Multimodal Technologies and Interaction

MDPI

Perspective

Challenges and Trends in User Trust Discourse in AI Popularity

Sonia Sousa [1,2,3,*,†], José Cravino [2,4,†] and Paulo Martins [2,3]

[1] School of Digital Technologies, Tallinn University, Narva Mnt 25, 10120 Tallinn, Estonia
[2] Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[3] INESC TEC-Institute for Systems and Computer Engineering, Technology and Science, 4000-008 Porto, Portugal
[4] CIDTFF—Centro de Investigação em Didática e Tecnologia na Formação de Formadores, Universidade de Aveiro, 3810-193 Aveiro, Portugal
* Correspondence: scs@tlu.ee
† Writting All and conceptual S.S.

Abstract: The Internet revolution in 1990, followed by the data-driven and information revolution, has transformed the world as we know it. Nowadays, what seam to be 10 to 20 years ago, a science fiction idea (i.e., machines dominating the world) is seen as possible. This revolution also brought a need for new regulatory practices where user trust and artificial Intelligence (AI) discourse has a central role. This work aims to clarify some misconceptions about user trust in AI discourse and fight the tendency to design vulnerable interactions that lead to further breaches of trust, both real and perceived. Findings illustrate the lack of clarity in understanding user trust and its effects on computer science, especially in measuring user trust characteristics. It argues for clarifying those notions to avoid possible trust gaps and misinterpretations in AI adoption and appropriation.

Keywords: human-computer Interaction; trust; human-to-human relationship; human-to-technology relationship

Citation: Sonia, S.; Cravino, J.; Martins, P. Challenges and Trends in User Trust Discourse in AI Popularity. *Multimodal Technol. Interact.* **2023**, *7*, 13. https://doi.org/10.3390/mti7020013

Academic Editors: Jan Auernhammer, Takumi Ohashi, Di Zhu, Kuo-Hsiang Chen and Wei Liu

Received: 8 December 2022
Revised: 26 January 2023
Accepted: 27 January 2023
Published: 31 January 2023

1. Introduction

Current digital transformation events forced new digital interaction patterns that did not exist 10 or 20 years ago, impacting how we play, work, and build communities. As a result, society, organizations, and individuals must rapidly adapt and adjust to these new digital norms and behaviours. In the short term, the division between online and physical activities diminished, increasing the capacity to act in a larger digital market and society. Consequently, these digital transformation events forced us to become more vulnerable to the actions of digital parties without adequately understanding or being able to assess their risk, competency, and intentions. Digital social media platforms like Facebook (2004), Twitter (2006), and Youtube (2005) or messaging services such as WhatsApp (2009) have promoted this new era of communication that resulted in continuous attempts to subvert their original purposes (i.e., malicious acts). Examples of this can be the generation of mistrust against vaccines, the creation of content supporting climate denial theories or disinformation campaigns, or the surge in data breaches [1–5].

This attempt of '*science and engineering of making intelligent machines*', as McCarthy [6] conceptualized Artificial intelligence (AI), resulted in the spread of software that uses computer technology to generate outputs such as content, predictions, recommendations, or decisions influencing the environment they interact. Moreover, adding to this notion of AI shared by the European Commission's Artificial Intelligence Act (https://artificialintelligenceact.com/, accessed on 6 December 2022) is the fact that nowadays, we can find AI systems that mimic, think and act like humans [7]. What highlights the potential of those AI mechanisms, not just as information revolution tools but as well as data threats, take the example reported in books like '*Weapons of math destruction: How big data increases inequality and threatens democracy*', or '*The AI delusion*'.

Therefore, what once was seen as science fiction has become a reality, emphasising people's fears and mistrust of the possibility of machines dominating the world. Those fears have led to the surge of new reports, articles, and principles that seek more trustworthy AI (TAI) visions that provide a Human-centered vision of users' trust and socio-ethical characteristics towards AI [5,8–12]. It also increased the discourse on AI and the need to complement existing data protection regulatory mechanisms (e.g., GDPR-ISO/IEC 27001) https://gdpr-info.eu/, accessed on 6 December 2022). It also highlights the need for seeking new 'responsible' AI ethical solutions that are less technical and profit-oriented.

This Human-centered vision of AI, however, is hard to understand, especially if, like Bryson [13], we try to make accountable the service providers, in fact, to the detriment of the default, mainstream attention is given to it for AI providers, developers, and designers, it is unclear how to ensure that the AI system designed by humans can be reliable, safe, and trustworthy [14]. More, AI's complexity makes it challenging to guarantee that AI is not prone to incorrect decisions or malevolent surveillance practices. Like the GDPR, as AI's popularity increase, it also increases its potential to create opaque practices and harmful effects and AI's unpredictability, making it hard to be audited, externally verified, or question (i.e., black box) [15]. Additionally, AI's unpredictability makes it difficult to avoid unforeseen digital transformations, harmful practices, and risks. It also makes it hard to predict behaviour and identify errors that can lead to biased decisions, unforeseen risks, and fears [10,16–19].

In conclusion, the increase in AI's popularity also increased its complexity, the number of decentralized and distributed systems solutions, increased as well the AI's opacity and unpredictability. When mixed with poor design practices, these AI characteristics can produce vulnerable interactions that lead to further breaches of trust (both real and perceived). With this work, we aim to share our vision regarding the challenges and trends in user trust discourse in AI popularity from a Human-Computer Interaction (HCI) perspective. Results presented are supported by the author's work in mapping the trust implications in HCI during the past decade and situated in the context of three recent systematic literature reviews performed on trust and technology [20–22]. Hoping that this clarifies the nature of user trust in recent AI discourse (RQ) and also avoids designing vulnerable AI artefacts that build on trust without understanding its influence in the uptake and appropriation of those AI complex systems. This work's main contribution is to link the previous trust in technology discourse with recent AI popularity and user trust trends. Then, it illustrates the importance of providing an HCI perspective of user trust discourse. Finally, establish a link between past trust in technology practices, current thoughts, and research gaps.

1.1. AI Popularity and the Discourse on Users' Trust

The recent waves of technology innovations are marked by AI popularity, the social network revolution, distributed data, automated decision-making, and the ability to trace and persuade behaviours [18,23–26]. These AI information-driven revolutions recently resulted in the spread of AI complex and distributed software solutions that generate automated and unpredictable outputs that cannot guarantee that they are not prone to provide incorrect content, predictions, or recommendations or mislead people into incorrect decisions that can have potentially harmful consequences in environments they interact, like malevolent surveillance practices and disinformation practices [18,27–32].

