

The Vice of Transparency A Virtue Ethics Account of Trust in Technology^a

Francesco Striano*

Abstract

Questo articolo esplora il rapporto tra fiducia, trasparenza e tecnologia da una prospettiva di etica delle virtù. Mette in discussione l'assunto che la trasparenza sia essenziale per la fiducia, distinguendo tra fiducia, affidamento e confidenza. La trasparenza viene poi esaminata sia come disponibilità informativa sia come processo sociale di negoziazione. L'articolo sostiene che la trasparenza nel primo senso può portare a un sovraccarico di informazioni e a dinamiche di controllo, sostenendo invece un rapporto equilibrato e virtuoso con la tecnologia che enfatizzi le capacità interpretative dell'utente. Propone che la fiducia nella tecnologia dipenda sia dagli atteggiamenti individuali sia dall'affidabilità degli oggetti. Infine, l'articolo critica il "culto della trasparenza" contemporaneo, proponendo che l'onestà, come virtù tecno-morale, sostituisca la trasparenza quale obiettivo progettuale. Le tecnologie oneste medierebbero e negozierebbero l'accesso degli utenti alle informazioni, promuovendo una fiducia autentica e sostenendo la fioritura umana.

Parole chiave: Confidenza, Affidabilità, Trasparenza, Fiducia, Etica delle virtù.

This article explores the relationship between trust, transparency, and technology from a virtue ethics perspective. It challenges the assumption that transparency is essential for trust, distinguishing between trust, reliance, and confidence. Transparency is then examined as both informational openness and a social process involving negotiation. The article argues that transparency in the first sense can lead to information overload and control dynamics, advocating instead for a balanced, virtuous relationship with technology that emphasizes user interpretative skills. It proposes that trust in technology depends both on individual attitudes and objectual

^a Questo saggio è un prodotto delle ricerche e delle collaborazioni condotte all'interno del progetto PRIN 2022 "Social Transformations & the Crisis of Expertise" (2022JR8Z8P) finanziato dall'Unione Europa – Next Generation EU. Saggio ricevuto in data 31/05/2024 e pubblicato in data 22/01/2025.

* Assegnista di ricerca, Università di Torino, email: francesco.striano@unito.it.

reliability. Finally, the article critiques the contemporary “cult of transparency,” proposing that honesty, as a techno-moral virtue, should replace transparency as the design goal. Honest technologies would mediate and negotiate user access to information, fostering authentic trust and supporting human flourishing.

Keywords: Confidence, Reliance, Transparency, Trust, Virtue Ethics.

1. Introduction

“Transparency builds trust,” claim company slogans or leadership development courses. A presumption of trust—or at the very least, a reliance, a distinction that will be revisited subsequently—is contingent upon the assumption of reciprocal transparency between the involved parties. However, this assertion requires further examination. In a sense, when one chooses to place trust in another individual, it does not necessarily entail that one has complete information regarding how that individual will behave in a given situation. Alternatively, trust is based on the character traits of the individual in question, or it is founded upon the reliability of a professional or tool due to a prior history of satisfactory performance, which provides the basis for the assumption that the same level of reliability will be maintained in the future. In this sense, even in the absence of complete information about the situation, the premises, and so forth, we nevertheless decide to place trust or reliance.

The question thus arises as to how this applies to our relationship with technology. The European Union has repeatedly emphasised the importance of transparency, particularly in relation to algorithms and artificial intelligence¹. Transparency appears to be a fundamental requirement of a trust relationship with technology. An opaque technology does not seem to be one that can be trusted. However, it is questionable whether it is even possible to “trust” technology in the first place. Furthermore, the relationship between trust and transparency in our interaction with technology is unclear. Does a virtuous relationship with technology, which promotes human flourishing, a good life with technology, and the use of technology towards a good life, require transparency?

This article will commence by examining the aforementioned inquiries, with the subsequent objective of determining the extent to which a methodology derived from virtue ethics—which appears to be optimally suited to the analysis of traits such

¹ As early as 26 January 2022, the *European Declaration on Digital Rights and Principles for the Digital Decade* already asserted the necessity of ensuring transparency about the use of algorithms and artificial intelligence, and that people are empowered and informed when interacting with them.” Consequently, the AI Act (Regulation - EU - 2024/1689) incorporates transparency (along with human agency and oversight; technical robustness and safety; privacy and data governance; diversity, non-discrimination and fairness; societal and environmental well-being and accountability) among the seven principles that must “ensure that AI is *trustworthy* and ethically sound” (*italic mine*). The two documents may be accessed at the following web addresses: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022DC0028> and https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689 (last accessed: 26 August 2024).

as trust, reliability, and transparency—can prove beneficial in resolving the identified issues.

The initial section will be dedicated to the development of a non-standard theory of trust, reliance, and confidence and to an examination of the applicability of these concepts to the relationship between humans and technological artefacts.

The second section will elucidate the existence of disparate accounts of transparency and illustrate the inherent ambiguity in defining a technology as “transparent.”

The third section will examine the conditions under which transparency is necessary for a human being to trust or rely on a technology.

The fourth section will review virtue-oriented approaches to the relationship between trust and transparency in relation to technologies.

In the conclusion, I will argue, following Shannon Vallor, that honesty is the techno-moral virtue to be cultivated in relation to trust issues in the techno-social sphere. Furthermore, I will argue the need to design technologies *that are themselves honest*.

