Communications of the Association for Information Systems

Volume 52 Article 16

2-23-2023

Algorithmic Transparency: Concepts, Antecedents, and Consequences – A Review and Research Framework

Tobias Bitzer
Technische Universitat Dresden, tobias.bitzer@mailbox.tu-dresden.de

Martin Wiener
Technische Universitat Dresden

W. Alec Cram
University of Waterloo

Recommended Citation

Bitzer, T., Wiener, M., & Cram, W. (2023). Algorithmic Transparency: Concepts, Antecedents, and Consequences – A Review and Research Framework. Communications of the Association for Information Systems, 52, pp-pp. https://doi.org/10.17705/1CAIS.05214

Communications of the Association for Information Systems

Research Article DOI: 10.17705/1CAIS.05214 ISSN: 1529-3181

Algorithmic Transparency: Concepts, Antecedents, and Consequences – A Review and Research Framework

Tobias Bitzer
TU Dresden
Dresden, Germany
tobias.bitzer@mailbox.tu-dresden.de

Martin Wiener
TU Dresden
Dresden, Germany

W. Alec Cram
University of Waterloo
Waterloo, Canada

Abstract:

The widespread and growing use of algorithm-enabled technologies across many aspects of public and private life is increasingly sparking concerns about the lack of transparency regarding the inner workings of algorithms. This has led to calls for (more) algorithmic transparency (AT), which refers to the disclosure of information about algorithms to enable understanding, critical review, and adjustment. To set the stage for future research on AT, our study draws on previous work to provide a more nuanced conceptualization of AT, including the explicit distinction between *AT as action* and *AT as perception*. On this conceptual basis, we set forth to conduct a comprehensive and systematic review of the literature on AT antecedents and consequences. Subsequently, we develop an integrative framework to organize the existing literature and guide future work. Our framework consists of seven central relationships: (1) AT as action versus AT as perception; factors (2) triggering and (3) shaping AT as action; (4) factors shaping AT as perception; as well as AT as perception leading to (5) rational-cognitive and (6) affective-emotional responses, and to (7) (un-)intended behavioral effects. Building on the review insights, we identify and discuss notable research gaps and inconsistencies, along with resulting opportunities for future research.

Keywords: Algorithmic Transparency (AT), AT as Action, AT as Perception, Literature Review, Research Framework.

This manuscript underwent editorial review. It was received 01/05/2022 and was with the authors for six months for one revision. Mazen ElMasri served as Associate Editor.

1 Introduction

> *It was a curious crowd of protestors that gathered outside the United Kingdom's (UK) Department of Education (DoE) in London in August 2020. Mostly comprised of high school graduates, the group carried signs stating: "trust our teachers", "the algorithm stole my future", and the most famous statement, which would lend its name to an entire protest movement, "f*** the algorithm" (Amoore, 2020). The students were rallying against an algorithm that adjusted the grades of 2020 graduates for effects of the COVID-19 pandemic and had led to grade reductions for almost 40% of the students. However, these downticks had not affected all students equally. As the algorithm was primarily based on past performance in the respective schools, the adjustments disproportionately impacted students from schools where results had traditionally been poorer, while students at top schools hardly suffered any reductions. The resulting public outrage was massive, eventually leading the DoE to scrap the algorithm and return to the original results (Maak, 2021). Media outlets came to call these demonstrations "the first true protest movement of the digital age" (Maak, 2021) and the "future of political protest" (Amoore, 2020), while another commentator bluntly stated that "the fact that an algorithm created with such a lack of transparency and accountability was trusted with the futures of hundreds of thousands of schoolchildren is deeply concerning."* (Ammara, 2020)

Until recently, few would have expected that algorithms—soberly defined as "*encoded procedures for transforming input data into a desired output based on specified calculations*" (Gillespie, 2014, p. 167)—could cause such a public outcry. The above example of the UK DoE is only one of many recent instances where humans have clashed with algorithmic judgements and demanded that algorithms be laid open, changed, or abolished altogether. In short, algorithms have transformed from a purely technical artifact to a subject of heated public debate.

As illustrated by the comments on the DoE case, a recurring theme in the discussions on algorithm-based decision making has been the call for a*lgorithmic transparency* (AT), which refers to the provisioning of information about algorithms to enable understanding, review, and adjustment (Diakopoulos & Koliska, 2017). Following the recent technological and social developments around the emergence of (big) data-driven decision-making (Galliers et al., 2017; Markus, 2017; Newell & Marabelli, 2015), paired with the outspoken criticism of the inscrutability of algorithms (Danaher, 2016; Marjanovic et al., 2018; Pasquale, 2015), AT has attracted growing academic research interest. A recent survey found a more than ten-fold increase in academic publications containing the term "algorithmic transparency" between 2016 and 2020, from about 45 to almost 500 (dimensions.ai, 2021[1]). While the study of AT originally stems from the field of digital journalism (Diakopoulos & Koliska, 2017), a growing interest from other disciplines has extended inquiries into topics including AT in recommendation agents for demand forecasting (Lehmann et al., 2020), AT in behavioral online advertising (Eslami et al., 2018), and AT in criminal intelligence profiling (Zouave & Marquenie, 2017), to name but a few.

From a theoretical perspective, authors have argued that transparency, including AT, may refer both to an *act* of making information available and to a *state* of information being available (Rader et al., 2018). This suggests that there is always a transparency-disclosing and a transparency-receiving party (Bernstein, 2017). There have been arguments that transparency can (and should) be part of a larger effort to educate the public on the "true nature and possible shortcomings of algorithmic governance" (Zarsky, 2016, p. 121). This has also raised questions regarding the "utopian hopes that transparency will reduce misconduct and cure problems in regulation, management, and organizations" (Flyverbom, 2016, p. 111). In the context of information systems (IS), H. J. Watson and Nations (2019) recently published a paper on "Addressing the Growing Need for Algorithmic Transparency", highlighting the relevance of the concept. Still, notwithstanding the growing interest in AT, research to date lacks an in-depth conceptual foundation and empirical results on the antecedents and consequences of AT have been inconclusive. For example, it is unclear whether AT has a neutral (Cramer et al., 2008), positive (Lai & Tan, 2019; Wang & Benbasat, 2016), or curvilinear (Kizilcec, 2016) relationship with trust. Similar inconsistencies can be observed for other relationships of interest, such as the influence of AT on adoption behavior (Eslami et al., 2019).

Bearing this ambiguity in mind and considering the evolving practical and academic advancements on the topic, as well as the groundwork laid by H. J. Watson and Nations (2019), we argue that it is time for a

[1] Using the free version of this bibliometric tool, we searched for "algorithmic transparency" in the full data.

review of the literature on AT. Not only has such an endeavor, to our knowledge, not yet been undertaken; but more importantly, we expect that such a review will reveal critical research gaps and inconsistencies and inform future research in the area. In particular, we aim to extend past work and contribute to the extant body of knowledge by (1) strengthening the conceptual foundations of AT; (2) reviewing the AT literature in a systematic manner; and (3) deriving a research framework. In doing so, our study offers a more nuanced conceptualization of AT, including the explicit distinction between *AT as action* and *AT as perception*, and delineates AT from related concepts. Moreover, following established guidelines (e.g., vom Brocke et al., 2015; Okoli & Schabram, 2010; Paré et al., 2015; Webster & Watson, 2002) and drawing on a review sample of 50 studies (published between 2002 and 2021), our study offers a comprehensive review and synthesis of prior research on the antecedents and consequences of AT. Finally, we propose an integrative framework to organize existing and guide future work.