This technological revolution wave also resulted, in an increased discourse toward trust in AI, seeking solutions to regulate, diminish people's fears, and guarantee a user trust approach to the topic. Take the example of the European Commission draft EU AI act (https://artificialintelligenceact.com/, accessed on 6 December 2022), the Organisation for Economic Co-operation and Development (OECD), the Business Machines Corporation (IBM) and their efforts to clarify the Trustworthy AI (TAI) principles and practices [11,33,34].

This increase and new TAI discourse challenge HCI practitioners, a need to establish new trust boundaries (e.g., regulations, socio-ethical practices, etc.) to ensure Humans' abil-

ity to monitor or control their actions [16,35]. However, like AI, addressing trust in technology can be a complex subject for non-experts, as it acknowledges the deterministic models (that aggregate system technical characteristics) and the human-social characteristics that envision trust through a set of indirect parameters. This has raised another challenge to AI popularity, seeking solutions to trigger users' trust in AI [10,16–19]. However, with the increased popularity of AI software, it is unavoidable for society to be susceptible to its opaque practices, which can lead to further breaches of trust. Adding to this, the fast spread and dependency on AI prevent individuals from fully grasping the intricate complexity of those machines' capabilities, which can lead to potentially harmful consequences. Consequently, we believe that a new trend in user trust research will be revealed, similar to the rise of the Special Interest Group in 1982, to address the need for the design of Human-Computer Interactions (e.g., SIGCHI). This will lead to new international standards, expert groups, and international norms to tackle this problem. Take the example of the high-level expert group (AIHLEG) (https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai, accessed on 6 December 2022) [34]. Or the Working group 3—trustworthiness referred to in the international standards for Artificial intelligence (ISO/IEC JTC 1/SC 42/WG 3) (https://www.iso.org/committee/6794475.html, accessed on 6 December 2022).

The above initiatives attempt to defy the surveillance capitalism practices and fight the corporate approach to data as the new oil of this century, seeking short-term profits without considering the ethical consequences of their actions. However, their focus is on tackling the AI problem and not so much focus on understanding or mapping the influence or consequences of user trust in its adoption and appropriation practices. The recent EU's AI act (https://artificialintelligenceact.com/, accessed on 6 December 2022) shares a broader vision of this problem and represents an attempt to incorporate the notions of risk and trust within AI's characteristics like complexity, opacity, unpredictability, autonomy, and data-driven qualities. highlighting the need for finding new user trust solutions that foster feelings of safety, control, and trustworthiness in current AI solutions.

In their regulatory scope (i.e., ethical guidelines for trustworthy AI), the EU encourage building public trust as an investment for ensuring AI innovation and respecting fundamental rights and European values. It also classifies trust as a need to be understood within four potential AI risks to health, safety, and fundamental rights from minimal or no risk, AI with specific transparency obligations (e.g., 'Impersonation' (bots)), High risk (e.g., recruitment, medical Devices), and an unacceptable risk (e.g., social scoring).

Those demand different Trustworthy AI (TAI) frames to ensure public trust in AI computing speech or facial recognition techniques in applications like social chatbots, human-robot interaction, etc. For AI providers and non-expert in trust, however, it is challenging to fully understand the user trust influence in AI acceptance and appropriation, as current, trustworthy AI principles provide a set of requirements and obligations with an unclear trust notion, sometimes associated with notions of ethics and accountability. In sum, for now, the EU's AI act is a very recent regulatory framework, but those principles are likely to be extended to the world, similarly to the GDPR. If so, it becomes unavoidable to clarify the nature of user trust in recent AI discourse (RQ1). Including clarifying the link between past trust in technology practices, current thoughts, and research gaps.

1.2. TAI Conceptual Challenges

The above-described misconceptions and malevolent practices, followed by the EU AI Act draft and adopting a risk-based approach (unacceptable risk, high risk, & limited or minimal risk), raised the need for addressing the challenges and trends in user trust discourse in AI. As well as for providing further conceptual clarifications and strategies that demystify the discourse of trust and socio-ethical considerations, user characteristics when facing risk-based decisions, and design and technical implementations of trust [22,36,37]. Avoiding trust in AI solutions that are marrow framed from technical or single constructs like explainable AI (XAI), privacy, security, or computational trust. That eventually cannot

guarantee that humans do not misinterpret the causality of complex systems with which they interact and lead to further breaches of trust and fears [27,38].

Take, for instance, the following socio-ethical considerations design toolkits, guidelines, checklists, and frameworks whose goal is to bring more clarity to the problem. Like the IDEO toolkits to established by the entitled trust Catalyst Fund (https://www.ideo.com/post/ai-ethics-collaborative-activities-for-designers, accessed on 6 December 2022), and the IBM Trustworthy AI, a human-centered approach (https://www.ibm.com/watson/trustworthy-ai, accessed on 6 December 2022), or [39], an agile framework for producing trustworthy AI was detailed in [40], and an article entitled Human-centered artificial intelligence: Reliable, safe & trustworthy was presented by Shneiderman (2020) [16]. Similarly the Assessment List for Trustworthy Artificial Intelligence (ALTAI) (https://altai.insight-centre.org/, accessed on 6 December 2022), and the EU Ethics guidelines for trustworthy AI (https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1, accessed on 6 December 2022). They neither offer clarity on trust notions nor explain how the proposed practices leverage user trust in AI.

As Bach et al. [22] and Ajenaghughrure et al. [20] findings confirm, measuring trust remains challenging for researchers. Currently, there is more than one way to define and measure trust. According to Bach et al. [22], Out of the 23 empirical studies explored, only seven explicitly define trust. At the same time, eight conceptualize it, and the remaining nine provide neither. Therefore, user trust is still an underexplored topic. According to Ajenaghughrure et al. [20], there is still a lack of clarity on measuring trust in real-time, and few solutions provide stable and accurate ensemble models (results from 51 publications). Those that exist are narrow, subjective, and context-specific, leading to an oversupply of models lowering the adoption. Especially when using psychophysiological signals to measure trust in real time.

This phenomenon happens despite computational trust research emerging in 2000 as a need to provide a technical-centred and automated way to tackle the trust factors in system design, i.e., to authenticate trustee's reputation and credibility [41]. Ultimately, and to avoid the past mistake of forward-push regulations, trust measures that ultimately are technically implemented without considering the tensions between the current state of creating new technical innovations, profit-oriented deployment, and its socio-technical implications across societies. Researchers need to look beyond the technical-centred vision of trust in computing and produce new user trust notions that help clarify the role of trust in these new technical profit-oriented AI innovations. As Rossi [42] (p. 132) argues, to fully gauge AI potential benefits, trust needs to be established, both in the technology itself and in those who produce it, and to put such principles to work, we need robust implementation mechanisms. Yet, researchers need to ensure its proper application by providing frameworks that clarify its implementation and avoid misinterpretation and misguided implementations [16,19]. Claiming once more for a shift from emphasis system's technical qualities toward human-centred qualities, similar to the move between usability and user experiences, i.e., from a focus on design features towards a focus on experiences of use [43–45].