2. Trust or Reliance: How do we relate to technology?

“Imagine a society in which there is no trust in doctors, teachers, or drivers,” writes Mariarosaria Taddeo: “This would require that all the members of the society to spend a significant portion of time and resources controlling the way others perform their tasks, at the expenses of their own tasks.”²

A technologically advanced society necessitates that all its constituent elements, whether human, environmental, mechanical, or informational, operate with the greatest possible autonomy. However, the willingness to confer agential autonomy³ upon a given component hinges upon the existence of trust.

The question thus arises as to how this concept of trust is to be defined. Furthermore, it is pertinent to inquire whether it is indeed feasible to repose trust in artificial agents. This issue is the subject of considerable debate, and there is no consensus in the literature. If we adhere to Taddeo’s original definition, trust can be conceptualised as a second-order property pertaining to primarily binary and purpose-oriented relationships. In order to achieve a desired outcome, a trustor decides to do so through an ability possessed by a trustee, who is perceived as a trustworthy agent and therefore the relationship between them will have the property of being beneficial to the trustor. “Such a property is a second-order property that affects the first-order relations occurring between agents and is called trust.”⁴

² M. Taddeo, *Trusting Digital Technologies Correctly*, in «Minds and Machines», n. 27, 2017, pp. 565-568: 566.

³ For a discussion on the possibility of artificial agents with proper agency, I would direct your attention to my *Can Artificial Agents Act? Conceptual constellation for a de-humanised theory of action*, in «S&F_scienzaefilosofia.it», n. 31, 2024, pp. 224-244.

⁴ M. Taddeo, *Modelling Trust in Artificial Agents, A First Step Toward the Analysis of e-Trust*, in «Minds and Machines», n. 20, 2010, pp. 243-257: 249.

This definition has the advantage of being readily applicable to trust in artificial agents; however, it may be inadequate in other respects. Firstly, due to its binary nature, it appears to present challenges in extending this definition to encompass more complex relationships or forms of trust, such as that which may be placed in institutions or decision-making processes. Furthermore, the definition may be perceived as somewhat circular: trust is defined as a quality inherent to relationships between individuals who are perceived as trustworthy. However, this raises the question of what factors contribute to an individual being regarded as trustworthy. What is this property that seems to be of first-order, in that it is possessed by the trustee, but must be defined by a second-order property?

Moreover, trust seems to be a more vague and undefined state. It can be understood as an attitude held by some individuals towards others or institutions that are perceived as trustworthy. The concepts of trust and trustworthiness appear to persist as attitudes and traits even when there is no relationship based on immediate utility at stake. For example, one may trust their partner both in the sense that they believe the partner will not lie to them and in the sense that they believe the partner will be able to provide support when needed. In such a case, the trustworthiness of the partner will be described as such even when there is no need to establish whether the partner is lying or when they are not providing support. Similarly, if the government of a State is described as untrustworthy, it is because, even if it is not currently causing harm or disseminating false information, there is a possibility that it may do so in the future. This distrust therefore translates into a state of heightened awareness and scrutiny of its actions.

The trust based on immediate usefulness described by Taddeo can be more accurately defined as “reliance.” Some literature suggests that reliance is a broader, more neutral, and more general concept than trust. It posits that trust is a special case of reliance⁵.

In line with this conception are de Fine Licht and Brülde, who propose the following definitions for reliance and trust.

They define reliance “as a three-place relation, where one agent (A) relies on an agent or some other object (B) to do something, to maintain some state, or the like (C).”⁶ Reliance “can be both voluntary and involuntary”⁷, so it can be applied to humans as well as objects: I rely on my bicycle to get me to the city centre because I judge that using it increases my likelihood of arriving on time for my appointment, because my bicycle is able to cover that distance without damage, because it is

⁵ Cf. S. Blackburn, *Trust, cooperation, and human psychology*, in Id., *Practical Tortoise Raising and other philosophical essays*, Oxford Academic, Oxford 2010, pp. 90-108.

⁶ K. de Fine Licht and B. Brülde, *On Defining “Reliance” and “Trust”: Purposes, Conditions of Adequacy, and New Definitions*, in «Philosophia», n. 49, 2021, pp. 1981-2001: 1989.

⁷ *Ibid.*, p. 1990.

specially designed to move nimbly in city traffic, and because it has no misplaced parts and is at my immediate disposal at this moment⁸.

Trust, on the other hand, is a special form of agential reliance in which “a trusting agent attributes a certain kind of moral motivation to the trustee, namely a perceived normative responsibility to care about something.”⁹ The attribution of this moral value is what makes it possible for trust to be *betrayed*, whereas a failed reliance can at best result in *disappointment*¹⁰. This is why, according to Deley and Dubois, a technology cannot be trusted, but, at best, can be relied upon: “technologies cannot possess a good will,”¹¹ and therefore cannot genuinely *betray* us.