2 Conceptual Foundations

As part of the larger discussion on topics like big data and datification, scholars of IS and related disciplines have been critical observers of the growing influence of algorithm-enabled technologies on all aspects of human life (Markus, 2017; Markus & Topi, 2015; Newell & Marabelli, 2015). Thus, the lack of information on the inner workings of algorithms has been problematized from an early stage and proposals for remedies have been put forward within the concepts of algorithmic transparency, explainable artificial intelligence (xAI), and algorithmic accountability. However, in doing so, researchers often employ inconsistent terminology (Rosenfeld & Richardson, 2019). In response, we first provide a more nuanced conceptualization of the AT concept and then delineate this concept from related ones.

2.1 Algorithmic Transparency

Despite prior research, it was not until 2017 that the concept of AT was coined by Diakopoulos and Koliska (2017) in the online journalism context. The use of the concept first spread to adjacent areas, such as social media (Rader et al., 2018), online review platforms (Eslami et al., 2018), and online advertising (Eslami et al., 2018). More recently, AT has also been applied in different contexts, such as energy systems (Cech, 2020) and demand planning (Lehmann et al., 2020). Diakopoulos and Koliska (2017) defined AT as the

> ***disclosure*** *of information about algorithms to enable monitoring, checking, criticism, or intervention by interested parties*. (p. 811, emphasis added)

This meaning is also reflected in other recent literature on the relationship between transparency and technology, which emphasizes the importance of “disclosure devices” (Hansen & Flyverbom, 2015) and highlights the need for “accessible” and “explainable” information (Giest & Grimmelikhuijsen, 2020). Also, some authors see a formal/legal foundation for AT in the European Union’s (EU) general data protection regulation (GDPR) that affords EU citizens a “right to explanation” on the functioning of algorithm-enabled technologies (Selbst & Powles, 2017). Similarly, in a comprehensive review of the transparency concept in general, Bernstein (2017) describes disclosure as one prominent meaning of transparency and defines it as the provisioning of new or previously unknown information. This information can have material effects on, and/or consequences for, both *information recipients* (e.g., increased market efficiency or improved relationships) and *information disclosers*.

By distinguishing the role of information disclosers and recipients, Bernstein (2017) notes an important aspect of (algorithmic) transparency: There are always (at least) two parties involved in transparency, one actively creating transparency and one perceiving transparency. For either of the two parties, the antecedents and consequences of transparency may be different. In a slight variation of the famous proverb, one might summarize that “(the beauty of) transparency is in the eye of the beholder.” AT can therefore be separated into the two components, *AT as action* and *AT as perception*. **AT as action**, on the one hand, refers to the act of provisioning information about an algorithm to its users or interested third parties (e.g., through publishing a manual on an algorithm). The subjects of this act are the developers of an algorithm and/or the parties deploying an algorithm toward the user base (hereafter referred to as “the disclosing party”). **AT as perception**, on the other hand, relates to the information received about an algorithm and how this information is observed by users or interested third parties (hereafter referred to as “the recipient”). Depending on the specific context or recipient, the information received may be perceived as largely identical with or materially distinct from the information provisioned. For example, a recipient unacquainted with terms like “training data” or “machine learning” may perceive only a small fraction of the information provided on an algorithm-enabled AI model, whereas an expert algorithm developer may perceive it as intended by the disclosing party. Similar notions are framed by past commentators (e.g., Zhao et al., 2019a), who distinguish between “objective” and “subjective” transparency. Rader et al. (2018) explain that:

> *Algorithmic transparency often refers to the **act** of making a system knowable or visible. This conceptualizes transparency as a mechanism or process […]. However, transparency is also sometimes treated as a **state** that is the outcome of a process [...]. It can be difficult to disambiguate in the literature whether 'transparency' refers to the mechanism or the outcome, the cause or the effect*. (p. 2-3, emphasis added)

Although AT as action and AT as perception are closely related through an antecedent-consequence relationship, it is through distinguishing between the two that additional consequences and antecedents of AT can be made visible. For instance, past literature points to the (un)intended consequences of AT as action, such as loss of intellectual property (Diakopoulos & Koliska, 2017; Hosanagar & Jair, 2018), which arguably have no implications for the way AT is perceived. Also, AT as perception is not only influenced by the information made available, but also by recipient characteristics, such as prior experience, profession, and tenure (Jarrahi & Sutherland, 2019). Therefore, notwithstanding the difficulty, we argue that a distinction between AT as action and AT as perception is critical for an accurate understanding of AT and that research can benefit significantly from it. Accordingly, in this study, we explicitly distinguish between *AT as action* and *AT as perception* and posit that their antecedents and consequences can be expected to be distinct as well.

Another aspect of AT concerns the actual object of transparency: the algorithm. Algorithms can be defined as “encoded procedures for *transforming input* data into a desired *output* based on specified calculations” (Gillespie, 2014, p. 167, emphasis added). This definition highlights that algorithms are procedural in nature and follow a sequence of input, transformation, and output. Accordingly, AT can be provided (*AT as action*) and perceived (*AT as perception*) with regard to the input, transformation, and output of the underlying algorithm. This process aspect is reflected in a statement on algorithmic transparency and accountability published by the Public Policy Council of the U.S. Association for Computing Machinery (USACM) (ACM US Public Policy Council, 2017): The seven principles for algorithmic transparency and accountability included in the statement cover the entire algorithmic process from the input data to the validation and testing of outputs.

2.2 Algorithmic Transparency and Related Concepts

Studies on the (lack of) transparency of algorithms have accompanied the research on (big) data-driven decision making from the outset. In this context, researchers have also been referring to AT-related concepts such as **explainable AI (xAI)** (Asatiani et al., 2020; Preece, 2018; Rosenfeld & Richardson, 2019; Rai, 2020) and **algorithmic accountability** (Caplan et al., 2018; Kroll et al., 2017) to problematize the (lack of) information available on the inner workings of algorithms. In the following, these concepts are discussed and compared to AT to delineate the three from one another (see Table 1).

Not surprisingly, research on the broader notion of AT has already existed prior to the concept’s ‘formal introduction’ in 2017. Examples include conceptual studies discussing algorithmic opacity (Burrell, 2016; Stohl et al., 2016) and the associated black box problem (Castelvecchi, 2016; Pasquale, 2015). In these studies, AT is frequently treated as one of the multiple aspects of transparency, which underscores the *conceptual origin* of AT in transparency. This origin is also manifested by researchers studying AT but using terms such as “transparency” in general (Cramer et al., 2008), “information transparency” (Awad & Krishnan, 2006), or “intelligibility” (Lim & Dey, 2011). Further, in the case of AT, “algorithmic” is the attribute defining the transparency object; as such, the *scope of the underlying technology* (that is made transparent) is narrowly focused on the algorithm. The majority of the literature discusses AT as a potential solution to address the perils of a broad range of unchecked algorithm-enabled applications (Weller, 2019). Issues related to AT range from risks to democracy and freedom of the press through filter bubbles (O'Neil, 2016) to discrimination through lack of due process (Citron, 2007; Crawford & Schultz, 2014) and unfair, inefficient governmental services (Giest & Grimmelikhuijsen, 2020). Here, most researchers take a positive view of AT; although there are exceptions, where AT is considered less helpful (e.g., Ananny & Crawford, 2018). In line with the definition of AT (see section 2.1), the *scope of information provisioning* focuses on the disclosure of information only (contrary to other concepts that also aim for understanding). Meanwhile, the *process steps covered* are comprehensive and encompass the algorithmic input, transformation, and output. Finally, a common characteristic shared by studies on AT is the inherent or explicit assumption that the *information provided* on algorithms is usually known to the algorithm authors, owners, or deployers. This is an important distinguishing characteristic compared to xAI.