2. Discussion

The challenges mentioned above have shifted current literature discourse towards Human-centered design (HCD) as a strategy to address the socio-technical characteristics of design features and lessen misinterpretation gaps in regulatory practices. These needs are followed by a need to clarify the current trust lenses of analysis, as trust can be a key to lessening the risks to the development, deployment, and use of AI in the EU or when it will affect people in the EU. However, as seen in the literature, trust divergent notions can prevent non-experts from adequately addressing this need from an HCD perspective, which can lead to an increased risk of failure, increasing its misinterpretation gaps that can be more harmful than good [46].

Therefore, needs and challenges that were not recognized 10 to 20 years ago are now a reality, which can create gaps in IT education. Currently, few curricula contemplate this socio-technical view or Human-centered design vision nor the ethical focus on measuring the risks of their potential misuse. As a result, IT and AI specialists might not be equipped with the necessary skills to address the challenges mentioned above, let alone know how to deal with this topic's complexity and application challenges, i.e., the Trustworthy AI (TAI) risk-based approach promoted by the EU. In that regard, despite agreeing that HCI researchers can contribute to broadening this analysis and helping IT, specialists adopt more user-centred strategies to prevent building systems that can lead to risky, unsafe life or long-term threatening social ramifications like the examples presented above [16,47]. They also need novel theories and validated strategies to address the socio-technical effects of trust in System complexity.

Like in the past, the focus shifted from measuring the usability characteristics of a system (e.g., efficiency and effectiveness) towards or focusing on hedonic characteristics (e.g., emotion and satisfaction), and now to a risk-based approach where trust is part of users' experiences. However, this needs to be followed by clear notions of trust, psychologically validated analysis, and associated underlying theories in context [36,42,43,48]. Trust, like satisfaction, is a human characteristic, not a machine characteristic. Past narrow views and assumptions on trust in technology might not fit in current Human-centered TAI applications [36]. A vision highlighted and shared in Figure 1, based on a culmination of various works performed in the past ten years (e.g., literature reviews, participatory research, teaching, supervising, etc.) to understand the nature of user trust in Human-Computer Interaction (HCI) [35,49–56].

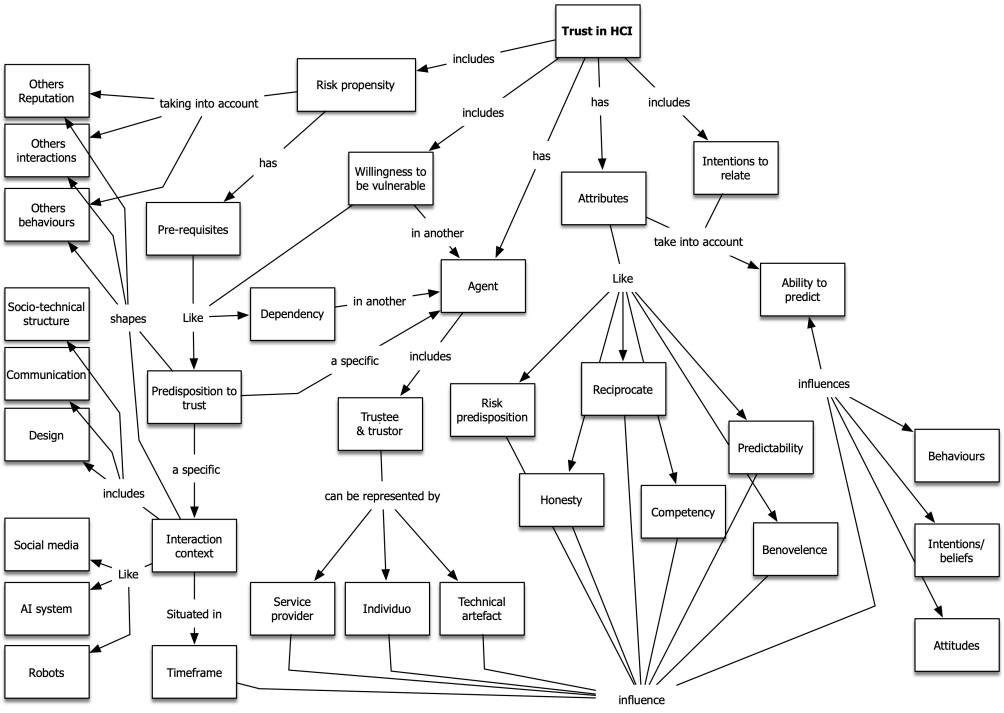

Figure 1. The nature of trust in HCI: Conceptualization.

The Nature of Trust Research in HCI

Trust in HCI, as illustrated in Figure 1, is a social-technical quality that supports the interrelationship between a trustee (the entity being trusted) and the trustor (the person who trusts). Trusting is a will to be vulnerable. Note that vulnerability implies risk acceptance based on how well the 'trustor' can perceive the 'trustee' as trustworthy, as Mayer et al. [35]. However, past views of trust tend to associate it with single constructs, like 'credibility', 'privacy', and 'security'. Associating user trust as a characteristic of disclosure of certain

types of information (i.e., privacy) or preventing cybersecurity threats to ensure system trustworthiness [57].

Maybe a reason why trust notions and applications in computer science literature provide an oversupply of trust visions, solutions, and applications. Take the example of Sousa et al. [21] findings (results from 69 publications) that reveal that trust notions and solutions can differ and depend on the application context. Mainly trust is addressed as a quality to influence technology adoption, data sharing credibility, and positively influencing user's intentions and behaviours. Take, for example, how trust is addressed within the privacy and security research topic. Herein, researchers see trust as avoiding potential misuse and access to personal data. It is sometimes mentioned as an information credibility attribute. Trust visions in e-Commerce, eHealth, or eGovernment are connected with 'risk', 'reputation', and 'assurance' mechanisms to establish loyalty, minimize risk and uncertainty and support decision-making. Solutions range from data-driven trust models to observing the impact of trust in encouraging decision-making and encouraging technology adoption (e.g., commercial transactions, diagnostic tools, adoption of services, etc.). In social networks, trust emerged as a way to sustain interaction processes between members of actor networks in emerging scenarios and argue that trust contributes to promoting the regulation of interaction processes. Trust is also useful in creating sustainable computer support collaborative work to support interpersonal interactions online.