Similarly, Christopher Thompson asserts that, in the absence of artefacts exhibiting mental states, it is not possible to place trust in technologies. He posits that when we appear to repose trust in the artefacts themselves, we are in fact confusing trust with reliance¹². Furthermore, if a form of trust in artefacts exists, it is the trust that we accord to their designers or developers when we use them¹³. Likewise, Deley and Dubois maintain that “reliability is a *mediator* of trust toward the makers of a technology and that relationships of reliance mediate our relationships of trust.”¹⁴

In contrast to Taddeo’s conceptualisation of trust as a form of property, these alternative perspectives view it as a relationship, although sometimes it seems that trust is a state of mind of the trustor, at other times a judgement on trustworthiness. In light of the aforementioned considerations, it seems pertinent to direct attention to an alternative definition that merits consideration. This definition, proposed by Shionoya, characterises trust as an “evaluative act of one individual directed toward another with regard to whether that other person is or is not trustworthy in the relevant circumstances.”¹⁵ The definition presented is based on the concept of trust between two individuals. However, Shionoya also discusses the notion of trust as a

⁸ These conditions correspond to what de Fine Licht and Brülde call *probability condition*, *competence condition*, *motivation condition*, *opportunity condition*. We could add the *value condition*, i.e., that I assign a positive value to my goal of getting to the city centre to arrive on time for my appointment and this leads me to judge my act of reliance positively as well. Cf. *ibid.*

⁹ *Ibid.*, p. 1991. In order to attempt to decouple the notions of intentionality and motivation on the one hand and responsibility on the other, I will refer once more to F. Striano, *Can Artificial Agents Act?*, cit., pp. 241-244, in which I attempted to extend the concept of responsibility (distinguishing it, however, from accountability) to artificial agents as well, and to broaden the concept of distributed responsibility.

¹⁰ Cf. A. Baier, *Trust and Antitrust*, in «Ethics», vol. 96, n. 2, 1986, pp. 231-260: 235.

¹¹ T. Deley and E. Dubois, *Assessing Trust Versus Reliance for Technology Platforms by Systematic Literature Review*, in «Social Media + Society», April-June 2020, pp. 1-8: 2.

¹² Cf. C. Thompson, *Faire confiance aux artefacts – Faire confiance à distance*, in M. Doueïhi and J. Domenicucci, *La confiance à l'ère numérique*, Éditions rue d'Ulm, Paris 2018, pp. 97-111: 108.

¹³ Cf. *ibid.*, pp. 105-107.

¹⁴ T. Deley and E. Dubois, *Assessing Trust Versus Reliance for Technology Platforms by Systematic Literature Review*, cit., p. 2, *italic mine*.

¹⁵ Y. Shionoya, *Trust as a Virtue*, in Y. Shionoya and K. Yagi, *Competition, Trust, and Cooperation. A Comparative Study*, Springer, Berlin-Heidelberg 2001, pp. 3-19: 10.

shared virtue, as well as its community-based aspects within the context of trust networks.

Furthermore, Shionoya posits that the evaluative act of “trust” is contingent upon a willingness to trust and thus upon a disposition to trust or *confide*. Shionoya employs the term “confidence” as a synonym for trust, whereas de Fine Licht and Brülde understand it as a synonym for reliance.¹⁶ I propose, however, that it be used to denote precisely that disposition which, according to Shionoya, underlies an act of trust. The result of my re-interpretation of Shionoya non-standard definition could be represented by a scheme (fig. 1) that sees trust as a disposition of the trustor leading to trust, understood as an evaluative act on being trustworthy, the latter being understood as a characteristic disposition of the trustee.

The latter disposition of the trustee, however, must have been earned in the past, probably on the basis that the trustee proved to be *reliable*. In this sense, we can describe reliance through an inverse scheme (fig. 2) to the confidence-trust-trustworthiness scheme derived from Shionoya. We can posit that reliability is a characteristic disposition of an individual or object, which inspires in a subject an evaluative act of reliance, which in turn leads to the evaluative subject’s disposition of confidence.

In this sense, even an artefact, given proof of reliability, could become the object of an evaluative act of trust, not because it is endowed with a will or the possibility of betrayal, but because it is invested with trustworthiness by the confiding subject.

The argument put forth is that a person may confide in another person as well as in a technology on the basis of an evaluative act of trust – eventually based on

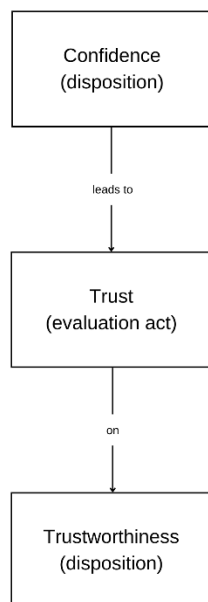


Fig.1

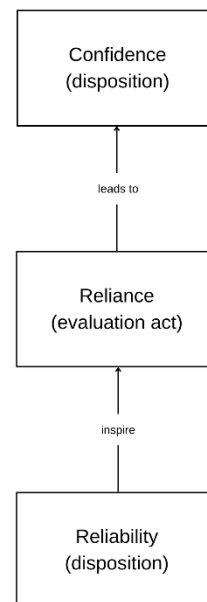


Fig. 2

¹⁶ Cf. K. de Fine Licht and B. Brülde, *On Defining “Reliance” and “Trust”*, cit., p. 1981.

precedent or simultaneous evaluative acts of reliance. This evaluation may be based on certain conditions that must be fulfilled by the subject of the evaluative act, its object, or the act itself¹⁷. At this juncture, the research question of this paper can be rephrased as follows: do these conditions of confidence include the transparency of technologies?

