UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Governance by trust mediators in the digital society

The redistribution of risk and vulnerability

Bodó, Balázs; Weigl, Linda; Araujo, Theo

DOI
10.1080/21515581.2025.2571505

Publication date
2026

Document Version
Final published version

Published in
Journal of Trust Research

License
CC BY

Link to publication

Citation for published version (APA):
Bodó, B., Weigl, L., & Araujo, T. (2026). Governance by trust mediators in the digital society: The redistribution of risk and vulnerability. *Journal of Trust Research*, *16*(1), 6-30. https://doi.org/10.1080/21515581.2025.2571505

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:17 Nov 2026

JOURNAL OF TRUST RESEARCH
2026, VOL. 16, NO. 1, 6–30
https://doi.org/10.1080/21515581.2025.2571505

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Governance by trust mediators in the digital society: The redistribution of risk and vulnerability

Balázs Bodó[a], Linda Weigl[a] and Theo Araujo [b]

[a]Institute for Information Law, University of Amsterdam, Amsterdam, The Netherlands; [b]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, The Netherlands

Online intermediaries and algorithmic technologies, such as platforms or AI, remediated and with it, transformed many of our social, economic, and political relations. This transformation of relationships also affects the trust which makes these relations possible, and which emerges through these interactions and beyond. We argue that digital technologies remediate trust, and in doing so, they produce new risks and vulnerabilities while redistribute old ones. We use six domains to illustrate how interpersonal and institutional trust are transformed through digital remediation. By reviewing contemporary literature on the impact of digital technologies on institutional and interpersonal trust, this article offers a new approach to conceptualising how trust mediators operate, how they succeed and fail to absorb and redistribute risks. The findings serve as an important stepping stone to identify new directions for a theoretical framework on trust in the digital society.

ARTICLE HISTORY
Received 18 November 2024
Accepted 30 September 2025

AREA EDITOR
Amanda Osuna

KEYWORDS
Digital trust; trust mediators; trust theory; risk governance; vulnerabilities

1. Introduction

Algorithmic systems increasingly shape how we connect, communicate, inform and consume. Thereby, we are inevitably exposed to known and unknown risks and vulnerabilities. This reality also gives trust in the digital society a new meaning, because our capacity to reason is not always sufficient to manage complexity and choose the least risky option in an algorithmically interconnected and opaque environment. It is virtually impossible and too complex to consider and assess all possible risks, as the changing nature of time and space introduces many more actors and possibilities into the equation. Thus, in any relationship that is in some way channeled through algorithmic technologies, trust, personal or impersonal, is necessary.

The very same actors that provide algorithmic technologies to keep up and use this trust, are also the ones actively trying to signal their trustworthiness and produce trust. They do so by attempting to manage both the risks of those whom they connect, as well as their own risks.

CONTACT Linda Weigl l.weigl@uva.nl

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

In this paper, we argue that technology mediates trust by becoming a risk manager (not necessarily a risk mitigator), and thereby actively shapes trust relations at an interpersonal and institutional level. This role comes with its own risks. For instance, platforms, while managing certain risks (friction or declining user experience, for instance), may amplify misinformation online. The unintended side-effect of algorithmic content moderation may be the growing distrust in otherwise trustworthy actors, such as public servants, scientists, journalists, political actors (Berinsky, 2017), and scientific experts (Krause et al., 2022). Platform rules and their enforcement can cast doubt on procedural fairness, and challenge the legitimacy of policy outcomes (Abiri & Buchheim, 2022; Martin et al., 2022). A similar observation applies to online dating, where platforms have normalised trust in digitally mediated connections, even though users are aware these may be manipulated, embedding emotional risk into dating (Haywood, 2018).

This article draws the first outlines of a theoretical framework to understand how digital infrastructures challenge existing interpersonal and institutional trust mechanisms to cope with uncertainty. As a foundation, we use the lens of the theoretical concept of 'digital trust mediators' (Bodó, 2021). We argue that this mechanism of trust production, in its effort to manage risks simply redistributes – in yet to be fully understood ways – risks and produces new vulnerabilities, or amplifies existing ones.

To do so, we analyze existing literature on trust and digitalisation. We conduct an exploratory literature review, focusing on interdisciplinary work that addresses how digital infrastructures intersect with trust. We identified six broader socio-economic dimensions that emerged repeatedly across different texts: (1) self-representation, (2) information flows (3) social relations (4) economic transactions (5) epistemic frameworks, (6) and institutional decision-making. These categories were developed inductively: after reviewing the literature, we asked ourselves how the various dynamics could be meaningfully grouped to reflect key domains in which trust is mediated through digital means. These dimensions, although not yet exhaustive, mark a starting point for our objective and give evidence to our theoretical approach.

2. Interpersonal, institutional and mediated trust

For this paper, we define trust as a shared social resource which enables social co-existence and collaboration in the face of risks and potential harms. Seeing it as resource allows us to look beyond trust as an individual attitude, or as an institutional arrangement, and opens the way to ask how this resource comes to be, and how, actors and processes is produced, transformed, amplified or diminished, with a special focus on the role of the newly emerging sociotechnical infrastructures of mediated communication in the digital space.

We can also ask these questions in relation to both interpersonal and institutional trust. On the individual level, the literature identifies several possible sources of trust, such as experience and familiarity (strategic trust), the general belief that most people can be trusted even without much prior knowledge (moralistic trust) which generate social cohesion and civic life (Uslaner, 2002), or the other way around, as in Putnam et al. (1993) who argue that civic engagement helps creating trust. We assume that both dynamics can be present in our social trust relations.

The role of institutional trust is well mapped out too (Möllering, 2006; Shapiro, 1987; Sztompka, 1999; Zucker, 1986). Zucker (1986) was the first to note that trust can be

institutionally produced, in the absence of other forms, such as interpersonal, and group-based trust sources. Giddens (1990, 1999) highlights the stability, predictability, and transparency created by institutions and abstract systems as a source of societal trust. Bachmann (2001) highlights the role of inter-organisational power relations (a form of control) which complements other forms, such as facework (interpersonal trust relations between organisations), and systemic trust (binding norms and standards) which supports trust within institutional arrangements. Sztompka (1999) describes the mechanisms which build trust in the institutional system as a whole by strategically managing distrust among the trust producing institutions. The ways these institutional arrangements can redistribute risks and harms in society (Beck, 1992) and reduce the vulnerabilities of individuals (Hamm & Banner, 2025) are more complex and numerous to do justice in this short review, but they point to the fact all parties concerned are locked in a strong interdependent network of relations in which everyone participates in the production of trust as a shared resource, usually in both their capacities of trustor and a trustee.

The relationship between trust and institutions has also been explored in the digital context. Putnam was cautious about whether online platforms could genuinely support trust, while others argued that digital media use has little influence on trust, since it stems from deeper, more personal values (Håkansson & Witmer, 2015). Still, it would be naïve to assume that digital environments have no influence on how and whom we (dis)trust. In this context, a new wave of trust scholars started to review the characteristics of this relationship (Bodó, 2021; Bodó, 2021; Botsman, 2017; Keymolen, 2016; Werbach, 2018). In the following, we summarise the main points of this latter thread of scholarship.

First, *remediation*. Existing trust relations are remediated when new techno-social actors compete with existing trust producing institutions and the trust relations they facilitate. The iconic example is how digitisation transformed our information landscape. Traditional news organisations are challenged by social media companies and AI-generated content factories. Such disruption often comes with the replacement of trusted (or at least familiar) institutional trust production logics with untested ones. Second, *expansion*. Digital technologies also produce trust in ways that were not possible before. The instruments and the scale at which e-commerce or platforms such as Uber or Airbnb enable trust between individuals (by maintaining planetary-scale reputation accounting systems) is unprecedented. Last but not least, *destabilisation*. As Beck noted (1992), new technologies bring new complexities to social relations. Trust is a source of innovation, but like any other, digital innovation generates new risks and uncertainties. These future uncertainties need to be dealt with somehow (Beckert, 2016), but the main instrument to deal with uncertainty – trust – is the very thing being disrupted. In other words, digital innovation disrupts the same trust relations and trust-producing mechanisms needed to deal with the uncertainty of innovation itself, not by removing trust entirely, but by replacing it with potentially misleading proxies, making genuinely trustworthy mechanisms even more urgent.

As we argue in the rest of the paper, most of the current theoretical, sociological, philosophical frameworks of trust do not fully consider the trust dynamics of the digital society. In the following we offer the first outlines of such a framework to better account for the digitisation-specific questions around the theory of (institutional) trust, by weaving in these three characteristics of digitally mediated trust (remediation,

expansion and destabilisation) together with what we know about interpersonal and institutional trust relationships.

3. The impact of digital infrastructures on trust

Digital infrastructures now touch almost all aspects of modern life, and almost all trust relations. Based on a review of the literature, drawing mainly on journal articles from fields such as STS, sociology, political science, communication science, and media studies, we provide an exploratory overview rather than a fully systematic review. We began with the following keyword search combinations: 'trust citizen digital technology', 'trust decentralized technology', 'trust blockchain', 'trust decentralized infrastructures', 'trust AI', 'trust machine learning', 'trust algorithmic technology', and 'trust online platforms'. This search was complemented with the snowballing method of key, well-cited articles on both interpersonal and institutional trust in the context of digital technologies. We retained only those articles that, in our assessment, made a substantial contribution. In our collection of articles, we identified what emerged as the most important dimensions in which digital innovation mediates interpersonal and institutional trust relations: (1) the affordances of online *self-representation*, (2) and the nature of *social relations* emerging via online interactions; (3) the fragmentation of *epistemic frameworks*, and (4) the characteristics of digital *information flows*; (5) the digital trust infrastructures supporting *economic transactions*, and (6) algorithmic *institutional decision-making*. We identified these dimensions as consistently discussed themes across the literature that capture both interpersonal and institutional trust dynamics in distinct but interrelated domains.