Regarding its associated concepts, trust is associated with transparency, assurances, data credibility, technical and design feature, trustworthiness, users' predispositions to trust, explicability, etc. Mainly ways to reduce the uncertainty and risk of unpredictable and opaque systems, e.g., speech and facial recognition systems, crewless aerial vehicles (e.g., drones), IoT applications, or human-robot interactions (HRI). However, most trust studies present a narrow and simplified view, focusing on data-driven computational trust mechanisms to rate a system or a person as reliable. Presenting a view of trust as rational expectation, person, object, or good reliability or credibility when a first encounter occurs and no trust has been established, i.e., establish trust between two strangers. Discarding more complex aspects o trusted relations through time, the Human-system relationship is established through various indirect attributes like risk perceptions, competency, and benevolence [52,58–61].

The above paragraph illustrates the pertinence of providing new user trust visions that can be adjusted to new digital AI regulations, behaviours, and innovations. It also illustrates the complexity of both subjects, AI and user trust. On the one hand, the new EU AI act sees public trust as a way to guarantee AI innovations, guaranteeing that it is not prone to high risks like leading users to incorrect decisions or malevolent surveillance practices. On another, the AI providers are not experts in trust in technology, which make it hard for them to acknowledge the deterministic models (that aggregate system technical characteristics) and the human-social characteristics that envision trust through a set of indirect parameters. In literature, for instance, trust is associated with narrow views like 'reputation', 'privacy', and 'security'. Literature on trust and computing also comes associated with computational trust model [62].

With the same regard to security and privacy measures and their role in fostering AI trustworthiness, recent malevolent use demonstrates that new visions need to be adjusted to prevent mistrust in technology. Just addressing trustworthy AI measures as a way of preventing intrusion, allowing the individual the right to preserve the confidentiality, integrity, and availability of information might not be enough within today's socio-technical complexity [63,64]. Privacy refers to the individual's right to determine how, when, and to what extent they will disclose their data to another person or an organization [65]. As Figure 2 illustrates, user trust considers Socio-ethical considerations, Technical artefact, Application context, and Trustee & trustor characteristics [21,22,66–69].

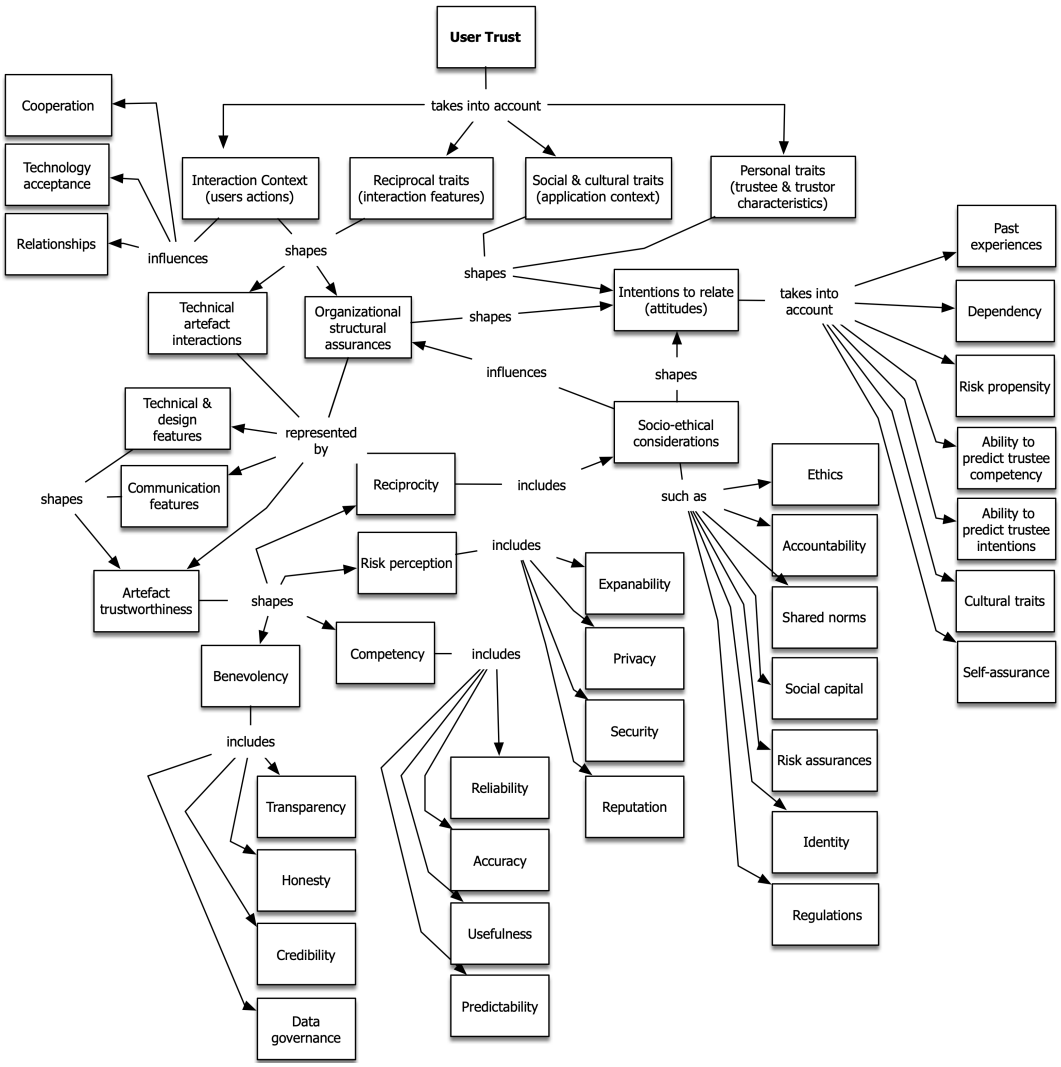

Figure 2. The user trust influence in shaping the interaction context: Conceptualization.

A trustful relationship requires assessing the risks (i.e., gains and losses). Requires evaluates the tool's ability (e.g., competence, reliability, accuracy, etc.) to perform the desired outcomes; and assesses if an entity (i.e., company, organization, institution, etc.). Requires individuals exceptions that digital relationships follow expected social norms and ethical principles. For instance, trustworthiness is an attribute or quality of a system, a person, or an organizational structure. As the Oxford dictionary describes it, **Trustworthiness** is the quality or fact of being trustworthy (=able to be trusted). Work like the Assessment List for Trustworthy Artificial Intelligence (ALTAI), or Trustworthy AI (TAI), is human-centred, or even Human-centered artificial intelligence: Reliable, safe & trustworthy do not address it or only address it from a shallow view.