3. *What does transparency mean?*

Prior to responding to this question, it is necessary to pose a preliminary inquiry: what is the precise meaning of the term “transparency?” If, as we have seen, the notion of trust is not entirely unambiguous, the notion of transparency also lends itself to ambiguity or different interpretations. Further difficulties arise when transparency is to be applied to the technological sphere. In order for transparency to be achieved with regard to technology, what, exactly, must be transparent in a technology?

Bridget O’Brien notes that there are at least two ways of framing the question of transparency, two different accounts that, as we shall see, can have different reflections on the question of trust¹⁸.

The first account is that of transparency as informational openness. In this view, transparency is to be understood as comprehensive disclosure, as a free, continuous, and total flow of information, without restriction or censorship of any kind. In the context of medicine, for instance, this account entails that patients are granted comprehensive and unimpeded access to information pertaining to their care. This encompasses details about the treatments they are receiving, the diagnostic procedures scheduled, and the associated financial and social costs. In the context of the relationship between institutions and citizens, it is imperative that no information be withheld. In the case of a decision-making support algorithm the source code must be open and searchable, and each step must be explainable.

A second account is that which views transparency as a social process, specifically as a communicative act that is not without tensions and negotiations. This account challenges the assumption that access to comprehensive information is inherently beneficial. It posits that the mere availability of information does not guarantee its objectivity or completeness, and that individuals may not always possess the capacity to receive and interpret it accurately. This is because, according to this account, transparency is dependent on three interrelated elements: the *content*, the *receiver* (“the viewer,” as defined by O’Brien) and the *medium*. In her analysis, O’Brien employs the metaphor of the window to illustrate the three components of transparency in relation to information. She posits that the scene outside the window, which represents the content, may fluctuate due to external factors beyond human control, such as weather, season, and the mobility of objects. The individual who

¹⁷ These conditions may be consistent with those proposed by de Fine Licht and Brülde. Please refer to footnote 9 for further details.

¹⁸ See B. C. O’Brien, *Do You See What I See? Reflections on the Relationship Between Transparency and Trust*, in «Academic Medicine», vol. 94, n. 6, 2019, pp. 757-759.

receives the information, on the other hand, can be conceptualised as the viewer, situated on the internal side of the window, influences the scene observed through the lens of their senses, prior knowledge, emotions, and interpretative abilities. The medium, in this case the window itself, shapes the scene in a particular manner, influencing the viewer's gaze and potentially distorting or deforming it.

The second account does not directly contradict the first. It may be the case that, under ideal conditions, a direct and total flow of information is possible. This will be discussed further in the next session, along with the question of whether it is also desirable to build a relationship of confidence with the technology. However, the social account appears to provide a more accurate representation of the intricacies associated with information transmission mechanisms. This approach facilitates comprehension of the fact that transparency can be conceptualised as a negotiation between these elements. The concept of transparency, in this context, refers to the optimal access to information, which may not be comprehensive but is tailored to the specific needs of the moment. This represents a goal, and the effort of negotiation should be directed towards its realisation. Furthermore, it seems desirable that the process of selecting the portion of information of interest to the receiver should also be transparent. Furthermore, the degree of deformation that a medium implements on the content should also be transparent and known. It is also essential that the receiver is equipped with the requisite knowledge and skills to interpret the information in an appropriate manner. The aforementioned measures must be implemented in order to achieve transparency, according to the social account.

It would appear that the focus of this paper should be on interventions on the medium. The question thus arises as to how one might intervene in order to make technologies (media) more transparent, thereby increasing user trust. Nevertheless, this raises further questions. Firstly, it is necessary to define what is meant by the term "transparent technology." It is necessary to determine whether the discussion is focused on support transparency or interface transparency.

The term "transparency of the medium" is used to describe a situation in which the medium presents itself as transparent and accessible. In the context of computer-based media, this could include open-source technologies or free software. It is not my intention to provide a detailed analysis of the differences between the two movements. However, it is important to note that a key aspect that unites them is the accessibility of the source code (and, in the case of free software, the option of modifying it according to one's own requirements). This could be regarded as an instance of transparency of the support, but it is also important to consider that, according to this account, the support can only be genuinely transparent to those individuals (or other machines) who are able to read, interpret, and potentially modify the code. For the average user, this form of transparency (which, according to the informational account of transparency, would appear to be transparent) may not satisfy the social requirements of transparency.

To address this challenge, significant efforts have been made throughout the history of personal computing to develop interfaces that are perceived as

“transparent” in terms of user-friendliness. Nevertheless, this transparency assumes an illusory quality. The interface offers a selective view, translating computational elements into human-experienced forms while simultaneously shaping perception by obscuring functional and operational aspects. Despite this partial vision, digital interfaces present themselves as *transparent totalities*, claiming to reveal all of being and provide unprecedented access to the world¹⁹. This creates the *illusion* of full access to information and a virtual world at the user’s disposal. The epistemological and ethical consequences are significant: users may believe they fully understand the real world while feeling detached from responsibility, acting as if in a fictional realm²⁰. The crucial mediation process is overlooked, as the design promotes the illusion of immediate access to both the real and virtual worlds.