3.1. Self-representation

In interpersonal relations, (dis)trust decisions are based on what we (do not) know about the other. The bulk of this knowledge stems from how others present themselves in social interactions. Self-(re)presentation thus refers to how individuals display themselves, aiming to shape others' perceptions (Krämer & Haferkamp, 2011; Leary & Kowalski, 1990).

In self-representation through digital interfaces, digital technologies serve both as a tool for communication and they enable – and require – the creation of a digital 'version' of the self. On the one hand, we use stories and narratives to shape our presentation online, as we attempt to make our lives meaningful and legible to others (Poletti, 2020). On the other hand, users share or tolerate the collection of information about themselves to reduce uncertainty and anonymity, and to appear authentic and trustworthy to other individuals or groups (Bryce & Fraser, 2014; Zhao et al., 2008). The public representations of the self are key assets for the way the trustworthiness of oneself and of others is displayed and assessed. In the digital context, these two dynamics face significant, media-, or technology-specific challenges.

Context-constraints on self-representation. Trust requires predictability, familiarity, and an expectation of some sort of 'normalcy'. Since the expectations of 'normalcy', and integrity are context-specific (one values different qualities in the same person in their roles as a parent, a lover, or a professional), ways of self-representation try to maintain multiple locally coherent narratives of the self. However, online interactions may collapse contexts usually kept distinct in offline interactions (Chambers, 2017;

Marwick & boyd, 2011). Even though online platforms are often organised around particular interests or roles (compare, for instance, different accounts of the same person on Facebook, LinkedIn, Instagram, and X), the boundaries within these spaces are vague. Content and interactions can easily move across contexts, and different audiences (family, colleagues, strangers) may be exposed to the same image. Second, some features of mediated online communication, namely the dynamics of communication, platform design, and the public nature of social engagement, play an important role in the nature and scope of personal information shared in the engagement with strangers. The sheer speed and amount of interactions a person can have during just one day may gradually, but subtly, normalise a more translucent online environment (Baker, 2005; Jamieson, 2013). The collapse of contexts in the digital space complicates how integrity is signalled. Though in the literature (Mayer et al., 1995) integrity (or value congruence between trustor and trustee) is discussed as context-independent property of the trustee, there might be substantial differences between values in one domain of life and values exhibited in others. The most straightforward such a division can occur between one's professional and personal identities: the serious, focused, highly respectful professional, and the private person who engages in transgressive consensual kinks in the bedroom. Online environments make it easier to selectively display aspects of the self, or to shield certain audiences from other aspects. Integrity can be convincingly performed even when it is not fully grounded, which undermines one of the most important, and the only non-task-specific component of trustworthiness. This connects directly to the point on selective self-presentation.

Selective and strategic self-presentation. A large amount of research examines factors and reasons behind strategic, or inauthentic self-representation online. In the social media influencer economy, these vulnerabilities can be beneficial for those who can strategically turn their self-representation into a particularly aesthetic and popular image (Harris & Bardey, 2019; Khamis et al., 2016; Kristinsdottir et al., 2021). While most media misportrayals come with relatively limited risks,[1] platforms acting as self-representation mediators can allow deceptive and criminal users to create entirely fake digital impressions, leading others into trusting fabricated personae, resulting in cases of catfishing, imposter scams, or sextortion (Bates, 2017; Paat & Markham, 2021).

In all these instances, digital platforms provide the mediating tool, the 'identity technology' (Poletti & Rak, 2014) for how individuals shape their own narrative. Specifically, the interfaces and software design through which users portray themselves to the world are defined by the interests of commercial entities that impose their rules, business models, and political perspectives, suppressing some facets of the self, mandating and highlighting others (Are et al., 2024; Bodó, 2021). Moreover, the open nature of engagement between users, coupled with the pursuit of social validation and social gratification also impacts the nature of personal information disclosed about oneself (Turkle, 2012). Any trust towards another party is thus also connected to the intermediary to allow for trustworthy – or prohibit untrustworthy – digital representation of the individual on the platform. Currently, this is a giant leap of faith, or rather, a misplaced trust, if we look at the countless failures of digital intermediaries to prevent fraud, manipulation, and other online harms.

In all these cases, the object of trust has shifted: users are forced to read the other through their mediated representation, as shaped by the intermediary. The intermediary, however, remains an invisible trustee. Yet with the parallel growth of their power and

legal accountability, intermediaries increasingly shape, police, and alter the self-representation of users. This turns bipartite trust relations between two users into a trust triangle with a powerful, technological agent in between.

3.2. Social relations

The second impact of digital remediation is related to the trust formation process. Kramer et al. (1996) deconstruct the interpersonal trust process into three different stages. First, the relationship starts off with calculus-based trust where trustors weigh the costs and benefits of placing 'blind faith' in the other person, prior to establishing a trust relationship. In the second, knowledge-based trust stage, trustors try to gain more knowledge about the context and the trustee to predict the trustee's behaviour and evaluate their trustworthiness. Finally, at the identification-based trust stage the decision to trust is rooted in shared values, and a strong personal and emotional connection.

By mediating 'new ideas and experiences of intimacy, friendship and identity through new forms of social interaction and new techniques of public display' (Chambers, 2013, p. 1), digital technologies offer an alternative path for these analogue trust-formation processes. It has become commonplace for many to engage with complete strangers. In doing so, technology mediates a new form of relationship, latent ties.

Latent ties. Latent ties lie somewhere between weak ties (typically characterised by less frequent exchanges, shorter periods of contact, less intimacy and little information sharing), and strong ties (requiring frequent and durable contact, higher levels of intimacy, more self-disclosure, and reciprocal exchanges). A notable feature of latent ties (Haythornthwaite, 2002) is that they are contingent upon the structures in which the ties form, imposed by platform designers and engineers. For example, in social media environments, algorithms curate with whom we interact, setting the path for the evolution of latent ties. A latent tie can exhibit high levels of intimacy, extensive information sharing, and emotional involvement, and may evolve into stronger forms of connection (Haythornthwaite, 2002). Algorithmic amplification can accelerate the perception of shared values and familiarity between users by selectively highlighting or suppressing interactions, content or behaviours. Ma et al. (2011) have shown, however, that there is no reason to assume that value or taste based similarities between users (something most person-recommendation systems rely on) would correlate with the strength of trust between them. Users may feel an immediate emotional connection of identification with others, even in the absence of the traditional stages of trust development articulated by Kramer et al. (1996), such as information-gathering or information verification to build calculus- and knowledge-based trust. This does not per se mean that such trust is always misplaced, but it introduces a potential risk: trust may form before a real understanding and justified assessment of the other's reliability, capability, benevolence, etc. is established, and encourage users to prematurely jump to identification-based trust. We cannot conclude that algorithms force users to bypass trust development stages by default, but the environment that nudges users toward identification-based trust more quickly increases risks such as over-disclosing information to strangers, and on the other, increased vulnerability towards deception, fraud (Bates, 2017; Paat & Markham, 2021).

3.3. Epistemic frameworks

A precondition for trust is a shared epistemic framework which creates familiarity for strangers needing to enter a trust relationship (Zucker, 1986). Shared epistemic frameworks may take the form of myths, customs, religious texts, language, shared narratives, or their modern, institutionalised, rational counterparts, such as laws, science, (public service or mass) media, mainstream and niche cultural works, public education curricula, which all produce 'grand narratives' or metanarratives (Lyotard et al., 2001). The dynamics of epistemic frameworks are inherently tied to the larger institutional frameworks which maintain them (Foucault, 1981). In high modernity, the combination of the centralised, heavily gatekept, top-down mass media, and strong public institutions has been shown to create and stabilise cultural norms, expectations, and interpretative frameworks (Herman & Chomsky, 2002). In contrast, since the emergence of digital media, we witness a breakdown of existing knowledge paradigms, as well as the rise of smaller, counterhegemonic epistemologies.

Fragmentation of epistemic frameworks. The emergence of digital technologies upset epistemic frameworks of high modernity. In practice this change is captured by the rise of epistemic fringes through filter bubbles and echo chambers, personalisation, and behavioural modification, 'post-truth politics', the spread of misinformation, disinformation, and conspiracy theories (United Nations, 2017). Though this process is the easiest to identify in the case of media, it encompasses all other institutions tasked with producing shared epistemologies. The increasing numbers of home-schooled children removed from public education systems; or a growing distrust in scientific institutions and science also point to the slower or faster evacuation of grand narratives.