On the other hand, if **Trust** is to believe that someone is good and honest and will not harm you or that something is safe and reliable. **Trustworthy**, on the other hand, is the ability to be trusted, and trustworthiness is a judgment of trust. Trusting reflects the strength of a person's belief that someone else will help them to achieve an acceptable outcome [70,71]. Trustworthiness and trustworthy are characteristics of trust, and in any complex construct, both (qualities and characteristics) are measured through indirect and direct interrelated factors TAI regulations are one example.

Measuring the attribute or quality of a system (e.g., privacy or security) might not be enough to address it. Or, for instance, take the example of system explainability (XAI) or computational trust models called by some reputation mechanisms [62,72]. As Davis et al. [27] claim, some technical-centred explanations might mislead individuals'

understanding of the subject. As Dörner [38] work illustrates, in some cases, humans' limitations to understanding a complex subject might prevent them from misunderstanding their work. As Sousa [73] result revealed, the interrelations between trust and performance can be negative, i.e., the higher the trust, the lower the performance. Yet, some limited-risk applications do not prevent them from using and benefiting from these tools. I do not need to understand a car's or aeroplane's mechanics to trust and use it. Individuals already (successfully) interact with complex AI systems daily. But I should be aware of its potential threats to making knowledgeable decisions, especially when adopting an AI system leads to an unacceptable or high-risk approach as the EU act describes it. Thus, to successfully maintain trust in specific events, we should not look at it from a narrow technical perspective. User trust in AI (i.e., technical artefact) can also be influenced by users' cultural traits, past experiences, and applications context [36,42].

Therefore, it is important to include both visions: **Trust as a personal trait**, understood as a perceived skill set or competencies of trustee characteristics (e.g., how teachers are perceived in a school system); **Trust as a social trait**, understood as the mutual "faithfulness" on which all social relationships ultimately depend. **Trust** reflects an attitude, or observable behaviour, i.e., the extent to which a trustee can be perceived in society. For instance, to want to do good, be honest – 'benevolence.' Follow privacy and security regulations. **Trust as reciprocal trait**, closely related with ethical and fair. For instance, the extent to which the trustee adheres to a set of principles that the trustor finds acceptable—for instance, an economic transaction.

This led to another application challenge, trust measurements [21,22]. Current trust misconceptions lead to an oversupply of computational trust assessment models that can only be used in narrow AI applications. Han et al. [74] recommendation trust model for P2P networks and Hoffman et al. [75] trust beyond security: an expanded trust model is an example of that. Some, however, measures of trust across gender stereotyping and self-esteem indicate that trust can be measured in a broader socio-technical perspective [76,77]. The EU self-assessment mechanisms for Trustworthy Artificial Intelligence created by the AIHLEG expert group is another example [78] of broadening the view. The same regards the Human-Computer trust (HTC) psychometric scales proposed by Madsen and Gregor [79], SHAPE Automation trust Index (SATI) [80]. We need new HCI mechanisms to measure potentially faulty TAI design practices, which can lead to risky, unsafe life or threatening social ramifications [16,47]. If not, HCI researchers might continue looking for specific and narrow solutions that can fail when applied in broader contexts. An example is the latest pursuit of AI system explainability (XAI) or computational trust models might not be enough to foster trust in users. On the contrary, some technical-centred explanations might mislead individuals' understanding of the subject. Or, some computational trust models are so narrow in their application that they might successfully maintain the initial trust formation in e-commerce but are not valuable for e-health.

Generally, looking at the above concepts, many researchers understand trust as a specific quality of the relationship between a trustee (the entity being trusted) and the trustor (the person who trusts). In other words, the trust dynamic contemplates a subtle decision based on the game's complexity that they find themselves playing, as Bachrach and Zizzo [81] and Luhmann [56] describe it. However above paragraph also represents the need for the trustor (i.e., human) to perceive the trustee as trustworthy. For instance, this system can have all the technical mechanisms to be secure, but if the trustor cannot see those who see these mechanisms in action, they might perceive that they are not in place. So, trustworthiness and being trustworthy are two complementary aspects of trust. Perceived trustworthiness is an individual's assessment of how much an object, a person, or an organization, can be trusted. People assess trustworthiness based on information they perceive and or receive influenced by cultural and past experiences, and both qualities can evolve through time. In conclusion, in the Socio-ethical AI context, trust notions still need further clarification to ensure that solutions foster public trust and fundamental rights for minimal or no risk in the AI data protection process and non-bias

(see EU's AI act). Same regards how we connect trust with information credibility and ethical practices. As well as studying the socio-ethical AI implications (i.e., explainable AI) in the acceptance and use of the technology. Or even when using a trust to seek more control in automated and intelligent system predictions. Or, provide socio-ethical AI as transparency and responsibility solutions through trusted agencies and other audition mechanisms [21,82].

3. Conclusions

The first digital revolution, i.e., Internet revolution in 1990, has brought big changes to how we communicate and interact across-country. However, recent digital revolutions characterised by the data-driven and information revolutions transformed the world and society as we know it. AI systems enabled by the social network, followed by the ability to trace and persuade behaviours, have altered social democratic practices and applications. The challenge nowadays is finding ways to adjust current regulatory practices to these new digital practices. Including looking for ways to fight the advancement of potential AI malpractices and minimize the risk of malevolent use.

The above findings reveal the importance of clarifying the user trust notions in recent AI discourse. This is to lessen possible misinterpretation of trust and notion gaps in these new ethical guidelines for trustworthy AI regulatory practices. This is to avoid misconceptions about user trust in AI discourse and fight the tendency to design vulnerable interactions that lead to further breaches of trust, both real and perceived. Provide also evidence of the lack of trust and understanding of computer science, especially in assessing trust user characteristics and user-centred perspectives Ajenaghughrure et al. [20], Sousa et al. [21], Bach et al. [22]. Also, frame the term 'trustworthy computing' as critical for technology adoption and complex system appropriations. As Shneiderman [16] stresses, we need to conceive a more Human-Centered Artificial Intelligence (HCAI) paradigm where human-machine relationships are perceived as safe, reliable, and trustworthy.