Alongside AT, the rising prominence of AI has drawn (renewed) attention to **xAI.** This concept dates back to the early days of AI in the 1970s and 1980s and, at least initially, has been primarily used in a technical sense (Preece, 2018). However, the proliferation of AI-enabled technologies into people's everyday lives and into critical fields, such as social media and criminal profiling, led to a significant rise in importance. As such, xAI was no longer considered a primarily technical concept and spread into a variety of fields (Gerlings et al., 2021; Goebel et al., 2018). While both xAI and AT share the objective of "open[ing] the black box" (Castelvecchi, 2016), they are conceptually distinct. Most importantly, and driven by its origins, xAI broadly covers any aspects of advanced AI, from algorithms to data, and is used to describe ways of making these aspects accessible and understandable (Rai, 2020; Vilone & Longo, 2021). Also, as indicated above, the xAI concept continues to show a strong focus on the technical characteristics of advanced algorithms used in present-day AI, which make it challenging, or even impossible, to explain 'what is going on' – even for experts (Burrell, 2016; Guidotti et al., 2019). Compared to AT, the *scope of* relevant applications is narrow, since xAI tends to consider technologies enabled by advanced AI (Pedreschi et al., 2019), whereas AT is often studied in the context of technologies that rely on comparatively simple algorithms (e.g., recommendation agents) (Zhao et al., 2019b). Further, xAI arguably shows an inherent focus on the output 'stage' of algorithms, as it is in this stage that results become visible to the user. In contrast, AT equally considers all three stages (input, transformation, and output).

Table 1. Conceptual Differentiation of AT, xAI, and Algorithmic Accountability

Dimension	Algorithmic transparency (AT)	Explainable AI (xAI)	Algorithmic accountability
Problem addressed	*Black box problem* (in terms of lack of clarity on the functioning of algorithms) (Goad & Gal, 2018; Pasquale, 2015).	*Black box problem* (in terms of lack of clarity on the reasoning of AI) (Rai, 2020).	Loss of legitimacy/lack of public control of algorithm-enabled technology due to *black box problem* (e.g., Caplan et al., 2018).
Conceptual origin	Transparency: 'algorithmic' refers to the transparency object (Diakopoulos & Koliska, 2017; Goad & Gal, 2018).	AI: 'explainable' refers to one (of multiple) AI technology attributes, such as accuracy or resource-need (Burrell, 2016).	Accountability: 'algorithmic' refers to the accountability object (Wieringa, 2020).
Scope of the underlying technology	Narrow: Focus on algorithms (Diakopoulos & Koliska, 2017).	Broad: any aspect related to AI (e.g., algorithms, training data) (Rai, 2020).	Narrow: Focus on algorithms (Desai & Kroll, 2017; Wieringa, 2020).
Scope of supported applications	Broad: Covering all algorithm-enabled applications/"advice-giving systems" (Zhao et al., 2019b), not only advanced ones, e.g., forecasting (Lehmann et al., 2020), or grading (Kizilcec, 2016).	Narrow: Focus on advanced AI-enabled applications (and the algorithms embedded in such applications) (Pedreschi et al., 2019; Rai, 2020).	Focused: Covering mainly algorithm-enabled applications with a critical societal role, e.g., algorithms used for criminal profiling or education, etc. (Shah, 2018; Wieringa, 2020).
Scope of information provisioning	Focused: Disclosure (Diakopoulos & Koliska, 2017).	Comprehensive: Explanation, i.e., generate understanding/ interpretability (Gerlings et al., 2021; Preece, 2018).	Mixed: May extend beyond disclosure, e.g., development of a critical audience (Kemper & Kolkman, 2019).
Covered process step(s)	Comprehensive: Coverage of input, transformation, *and* output (Diakopoulos & Koliska, 2017; Zhao et al., 2019b).	Focused: Emphasis on the output of AI-enabled technology ("*post-hoc* interpretations" (Preece, 2018, p. 68).	Comprehensive: Coverage of critical aspects relating to societal debates (requiring high scrutiny) (Kroll et al., 2017).
Characteristic(s) of information provisioned	Usually known to algorithm developers, and/or owners; influenced by multiple factors (e.g., cost considerations) (H. J. Watson & Nations, 2019).	Either known or unknown due to the unique characteristics of AI-enabled systems (e.g., deep learning) (Rai, 2020).	Often known to algorithm developers, and/or owners; potential differences across recipients (e.g., auditors vs. public) (De-Laat, 2018).

Another AT-related concept, **algorithmic accountability**, also emerged in response to the rise of big data and AI. Many of the early papers on algorithmic accountability discuss the topic alongside AT (e.g., Brill, 2015; De-Laat, 2018), and both concepts share the scope of the technology studied (algorithms) and the fact that information is usually known to the discloser. Still, the two represent distinct concepts since, as Ananny and Crawford (2018) explain, AT will not necessarily lead to algorithmic accountability. A similar position is taken by Diakopoulos and Koliska (2017) who—citing Dörr and Hollnbuchner (2017)—argue that "[t]ransparency is just *one* approach toward the ethics and accountability of algorithms" (p. 811, emphasis added). Even more importantly in our view, the two concepts address innately different problems: AT addresses the black box problem as a whole and studies which information should be disclosed to increase clarity on the functioning of algorithms. Conversely, algorithmic accountability focuses on one specific consequence of the black box problem, namely the loss of democratic legitimacy and seeks to address this problem through "the assignment of responsibility for how an algorithm is created and its impact on society" (Caplan et al., 2018, p. 10). Accordingly, AT studies cover a broad scope of applications, from recommendation agents to preventive policing, while algorithmic accountability studies tend to have a strong focus on technologies with a critical societal role (Kroll et al., 2017). Meanwhile, the scope of information provisioning for algorithmic accountability may extend beyond the mere disclosure of information (AT), for instance towards the development of critical audiences for algorithms (Kemper & Kolkman, 2019).

3 Review Methodology

The overarching objective of our study is to synthesize prior findings and point out future research directions by compiling, combining, and comparing the results of primary studies on AT as action and AT as perception. To reach this objective, we follow a systematic approach based on guidance provided by leading IS scholars (Boell & Cecez-Kecmanovic, 2015a; Boell & Cecez-Kecmanovic, 2015b; Chiasson, 2015; Oates, 2015; Okoli & Schabram, 2010, Schultze, 2015; Paré et al., 2015; R. T. Watson, 2015). That is, we use a comprehensive search strategy, focus on empirical studies, and use a narrative synthesis approach to analyze the findings of relevant studies (as opposed to statistical or critical interpretive methods). To ensure the overall reliability and trustworthiness of the review results, established guidelines on how to conduct literature reviews stress the importance of following a systematic, transparent, and repeatable approach (Cram et al., 2020; Paré et al., 2016; Rowe, 2014). Consistent with this guidance, we now elaborate on the boundaries of our review, as well as the literature search and analysis processes (please refer to Figure 1 for a visual representation of the review methodology).