Epistemic polarisation. The breakdown of shared epistemologies coincides with the rapid proliferation of counternarratives both factual and mythical on every subject from vaccines to politics. In that process, platforms and social media are perfect infrastructures to 'resist incorporation into the dominant discourse' (Baier, 2024, p. 97). These local and transient narratives can successfully challenge the discursive hegemony of epistemic institutions of modernity. This matters, as Fukuyama (1996) shows, if larger scale trust infrastructures and frameworks are distrusted, the scope of collaboration is restricted to the local trust networks. With the gradual collapse of societal-scale epistemic frameworks, these counterhegemonic frameworks emerge as precursors to trust relations within and between different groups (Nooteboom, 2003, p. 25). Through the stabilisation of expectations about the beliefs and behaviour of another, they create the foundations for in-group trust, and as a corollary, form the basis of mistrust in inter-group relations (Cook et al., 2009).

Epistemic uncertainty. Generative AI systems take this process even further by destabilising and destroying the chances of epistemic certainty both through their stochastic 'hallucinations', and through the unreliable efforts to control them (Milmo & Hern, 2024). In the absence of reliable signals of the trustworthiness of information the influx of AI-generated slop also destabilises trust in every piece of knowledge (Abiri & Buchheim, 2022). In sum, we witness a trajectory of epistemic fragmentation, which has led us from traditional (media) institutions thought to manufacture consent, through social media apparently manufacturing dissent, to AI manufacturing nonsense. Digital information communication technologies may not be the sole drivers of this process, but they apparently create conditions that make it possible.

These small, counterhegemonic epistemic niches can indeed offer an effective basis of particularised trust[2]: much-needed predictability, security, simplicity in an overwhelmingly complex world full of uncertainties and unmanageable risks. They come, however, at certain costs. First, epistemic niches make trusting across boundaries difficult. Second, the co-existence of competing epistemic frameworks seems to reduce systemic trust by destabilising trust in those institutional arrangements, whose primary task is to create some overarching trustworthy interpretative frame to the world (Nai et al., 2024; van der Meer et al., 2023). The prevailing problem is therefore not only social polarisation, but the parallel breakdown of the mechanisms to maintain trust between segments of society. One such mechanism consists of the institutional structures producing factual, scientific information necessary to underwrite policy. Distrust towards it 'threatens to alienate large parts of the population from the processes of social knowledge production and collective decision-making, thus threatening overall democratic legitimacy' (Abiri & Buchheim, 2022, p. 82), and the social cooperation necessary for maintaining liberal democratic order.[3]

3.4. Information flows

A main source of stable epistemic frameworks in society are truthful, accurate, and free-flowing information flows (Shapiro, 1987; Zucker, 1986). In the previous section we discussed multiple epistemic institutions, here we highlight the role of media organisations in societal trust production. News organisations manage flows of ground truths and discourses. Through their watchdog function, they are part of institutionalised distrust frameworks, which keep democratic institutions trustworthy (Sztompka, 1998). The introduction of digital technologies came with hopes to democratise participation in the creation, circulation, and access to information. Through logics such as the 'wisdom of crowds', or 'given enough eyeballs, all bugs are shallow', the promise of counterhegemonic narratives from that era was that digitisation would provide new, potentially more trustworthy information infrastructure alternatives to the existing ones. In some instances, this has happened, but the abundance and speed of information flows, and the dominant role of user metrics in news distribution let to relevant to the shifts in societal trust.

Abundance and speed of information flows. Both the abundant volume and speed of information are the byproduct of the digital media environments, which radically reduced the costs of producing and disseminating information. This poses a unique challenge for information consumers, intensified with the 'unbundling of news' through which the 'solid-state entity of news has disappeared' (Trilling, 2019, p. 298). Instead of reading a full newspaper, we now consume individual 'unbundled' pieces and snippets of information, selected for us by opaque algorithms. The speed of information sharing also poses a challenge. It takes time to establish the trustworthiness of a source and a piece of information through rational scrutiny. Under such conditions, trust can flow from irrational, emotional sources (Beckett & Deuze, 2016), particularly in times of crisis (Kim & Cameron, 2011). Selective exposure theory suggests that people tend to focus on news, opinions, information which already aligns with their pre-existing personal convictions (Zillmann & Bryant, 1985). Under constant time pressure it may be easier to trust information that explains phenomena in a simple, plausible and affective way (the self-confirmation heuristic), than engaging in the tedious cognitive task to unpack

inexplicable or even illogical events and verify them individually (Metzger et al., 2020; Yeo et al., 2015).

User metric driven news. Platforms and search engines employ algorithms to disseminate news (Entman & Usher, 2018). User metric driven recommender algorithms, combined with the self-selection of information may generate 'protective filters' and 'echo chambers' for a subset of (oftentimes highly partisan) individuals (Dahlgren, 2018; Ross Arguedas et al., 2022). For some consumers, personalised and self-reinforcing content, and connections to like-minded users enhances isolation, discursive divides, and political polarisation. It takes substantial effort to push back against this self-reinforcing logic, and design recommenders which inject some kind of diversity into recommendations (Helberger, 2019).

As it stands now, digital intermediaries in the business of producing trust through the organisation of information flows offer to reduce of the complexity of the real world by providing easily accessible information with quick explanatory potential for our fast-moving world, which tries to capitalise on, rather than counteract the individual, cognitive limitations and shortcomings of trusting and verifying the information landscape around us.

3.5. Economic transactions

Trust is a key component of economic relations (Aghion et al., 2010; Bachmann & Kroeger, 2017; Greif et al., 1994; Zucker, 1986). Digitisation brought two important changes to economic relations: it changed the scale and scope on which economic transactions now take place, and it established novel frameworks to produce trust and manage distrust in both the preexisting and the new loci of trade.

Scale and scope. Up until the emergence of global e-commerce platforms, global trade was the privilege of large, corporate entities. Platforms like Amazon or Alibaba, Uber or AirBnB 'democratized' trade by allowing individuals to participate; and expanded the distance across which economic transactions can take place (Bamberger & Lobel, 2017; Van Dijck et al., 2018).

New trust supports. These new economic transactions need corresponding trust support, which the pre-existing institutional order was unprepared to provide. The post-war institutional trust production frameworks (such as the World Trade Organisation) were designed to serve big, corporate actors, not billions of individual users trying to establish economic relations across huge cultural, geographic distances. Platforms emerged as the trust frameworks for global e-commerce, as private, for-profit trust infrastructures (Botsman, 2017; Keymolen, 2016). The trust-relevant e-commerce platforms are organised around two different trust production approaches: the production of trustworthiness signals; and the infrastructures of fulfilment.

Privately produced trustworthiness signals. As we discussed earlier, online knowledge-based trust is complicated by the fact that digital intermediaries control the resources to build one's own, and assess the reputation of others. Platforms offer context-independent, platform-specific reputation signals (such as ratings, stars, reviews), through the aggregation of their users' activities (Bolton et al., 2013; Hesse & Teubner, 2020; Tadelis, 2016). Unlike traditional reputation information, such platform-produced reputation signals are standardised, cut across, and are disembedded from all the different local,

culturally and socially specific contexts in which the platform operates. Despite their widespread use, we know little about the trustworthiness of these signals. It is unclear what information gets collected and aggregated into the reputation signals, and it is even less known how these data points are processed through the algorithmic black boxes. It is also opaque how these trust producers set the thresholds of untrustworthiness, and with what effectiveness and rigour they filter out untrustworthy actors (Houser, 2020; Kerr, 2019). And even when all these steps are transparent, it is still highly problematic to ensure that the output of such a reputation production process is unbiased (Guo et al., 2025).

Fulfilment infrastructures and enforcement. Platforms also produce trust and manage distrust through various mechanisms of control and insurance to ensure that transactions take place quasi-independently of the reputation of the parties. These mechanisms range from fully automated smart contracts and blockchains, via private arbitration systems to resolve conflicts between users, to online payment, delivery, and insurance systems. By enforcing the fulfilment of transactions, blockchain systems provide trust infrastructures to anonymous parties where reputation is unobservable. Private (sometimes algorithmic) arbitration manages the outfall of breaches of trust much faster than courts. Online payment systems ensure payment, or a refund in case of a breach. Platform specific safeguards can ensure the delivery of goods or services, or provide some form of insurance in case of a breach. These trust safeguards, even if they are not fully algorithmic, are in many aspects similar to traditional, contract-based guarantees of fulfilment. They are also often quite removed and insulated from democratic oversight and accountability, ranging from technologically (and by some definition, legally) sovereign blockchain based systems (Filippi et al., 2021; Herian, 2021) to the private, and oftentimes only weakly regulated enforcement of the terms and conditions of social media and e-commerce platforms.

While these new trust production approaches may seem to complement existing ones, the fact that these are provisioned (algorithmically) by large private companies raise issues of commodification, enclosure, and obscurity. The commodification of trust production refers to the process in which trust is transformed from a socially produced and managed resource to a market commodity. Platforms are infrastructures that challenge, and in many relations, replace social, communal, or democratically accountable ways to maintain standards of reputation, trustworthiness and untrustworthiness, and turn trustworthiness signals into a product to be sold on the marketplace (Bodo 2021). Commodification turned such reputation information into indispensable economic resource for the subject of those signals. For example, Gandini (2016) highlighted the highly asymmetric power relations these signals produce. Platform workers depend on the signals to maintain their living, and a loss of access to these systems can mean a loss to one's livelihood. Meanwhile constant labour is necessary to maintain the signal, without any guarantee that such labour will not be suddenly lost due to some obscure, algorithmic decision. Similar logics apply to every social media user. Commodified trustworthiness signals are highly data intensive, algorithmic, yet opaque to the outside. Despite such scores being universal and widely used, little is known about what they reference, and what relationship they bore to our local, context-specific understanding of someone else's trustworthiness.

