimensions (technical, commercial, and societal) requires moving beyond mere technical and individual transparency to embrace a normative conception of explanation, which is inevitably tied to justification and accountability. True accountability requires that agents, whether developers or the AI systems themselves, offer normative reasons when justifying their actions; these are reasons that are contextually appropriate and capable of withstanding reasonable rejection, ensuring that decisions are not only explained but also legitimately defensible. Seen through this lens, the three dimensions of opacity correspond to distinct failures of justification.

Some might argue that distinguishing between technical, commercial, and societal opacity risks overcomplicates an already dense debate, making it harder to formulate actionable governance strategies. Yet, the purpose of this tripartition is precisely to clarify rather than complicate: by analytically separating these dimensions, we can better identify which forms of intervention — technical or institutional (both specific for some companies or for the population) — are most appropriate in each case. Far from multiplying categories, this approach allows for a more targeted and integrated understanding of the different mechanisms through which opacity undermines accountability and trust. By redistributing epistemic responsibility among all stakeholders (developers, regulators, and affected individuals), AI governance can genuinely promote fairness, legitimacy, and trust in socio-technical decision-making.

Declarations

*The author declares no conflict of interest or ethical issues for this work.

Relevant Literature

1. ALI, S. n.d.n. “Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence.” [Information Fusion], pp1-52, 2023. Vol 99. No101805.
2. BOGE, F. J. “Two Dimensions of Opacity and the Deep Learning Predicament.” [Minds and Machines], pp43–75, 2022. Vol 32. No1.
3. CASTELVECCHI, D. “Can We Open the Black Box of AI?” [Nature News], pp20-23, 2016, Vol 538. No7623.
4. CREEL, K. A. “Transparency in Complex Computational Systems.” [Philosophy of Science], pp568–589, 2020. Vol 87. No4.
5. DANZIGER, S., LEVAV, J. & AVNAIM-PESSO, L. “Extraneous Factors in Judicial Decisions.” [Proceedings of the National Academy of Sciences], pp6889–6892, 2011. Vol 108. No17.
6. DIAKOPOULOS, N. “Algorithmic Accountability: Journalistic Investigation of Computational Power Structures.” [Digital Journalism], pp398–415 2015. Vol 3. No3.

7. FACCHINI, A. & TERMINE, A. “Towards a Taxonomy for the Opacity of AI Systems.” [Philosophy and Theory of Artificial Intelligence], pp73–89, 2022. Vol 63.
8. FORST, R. [The Right to Justification: Elements of a Constructivist Theory of Justice]. 2012. Columbia University Press: New York.
9. HARESAMUDRAM, K., LARSSON, S. & HEINTZ, F. “Three Levels of AI Transparency.” [Computer], pp93–100, 2023. Vol 56. No2.
10. KALASAMPATH, K. n.d.n. “A Literature Review on Applications of Explainable Artificial Intelligence (XAI).” [IEEE Access], pp.41111-41140, 2025. Vol13.
11. LU, S. “Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence.” [Vanderbilt Journal of Entertainment & Technology Law], p.99-159. 2020. Vol 23. No1.
12. NECHES, R., SWARTOUT, W. & MOORE, J. “Explainable (and Maintainable) Expert Systems.” [Proceedings of the 9th International Joint Conference on Artificial Intelligence], p.382-389, 1985.
13. RETZLAFF, C. O. et al. “Post-Hoc vs Ante-Hoc Explanations: xAI Design Guidelines for Data Scientists.” [Cognitive Systems Research], pp1-17, 2024. Vol 86. No101243.
14. SHORTLIFFE, E. H. “Computer-Based Medical Consultations: MYCIN.” [Journal of Clinical Engineering], 1976. Elsevier: New York.
15. VAN LENT, M.; FISHER, W.; MANCUSO, M. “An Explainable Artificial Intelligence System for Small-unit Tactical Behavior.” [Proceedings of the Nineteenth National Conference on Artificial Intelligence], 2004. American Association for Artificial Intelligence (AAAI) Press: California.
16. VREDENBURGH, K. “The Right to Explanation.” [Journal of Political Philosophy], pp209–229, 2022. Vol 30. No2.
17. ZEDNIK, C. “Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence.” [Philosophy & Technology], pp.265–288, 2021. Vol 34. No2.