Two additional conceptualisations of transparency in technology are transparency in operations and transparency in design²¹. The former concerns the utilisation of technologies and the notion of information sharing and accountability on the part of the entity employing specific technologies or processes, particularly in relation to user data. The latter emphasises the necessity for public regulation of the explicability of technological processes and the importance of source code auditability. These two conceptualisations of transparency can be situated within the informational account.

4. *What kind of transparency is necessary to have confidence?*

As illustrated at the outset with reference to Taddeo, trust is a crucial element in intricate social interactions, as it enables the conservation of energy—both attentional and otherwise—that would otherwise be expended on the constant monitoring of others’ behaviour. This also applies to the energy we would expend in the constant supervision of technologies. The development of artificial intelligence and the constant research into the design of autonomous agents that can assist us in activities,

¹⁹ A critique of this simplistic notion of transparency, which fails to acknowledge the inherent limitations of the interface and the inevitable trade-off between visibility and obscurity, can be found in M. Carbone and G. Lingua, *Toward an Anthropology of Screens. Showing and Hiding, Exposing and Protecting*, Palgrave Macmillan, Cham 2023, pp. 107 ff.

²⁰ This discussion is not intended as a comprehensive development of this topic. However, I have previously discussed it in greater depth in a few publications, the most recent of which are F. Striano, *The dangerous liaison between rape culture and information technologies. Reality, virtuality, and responsibility in cyber-rapes*, in M. L. Edwards and O. Palermos (eds.), *Feminist Philosophy and Emerging Technologies*, Routledge, London 2024, pp. 74-94 and Id., *Violenza virtuale. Vita digitale e dolore reale*, il Saggiatore, Milano 2024.

²¹ Cf. D. Kwan, L. M. Cysneiros and J. C. S. do Prado Leite, *Towards Achieving Trust Through Transparency and Ethics*, in «2021 IEEE 29th International Requirements Engineering Conference Proceedings», pp. 82-93: 88-90. These two conceptualisations of technological transparency can be linked to what David Heald in political theory calls “event transparency” (transparency of inputs, outputs, and outcomes) and “process transparency” (transparency of procedures and decision-making processes): cf. C. Hood, D. Heald (eds.), *Transparency: The Key to Better Governance?*, The British Academy, Oxford 2006, pp. 30-32.

decisions, and even artistic productions is precisely in the direction of relieving us of certain tasks and ceding control to these agents, which we must be able to trust.

It is imperative that any entity designated as an artificial agent, or indeed any person or object in which we place trust, is able to demonstrate reliability and earn our trust. The objective is to instil a disposition of confidence in the subject. Both institutionally and in the scientific literature²², transparency is regarded as a means of reinforcing this disposition. In light of the preceding discussion on transparency, it is now necessary to consider which of the various accounts presented should be adopted and whether an increase in transparency truly corresponds to a strengthening of the confidence disposition.

It can be argued that an informational account is a fundamental requirement for fostering confidence in technologies. A technology that enables comprehensive access to data, performance, and metrics appears to be one that has no hidden aspects, a reliable and trustworthy technology. This level of access can potentially enhance the quality of performance and ethical use due to the transparency not only of the support but also of the entire process.

Nevertheless, it is not necessarily the case that mere equal access to information is in itself beneficial for the establishment of trust. The lack of guidance to transform data from disparate sources into meaningful information, or the overflow of information not accompanied by interpretative support, could potentially lead to a state of stress for the user, resulting in a loss of trust in a system that, while sharing access to its algorithm with any individual, employs a language that is opaque to those lacking the requisite expertise.

This is why the social account of transparency places emphasis on the involvement of a number of factors in order to guarantee effective transparency. The construction of meanings and the interpretation of information are collaborative processes that involve the interplay between content, viewer, and medium. Therefore, transparency can be conceptualised as a virtuous relationship between these three poles, which is a crucial aspect in the establishment of trust. However, even within this account, it is unclear whether the disposition of confidence within the human user (the viewer) is exclusively contingent on the establishment of a transparent relationship²³. It is not necessarily the case that the medium must be transparent in order for it to be trusted. Indeed, it may have to conceal certain functions in order to make itself comprehensible. It is similarly unproven that the viewer desires transparency of content in order to be able to trust it. Sometimes, we elect to place our trust in a given entity precisely because this relieves us of the necessity of continually monitoring the process in all of its parts and of bearing the responsibility (however shared with the medium and other actors in the process) of interpreting and managing a substantial amount of information.

²² See both the EU documents and the articles that have been previously cited.

²³ Transparency can be seen as the opposite of secrecy, but it does not necessarily limit deception, manipulation, or disinformation. See, in this regard, C. Birchall, *Radical Secrecy: The Ends of Transparency in Datafied America*, University of Minnesota Press, Minneapolis-London 2021, pp. 71-74.

We were confronted with a dilemma: on the one hand, transparency seems to be a crucial factor in establishing confidence. If a technology were entirely opaque and its designers were to refuse to share information about its development, we would likely be inclined to distrust it. This is evident in the case of AI and decision-support algorithms, where the fear of these systems replacing human decision-making stems from our lack of understanding of their functions and processes. On the other hand, complete and total transparency does not necessarily foster confidence. Constant monitoring of processes suggests a shift from trust to user control, undermining confidence. Moreover, the availability of an overwhelming amount of information that users are unable to interpret can lead to increased opacity and mistrust rather than clarity²⁴. Conversely, there are instances when we appear to repose trust or reliance in it, specifically to avoid the burden of interpretation that transparency (at least as understood according to the social account) places on us.