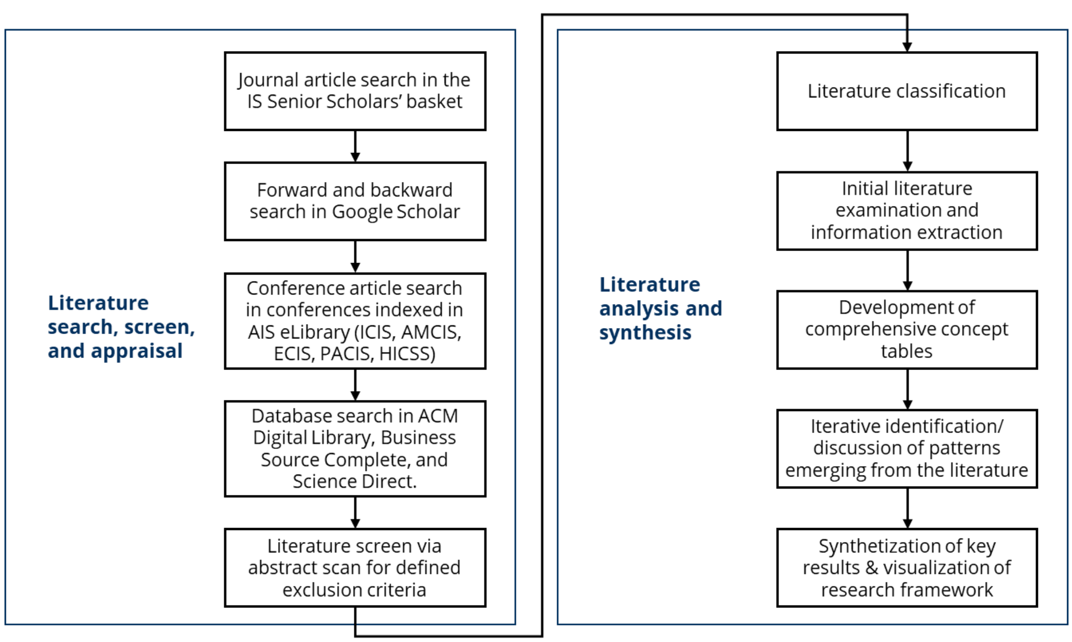

Figure 1. Visualization of Literature Review Activities

3.1 Review Boundaries

Before searching for relevant literature items, we defined the purpose and boundaries of our literature review. To do so, we drew on the AT definition provided in the Conceptual Foundations section. That is, the focus of our review lay on synthesizing studies that combine the topics *algorithm* and *transparency* (or close synonyms). More specifically, to be included in our review sample, a study had to examine an algorithm, or algorithm-enabled technology, as the *object of transparency* (i.e., as the object about which information is disclosed) and foreground the social antecedents and/or effects of AT, rather than the question how to accomplish transparency in purely technical terms. Consistent with this thinking, we did include studies that discussed the transparency of algorithm-enabled technologies, even if these studies did not make explicit reference to the term "algorithmic transparency". One example is Awad and Krishnan's (2006) discussion of information transparency in algorithm-enabled e-commerce personalization. Conversely, we excluded studies from our sample when their focus was on xAI or algorithmic accountability, which we deem distinct concepts that tend to address technical or legal aspects (see section 2.2 above). Dedicated literature reviews already exist for these topics (e.g., Gerlings et al., 2021; Wieringa, 2020). We also excluded studies that do concern transparency (in terms of disclosure), but without a clear link to algorithms. For example, Strich, Mayer, and Fiedler's (2021) study "disclos[ed] mechanisms through which employees strengthen and protect their professional role identity" (p. 304) when interacting with AI. In this study, mechanisms of AI adaptation are made transparent, but not the actual algorithms behind the AI. Moreover, in line with common guidelines (Paré et al., 2015; vom Brocke et al., 2015), we oriented our search toward qualitative and quantitative empirical studies in peer-reviewed journals and conferences but excluded purely conceptual studies. A similar approach has also been used in other literature reviews published in leading IS outlets (e.g., Wiener et al., 2016). Following the guidance of Okoli and Schabram (2010), the review purpose and boundaries, as well as the intended review procedure were documented in a review protocol for reference throughout the process.

3.2 Literature Search, Screening, and Appraisal

Given the relative novelty of the AT concept, we decided to use a sequential search approach, whereby we first applied a keyword search to the IS senior scholars' basket of journals, then extended our search to a broader set of publication outlets using forward and backward searches (vom Brocke et al., 2015; Webster & Watson, 2002), and finally used a conference and database search to ensure coverage of key studies.

For the (initial) keyword search, we anticipated that extant literature would use a range of different terms to refer to AT as action and AT as perception. We thus deliberately employed a comparatively broad search string that included synonyms of both "algorithm" and "transparency." This search string was tested and refined in multiple iterations. In our final string, we combined the terms "algorithm*" or "dat*" (e.g., for big data or datification) with the terms "transparen*" or "disclos*". Based on this search we identified an initial set of 45 potential studies, whose abstracts were scanned for exclusion criteria (see below).

Next, to expand our search, we used *Google Scholar* to conduct backward and forward searches, which resulted in the identification of an additional 26 studies. The high number of relevant studies from outlets outside the canon of top IS journals is indicative of both the interdisciplinarity and relative novelty of the AT concept. In addition, we searched through completed research papers from conferences included in the *Association for Information Systems (AIS) eLibrary*, adding another four studies. Finally, to ensure that we had not missed any important studies we conducted a complementary database search. Here, to ensure consistency and broad coverage, we deployed the same search string ("algorithm*" OR "dat*" AND "transparen*" OR "disclos*) in the publication title and abstract in the *Science Direct* and *EBSCO Business Source Complete* databases, as well as in the *ACM Digital Library.*

Following established practice (Okoli & Schabram, 2010; vom Brocke et al., 2015), we then screened the abstracts of the identified studies and discarded those that met specific exclusion criteria based on our review boundaries (please refer to Table 2 for a summary). Specifically, we excluded: 1) non-empirical studies; 2) publications that referred to our search term "dat*" in unintended ways (such as studies on the disclosure of "personal data", e.g., Miltgen & Peyrat-Guillard, 2014, or studies using the expression "to date", e.g., Durcikova & Gray, 2009); studies where the term "transparency" was 3) used without relation to algorithms (e.g., with respect to optics); or 4) only generically; and 5) studies where the term "disclosure" was not used in relation to an algorithm (such as in open source research, e.g., Shaikh and Vaast (2016), or in research on the disclosure of personal information, e.g., Karwatzki et al. (2017)). In total, we identified a final review sample of 50 studies from 42 distinct publication outlets (see Appendix A).

Table 2. Overview of Exclusion Criteria

Exclusion criterion	Details on typical studies in this category	Example reference(s)
Non-empirical study, i.e., a study that does not present (qualitative or quantitative) empirical data.	Conceptual studies that treat questions of algorithmic transparency in relation to ethical and/or legal questions, including questions on ethical treatment of (input) data.	Allen, Burk, and Davis (2006).
Unintended use of the term “dat*”.	Studies using “dat*” in the combinations “to date”, “empirical data”, “data set”, “experimental data”, “survey data”, “financial data”, “open data”, “panel data”, or “research data”.	Durcikova and Gray (2009), Miltgen and Peyrat-Guillard (2014).
Use of the term “transparency” without relation to algorithms.	Studies of transparency in a variety of contexts, such as open-source communities, or organizations.	Shaikh and Vaast (2016).
Generic use of the term “transparency” in relation to algorithms.	Numerous (often purely technical) studies stating “transparency” as a general objective for algorithms in their introductory statements and/or exploring ways to make particular (types of) algorithms “more transparent”.	Mark and Anya (2019); Shi and Hurdle (2018).
Use of the term “disclosure” without relation to algorithms.	Studies discussing information disclosure in digital services, websites, or online platforms.	Karwatzki et al. (2017)

To obtain a high-quality review sample, we also used a set of formal quality criteria throughout our literature search and screening. First, we applied our search only to completed (i.e., no research-in-progress) studies. Second, we confirmed that all articles had undergone peer review to ensure that their methodological and analytical rigor have been checked. Given the comparative novelty of the topic of AT we opted for this comparatively liberal set of criteria, in line with the methodological guideline that “newer and emerging areas of research might call for more lenient methodological standards for inclusion in a literature review in order to not prematurely exclude work in areas that are not yet well understood” (Okoli & Schabram, 2010, p. 26).