3.6. Institutional decision-making

As we discussed above, public institutions play a key role in producing trust as a social resource. With the rapid proliferation of private, technological forms of trust production, and the fragmentation of epistemic frameworks, the fate of public trust infrastructures, available (at least in theory) freely and unconditionally for every citizen, seems to be more consequential than ever. Rothstein (2005) argues that impartial political institutions that ensure social and economic equality are key to overcoming social traps, situations where cooperation breaks down due to mutual distrust. On the trustors' side, as Rothstein (2013) argues, individuals often form beliefs about whether 'people in general' can be trusted based on their interactions with public officials and institutions. Public institutions are designed to be part of a network of accountability, an institutionalised system of distrust, which keeps every constituent institution ultimately trustworthy (Sztompka, 1998).

In today's digital society, under the conditions of heightened uncertainties, such as political unrest, economic crises, or rapid technological disruption, the task of providing a universal narrative of the future becomes all but impossible, and this is reflected in the growing distrust in those public institutions whose task it would be to do so (Beckert, 2016; Hetherington, 1998; Kim, 2005). On top of external pressures, there are processes internal to the body of the state, which jeopardise trust based on the accountability, and fairness of public institutions. In the past decades, the integration of algorithmic systems increasingly shaped the way how judicial decisions are made (Fabri, 2024; Finck, 2019; Qiao & Metikoš, 2025), or how welfare and benefits are distributed (Alon-Barkat & Busuioc, 2023). Doubts and concerns about the implementation of algorithmic systems in this sector are not new. There is already a broad scholarly agreement that the integration of AI into public decision-making processes raises significant political and ethical concerns and moral hazards. Real-world disasters, such as the childcare benefits scandal in the Netherlands (van Dam & Freriks, 2020), Australia's Robodebt scheme (Mao, 2023), or Spain's VioGén algorithm's failure to rate the risk of domestic violence (Satariano & Pifarré, 2024), underscores both the importance of accuracy, fairness, empathy and accountability in public decision-making, and the challenges in achieving this within automated processes (Eubanks, 2018; O'Neill, 2020). Risks and failures like these can be anchored in two main issues that algorithmic decision-making creates, and which in turn increase popular distrust and mistrust in public institutions deploying these systems. The first one concerns human agency and accountability (Eubanks, 2018). The second concern is transparency (Danaher, 2016; von Eschenbach, 2021).

Declining human agency, and incomplete algorithmic accountability. The main driver to insert algorithmic decision-making systems in public administration is to replace a subjective, often messy, arbitrary, inconsistent system of human decision-making with a predictable and quantifiably reliable one. Yet, when human discretion and oversight in, for instance, law enforcement or welfare administrations is replaced with automated and algorithmic decision-making mechanisms, accountability mechanisms become blurred (Eubanks, 2018; Lee, 2018; von Eschenbach, 2021). Already in 1996, Helen Nissenbaum (1996) was able to identify four main barriers to accountability in the 'computerized society': (1) the problem of many hands, which translates into the complex, multi-actor supply chains of algorithmic systems (Cobbe et al., 2023) (2) the problem of bugs, which for AI systems, would include various technological issues ranging from bad

data, to incorrect inferences, and lax engineering (Diakopoulos, 2015), (3) blaming the computer, whereby we distance human action from their causal impacts, and (4) ownership without liability, whereby 'software is released in society, for which [citizens] bear the risks' (Nissenbaum, 1996, p. 12). Algorithmic systems that aim to replace human decisions diffuse accountability mechanisms. This happens through replacing the risks we recognise and attempt to mitigate with institutional mechanisms (e.g. anti-discrimination law, professional standards, liability regimes) with risks that are less familiar and less visible. In this sense, algorithms do not resolve the accountability problems of human decision-making, but they tend to amplify them.

Lack of transparency. One of the main sources of algorithmic uncertainty, and problems of accountability, is the opacity of algorithmic institutional decision-making procedures. Proprietary algorithms and datasets are often kept confidential. This lack of transparency limits public insight and scrutiny, and also has an impact on trust (Busuioc et al., 2023; Diakopoulos, 2015; Kroll et al., 2017). However, the relationship between transparency and trust is complex, and the two do not automatically correlate. Citizens evaluate institutional trustworthiness not only on the basis of access to information, but also through pre-existing beliefs about government (Grimmelikhuijsen, 2012). Moreover, transparency that takes the form of explainability (rather than mere accessibility) has been shown to strengthen perceptions of procedural fairness and trust in both algorithms and the people operating them (Grimmelikhuijsen, 2023). With procedural fairness being one of the main drivers of trust in public institutions when algorithmic elements remain a black box and produce outcomes experienced as unfair, institutional trust will ultimately be affected.

4. Theoretical implications: Towards a trust theory for the digital age

In the sections above we discussed some of the most important ways digital technologies may upset and alter pre-existing trust relations, trust building processes, and the institutional mechanisms to safeguard trust. Based on this, we now lay out our proposal on how to include these developments in the theoretical discussions on trust to take into account the specificities of digital intermediaries at the heart of these relationships and processes.

The common development which connects the changes in these six domains is that a new, technological entity steps into the multitude of trust relations with a simple business proposal: 'Trust me, I will take care of all the complexity that you have been struggling with in the past'.

Trusting as an individual attitude, trust development as an everyday situated practice (Eyal et al., 2024), trust-producing public and private institutions (Zucker, 1986), and the production of trust by technological intermediaries (Bodó, 2021) offer the same thing, namely the reduction of future uncertainty and complexity. The four approaches, however, achieve these in radically different ways. Individual trusting attitude is best characterised as a 'leap of faith'. Everyday trust-requiring practices rely on complex support networks for trust, i.e.: context-specific mixes of familiarity, control, insurance and distrust, which amalgamates faith, calculative, reflexive trust, risk management, and some level of control, which form trust over time through contexts-specific interactions. Public trust is a complex and costly, formalised and heavily regulated institutional distrust framework, in which institutional actors form a network of oversight and

accountability. In other words, trust is a labour-, and resource intensive practice in constant search for some anchor, or grounding in some internal or external support. Eyal at al. refers to this process a sort of 'infinite regression of reasons and counter-reasons that precludes explaining trusting by reference to any 'decisive grounds'' (2024, p. 182) Technological intermediaries in the business of trust production propose to simplify this complex search for a reason for trusting, and allow their users to not descend too far into the rabbit hole of who and why to (not) trust, by offering up themselves as the first and last guarantor of trustworthy interactions. Trust technologies offer to absorb many of the risks and harms related to digitally mediated interpersonal and institutional relations and promise to deal with those risks and uncertainties for their users – so they don't have to. They take care of the careful balancing of different trust safeguards, manage the risks, control the counterparties, insure against bad outcomes, create the epistemic frameworks, enforce the rules, etc.

They do so through promising a safe and predictable environment for the relations they (re)mediate. They measure and manage the reputation of what otherwise unknown counterparties, as Airbnb, eBay or Uber do: rather than having to negotiate with a random stranger, in a foreign language over room quality, delayed delivery, or the shortest path taken in a foreign city, platforms provide uniform and impersonal rules and conflict resolution systems which apply across all interactions and localities. Rather than having to guess culture and context-specific cues of trustworthiness, intermediaries force users into the same matrix of legibility, and promise to filter out bad actors. Those intermediaries which disseminate knowledge, news, information, such as YouTube, Facebook, news recommenders, and other forms of social media boast how they allow users to have access to the information they like. Irrespective of whether this takes the form of 'free speech maximalisms' offering a liberation from 'institutional censorship', or manifests as elaborate content moderation and fact checking, the underlying promise is the same: 'you can trust us more to select information for you than anyone else'. A similar logic is observable when we look at the reasons why institutions choose digital intermediaries to assist in their decision-making: data-driven, software-encoded-rules based systems may be more transparent, verifiable and unbiased than humans (Eubanks, 2018; Guo et al., 2025; O'Neil, 2017).

There is, of course, extensive evidence that many of these promises around risk-absorption are far-fetched or simply inaccurate. This tension – between the promises to protect users from risks and harms by taking over the task of dealing with complexity, and the reality of these promises being unfulfilled – highlights the precise point at which we can develop new approaches to integrate the role of these intermediaries into the current trust theory frameworks.

First, we propose to start upgrading Beck's risk framework with the individual and societal uncertainties, risks and harms produced by the third industrial revolution. Even though Beck's theories are based on examples from the era of mass industrial production, risks haven't gone away, but instead accompanied by a new class of risks, which digitisation now produces. We haven't fully integrated these immaterial vulnerabilities in risk theory. The work done on issues such as cybersecurity (Searle et al., 2024; Searle & Renaud, 2023), or e-commerce (Chen & Dhillon, 2003) may address specific issues, but do not necessarily operate at the holistic, societal level. While Beck correctly identified some of the challenges, such as the crisis in expertise, and the inadequacies of the

welfare state to reallocate risks in general, it was written before the emergence of trust-related institutional challenges with regards to the risks and harms brought about by digitalisation, including, the extreme epistemic fragmentation, or the psychological and social effects of algorithmically mediated mass communication.