The social account of transparency, however, appears to indicate that rather than seeking a prescriptive formula for achieving the optimal degree of transparency to foster confidence in the socio-technical system, the most promising approach is to strike a balance between attitudes. This entails cultivating in the human subject the capacity for interpretative engagement and in the medium a certain degree of transparency, which should be balanced with explicability and usability. The objective, therefore, should be to foster a *virtuous* relationship with technology.

5. *Virtue-Oriented Approaches*

A core tenet of virtue ethics is the assertion that a virtuous individual is capable of making decisions that are morally sound. The question thus becomes not “how should I behave?” but “what kind of person do I want to be?” The cultivation of virtues—being a virtue defined as “an acquired human quality the possession and exercise of which tends to enable us to achieve those goods which are internal to practice and the lack of which effectively prevents us from achieving such goods”²⁵—and the knowledge of when to exercise them through *phronesis* (practical wisdom) enables the betterment of the individual. It is not the case that we are good people if we perform just and moral actions; rather, it is the case that if we are virtuous, we will behave correctly. This perspective also extends to the relationship between humans and technologies: to use the technologies at our disposal virtuously (i.e. to be people who use them to behave virtuously, but also to use them to make the most of their potential) is to achieve the goods which are internal to techno-social practices.

²⁴ Cf. M. Ananny, K. Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, in «New Media & Society», vol. 20, n. 3, 2019, pp. 973-989: 979. This is the so-called “data smog” phenomenon. See also C. Birchall, *Radical Secrecy*, cit., p. 71.

²⁵ A. MacIntyre, *After Virtue: A Study of Moral Theory*, University of Notre Dame Press, Notre Dame 1984, p. 191.

Nevertheless, often “the technologies’ users lack such dispositions and habits of a virtuous person.”²⁶ Furthermore, the technologies themselves do not always appear to be designed with the intention of fostering or cultivating virtues. Firstly, it is necessary to consider which virtues should be cultivated in order to establish a more positive relationship with the contemporary socio-technical system.

In certain theoretical frameworks, trust is conceptualised as a virtue²⁷ that can inform our engagement with technology. In accordance with a virtue-oriented approach, a trust evaluative act is not contingent upon objective and unambiguous data; rather, it is shaped by an individual’s attitude towards technology and a number of factors such as “our beliefs, values, the nature of the problem, how we perceive the problem, prior decisions, availability of alternative options, etc.”²⁸ Nevertheless, this attitude must be founded upon the reliability of the technology in question²⁹. This perspective aligns with the account I previously outlined in section 1, which posits that confidence is a disposition shaped by reliability, and that it motivates the expression of evaluative trust based on the disposition itself, rather than on rational arguments.

Accordingly, a virtue-oriented approach should encourage the utilisation of technologies by prioritising the cultivation of virtue traits and fostering a robust disposition towards confidence. Confidence should, in turn, be nurtured by trustworthy technologies, as they too are, in a sense, virtue-oriented. Value Sensitive Design (VSD), for instance, could be taken as a design model, as its tripartite model reserves a prominent place for the incorporation of values useful for human flourishing within the technology.

VSD is a design theory that views techno-social processes as dynamic and iterative, engaging researchers, designers, engineers, and policy-makers in the design

²⁶ A. Bilal, S. Wingreen, R. Sharma, *Virtue Ethics as a Solution to the Privacy Paradox and Trust in Emerging Technologies*, in «Proceedings of the 3rd International Conference on Information Science and Systems (ICISS 2020)», pp. 224-228: 225.

²⁷ McCraw argues that epistemic trust, understood as an attitude of dependence and confidence in another subject considered trustworthy and authoritative, is a virtue that enables knowledge to be gained through witnessing (cf. B. W. McCraw, *Virtue Epistemology, Testimony, and Trust*, in «Logos & Episteme», vol. V, n. 1, 2014, pp. 95-102 and Id., *Proper Epistemic Trust as a Responsibilist Virtue*, in K. Dormandy (ed.), *Trust in Epistemology*, Routledge, London 2019, pp. 189-217). Hills, instead, argues that *trustworthiness* is a moral virtue in that it is a disposition to take responsibility for what is entrusted, requiring correct values, motives, and judgements, and contributing to a good life and a well-functioning society (cf. A. Hills, *Trustworthiness, Responsibility and Virtue*, in «The Philosophical Quarterly», vol. 73, n. 3, 2023, pp. 743-761). On trustworthiness as a virtue see also N. N. Potter, *How Can I Be Trusted?: A Virtue Theory of Trustworthiness*, Rowman & Littlefield, Lanham 2002. For a conception of *reliability* as an intellectual virtue – insofar as it implies a combination of cognitive ability, stability of character, and a response to reasons and foundations of justification, not merely an ability to arrive at true beliefs – see R. Audi, *Rational Belief: Structure, Grounds, and Intellectual Virtue*, Oxford University Press, Oxford 2015, pp. 85-97.

²⁸ A. Bilal, S. Wingreen, R. Sharma, *Virtue Ethics as a Solution to the Privacy Paradox and Trust in Emerging Technologies*, cit., p. 225.