Next, to describe our review sample, we classified all studies along ten dimensions, namely, publication year and outlet, research approach and methodology, level of analysis, scientific discipline, industry and application domain, theoretical underpinnings, and (first) authors’ country of origin. The 50 studies in our final sample were published between 2002 and (June) 2021 in a total of 19 distinct journals and 23 distinct conferences. 86% (43) of the studies were published from 2015 onward. While most studies were quantitative in nature (30), and often based on online experiments (16), there was also a considerable number of qualitative studies (18). Two studies used a mixed method approach. The vast majority of the reviewed studies (88%) utilized an individual level of analysis. Although this finding may be somewhat surprising, given the equally important relevance of AT for organizations and society, it can be partly explained by the relative difficulty of gathering empirical data on AT at these levels. A similar emphasis can be observed for the application domain, where studies on recommendation agents (16) and the related domain of decision support systems (9) account for half of our review sample (see Appendix B for detailed descriptive statistics).

3.3 Literature Analysis and Synthesis

After the literature search (and classification), we began our analysis by thoroughly reading through each study and taking detailed notes of its central arguments (the process referred to as “data extraction” by Okoli & Schabram, 2010). To do so, we used a spreadsheet with separate columns for each analysis dimension, such as a study’s conceptual focus (i.e., AT as action versus AT as perception), AT antecedents and consequences, as well as supporting evidence (Bandara et al., 2015), thereby following a concept-centric logic (Webster & Watson, 2002). On this basis, we expanded our analysis beyond the contents of individual studies and systematically searched for shared patterns and themes across studies. Here, the focus of our analysis was on identifying and describing the nature of central relationships linking key concepts. Throughout this process, an iterative approach was followed, whereby the author team periodically discussed emerging patterns and themes, as well as different organizing options. Team discussions were based on comprehensive concept tables compiled and continuously refined by the first author. We also made use of visual tools, such as graphical representations in Microsoft PowerPoint, to identify key patterns between the identified categories and concepts and to develop an integrative framework. It was at this stage in the analysis process that the critical importance of the conceptual distinction between AT as action and AT as perception became apparent. In total, it took nine iterations of in-depth discussions among the authors before a stable research framework emerged, consisting of eight concepts and seven relationships (see Figure 2 below). We present our review results along these relationships in the next section.

4 Review Results

At the heart of our research framework is the distinction between AT as action and AT as perception. The subsequent presentation of the review results thus begins with the link between these two subconstructs of AT. After this, we follow a process logic and first present the antecedents of AT as action and perception and then the consequences of AT as perception. As noted above, our framework encompasses a total of seven key relationships: (1) AT as action versus AT as perception; (2) factors triggering AT as action; (3) factors shaping AT as action; (4) factors shaping AT as perception; as well as AT as perception leading to (5) rational-cognitive responses; (6) affective-emotional responses; and (7) (un-)intended behavioral (side) effects. In this regard, we would like to reiterate that our research framework and its constitutive relationships resulted from our concept-centric analyses and review results. Still, using this framework as an integrative organizing device, we present it at the outset of this section (see Figure 2).

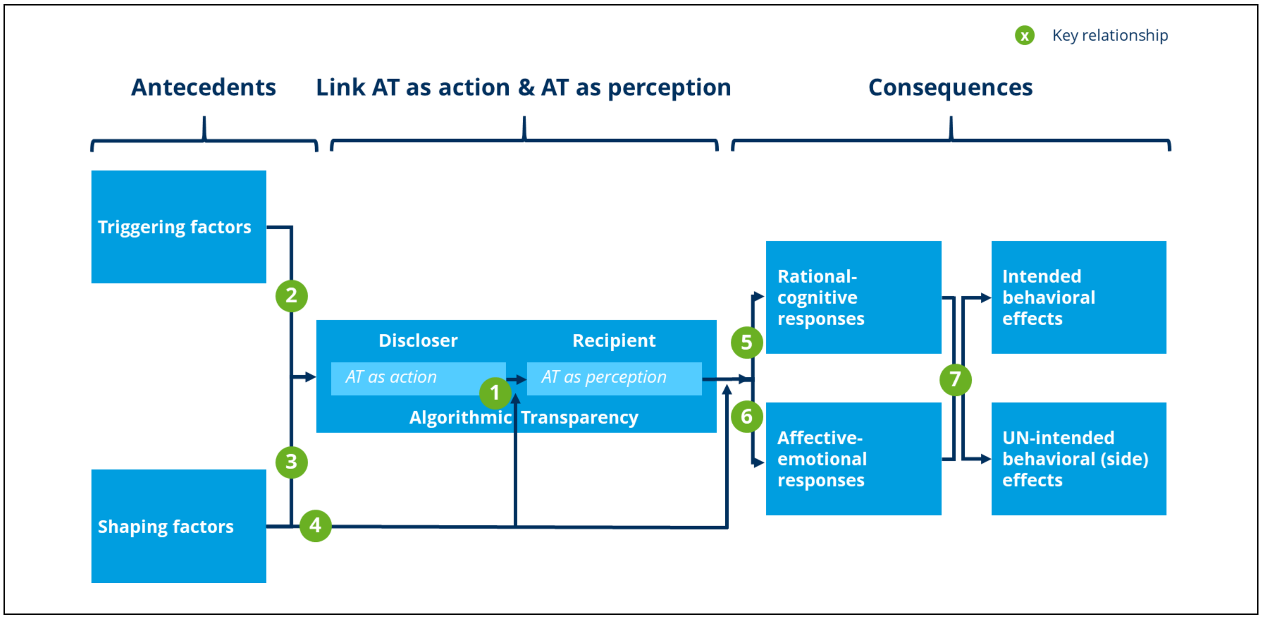

Figure 2. Integrative Research Framework (Derived from Literature Review Results)

4.1 Relationship 1: AT as Action Versus AT as Perception

At least implicitly, numerous studies examine the link between AT as action and AT as perception (i.e., the question of how much and what kind of disclosed information on algorithms is also perceived by recipients). Unsurprisingly, studies offer strong evidence that increasing AT as action does indeed lead to increased AT as perception (Cramer et al., 2008; Lehmann et al., 2020; Rühr, 2020; Salminen et al., 2020; Wang & Benbasat, 2007, 2016), with some studies adding interesting nuances. For example, even though most existing work does not explicitly distinguish between AT as action and perception, some studies examine how different types of information (in terms of AT as action) influence the degree of AT perceived by recipients. These studies find that the effect of AT as action on AT as perception tends to be stronger when *multiple* types of information are disclosed (e.g., data used, rules applied, and confidence scores); in contrast, when *only one* of these types of information is disclosed, the effect tends to be weaker (Cramer et al., 2008; Lim & Dey, 2011; Rühr, 2020; Wang & Benbasat, 2007, 2016). Further, some studies also provide technical guidance on how to increase perceptions of AT (e.g., by developing a rule-based simulator), even if the amount of information disclosed (AT as action) cannot be increased.

This question is of high relevance when the model itself or its training data cannot be disclosed due to intellectual property protection or privacy law (Domingo-Ferrer et al., 2019).

4.2 Antecedents of AT as Action and AT as Perception

Antecedents of AT as action can be divided into *triggering* factors, which prompt disclosing parties to make information on algorithms available (relationship 2), and *shaping* factors, which influence the specific way in which AT as action is performed, what kind of information is made available, and how this information is presented (relationship 3). Further, in terms of antecedents of AT as perception, the reviewed studies also point to factors *shaping* the way in which disclosed information is perceived by recipients (relationship 4).