Second, trust theory and empirical research can build muscle to grapple with how digital intermediaries build trust, shape and interact with the trust formation processes at the individual and societal levels. We have sketched out some of the dimensions in which research can do more: from the design and shaping of interactions, via their influence one's self-representation and legibility, to how they prioritise certain types of social relations, information, or even epistemic structures over others. Though much research is available on these individual points of interference, theory needs to also include how these design decisions, taken together as a whole influence broader economic transactions and institutional decision-making.

In addition to the internal design of trust intermediaries, theory needs to include data as the raw material for trust production, and a source of trust in the digital society. Data, and data driven-decisions are envisioned as tools of transparency and accountability and sources of algorithmic objectivity. We do not question that under certain conditions data can successfully fulfil these roles. Yet, it is also clear that the role of data is more complex than that: it reduces life worlds into what can be measured and what is included in the models, and paradoxically can lead to more, rather than less complexity and uncertainty (Bridle, 2018; Hong, 2020). Data(fication) and trust seems to be a promising domain for the next generation of trust theories.

The third space where we see opportunities for theoretical development concern the very process of how private digital intermediaries deal with the very risks they produce (Griffin, 2025). We argue that digital trust production, in the context of how these intermediaries are regulated (Palumbo & Ducuing, 2025); how they are (not) embedded in the network of other institutions; how they are built as hybrids of intransparent algorithmic systems and exploitative low skilled labour; and their dominant positions in the production of social relations and epistemic systems is problematic irrespective of whether they are successful in absorbing risks or not, irrespective of whether they can be trustworthy trust producers or not.

In the 'ideal' case, some digital trust mediators can be expected to become trustworthy trust producers by successfully absorbing the risks they offer to take care of. If they succeed, online trust production becomes a radically centralised economic activity. This raises the spectre of having large, private, loosely regulated, algorithmic, data-intensive corporate monopolies as the central trust producers in key societal domains, which, due to their position will pose systemic risks for our societies. Just like some corporate entities in other (also trust related) sectors (such as banking, or food, or energy), a few, entrenched monopolies will wield unprecedented amounts of power: having undue, undemocratic, unaccountable influence on key aspects of human life, the environment, and social relations, and take decisions that prioritise short term business interests at the expense of long-term public interest. On the flipside, we also need to acknowledge that while these concentrated trust-producing powers pose systemic risks to societies, in contexts where democratic institutions are weak or absent, such digital intermediaries can paradoxically provide reliability in ways that might otherwise be unattainable. All this opens up the challenge for theory to develop the political economy of trust production,

focusing on the process, conditions and consequences of turning trust into a commodity (Bodo 2021; Polanyi, 1944).

In a more realistic scenario, these digital trust producers ultimately fail to deliver on their promises, and cannot reduce the complexity, cannot absorb and manage all the risks, and therefore simply redistribute them – sometimes in different forms of harms and vulnerabilities. No harm is done if – like blockchains – they fail before they can truly disrupt and crowd out the traditional trust infrastructures. Yet, in many other cases we may have passed that point. Social media disrupted (local) journalism; algorithm-fueled echo chambers successfully question the trustworthiness of public institutions, scientists, experts; and there are real harms suffered by citizens due to algorithmic welfare decisions. In these instances digital trust producers managed to convince a substantial share of citizens, consumers, public servants, entrepreneurs that they can reduce the complexity and deal with the risks on their behalf, while in reality they cannot. This comes as no surprise. It is easier to mask that complexity with technology than to actually 'solve' it. In other words, the trust production activities of digital intermediaries at the moment are misleading. Digital intermediaries promise to manage risks that are impossible to mitigate, and promise to reduce a complexity, while they are actually aggravating it through their own (conflicting) interests (Bodo, 2025). In these scenarios trust theory needs to pay more attention to the role of untrustworthy middlemen and the breakdown of societal-scale trust relations.

Lastly, these trust production infrastructures, developed and controlled by private companies, often aim to supplant traditional public institutions (Bodo 2021). In all above-mentioned cases we observe a rearticulation of trust logics through weakly regulated private industries. Platforms have their own, self-governed entities ensuring 'Trust and Safety' for users with their own policies and content moderation rules. This shift means that trust, which, in the institutional context was once fostered through democratic processes and public accountability, is now generated by proprietary technologies designed primarily for profit and private power rather than public value. Recent changes in regulatory regimes, especially in the European risk-based technology regulation frameworks digital trust producers 'are increasingly made legally responsible for the assessment of the risks that their activities generate, for the determination of the appropriate safeguards and for the implementation of such safeguards' (Palumbo & Ducuing, 2025, p. 5). While it is not uncommon to witness a constantly shifting balance between public and private entities in the production of societal resources, we argue that it is the first time where we can observe such a large scale reallocation of the task of trust production from the institutions of the modern state, and from traditional communal practices towards planetary-scale private infrastructures. While it may be way to early to assess the success and the impact of this transition, we can prepare trust theories to conceptualise it.

5. Acknowledging trust related practices in the digital age

We would like to acknowledge that there is already a lot of research and practical work being done in – among others – policy, information law and platform regulation, as well as platform design, in response to all the problems and challenges in the six domains we identified above. Without the hope of giving a complete picture, we would like to point out that there is, for instance, a tedious discussion on the challenges

and mechanisms to increase online authenticity (and thereby trustworthiness) by means of privacy-protecting, identity management and age verification mechanisms (Beduschi, 2019; Norcie et al., 2013; Stets & Serpe, 2019; Wang et al., 2023). This also extends to practice-oriented research on the questioned effectiveness of the flagging of edited or AI-generated content on users' trust (Burrus et al., 2024; Gamage et al., 2025). For information flows, work on evolving platform policies and (automated) practices of content moderation and fact checking to combat mis- and disinformation should also be noted (Gorwa et al., 2020; Morrow et al., 2022; Persily & Tucker, 2020; Tulin et al., 2023). The fairness and diversity in recommendation systems is also a very active field (Helberger et al., 2018; Vrijenhoek et al., 2024).

There is also much research in the legal and regulatory domains. Tackling user harms, increasing the trustworthy use of AI and countering the power of large digital monopolies through regulations like the Digital Services, the Digital Markets Act or the AI Act has been in the forefront of EU rulemaking for more than a decade now (Laux et al., 2024; van Hoboken et al., 2023; Witt, 2023). This research encompasses a wide range of issues from the focuses on the efficiency of transparency measures (Araujo et al., 2020) via the problems raised by the risk-based regulation approach towards technology (Paul, 2024), to liability rules for disseminating factually incorrect, deliberately deceptive, or manipulative content and other forms of algorithmic bias and discrimination (Fahy et al., 2025). With regard to economic transactions online, research tackled the transparency around reputation systems and the portability and interoperability of trustworthiness scores (Hesse & Teubner, 2020), while for the problems within automated and AI-driven institutional decision-making, legal requirements for meaningful human oversight to preserve public trust and procedural legitimacy are also active fields of research and practical interventions (Green, 2022; Langer et al., 2024; Laux, 2024).

Lastly, research related to the architecture of these technical systems is also looking into interface and platform design solutions which play a role in the trustworthiness perceptions and risk exposure. Issues such as fairness in recommender systems, bias and hallucinations in AI systems, friction-by-design, or better interface choices and controls inform more trustworthy infrastructures of digital trust production (Burke, 2017; Ekstrand et al., 2018; Fabbri, 2023; Kay & Kummerfeld, 2013; Mathur et al., 2021)

All this work is relevant for the trust research discourse and community, and vice versa, all work in the domain of trust research could and should inform the practical responses. Our engagement with the latter body of research allowed us to make two noteworthy observations. First, much of this work references the concept of trust one way or another, in terms like trusted AI, or increasing trust in technology, or aiming to create a more safe and secure digital society. Despite this prevalence of the concept of trust, the majority of the practical responses remains somewhat distant and isolated from the trust literature. It is hard to say whether this is due to the latter's complexity and often abstract nature and language, or its focus on individual trusting attitudes, and/or organisational structures, as opposed to sociotechnical assemblages. But this gap points to the need we also identified in our paper, that trust theory and trust related practices in the digital domain need a closer alignment.

Second, all this practical work focuses on quite narrowly defined problems and loci of interventions. In contrast, we argue is that all these local, and often isolated responses are part of a closely knit continuum. The labels news consumers may or may not see next to

AI generated content, and the liability of online intermediaries for their potential failure to protect human rights are all essential components of the complex and closely interrelated mechanisms which allow (or not) individuals as users, consumers or citizens to trust each other in all those social, economic, or intimate relations they encounter with each other through digital infrastructures. Trust research can provide the overarching framework which brings together all the currently rather isolated efforts and exposes them to each other.

6. Conclusion

From our review we saw that trust mediation through digital actors is happening in all socio-economic areas. Having outlined trust dynamics in six distinct domains, we found that we lack a comprehensive framework to unify these developments. We introduced the idea that what connects digital trust mediators is their efforts to absorb and redistribute external risks by promising to simplify the complex landscape of reality and offering to handle complexities on behalf of users and consumers. Unfortunately, this simplification of risks is creating a potentially false, and perilously misguided sense of trustworthiness and security. Additional systemic or domain-specific risks emerge. Our theory posits that, though different, these risks stem from the same mechanism of trust being increasingly centralised in the hands of global hegemonies through the design, data and usage of data, often sidelining local practices and leading to a homogenised, less transparent system of trust. Our paper calls for theory development to understand the implications of these dynamics and address the resulting trust disruptions. In particular, future research could examine the conditions under which digital trust mediation alters existing trust institutions, and societal trust dynamics, and how practical responses contribute to that process.