²⁹ *Ibid.*

process. VSD follows a threefold methodology: conceptual investigation to identify stakeholders, their values, and potential conflicts; empirical investigation using social science methods to study stakeholders' values and motivations; and technical investigation to assess how technologies support or hinder these values. These investigations are interrelated, often conducted together, and iterated throughout the design and testing phases³⁰.

The concept of transparency also appears to be pertinent in this context. It is noteworthy that transparency is regarded as a socially oriented value in VSD, particularly in the development of responsible artificial intelligence³¹.

In this sense, particularly when considered within the context of the social account, transparency can be conceptualised as a virtue, or at the very least, as a means of exercising it, or of gaining trust and strengthening the disposition of confidence. It seems reasonable to posit that a certain degree of transparency on the part of the medium and an effort to increase interpretative capacities are necessary for the establishment of a virtuous and trusting relationship between humans and technology. However, there is also an alternative approach to transparency from the perspective of virtue ethics.

Shannon Vallor, for instance, links the new cult of transparency to *sousveillance*, understood as “contemporary culture of expanding, reflexive, and manifold forms of watching and being watched.”³² In this context, transparency is understood to encompass not only the transparency of institutions, instruments, and means, but also the transparency of individuals with regard to their own selves. The processes of datafication and self-tracking afford individuals access to information about themselves that they might otherwise be unaware of. In this way, proponents of the Quantified Self Movement might argue that these processes enable individuals to cultivate and enhance their own self.

While transparency can act as a deterrent to illegal or immoral behaviour, particularly if this transparency is imposed on institutions or companies that develop, for example, decision-making technologies or artificial intelligence models (increasing trust in such institutions or companies), it is also true that the demand for absolute transparency can lead to a loss of space for free moral and cultural play. Furthermore, the utilisation of technologies to perpetuate and reinforce moral and political practices that are perceived as virtuous can impede and obstruct moral and political experimentation. This, in turn, may contribute to a decline in confidence in the capacity to adapt³³.

³⁰ Cf. B. Friedman and D. G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*, The MIT Press, Cambridge (MA) 2019.

³¹ Cf. J. Dexe, U. Franke, A. A. Nöu, A. Rad, *Towards Increased Transparency with Value Sensitive Design*, in H. Degen, L. Reinerman-Jones, (eds.), *Artificial Intelligence in HCI. HCII 2020. Lecture Notes in Computer Science*, Springer, Cham 2020, pp. 3-15.

³² S. Vallor, *Technology and the Virtue: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press, Oxford 2016, p. 190.

³³ Cf. *ibid.*, p. 191.

The contemporary cult of transparency, driven by the ideas of the neutrality of technology and a certain degree of techno-fatalism³⁴, has the potential to become a significant vice, which could limit our ability to flourish. This is because it is based on a quantitative and supposedly universal concept of the self, which is assumed to be objective and devoid of the particular characteristics of individuals. Vallor does not oppose the examined life or the cultivated self, which are fundamental and preparatory to securing a good life, providing it with the possibility of enrichment and enlargement. Nevertheless, she points out that although every attempt to cultivate one's self is also shaped by social and cultural forces, this does not mean that a single universal model valid for all and a single technique for achieving the goal of a good life can be imposed on cultivation³⁵. The Quantified Self Movement³⁶, however, appears to pursue this objective through the means of a supposed objectivity and a quantitative model.

In this sense, transparency is not a virtue; rather, it is a vice to be avoided on the road to developing a virtuous relationship with technologies. Furthermore, the notion of transparency, particularly when interpreted through the lens of mere information disclosure, has the potential to disempower users. This is because it subjects them to an overwhelming influx of data, which may nudge their actions in ways that align with the agendas of policymakers or technology companies, without prompting them to reflect on their own intentions and goals³⁷. Such circumstances may result in a reduction of trust and an increase in suspicion of manipulation.

A further, more radical perspective on the relationship between trust and transparency is presented by C. Thi Nguyen³⁸. The author posits that, in fact, transparency and trust are radically alternative concepts. Vallor's concept of the "cult of transparency" (which can be seen as a cult of surveillance) impose to experts to effectively communicate their reasoning to non-experts, as Nguyen argues. However, it is inherent to the nature of expertise that the reasoning employed by experts is not always accessible to non-experts. This would result in experts communicating only that which can be publicly justified, undermining the practical application of expertise itself (what Nguyen calls the *epistemic intrusion argument*)³⁹.

Similarly, the imposition of transparency on individuals or communities will, paradoxically, result in an increase in the prevalence of untruths. This is because some individual or community deliberations are based on intimate reasons and discourses that are challenging to articulate outside the group to which they belong. Consequently, there is a tendency to fabricate reasons that are not entirely accurate in

³⁴ Cf. *ibid.*, p. 193.

³⁵ Cf. *ibid.*, p. 198.

³⁶ Cf. *ibid.*, pp. 198-202.

³⁷ Cf. *ibid.*, pp. 202-204.

³⁸ Cf. C. T. Nguyen, *Transparency is Surveillance*, in «Philosophy and Phenomenological Research», vol. 105, n. 2, 2022, pp. 331-361.

³⁹ This argument is similar to Birchall's argument on self-censorship: cf. C. Birchall, *Radical Secrecy*, cit., p. 71.

order to justify certain decisions in the name of transparency. This suggests that it is not possible to place trust in any individual or deliberating community that claims to be transparent.