4.2.1 Relationship 2: Factors Triggering AT as Action

Triggering factors refer to the concrete legal, human/social, or technological conditions that prompt disclosing parties to make information on algorithms available (see Table 3 for an overview). While only one study focuses exclusively on such factors (Hepenstal et al., 2020), a variety of papers discuss triggering factors alongside other questions about AT as action. Legal factors are the most prominent category of triggers. These refer to **(semi-)legal requirements**, such as the European General Data Protection Regulation (GDPR), that demand AT as action from businesses or public entities. These factors are relevant in different national contexts, such as Europe (Ahonen & Erkkilä, 2020; Criado et al., 2020), the United States (Fink, 2018), and South Korea (Kim & Moon, 2021), yet the degree to which regulations are binding differs. Although researchers agree that a legal regulation *can* trigger AT as action, some argue that in practice the influence of regulation is lower than one might expect, as most laws do not specify concrete requirements (Criado et al., 2020). Additional nuances are added through the discussion of other semi-legal requirements of AT as action, such as public audits (Criado et al., 2020) and court rulings (Fink, 2018).

Table 3. Factors Triggering AT as Action

Factor category	Factor	Short description (References)
Legal factors	(Semi-)legal requirements	Laws and regulations of (supra-)national bodies, commission proposals, public authority statements/measures, court rulings, and external audits that spell out more or less explicit and binding requirements for AT as action (Ahonen & Erkkilä, 2020; Criado et al., 2020; Domingo-Ferrer et al., 2019; Fink, 2018; Kim & Moon, 2021; H. J. Watson & Nations, 2019).
Human/social factors	Public pressure	Public criticism due to suspected or confirmed discrimination when algorithms use protected class attributes, such as gender or ethnicity (Asatiani et al., 2020; Kim & Moon, 2021) or appear to be “creepy” (H. J. Watson & Nations, 2019).
	Risk reduction for disclosing party	Businesses’ intent to reduce safety, reputational, or commercial risks through early detection of undesirable algorithmic behavior (Asatiani et al., 2020; Hepenstal et al., 2020) and incorporation of user feedback on improvement opportunities (Domingo-Ferrer et al., 2019).
	(Desire for) trust building	Disclosing parties’ intent to provide information in order to build trust among algorithm users and to stimulate engagement (e.g., in digital journalism); however, mixed results regarding the use of AT as action as trust-building measure (Diakopoulos, Cass, & Romero, 2014; Diakopoulos & Koliska, 2017; H. J. Watson & Nations, 2019).
	Desire for control	Use of AT as action as a means to reduce recipients’ desire for control over a given algorithm (e.g., under certain conditions, investment algorithms may perform best without any user interference, leading to a performance-control dilemma) (Rühr, 2020).
Technology factors	Incorrect input data	Users or affected parties providing potentially false input data for an algorithm and then requesting additional information that enables them to validate and correct causes of implausible results (Bunt et al., 2012; Cech, 2020).
	Questionable algorithmic output	Recipients questioning algorithmic output, due to (perceived) inaccuracy or feelings of exploitation, and requesting explanations and validation from disclosing parties (Asatiani et al., 2020; Bunt et al., 2012; Cech, 2020).

Often going hand-in-hand with legal requirements are human/social factors, as another important category of triggers. An often touched-upon aspect in this context is **public pressure** due to perceived discrimination (Asatiani et al., 2020; Kim & Moon, 2021). A closely related aspect is a high degree of user mistrust resulting from perceived “creepiness” of an algorithm, which can lead to material public/user pressure for AT as action (H. J. Watson & Nations, 2019). On the other hand, to avoid additional regulations and/or public concerns, disclosers themselves may decide on AT as action to **reduce discloser risks**. For example, studies find that disclosers intend to detect suspected harms, such as discrimination, proactively, as they wish to minimize the risk of hazards for human stakeholders (Asatiani et al., 2020). Also, disclosers may intend to gather recipients’ suggestions for improvements (Domingo-Ferrer et al., 2019) by disclosing information on their algorithms. In addition, multiple studies discuss a range of business rationales that may trigger AT as action. In this regard, there is strong consensus that AT as action can be triggered by a **desire to build user trust**, which can in turn have desirable secondary effects, such as increased engagement, higher usage rates, or greater willingness to share (more) personal data (Diakopoulos et al., 2014; Diakopoulos & Koliska, 2017; H. J. Watson & Nations, 2019). An interesting additional finding comes from one study (Rühr, 2020), which finds a significant negative effect of AT as action on users’ **desire for control** and suggests that AT as action may be triggered by disclosers being reluctant to give control to users.

Finally, some studies also find that AT as action may be triggered by technology factors. On the one hand, technology motives may play a role in the case of **incorrect input data,** when there is a need to troubleshoot errors and to identify what data inputs provided by users are leading to unexpected results (Bunt et al., 2012; Cech, 2020). On the other hand, when the technology leads to **questionable algorithmic output** that may appear inaccurate or counter-intuitive**,** disclosers may want to leverage AT as action to address user inquiries (Asatiani et al., 2020; Bunt et al., 2012; Cech, 2020).

4.2.2 Relationship 3: Factors Shaping AT as Action

Shaping factors are conditions influencing the specific way in which AT as action is performed, such as the type, quantity and quality of information disclosed (Table 4). Relevant studies identified in the review sample cover a range of different factors relating to the same categories of legal, human/social, and technology factors identified for factors triggering AT as action.

First, with regard to the legal factors category, both **legal and political requirements** are not only found to trigger AT as action but also to shape it. Related studies note that the provision of clear regulations on AT as action is still in its infancy (meaning there is very little explicit guidance on what information needs to be disclosed and how). This can create inconsistencies even under the same legal framework (Fink, 2018). For example, different U.S. government agencies appear to disclose highly variable amounts of information in response to freedom of information requests on algorithms, ranging from full source code disclosure to no disclosure at all. Accordingly, multiple studies conclude that AT as action will be shaped by the emergence of more explicit and specific legal action that aims to resolve current ambiguities and ensures that the principal right to AT as action is attained in practice (Ahonen & Erkkilä, 2020; Kim & Moon, 2021). Some authors also suggest that even the prospect of more regulation will prompt disclosing parties to take preemptive action and change their approach to AT as action (H. J. Watson & Nations, 2019).

In terms of human/social factors, studies point to the importance of the **cultural background** of receiving and disclosing parties. That is, the degree to which transparency is an established value in the culture of recipients (Ahonen & Erkkilä, 2020), or disclosing parties (Khovanskaya et al., 2016), shapes the information disclosed. Another key factor is **algorithmic literacy:** when recipients are system developers or other expert users who are knowledgeable about algorithms (i.e., have *high* algorithmic literacy), the disclosing party tends to provide *more* (varied) information. On the other hand, when recipients are lay users without deeper knowledge about algorithms (i.e., with *low* algorithmic literacy), they may easily become overwhelmed by detailed information. Thus, disclosing parties are found to provide *less* information (Cech, 2020). In addition, two studies (Rader et al., 2018; Springer & Whittaker, 2020) suggest that recipients’ **disposition to trust** an algorithmic technology – and related individual traits like the need for control or cognition – can influence AT as action. Recipients with a high disposition to trust tend to request less information, and may respond *negatively* to greater levels of AT as action (e.g., lose confidence in the technology). In line with this logic, a discloser may prefer *lower* levels of AT as action for such recipients (i.e., with a high disposition to trust) to not reduce their innate trust.