We are now acknowledging that despite the attention given to technical characteristics like 'privacy', 'security', or 'computational trust and reputation', malevolent technological practices still prevail. Also, widespread AI-driven applications and their associated risks bring new challenges to distrust and fear across-country discourse. Take the examples of persuasive malevolent design, deceptive designs, unethical business decisions associated with increasing concerns on socio-ethical AI, technical and design features, and user characteristics that Bach et al. [22] work to mention. Another challenge addressed is the human-likeness that misguides users to misplace the attributes of a trusted human, human-to-human exchanges mediated through technology and their trust in a human-artefact relationship [83–86]. Even though some researchers claim that 'people trust people, not technology, as technology does not possess moral agency and the ability to do right or wrong [56,87–90]. Researchers fail to acknowledge trust complexity and how its indirect measures affect users' trust perceptions in the system's adoption and appropriation.

After two decades of investment and advances, we now change the discourse towards a human-centred view and the need to develop, deploy and measure the quality of trust perceptions from an HCD perspective. Addressing AI-related attributes like reliability, safety, security, privacy, availability, usability, accuracy, robustness, fairness, accountability, transparency, interpretability, explainability, ethics, and trustworthiness.

Funding: This work was supported by the Trust and Influence Programme [FA8655-22-1-7051], the European Office of Aerospace Research and Development, and the US Air Force Office of Scientific Research. This study was partly funded by AI-Mind, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964220.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Appari, A.; Johnson, M.E. Information security and privacy in healthcare: Current state of research. *Int. J. Internet Enterp. Manag.* **2010**, *6*, 279–314. [CrossRef]
2. Oper, T.; Sousa, S. User Attitudes Towards Facebook: Perception and Reassurance of Trust (Estonian Case Study). In *HCI International 2020-Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, 19–24 July 2020, Proceedings, Part III 22*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 224–230.
3. Sousa, S.; Kalju, T. Modeling Trust in COVID-19 Contact-Tracing Apps Using the Human-Computer Trust Scale: Online Survey Study. *JMIR Hum Factors* **2022**, *9*, e33951. [CrossRef] [PubMed]
4. Sousa, S.; Bates, N. Factors influencing content credibility in Facebook's news feed. *Hum.-Intell. Syst. Integr.* **2021**, *3*, 69–78. [CrossRef]
5. Sundar, S.S. Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *J. Comput.-Mediat. Commun.* **2020**, *25*, 74–88. [CrossRef]
6. McCarthy, J. *What Is Artificial Intelligence?*; Springer: Dordrecht, The Netherlands, 2007.
7. Russell, S.; Norvig, P. A modern, agent-oriented approach to introductory artificial intelligence. *Acm Sigart Bull.* **1995**, *6*, 24–26. [CrossRef]
8. Xu, W.; Dainoff, M.J.; Ge, L.; Gao, Z. From Human-Computer Interaction to Human-AI Interaction: New Challenges and Opportunities for Enabling Human-Centered AI. *arXiv* **2021**, arXiv:2105.05424.
9. WIRED. AI Needs Human-Centered Design. 2021. Available online: https://www.wired.com/brandlab/2018/05/ai-needs-human-centered-design/ (accessed on 6 December 2022).
10. Hickok, M. Lessons learned from AI ethics principles for future actions. *AI Ethics* **2021**, *1*, 41–47. [CrossRef]
11. Watson, I. Trustworthy AI Research. 2021. Available online: https://research.ibm.com/topics/trustworthy-ai (accessed on 6 December 2022).
12. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. In *Ethics, Governance, and Policies in Artificial Intelligence*; Springer: Cham, Switzerland, 2021; pp. 19–39.
13. Bryson, J.J. The artificial intelligence of the ethics of artificial intelligence. In *The Oxford Handbook of Ethics of AI*; Oxford University Press: Oxford, UK, 2020; p. 1.
14. Shneiderman, B. Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy Human-Centered AI systems. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2020**, *10*, 1–31. [CrossRef]
15. Zhang, Z.T.; Hußmann, H. How to Manage Output Uncertainty: Targeting the Actual End User Problem in Interactions with AI. In Proceedings of the IUI Workshops, College Station, TX, USA, 13–17 April 2021,.
16. Shneiderman, B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int. J. Hum. Interact.* **2020**, *36*, 495–504.
17. Araujo, T.; Helberger, N.; Kruikemeier, S.; De Vreese, C.H. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI Soc.* **2020**, *35*, 611–623. [CrossRef]
18. Lopes, A.G. HCI Four Waves Within Different Interaction Design Examples. In Proceedings of the IFIP Working Conference on Human Work Interaction Design, Beijing, China, 15–16 May 2021; Springer: Cham, Switzerland, 2022; pp. 83–98.
19. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
20. Ajenaghughrure, I.B.; Sousa, S.D.C.; Lamas, D. Measuring Trust with Psychophysiological Signals: A Systematic Mapping Study of Approaches Used. *Multimodal Technol. Interact.* **2020**, *4*, 63. [CrossRef]
21. Sousa, S.; Cravino, J.; Lamas, D.; Martins, P. Confiança e tecnologia: Práticas, conceitos e ferramentas. *Rev. Ibérica Sist. Tecnol. InformaÇ* **2021**, *45*, 146–164.
22. Bach, T.A.; Khan, A.; Hallock, H.; Beltrão, G.; Sousa, S. A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *Int. J. Hum. Interact.* **2022**, *38*, 1095–1112. [CrossRef]
23. Li, L.; Ota, K.; Dong, M. Humanlike driving: Empirical decision-making system for autonomous vehicles. *IEEE Trans. Veh. Technol.* **2018**, *67*, 6814–6823. [CrossRef]
24. Haigh, T. Remembering the office of the future: The origins of word processing and office automation. *IEEE Ann. Hist. Comput.* **2006**, *28*, 6–31. [CrossRef]
25. Harrison, S.; Tatar, D.; Sengers, P. The three paradigms of HCI. In Proceedings of the Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 28 April–3 May 2007; pp. 1–18.
26. Bødker, S. When second wave HCI meets third wave challenges. In Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, Oslo, Norway, 14–18 October 2006; pp. 1–8.
27. Davis, B.; Glenski, M.; Sealy, W.; Arendt, D. Measure Utility, Gain Trust: Practical Advice for XAI Researchers. In Proceedings of the 2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX), Salt Lake City, UT, USA, 25–30 October 2020; pp. 1–8. [CrossRef]
28. Páez, A. The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds Mach.* **2019**, *29*, 441–459. s11023-019-09502-w. [CrossRef]
29. Ashby, S.; Hanna, J.; Matos, S.; Nash, C.; Faria, A. Fourth-wave HCI meets the 21st century manifesto. In Proceedings of the Proceedings of the Halfway to the Future Symposium, Nottingham, UK, 19–20 November 2019; pp. 1–11.