Notes

1. It is important to highlight that photo editing and the emphasis and glorification of certain physical appearances can, for instance, lead to disordered eating, orthorexia, body image issues, and depression (Tiggemann, 2022; Wick & Keel, 2020).
2. Particularized trust is understood as placing trust in particular targets like relatives or acquaintances and is built on pre-existing ties, such as kinship, geographic, or interest-based ties (Uslaner, 2002; Zheng et al., 2023).
3. Knowledge produced by scientific institutions may conflict with local/indigenous modes of knowledge production, and the latter has been shown to be harmed by the former (Krikorian & Kapczynski, 2010). That being said, for the better or worse, modern scientific disciplines based on the scientific method have been accepted as a bridge between the various pockets of local, communal knowledge contained under the umbrella of the modern state.

Acknowledgements

This publication has been supported via the NWO Grant A Law-compliant Trustworthy Blockchain Economy (BLOCK.2019.004), as well as the UvA Trust in the digital society Research Priority Area.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Balázs Bodó is a Professor of Information Law and Policy with a special emphasis on technology governance at the Institute for Information Law (IViR) at the University of Amsterdam. He is the founding (co)director of the University of Amsterdam's interdisciplinary research area on Trust in the digital society. His academic interests include digital piracy, decentralised techno-social systems, shadow libraries, informal media economies, regulatory conflicts around new technological architectures and trust.

Linda Weigl is a postdoctoral researcher at the University of Amsterdam's (UvA) Institute for Information Law (IViR) with a background in Political Science and European public policy. The goal of her research is to uncover the power dynamics and risks behind digital trust infrastructures by scrutinising the accountability of companies, governments' regulatory power and users' risk awareness.

Theo Araujo is a Professor of Media, Organizations and Society in the Department of Communication Science and Scientific Director of the Amsterdam School of Communication Research (ASCoR) at the University of Amsterdam. His research focuses on the increasing adoption of artificial intelligence and related technologies within our communication environment, including conversational agents and automated decision-making.

ORCID

Theo Araujo http://orcid.org/0000-0002-4633-9339

References

Abiri, G., & Buchheim, J. (2022). Beyond true and false: Fake news and the digital epistemic divide. *Michigan Technology Law Review*, *29*(1), 59–110. https://doi.org/10.36645/mtlr.29.1.beyond

Aghion, P., Algan, Y., Cahuc, P., & Shleifer, A. (2010). Regulation and distrust. *The Quarterly Journal of Economics*, *125*(3), 1015–1049. https://doi.org/10.1162/qjec.2010.125.3.1015

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, *33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Are, C., Talbot, C., & Briggs, P. (2024). Social media affordances of LGBTQIA+ expression and community formation. *Convergence*, *31*(4), 1401–1422. https://doi.org/10.1177/13548565241296628

Bachmann, R. (2001). Trust, power and control in trans-organizational relations. *Organization Studies*, *22*(2), 337–365. https://doi.org/10.1177/0170840601222007

Bachmann, R., & Kroeger, F. (2017). Trust, power or money: What governs business relationships? *International Sociology*, *32*(1), 3–20. https://doi.org/10.1177/0268580916673747

Baier, C. (2024). Narratives of post-truth: Lyotard and the epistemic fragmentation of society. *Theory, Culture & Society*, *41*(1), 95–110. https://doi.org/10.1177/02632764231162027

Baker, A. J. (2005). *Double click: Romance and commitment among online couples*. Hampton Press.

Bamberger, K. A., & Lobel, O. (2017). Platform market power (SSRN Scholarly Paper No. 3074717). https://papers.ssrn.com/abstract=3074717

Bates, S. (2017). Revenge porn and mental health: A qualitative analysis of the mental health effects of revenge porn on female survivors. *Feminist Criminology*, *12*(1), 22–42. https://doi.org/10.1177/1557085116654565

Beck, U. (1992). *Risk society: Towards a new modernity*. Sage Publications.

Beckert, J. (2016). *Imagined futures: Fictional expectations and capitalist dynamics*. Harvard University Press.

Beckett, C., & Deuze, M. (2016). On the role of emotion in the future of journalism. *Social Media + Society*, *2*(3), 2056305116662395. https://doi.org/10.1177/2056305116662395

Beduschi, A. (2019). Digital identity: Contemporary challenges for data protection, privacy and non-discrimination rights. *Big Data & Society*, *6*(2), 2053951719855091. https://doi.org/10.1177/2053951719855091

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, *47*(2), 241–262. https://doi.org/10.1017/S0007123415000186

Bodo, B. (2025). Humble tools of divine intervention – The misunderstood role of algorithms in public opinion formation. *Dialogues on Digital Society*, 29768640251369974. https://doi.org/10.1177/29768640251369974

Bodó, B. (2021). Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. *New Media & Society*, *23*(9), 2668–2690. https://doi.org/10.1177/1461444820939922

Bodó, B. (2021, May 11). The Commodification of Trust. Blockchain & Society Policy Research Lab Research Nodes 2021/1, Amsterdam Law School Research Paper No. 2021-22, Institute for Information Law Research Paper No. 2021-01. https://dx.doi.org/10.2139/ssrn.3843707

Bolton, G., Greiner, B., & Ockenfels, A. (2013). Engineering trust: Reciprocity in the production of reputation information. *Management Science*, *59*(2), 265–285. https://doi.org/10.1287/mnsc.1120.1609

Botsman, R. (2017). *Who can you trust? How technology brought us together and why it might drive us apart (illustrated edition)*. PublicAffairs.

Bridle, J. (2018). *New dark age: Technology, knowledge and the end of the future*. Verso.

Bryce, J., & Fraser, J. (2014). The role of disclosure of personal information in the evaluation of risk and trust in young peoples' online interactions. *Computers in Human Behavior*, *30*, 299–306. https://doi.org/10.1016/j.chb.2013.09.012

Burke, R. (2017). Multisided fairness for recommendation (No. arXiv:1707.00093). arXiv. https://doi.org/10.48550/arXiv.1707.00093

Burrus, O., Curtis, A., & Herman, L. (2024). Unmasking AI: Informing authenticity decisions by labeling AI-generated content. *Interactions*, *31*(4), 38–42. https://doi.org/10.1145/3665321

Busuioc, M., Curtin, D., & Almada, M. (2023). Reclaiming transparency: Contesting the logics of secrecy within the AI Act. *European Law Open*, *2*(1), 79–105. https://doi.org/10.1017/elo.2022.47

Chambers, D. (2013). Social media and personal relationships: Online intimacies and networked friendship.

Chambers, D. (2017). Networked intimacy: Algorithmic friendship and scalable sociality. *European Journal of Communication*, *32*(1), 26–36. https://doi.org/10.1177/0267323116682792

Chen, S. C., & Dhillon, G. S. (2003). Interpreting dimensions of consumer trust in E-commerce. *Information Technology and Management*, *4*(2), 303–318. https://doi.org/10.1023/A:1022962631249

Cobbe, J., Veale, M., & Singh, J. (2023). Understanding accountability in algorithmic supply chains. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1186–1197). https://doi.org/10.1145/3593013.3594073

Cook, K. S., Levi, M., & Hardin, R. (2009). *Whom can we trust? How groups, networks, and institutions make trust possible*. Russell Sage Foundation.

Dahlgren, P. (2018). Media, knowledge and trust: The deepening epistemic crisis of democracy. *Javnost – The Public*, *25*(1–2), 20–27. https://doi.org/10.1080/13183222.2018.1418819

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, *29*(3), 245–268. https://doi.org/10.1007/s13347-015-0211-1

Diakopoulos, N. (2015). Algorithmic accountability. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Ekstrand, M. D., Tian, M., Azpiazu, I. M., Ekstrand, J. D., Anuyah, O., McNeill, D., & Pera, M. S. (2018). All The cool kids, how do they fit in? Popularity and demographic biases in recommender evaluation and effectiveness. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 172–186). https://proceedings.mlr.press/v81/ekstrand18b.html

Entman, R. M., & Usher, N. (2018). Framing in a fractured democracy: Impacts of digital technology on ideology, power and cascading network activation. *Journal of Communication*, *68*(2), 298–308. https://doi.org/10.1093/joc/jqx019

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Publishing Group.

Eyal, G., Au, L., & Capotescu, C. (2024). Trust is a verb! A critical reconstruction of the sociological theory of trust. *Sociologica*, *18*(2), 169–191. https://doi.org/10.6092/issn.1971-8853/19316

Fabbri, M. (2023). Social influence for societal interest: A pro-ethical framework for improving human decision making through multi-stakeholder recommender systems. *AI & Society*, *38*(2), 995–1002. https://doi.org/10.1007/s00146-022-01467-2

Fabri, M. (2024). From court automation to e-justice and beyond in Europe. *International Journal for Court Administration*, *15*(3), 1–7. https://doi.org/10.36745/ijca.640

Fahy, R., Buijs, D., & van Hoboken, J. (2025). The regulation of disinformation under the digital services Act. *Media and Communication*, *13*, Article 9615. https://doi.org/10.17645/mac.9615

Filippi, P. D., Wray, C., & Sileno, G. (2021). Smart contracts. *Internet Policy Review*, *10*(2), 1–9. https://policyreview.info/glossary/smart-contracts

Finck, M. (2019). Automated decision-making and administrative law (SSRN Scholarly Paper No. 3433684). Social Science Research Network. https://papers.ssrn.com/abstract=3433684

Foucault, M. (1981). The order of discourse. In R. Young (Ed.), *Untying the text: A post-structuralist reader* (pp. 51–78). Routledge.