Furthermore, for commercially active disclosers, **business considerations** are found to play an important role. For instance, it is surprisingly unclear whether the benefits of AT as action truly outweigh the costs. In this regard, benefits are found to be rather abstract while costs appear to be very concrete (Diakopoulos & Koliska, 2017). The unfavorable business case for AT as action is found to be compounded by **intellectual property concerns**, such as whether AT as action could cause a loss of proprietary knowledge on algorithms, which are often treated as trade secrets (Brauneis & Goodman, 2018; Diakopoulos & Koliska, 2017; Jhaver et al., 2018). This concern is of even greater importance in the public sphere, where algorithms deployed for activities of public interest, such as preventive policing, were developed by private companies. In these situations, AT as action faces a dilemma. On the one hand, protecting intellectual property; on the other hand, responding to public pressure and broadly held **ethical beliefs** of a right to transparency on governmental decisions generally (Diakopoulos & Koliska, 2017; H. J. Watson & Nations, 2019), and in situations of (suspected or actual) discrimination more specifically (Asatiani et al., 2020). One paper even argues that such ethical considerations can lead to “a culture of disclosure” (Brauneis & Goodman, 2018, p. 137), whose influence on AT as action overrides that of other factors. The dilemma of disclosing parties is further complicated by **expected system gaming**. That is, disclosers are concerned that AT as action may give away the logic, criteria, or thresholds used in algorithmic decision-making and thus allow a deliberate circumvention of these decisions. While the general risk of system gaming is described in multiple studies (Diakopoulos & Koliska, 2017; Jhaver et al., 2018), there is little actual evidence of the extent to which this is an issue and under which conditions this issue is of (particular) relevance. Accordingly, one paper also argues that concerns about IP protection and system gaming are often used as a (sometimes questionable) blanket argument for low AT as action (Brauneis & Goodman, 2018).

Table 4. Factors Shaping AT as Action

Factor category	Factor	Short description (References)
Legal factors	Legal/political requirements	Concrete guidelines regarding the amount and type of information to be disclosed about an algorithm in some jurisdictions (Ahonen & Erkkilä, 2020; Asatiani et al., 2020; Brauneis & Goodman, 2018; Fink, 2018; Kim & Moon, 2021; H. J. Watson & Nations, 2019).
Human/social factors	Cultural background	Degree to which open/public access to information is an established principle or value in recipients’ cultures (Ahonen & Erkkilä, 2020) and/or disclosing parties’ cultures (Khovanskaya et al., 2016).
	Algorithmic literacy	Recipients’ knowledge of how algorithms work and how algorithms impact the function of a given system they use (Cech, 2020).
	Disposition to trust	Users’ innate willingness to accept technology and related information provided by a third party without challenging and questioning this technology/information (Rader et al., 2018; Springer & Whittaker, 2020).
	Business considerations	Perceived/calculated tradeoff between costs (e.g., increased effort) and benefits (e.g., user satisfaction and trust) of disclosing information on an algorithm (Diakopoulos & Koliska, 2017).
	Intellectual property concerns	Degree to which knowledge embedded in a given algorithm is ‘unique’ (e.g., contains corporate secrets) and thus warrants particular protection (Brauneis & Goodman, 2018; Diakopoulos & Koliska, 2017; Jhaver et al., 2018).
	Ethical considerations	Basic value premises and culturally instilled practices, especially in democratic societies, in relation to open government (processes), minority protection, and the right to appeal public decisions in court-rulings, etc. (Asatiani et al., 2020; Diakopoulos & Koliska, 2017 ;H. J. Watson & Nations, 2019).
	Expected system gaming	Disclosing parties’ concern that recipients will engage in evasion (e.g., keeping cash payments beyond an algorithm’s threshold to avoid tax investigations) when thresholds and decision criteria are disclosed (Diakopoulos & Koliska, 2017; Jhaver et al., 2018).
Technology factors	Technology intrusiveness	Degree to which a specific technology employs sensitive user data (e.g., personal data, facial recognition) (H. J. Watson & Nations, 2019).
	Application characteristics	Specific functionalities provided by a given algorithm-enabled

		technology, as well as the accuracy and complexity of these functionalities (Bunt et al., 2012; Cech, 2020; Khovanskaya et al., 2016; Shin & Park, 2019).
	Documentation	Amount and type of information on an algorithm made available by developers (Brauneis & Goodman, 2018).

In terms of technology-related factors shaping AT as action, H. J. Watson and Nations (2019) suggest that users judge algorithm-enabled technologies on a "creepiness scale" depending on perceived **technology intrusiveness** and potential harm. Companies can address this "creepiness" by shaping AT as action accordingly: "With care, companies can benefit from and avoid the penalties associated with using personal data and algorithms inappropriately" (p. 491). This perspective is expanded by other studies (Cech, 2020; Khovanskaya et al., 2016), which find that disclosers need to adjust AT as action depending on the **application characteristics** and functionalities offered by an algorithm-enabled application (e.g., data collection, data aggregation, or visualization). As well, extant literature suggests that different levels of AT as action are required for different recipient-technology combinations. For instance, AT as action towards lay users, such as administrative staff entering data, needs to differ from AT as action towards experts who interpret results. Similarly, Bunt et al. (2012) find that low-complexity applications require significantly less AT as action than more complex systems and Shin and Park (2019) find that recipients interacting with more algorithmic features rate higher in AT. Another interesting observation is made by Brauneis and Goodman (2018) who point out that many public entities (e.g., courts, councils), do not have enough **documentation** on the privately-developed algorithms they are using and are thus unable to disclose detailed information.

The large variety of shaping factors that have been identified by extant research indicates a relative focus of prior studies on this aspect. Yet, the findings also hint at a noteworthy inconsistency as there appear to be factors that can lead to more (e.g., legal requirements) and less (e.g., IP concerns) AT as action and it is unclear which of them prevail in the case of a tradeoff decision.

4.2.3 Relationship 4: Factors Shaping AT as Perception

Beyond factors triggering and shaping AT as action, studies also point to a set of recipient characteristics shaping their AT perceptions (refer to Table 5 for a summary). First, recipients' **prior experience** with an algorithm (i.e., when they can try it out and see the results), is found to play an important role. When an algorithm is newly released, recipients have not yet interacted (much) with it and the influence of experience on AT as perception is low. However, when an algorithm has been in existence for some time, recipients have been able to try it out and the experience they have thereby gained has a high influence on AT as perception. This influence can attenuate or even break the link between AT as action and AT as perception. In other words, recipients are no longer influenced by the information that is disclosed on an algorithm, but by their own experience from interacting with the algorithm (Hardin et al., 2017; Jarrahi & Sutherland, 2019). This effect may lead to recipients spending less time viewing the disclosed information (Springer & Whittaker, 2020), or even to recipients defending low levels of AT as action, as their engagement with the algorithm leads them to *perceive* high AT irrespective of the actual amount of information disclosed (Eslami et al., 2019). Relatedly, one study on algorithmic management finds that recipients' **specialization and expertise** can influence their perception of AT: When recipients are highly specialized or high-tenure experts, they tend to care less about the information disclosed on the algorithm managing them and AT as perception remains lower (Jarrahi & Sutherland, 2019).

Table 5. Factors Shaping AT as Perception

Factor	Factor description (References)
Prior experience	Degree to which users have had the chance to work with and try out an algorithm-enabled technology, thereby gaining transparency through "sensemaking" (Jarrahi & Sutherland, 2019, p. 583), through "repeated interactions with the system (Hardin et al., 2017, p. 1159; Springer & Whittaker, 2020), or through "engagement" (Eslami et al., 2019, p. 10).
Specialization and expertise	Degree of specialization of recipients (niche areas with limited labor supply vs. commodity work) and length of time that recipients have already worked in their job (Jarrahi & Sutherland, 2019), and the resulting demand for recipients' skills.