30. Zuboff, S.; Möllers, N.; Wood, D.M.; Lyon, D. Surveillance Capitalism: An Interview with Shoshana Zuboff. *Surveill. Soc.* **2019**, *17*, 257–266. [CrossRef]
31. Marcus, G.; Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust*; Vintage: Tokyo, Japan, 2019.
32. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608.
33. OECD. Tools for Trustworthy AI. 2021. Available online: https://www.oecd.org/science/tools-for-trustworthy-ai-008232ec-en.htm (accessed on 6 December 2022).
34. EU. *Ethics Guidelines for Trustworthy AI*; Report; European Commission: Brussels, Belgium, 2019.
35. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An integrative model of organizational trust. In *Organizational Trust: A Reader*; Academy of Management: New York, NY, USA, 2006; pp. 82–108.
36. Hilbert, M. Digital technology and social change: The digital transformation of society from a historical perspective. *Dialogues Clin. Neurosci.* **2020**, *22*, 189. [CrossRef]
37. Thiebes, S.; Lins, S.; Sunyaev, A. Trustworthy artificial intelligence. *Electron. Mark.* **2021**, *31*, 447–464. [CrossRef]
38. Dörner, D. Theoretical advances of cognitive psychology relevant to instruction. In *Cognitive Psychology and Instruction*; Springer: Berlin/Heidelberg, Germany, 1978; pp. 231–252.
39. Smith, C.J. Designing trustworthy AI: A human-machine teaming framework to guide development. *arXiv* **2019**, arXiv:1910.03515.
40. Leijnen, S.; Aldewereld, H.; van Belkom, R.; Bijvank, R.; Ossewaarde, R. An agile framework for trustworthy AI. NeHuAI@ ECAI. 2020. pp. 75–78. Available online: https://www.semanticscholar.org/paper/An-agile-framework-for-trustworthy-AI-Leijnen-Aldewereld/880049a16c8fea47dcfe07450668f5507db5e96d (accessed on 6 December 2022).
41. Seigneur, J.M. Trust, Security, and Privacy in Global Computing. Ph.D. Thesis, University of Dublin, Dublin, Ireland, 2005.
42. Rossi, F. Building trust in artificial intelligence. *J. Int. Aff.* **2018**, *72*, 127–134.
43. Hassenzahl, M.; Tractinsky, N. User experience-a research agenda. *Behav. Inf. Technol.* **2006**, *25*, 91–97. [CrossRef]
44. Lee, D.; Moon, J.; Kim, Y.J.; Mun, Y.Y. Antecedents and consequences of mobile phone usability: Linking simplicity and interactivity to satisfaction, trust, and brand loyalty. *Inf. Manag.* **2015**, *52*, 295–304. [CrossRef]
45. McCarthy, J.; Wright, P. Technology as experience. *Interactions* **2004**, *11*, 42–43. [CrossRef]
46. Akash, K.; McMahon, G.; Reid, T.; Jain, N. Human trust-based feedback control: Dynamically varying automation transparency to optimize human-machine interactions. *IEEE Control Syst. Mag.* **2020**, *40*, 98–116. [CrossRef]
47. Wogu, I.A.P.; Misra, S.; Udoh, O.D.; Agoha, B.C.; Sholarin, M.A.; Ahuja, R. Artificial Intelligence Politicking and Human Rights Violations in UK?s Democracy: A Critical Appraisal of the Brexit Referendum. In Proceedings of the The International Conference on Recent Innovations in Computing, Jammu, India, 20–21 March 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 615–626.
48. Alben, L. Quality of experience: Defining the criteria for effective interaction design. *Interactions* **1996**, *3*, 11–15. [CrossRef]
49. Sousa, S.C.; Tomberg, V.; Lamas, D.R.; Laanpere, M. Interrelation between trust and sharing attitudes in distributed personal learning environments: The case study of lepress PLE. In *Advances in Web-Based Learning-ICWL 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 72–81.
50. Sousa, S.; Lamas, D.; Hudson, B. Reflection on the influence of online trust in online learners performance. In Proceedings of the E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Honolulu, HI, USA, 13–17 October 2006; Association for the Advancement of Computing in Education (AACE): Morgantown, WV, USA, 2006, pp. 2374–2381.
51. Sousa, S.; Lamas, D. Leveraging Trust to Support Online Learning Creativity–A Case Study. *ELearning Pap.* **2012**, *30*, 1–10.
52. Sousa, S.; Lamas, D.; Dias, P. The Implications of Trust on Moderating Learner's Online Interactions—A Socio-technical Model of Trust. In Proceedings of the CSEDU 2012-Proceedings of the 4th International Conference on Computer Supported Education, Porto, Portugal, 16–18 April 2012; e Maria João Martins e José Cordeiro, M.H., Ed.; SciTePress: Setúbal, Portugal, 2012; Volume 2, pp. 258–264.
53. Lankton, N.K.; McKnight, D.H.; Tripp, J. Technology, humanness, and trust: Rethinking trust in technology. *J. Assoc. Inf. Syst.* **2015**, *16*, 1. [CrossRef]
54. McKnight, D.; Chervany, N. Trust and distrust definitions: One bite at a time. In *Trust in Cyber-Societies: Integrating the Human and Artificial Perspectives*; Falcone, R., Singh, M.P., Tan, Y., Eds.; Springer: Berlin, Germany, 2002; pp. 27–54.
55. Gambetta, D. Trust making and breaking co-operative relations. In *Can We Trust Trust?*; Gambetta, D., Ed.; Basil Blackwell: Oxford, UK, 1998; pp. 213–237.
56. Luhmann, N. Familiarity, confidence, trust: Problems and alternatives. *Trust. Mak. Break. Coop. Relations* **2000**, *6*, 94–107.
57. Cavoukian, A.; Jonas, J. *Privacy by Design in the Age of Big Data*; Information and Privacy Commissioner of Ontario: Mississauga, ON, Canada, 2012.
58. Gulati, S.; Sousa, S.; Lamas, D. Design, development and evaluation of a human-computer trust scale. *Behav. Inf. Technol.* **2019**, 1–12. [CrossRef]
59. Gulati, S.; Sousa, S.; Lamas, D. Modelling Trust: An Empirical Assessment. In Proceedings of the 16th IFIP TC 13 International Conference on Human-Computer Interaction—INTERACT 2017-Volume 10516, Mumbai, India, 25–29 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 40–61._3. [CrossRef]
60. Gulati, S.; Sousa, S.; Lamas, D. Modelling trust in human-like technologies. In Proceedings of the 9th Indian Conference on Human Computer Interaction, Bangalore, India, 16–18 December 2018; pp. 1–10.