Fukuyama, F. (1996). *Trust: The social virtues and the creation of prosperity*. Free Press.

Gamage, D., Sewwandi, D., Zhang, M., & Bandara, A. K. (2025). Labeling synthetic content: user perceptions of label designs for AI-generated content on social media. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–29). https://doi.org/10.1145/3706598.3713171

Gandini, A. (2016). *Reputation economy: Understanding knowledge work in digital society*. Palgrave Macmillan.

Giddens, A. (1990). *The consequences of modernity*. Polity Press.

Giddens, A. (1999). Risk and responsibility. *Modern Law Review*, *62*(1), 1. https://doi.org/10.1111/1468-2230.00188

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 2053951719897945. https://doi.org/10.1177/2053951719897945

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, *45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Greif, A., Milgrom, P., & Weingast, B. R. (1994). Coordination, commitment, and enforcement: The case of the merchant guild. *Journal of Political Economy*, *102*(4), 745–776. https://doi.org/10.1086/261953

Griffin, R. (2025). *A stakeholder mapping and research agenda. The politics of risk in the digital services act*. Weizenbaum Institute. https://doi.org/10.34669/WI.WJDS/5.2.6

Grimmelikhuijsen, S. (2012). Linking transparency, knowledge and citizen trust in government: An experiment. *International Review of Administrative Sciences*, *78*(1), 50–73. https://doi.org/10.1177/0020852311429667

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, *83*(2), 241–262. https://doi.org/10.1111/puar.13483

Guo, E., Geiger, G., & Braun, J.-C. (2025). Inside Amsterdam's high-stakes experiment to create fair welfare AI. *MIT Technology Review*, https://www.technologyreview.com/2025/06/11/1118233/amsterdam-fair-welfare-ai-discriminatory-algorithms-failure/

Håkansson, P., & Witmer, H. (2015). Social media and trust: A systematic literature review. *Journal of Business and Economics*, *6*(3), 517–524. https://doi.org/10.15341/jbe(2155-7950)/03.06.2015/010

Hamm, J. A., & Banner, F. (2025). Vulnerability: The active ingredient of trust in public governance. In F. Six , J. A. Hamm , D. Latusek , E. van Zimmeren , & K. Verhoest (Eds.), *Handbook on trust in public governance* (pp. 25–39). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781802201406/chapter2.xml

Harris, E., & Bardey, A. C. (2019). Do Instagram profiles accurately portray personality? An investigation into idealized online self-presentation. *Frontiers in Psychology*, *10*, 1–13. https://doi.org/10.3389/fpsyg.2019.00871

Haythornthwaite, C. (2002). Strong, weak, and latent ties and the impact of new media. *The Information Society*, *18*(5), 385–401. https://doi.org/10.1080/01972240290108195
Haywood, C. (2018). Men, masculinity and contemporary dating.
Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700
Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207. https://doi.org/10.1080/1369118X.2016.1271900
Herian, R. (2021). Smart contracts: A remedial analysis. *Information & Communications Technology Law*, *30*(1), 17–34. https://doi.org/10.1080/13600834.2020.1807134
Herman, E. S., & Chomsky, N. (2002). *Manufacturing consent: The political economy of the mass media*. Pantheon Books.
Hesse, M., & Teubner, T. (2020). Reputation portability – Quo vadis? *Electronic Markets*, *30*(2), 331–349. https://doi.org/10.1007/s12525-019-00367-6
Hetherington, M. J. (1998). The political relevance of political trust. *American Political Science Review*, *92*(4), 791–808. https://doi.org/10.2307/2586304
Hong, S. (2020). *Technologies of speculation: The limits of knowledge in a data-driven society*. New York University Press.
Houser, K. (2020, January 4). Airbnb claims its AI can predict whether guests are psychopaths. *Futurism*. https://futurism.com/the-byte/airbnb-ai-predict-psychopaths
Jamieson, L. (2013). Personal relationships, intimacy and the self in a mediated and global digital age. In K. Orton-Johnson, & N. Prior (Eds.), *Digital sociology: Critical perspectives* (pp. 13–33). Palgrave Macmillan UK. https://doi.org/10.1057/9781137297792_2
Kay, J., & Kummerfeld, B. (2013). Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *ACM Transactions on Interactive Intelligent Systems*, *2*(4), 24–42. https://doi.org/10.1145/2395123.2395129
Kerr, D. (2019, November 26). *Some Uber drivers aren't who you think they are*. CNET. https://www.cnet.com/news/uber-drivers-using-fake-identities-isnt-just-alondon-problem/
Keymolen, E. L. O. (2016). *Trust on the line: A philosophical exploration of trust in the networked era*. Erasmus University Rotterdam. hdl.handle.net/1765/93210.
Khamis, S., Ang, L., & Welling, R. (2016). Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies*, *8*(2), 191–208. https://doi.org/10.1080/19392397.2016.1218292
Kim, H. J., & Cameron, G. T. (2011). Emotions matter in crisis: The role of anger and sadness in the publics' response to crisis news framing and corporate crisis response. *Communication Research*, *38*(6), 826–855. https://doi.org/10.1177/0093650210385813
Kim, S.-E. (2005). The role of trust in the modern administrative state: An integrative model. *Administration & Society*, *37*(5), 611–635. https://doi.org/10.1177/0095399705278596
Kramer, R., Tyler, T., Lewicki, R., & Bunker, B. (1996). Developing and maintaining trust in work relationships. In R. Kramer & T. Tyler (Eds.), *Developing and maintaining trust in work relationships* (pp. 114–139). SAGE Publications, Inc., https://doi.org/10.4135/9781452243610.n7
Krämer, N., & Haferkamp, N. (2011). *Online self-presentation: Balancing privacy concerns and impression construction on social networking sites* (pp. 127–141). https://doi.org/10.1007/978-3-642-21521-6_10
Krause, N. M., Freiling, I., & Scheufele, D. A. (2022). The "infodemic" infodemic: Toward a more nuanced understanding of truth-claims and the need for (not) combatting misinformation. *The Annals of the American Academy of Political and Social Science*, *700*(1), 112–123. https://doi.org/10.1177/00027162221086263
Krikorian, G., & Kapczynski, A. (2010). *Access to knowledge*. Zone Books. http://www.tcrecord.org/Content.asp?ContentId=775
Kristinsdottir, K. H., Gylfason, H. F., & Sigurvinsdottir, R. (2021). Narcissism and social media: The role of communal narcissism. *International Journal of Environmental Research and Public Health*, *18*(19), 1–14. https://doi.org/10.3390/ijerph181910106
Kroll, J., Huey, J., Barocas, S., Felten, E., Reidenberg, J., Robinson, D., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*(3), 633–706.

Langer, M., Baum, K., & Schlicker, N. (2024). Effective human oversight of AI-based systems: A signal detection perspective on the detection of inaccurate and unfair outputs. *Minds and Machines*, *35*(1), 1. https://doi.org/10.1007/s11023-024-09701-0

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society*, *39*(6), 2853–2866. https://doi.org/10.1007/s00146-023-01777-z

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, *18*(1), 3–32. https://doi.org/10.1111/rego.12512

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(1), 34–47. https://doi.org/10.1037/0033-2909.107.1.34

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 2053951718756684. https://doi.org/10.1177/2053951718756684

Lyotard, J.-F., Seidman, S., & Alexander, J. C. (2001). The postmodern condition. In *The new social theory reader: Contemporary debates* (pp. 166–167). Routledge. https://books.google.nl/books?hl=en&lr=&id=mXgp5b8yvvkC&oi=fnd&pg=PA166&ots=p7kwdynvyD&sig=kwipOkxEV_TctyF5fjuVPyrqzr0

Ma, H., Zhou, D., Liu, C., Lyu, M. R., & King, I. (2011). Recommender systems with social regularization. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (pp. 287–296). https://doi.org/10.1145/1935826.1935877

Mao, F. (2023, July 7). Robodebt: Illegal Australian welfare hunt drove people to despair. *BBC News*. https://www.bbc.com/news/world-australia-66130105

Martin, A., Mikołajczak, G., & Orr, R. (2022). Does process matter? Experimental evidence on the effect of procedural fairness on citizens' evaluations of policy outcomes. *International Political Science Review*, *43*(1), 103–117. https://doi.org/10.1177/0192512120908874

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

Mathur, A., Kshirsagar, M., & Mayer, J. (2021). What makes a dark pattern … dark? Design attributes, normative considerations, and measurement methods. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). https://doi.org/10.1145/3411764.3445610

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

Metzger, M. J., Hartsell, E. H., & Flanagin, A. J. (2020). Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Communication Research*, *47*(1), 3–28. https://doi.org/10.1177/0093650215613136