When applied to technology, this can be illustrated as follows: (1) experts will only provide an explanation that is comprehensible to the public, therefore a complex technology can never be truly explicable; (2) despite the metaphors associated with AI, the technology itself “reasons” differently from a human and cannot therefore be totally transparent to the human, instead, it can at best deceive the human through supposedly transparent interfaces.

According to Nguyen, the inherent tension between transparency and trust is a genuine moral dilemma that cannot be easily resolved. Vallor presents a less pessimistic view, suggesting that flourishing is possible despite the challenges of transparency/surveillance⁴⁰. She proposes that this is contingent upon the cultivation of *technomoral virtues*⁴¹. In particular, the virtue that Vallor links to trust and reliability is *honesty*.

6. Conclusion: Towards Honest Technologies

Reliance on technology, while often measurable through performance outcomes, presents significant limitations. Its focus on results can detract from the importance of the processes involved, which are critical in establishing a trustworthy relationship with technology. Both trust and reliance place substantial expectations on the success of an interaction, but they do not guarantee it, which can lead to frustration. Moreover, the latter concept appears to depend heavily on transparency, while transparency seems to have at least an ambiguous relationship with transparency. This relationship, moreover, raises the unresolved issue of what conception of transparency should be applied. The specific understanding of transparency can significantly alter—or even undermine—the type of reliance or trust that is developed. Transparency, in this sense, can even negatively impact our relationship with technologies and those interactions mediated by them.

Instead, technology should aspire to be honest—maintaining and making evident the sense of interaction and encouraging hermeneutical attention⁴². This aligns more closely with a social account of transparency, where the relationship among content, viewer, and medium must be based on honesty. I propose that technologies fostering such honest interactions be termed “honest technologies.”

By this terminology I mean to refer to technologies that are not only (or perhaps not so much) transparent, but that do not conceal mediation or interaction.

⁴⁰ Cf. S. Vallor, *Technology and the Virtue*, cit., pp. 204-207.

⁴¹ In her taxonomy of technomoral virtues, Vallor identifies a total of 12 virtues: Honesty, Self-Control, Humility, Justice, Courage, Empathy, Care, Civility, Flexibility, Perspective, Magnanimity, Technomoral Wisdom (cf. *ibid.*, p. 120).

⁴² Cf. F. Striano, *Towards “Post-Digital”: A Media Theory to Re-Think the Digital Revolution*, in «Ethics in Progress», vol. 10, n. 1, 2019, pp. 83-93: 90.

The kind of honest technologies I envisage do not seek explicability at any cost, or at least not to the public. They must, however, make the processes of mediation and interaction evident and negotiate with the user the kind of information the user has an interest in accessing. Technologies designed to be honest, in short, must not sacrifice mediation awareness in the name of usability, and their honesty, combined of course with reliability, will be a far more powerful confidence booster than any ambiguous call for transparency.

Furthermore, when considering human relationships *mediated by technology*, honesty is once again central. The balance between community building and the right to privacy, including the right to be forgotten, is delicate. A blanket call for transparency may not be the solution; in fact, it may lead to an overexposure that contradicts an individual's desire for privacy. Here again, the virtue of honesty, as a technomoral and social virtue, becomes crucial. Drawing from Vallor's insights, honesty requires "*an exemplary respect for truth*" (aligned with the Buddhist concept of "Right View") and "*the practical expertise to express that respect appropriately in technosocial contexts*"⁴³ (akin to "Right Conduct"). The invitation to behave honestly thus provides a superior resolution to the tension between community ties or expert reasoning on the one hand and the demand for truthful information on the other, as highlighted by Nguyen, than the call for transparency.

This is not to suggest that there are no circumstances in which transparency should be demanded or that it is invariably a vice of the contemporary communication model. The practice of full disclosure can be perceived as a form of viciousness when it involves the indiscriminate sharing of information "without discernment or contextual sensitivity:"⁴⁴

an honest scientist must know her audiences and venues, and there are times when excessive precision or transparency will obscure the truth rather than reveal it. [...] Still there are circumstances in the life of government, academia, business, and other institutions—even the private lives of citizens—that justify opening the books for inspection, so to speak. We need the technomoral wisdom to make more intelligent and practically discerning use of the new technologies that can make opening the books in such circumstances easier.⁴⁵

Vallor invites us to reflect also on who our models of technomoral honesty might be⁴⁶. I extend this reflection by considering the influence that technology itself can have on our traits, conduct, and habits. Without succumbing to radical technodeterminism and admitting that the reasons for an act must be internal⁴⁷, we can acknowledge that, just as culture and examples shape our cultivation of virtues

⁴³ Cf. S. Vallor, *Technology and the Virtue*, cit., p. 122.

⁴⁴ *Ibid.*, p. 205.

⁴⁵ *Ibid.*, pp. 205-206.

⁴⁶ Cf. *ibid.*, pp. 122-123.

⁴⁷ Cf. S. van Hooft, *Caring: An Essay in the Philosophy of Ethics*, University Press of Colorado, Niwot 1995, pp. 140 and ff.

and internalization of motivations, so too can our habitual use of technology. If honest individuals design honest technologies, it is also true that habitual interaction with honest technologies can cultivate honesty in individuals.