61. Sousa, S.; Lamas, D.; Dias, P. Value creation through trust in technological-mediated social participation. *Technol. Innov. Educ.* **2016**, *2*, 5. [CrossRef]
62. Resnick, P.; Zeckhauser, R.; Friedman, E.; Kuwabara, K. Reputation systems: Facilitating trust in Internet interactions. *Commun. ACM* **2000**, *43*, 45–48. [CrossRef]
63. Renaud, K.; Von Solms, B.; Von Solms, R. How does intellectual capital align with cyber security? *J. Intellect. Cap.* **2019**, *20*, 621–641. [CrossRef]
64. Hansen, M. Marrying transparency tools with user-controlled identity management. In Proceedings of the IFIP International Summer School on the Future of Identity in the Information Society, Brno, Czech Republic, 1–7 September 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 199–220.
65. Buchanan, T.; Paine, C.; Joinson, A.N.; Reips, U.D. Development of measures of online privacy concern and protection for use on the Internet. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 157–165. [CrossRef]
66. Fimberg, K.; Sousa, S. The Impact of Website Design on Users' Trust Perceptions. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, San Diego, CA, USA, 25–29 July 2020; Springer: Berlin/Heidelberg, Germany, 2020, pp. 267–274.
67. Kim, Y.; Peterson, R.A. A Meta-analysis of Online Trust Relationships in E-commerce. *J. Interact. Mark.* **2017**, *38*, 44–54. [CrossRef]
68. Hancock, P.A.; Billings, D.R.; Schaefer, K.E. Can you trust your robot? *Ergon. Des.* **2011**, *19*, 24–29. [CrossRef]
69. Schmager, S.; Sousa, S. A Toolkit to Enable the Design of Trustworthy AI. In Proceedings of the HCI International 2021-Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence, Virtual Event, 24–29 July 2021; Stephanidis, C., Kurosu, M., Chen, J.Y.C., Fragomeni, G., Streitz, N., Konomi, S., Degen, H., Ntoa, S., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 536–555.
70. Hardin, R. *Trust and Trustworthiness*; Russell Sage Foundation: New York, NY, USA, 2002.
71. Bauer, P.C. Conceptualizing trust and trustworthiness. Political Concepts Working Paper Series. 2019. Available online: https://www.semanticscholar.org/paper/Conceptualizing-Trust-and-Trustworthiness-Bauer/e21946ddb6c3d66a347957d1e3cef434f63b22fb (accessed on 6 December 2022).
72. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
73. Sousa, S. Online Distance Learning: Exploring the Interaction Between Trust and Performance. Ph.D Thesis, Seffield Hallam University, Sheffield, UK, 2006.
74. Han, Q.; Wen, H.; Ren, M.; Wu, B.; Li, S. A topological potential weighted community-based recommendation trust model for P2P networks. *Peer- Netw. Appl.* **2015**, *8*, 1048–1058. [CrossRef]
75. Hoffman, L.J.; Lawson-Jenkins, K.; Blum, J. Trust beyond security: An expanded trust model. *Commun. ACM* **2006**, *49*, 94–101. [CrossRef]
76. Jensen, M.L.; Lowry, P.B.; Burgoon, J.K.; Nunamaker, J.F. Technology dominance in complex decision making: The case of aided credibility assessment. *J. Manag. Inf. Syst.* **2010**, *27*, 175–202. [CrossRef]
77. Muise, A.; Christofides, E.; Desmarais, S. More information than you ever wanted: Does Facebook bring out the green-eyed monster of jealousy? *CyberPsychology Behav.* **2009**, *12*, 441–444. [CrossRef]
78. EU, A.H. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment. 2021. Available online: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (accessed on 6 December 2022).
79. Madsen, M.; Gregor, S. Measuring human-computer trust. In Proceedings of the 11th Australasian Conference on Information Systems, Brisbane, Australia, 6–8 December 2000; Citeseer: Princeton, NJ, USA, 2000; Volume 53, pp. 6–8.
80. Goillau, P.; Kelly, C.; Boardman, M.; Jeannot, E. *Guidelines for Trust in Future ATM Systems-Measures*; European Organisation for the Safety of Air Navigation: Brussels, Belgium, 2003. Available online: https://skybrary.aero/bookshelf/guidelines-trust-future-atm-systems-measures-0 (accessed on 6 December 2022).
81. Bachrach, M.; Guerra, G.; Zizzo, D. The self-fulfilling property of trust: An experimental study. *Theory Decis.* **2007**, *63*, 349–388. [CrossRef]
82. Ajenaghughrure, I.B.; da Costa Sousa, S.C.; Lamas, D. Risk and Trust in artificial intelligence technologies: A case study of Autonomous Vehicles. In Proceedings of the 2020 13th International Conference on Human System Interaction (HSI), Tokyo, Japan, 6–8 June 2020; pp. 118–123.
83. Benbasat, I.; Wang, W. Trust in and adoption of online recommendation agents. *J. Assoc. Inf. Syst.* **2005**, *6*, 4. [CrossRef]
84. Söllner, M.; Hoffmann, A.; Hoffmann, H.; Wacker, A.; Leimeister, J.M. *Understanding the Formation of Trust in IT Artifacts*; Association for Information Systems: Atlanta, GA, USA, 2012.
85. Mcknight, D.H.; Carter, M.; Thatcher, J.B.; Clay, P.F. Trust in a Specific Technology: An Investigation of Its Components and Measures. *ACM Trans. Manage. Inf. Syst.* **2011**, *2*, 12:1–12:25. [CrossRef]
86. Söllner, M.; Leimeister, J.M. What we really know about antecedents of trust: A critical review of the empirical information systems literature on trust. In *Psychology of Trust: New Research*; Gefen, D., Ed.; Nova Science Publishers: Hauppauge, NY, USA, 2013.
87. Friedman, B.; Khan, P.H., Jr.; Howe, D.C. Trust online. *Commun. ACM* **2000**, *43*, 34–40. [CrossRef]

88. Zheng, J.; Veinott, E.; Bos, N.; Olson, J.S.; Olson, G.M. Trust without touch: Jumpstarting long-distance trust with initial social activities. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Minneapolis, MN, USA, 20–25 April 2002; pp. 141–146.
89. Shneiderman, B. Designing trust into online experiences. *Commun. ACM* **2000**, *43*, 57–59. [CrossRef]
90. Muir, B.M.; Moray, N. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* **1996**, *39*, 429–460. [CrossRef]