Milmo, D., & Hern, A. (2024, March 8). 'We definitely messed up': Why did Google AI tool make offensive historical images? *The Guardian*. https://www.theguardian.com/technology/2024/mar/08/we-definitely-messed-up-why-did-google-ai-tool-make-offensive-historical-images

Möllering, G. (2006). Trust, institutions, agency: Towards a neoinstitutional theory of trust. In *Handbook of Trust Research*. Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/9781843767541.00029.xml

Morrow, G., Swire-Thompson, B., Polny, J. M., Kopec, M., & Wihbey, J. P. (2022). The emerging science of content labeling: Contextualizing social media content moderation. *Journal of the Association for Information Science and Technology*, *73*(10), 1365–1386. https://doi.org/10.1002/asi.24637

Nai, A., Vermeer, S., Bos, L., & Hameleers, M. (2024). Disenchantment with political information: Attitudes, processes, and effects. In T. Araujo, & P. Neijens (Eds.), *Communication research into the digital society: Fundamental insights from the Amsterdam school of communication research* (pp. 69–86). Amsterdam University Press. https://doi.org/10.5117/9789048560592

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, *2*(1), 25–42. https://doi.org/10.1007/BF02639315

Nooteboom, B. (2003). The trust process. In *The Trust Process in Organizations*. Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/9781843760788.00008.xml

Norcie, G., De Cristofaro, E., Bellotti, V. (2013). Bootstrapping trust in online dating: Social verification of online dating profiles. In: A. A. Adams, M. Brenner, M. Smith (Eds.), *Financial cryptography and data security. FC 2013. Lecture notes in computer science*, Vol. 7862. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41320-9_10

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy (first paperback edition)*. B/D/W/Y Broadway Books.

O'Neill, O. (2020). Trust and accountability in a digital age. *Philosophy*, *95*(1), 3–17. https://doi.org/10.1017/S0031819119000457

Paat, Y.-F., & Markham, C. (2021). Digital crime, trauma, and abuse: Internet safety and cyber risks for adolescents and emerging adults in the 21st century. *Social Work in Mental Health*, *19*(1), 18–40. https://doi.org/10.1080/15332985.2020.1845281

Palumbo, A., & Ducuing, C. (2025). The blurring of the public-private dichotomy in risk-based EU digital regulation: Challenges for the rule of law (SSRN Scholarly Paper No. 5395729). *Social Science Research Network*. https://doi.org/10.2139/ssrn.5395729

Paul, R. (2024). European artificial intelligence "trusted throughout the world": Risk-based regulation and the fashioning of a competitive common AI market. *Regulation & Governance*, *18*(4), 1065–1082. https://doi.org/10.1111/rego.12563

Persily, N., & Tucker, J. A. (2020). *Social media and democracy*. Cambridge University Press.

Polanyi, K. (1944). *The great transformation: The political and economic origins of our time*. Beacon Press.

Poletti, A. (2020). *Stories of the self: Life writing after the book*. NYU Press.

Poletti, A., & Rak, J. (2014). *Identity technologies: Constructing the self online*. University of Wisconsin Press.

Putnam, R. D., Leonardi, R., & Nonetti, R. Y. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press. https://doi.org/10.2307/j.ctt7s8r7

Qiao, C., & Metikoš, L. (2025). Judicial automation: Balancing rights protection and capacity-building (SSRN Scholarly Paper No. 5125645). *Social Science Research Network*. https://doi.org/10.2139/ssrn.5125645

Ross Arguedas, A., Robertson, C. T., Fletcher, R., & Nielsen, R. K. (2022). *Echo chambers, filter bubbles, and polarisation: A literature review*. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-ETXJ-7K60

Rothstein, B. (2005). *Social traps and the problem of trust*. Cambridge University Press. https://doi.org/10.1017/CBO9780511490323

Rothstein, B. (2013). Corruption and social trust: Why the fish rots from the head down. *Social Research*, *80*(4), 1009–1032. https://doi.org/10.1353/sor.2013.0040

Satariano, A., & Pifarré, R. T. (2024, July 18). An algorithm told police she was safe. Then her husband killed her. *The New York Times*. https://www.nytimes.com/interactive/2024/07/18/technology/spain-domestic-violence-viogen-algorithm.html

Searle, R., & Renaud, K. (2023, January 6). Trust and vulnerability in the cybersecurity context. *Hawaii International Conference on System Science 2023. Hawaii International Conference on System Science 2023, USA*. https://strathprints.strath.ac.uk/82574/

Searle, R., Renaud, K. V., & van der Werff, L. (2024). Shaken to the core: Trust trajectories in the aftermaths of adverse cyber events. *Journal of Intellectual Capital*, *25*(5–6), 1154–1183. https://doi.org/10.1108/JIC-02-2024-0038

Shapiro, S. (1987). The social control of impersonal trust. *American Journal of Sociology*, *93*(3), 623–658. https://doi.org/10.1086/228791

Stets, J. E., & Serpe, R. T. (2019). *Identities in everyday life*. Oxford University Press.

Sztompka, P. (1998). Trust, distrust and two paradoxes of democracy. *European Journal of Social Theory*, *1*(1), 19–32. https://doi.org/10.1177/136843198001001003

Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, *8*(1), 321–340. https://doi.org/10.1146/annurev-economics-080315-015325

The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, Organization of American States (OAS) Special Rapporteur on Freedom of Expression, & African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. (2017). *Joint declaration on freedom of expression and "fake news", disinformation and propaganda (No. FOM.GAL/3/17)*. Organization for Security and Co-operation in Europe. https://www.osce.org/files/f/documents/6/8/302796.pdf

Tiggemann, M. (2022). Digital modification and body image on social media: Disclaimer labels, captions, hashtags, and comments. *Body Image*, *41*, 172–180. https://doi.org/10.1016/j.bodyim.2022.02.012

Trilling, D. (2019). Conceptualizing and measuring news exposure as network of users and news items. *Herbert von Halem Verlag*. https://dare.uva.nl/search?fac=fgw-ashms;f7-fulltext=yes;f13-organisation=Faculty%20of%20Social%20and%20Behavioural%20Sciences%20(FMG)::Amsterdam%20School%20of%20Communication%20Research%20(ASCoR);f14-typeClassification=Chapter;docsPerPage=1;startDoc=32

Tulin, M., Hameleers, M., de Vreese, C., Opgenhaffen, M., & Wouters, F. (2023). Beyond belief correction: Effects of the truth sandwich on perceptions of fact-checkers and verification intentions. *Journalism Practice*, *0*(0), 1–20. https://doi.org/10.1080/17512786.2024.2311311

Turkle, S. (2012). Alone together: Why we expect more from technology and less from each other.

Uslaner, E. M. (2002). *The moral foundations of trust*. Cambridge University Press. https://doi.org/10.1017/CBO9780511614934

van Dam, C., & Freriks, J. F. C. (2020). Ongekend onrecht Verslag – Parlementaire ondervragingscommissie Kinderopvangtoeslag (Nr 35510). Tweede Kamer der Staten Generaal. https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf

van der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies*, *24*(6), 803–823. https://doi.org/10.1080/1461670X.2023.2187652

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

van Hoboken, J., Quintais, J. P., Appelman, N., Fahy, R., Buri, I., & Straub, M. (2023). Putting the digital services act into practice: enforcement, access to justice, and global implications (SSRN Scholarly Paper No. 4384266). *Social Science Research Network*. https://papers.ssrn.com/abstract=4384266

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Vrijenhoek, S., Daniil, S., Sandel, J., & Hollink, L. (2024). Diversity of what? On the different conceptualizations of diversity in recommender systems. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 573–584). https://doi.org/10.1145/3630106.3658926

Wang, S., Pang, M.-S., & Pavlou, P. A. (2023). Cure or poison? Identity verification and the posting of fake news on social media. In A. R. Dennis, D. F. Galletta, & J. Webster (Eds.), *Fake news on the internet*. Routledge.

Werbach, K. (2018). *The blockchain and the new architecture of trust (illustrated edition)*. The MIT Press.

Wick, M. R., & Keel, P. K. (2020). Posting edited photos of the self: Increasing eating disorder risk or harmless behavior? *International Journal of Eating Disorders*, *53*(6), 864–872. https://doi.org/10.1002/eat.23263

Witt, A. C. (2023). The digital markets act: Regulating the wild west. *Common Market Law Review*, *60*(3), 625–666. https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2023047.pdf

Yeo, S. K., Xenos, M. A., Brossard, D., & Scheufele, D. A. (2015). Selecting our own science: How communication contexts and individual traits shape information seeking. *The Annals of the American*

Academy of Political and Social Science, *658*(1), 172–191. https://doi.org/10.1177/0002716214557782

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, *24*(5), 1816–1836. https://doi.org/10.1016/j.chb.2008.02.012

Zheng, J., Wang, T. Y., & Zhang, T. (2023). The extension of particularized trust to generalized trust: The moderating role of long-term versus short-term orientation. *Social Indicators Research*, *166*(2), 269–298. https://doi.org/10.1007/s11205-023-03064-2

Zillmann, D., & Bryant, J. (1985). Affect, mood, and emotion as determinants of selective exposure. In D. Zillman, & J. Bryant (Eds.), *Selective exposure to communication* (pp. 157–190). Hillsdale: Erlbaum.

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, *8*, 53–111.