4.3 Consequences of AT as Perception

Numerous studies describe the consequences of AT as perception.[2] Corresponding consequences can be divided into rational-cognitive responses (relationship 5) and affective-emotional responses (relationship 6) to AT as perception, as well as (un)intended behavioral (side) effects of AT as perception (relationship 7).

4.3.1 Relationship 5: Rational-Cognitive Responses to AT as Perception

Referring to recipients' conscious mental processes, rational-cognitive responses describe one category of direct consequences of AT as perception (see Table 6). Arguably, the most intuitive response is the recipients' **awareness** of an algorithm's existence, which extant research finds to be positively influenced by AT as perception (Rader et al., 2018). As well, AT as perception affects recipients' perception of algorithm **correctness**. For instance, one study finds that the more AT is perceived by recipients, the less confident they become that an algorithm is correct and start questioning its outputs (Rader et al., 2018). This finding is corroborated in two studies by Springer and Whittaker (2019; 2020), who find that recipients perceive *lower* algorithm accuracy with increasing AT perceptions. In contrast, (only) when recipients' expectations have previously been violated, perceived accuracy actually increases with greater AT perceptions.

Less surprisingly, AT as perception has been found to increase understanding. Interestingly, only one study measures recipients' **actual understanding** through control questions on the content provided and finds a positive effect of AT as perception (Cramer et al., 2008). The remaining studies only measure recipients' **perceived understanding** (also referred to as "subjective", "self-reported") and also consistently find positive effects (Lehmann et al., 2020; Lu et al., 2020; Springer & Whittaker, 2020; Yeomans et al., 2019). Relatedly, there is also empirical support for AT as perception having a positive effect on algorithm **interpretability**. That is, a recipient's intuitive understanding of the inputs an algorithm uses and of the reasons why it works in a specific way (Domingo-Ferrer et al., 2019; Lu et al., 2020; Rader et al., 2018). This kind of understanding helps recipients to calibrate the recommendations offered by an algorithm and thereby reduces their perceptual bias (Schaffer et al., 2015). Moreover, one study also finds that AT as perception reduces recipients' **perceived cognitive effort**, making it *feel* easier to follow the inner workings of an algorithm (Wang & Benbasat, 2016).

Furthermore, some studies find effects of AT as perception on the **perceived advice quality/value** of a given algorithm. However, evidence on the direction of these effects is mixed. While one study suggests a positive effect on perceived advice quality (Wang & Benbasat, 2016), another one finds a negative effect on the perceived value of algorithmic advice (Lehmann et al., 2020), and one study shows somewhat mixed results (Eslami et al., 2018). A fourth study finds that recipients can take both positive and negative stances towards an algorithm, depending on their engagement and personal gain from it (Eslami et al., 2019).

Echoing the theoretical debate on the link between AT and algorithmic accountability, **perceived fairness** is another prominent rational-cognitive response to AT as perception. There are a variety of constructs used to measure fairness and the empirical findings on the effect of AT as perception are far from conclusive. One study finds a negative effect of AT as perception on perceived fairness (Rader et al., 2018). However, most of the relevant studies in our review sample point to positive effects. For example, one study finds an increase in fairness when more AT is provided through different types of explanations (Lu et al., 2020). The same positive pattern can be observed in studies on discrimination perceived by recipients, which is central to perceived fairness (Domingo-Ferrer et al., 2019). Finally, one study (Lee et al., 2019) provides a more nuanced view and indicates that the effect of AT on fairness is influenced by recipients' ability to **control the outcome** of an algorithm's decisions, whereby only those recipients who are able to influence the algorithmic decision also show a positive effect of AT on their fairness perception. Also, when recipients' prior expectations of fairness (e.g., equality of distribution) have not been met by the information disclosed on the algorithm, this **expectation violation** leads to *lower* perceived fairness.

Table 6. Rational-Cognitive Responses to AT as Perception

Factor	Short description (References)

[2] While some of these studies examine *AT as action*, we still decided to include them here. This is because any AT-related actions will first lead to *AT as perception* (i.e., be perceived by the recipients of such actions) and only then give rise to consequences.

Awareness	Degree to which recipients are cognizant of an algorithm's existence and of the fact that a given tool's output is the result of an algorithm, not a human (system vs. user agency) (Rader et al., 2018; Salminen et al., 2020).
Perceived correctness	Degree to which recipients consider a system's outputs to be accurate and free of errors (Rader et al., 2018; Springer & Whittaker, 2019).
Actual understanding	Degree to which recipients truly comprehend the inner workings of an algorithm-enabled system (Cramer et al., 2008).
Perceived understanding	Degree to which recipients think that they comprehend the inner workings of an algorithm-enabled system (Cramer et al., 2008; Diakopoulos et al., 2014; Lehmann et al., 2020; Lu et al., 2020; Springer & Whittaker, 2020; Yeomans et al., 2019).
Interpretability	Degree to which recipients are able to intuitively follow the way an algorithm-enabled system works and can make sense of the information used, reasons and rules behind it, and type of results produced (Domingo-Ferrer et al., 2019; Lu et al., 2020; Rader et al., 2018; Salminen et al., 2020; Schaffer et al., 2015).
Perceived cognitive effort	Degree to which recipients perceive that they need to expend time and mental processing efforts when interacting with an algorithm (Wang & Benbasat, 2016).
Perceived advice quality/value	Degree to which recipients consider algorithmic recommendations to be valuable and matching their needs and preferences in a given decision context (Cramer et al., 2008; Eslami et al., 2018; Lehmann et al., 2020; Rader et al., 2018; Wang & Benbasat, 2016).
Perceived fairness	Degree to which recipients consider an algorithm-enabled system to be equal towards everyone, without bias, and free of discrimination (Domingo-Ferrer et al., 2019; Lee et al., 2019; Lu et al., 2020; Rader et al., 2018).
Perceived control	Degree to which recipients believe that their actions can affect the outcome of an algorithm-enabled system (Jarrahi & Sutherland, 2019; Rader et al., 2018).
Perceived social presence	Degree to which recipients experience an artificial, algorithm-enabled system as intelligent and human-like (B. Liu & Wei, 2021).
System acceptance	Degree to which recipients accept that an algorithm-enabled technology collects data on their behavior and derives personalized online user offerings (services or advertising) from these data (Awad & Krishnan, 2006), or intend to purchase items recommended by a given algorithm (B. Liu & Wei, 2021).

Similarly, evidence is mixed on the effect of AT as perception on **perceived control**. For example, while one study (in the context of recommendation agents) finds that recipients feel they have *less* control when AT as perception increases (Rader et al., 2018), another study (in the context of digital labor platforms) finds that recipients sense *more* control under such conditions (Jarrahi & Sutherland, 2019). Furthermore, one study finds that greater AT as perception also leads to a decrease in **perceived social presence.** This means that recipients experience an algorithm-enabled technology as less intelligent and human-like (B. Liu & Wei, 2021). Finally, greater AT as perception can also affect recipients' attitude toward, and in particular their **acceptance** of, an algorithm-based system. However, the direction of this effect remains unclear. On the one hand, recipients perceiving greater AT tend to show greater acceptance of technologies that use algorithms for providing personalized services or personalized advertising (Awad & Krishnan, 2006). On the other hand, in the case of online behavioral targeting, greater AT as perception appears to lead to reduced purchase intention for recipients with low anthropomorphism tendency through lower social presence (B. Liu & Wei, 2021). To summarize, the present state of research offers strong evidence for a positive effect of AT as perception on some rational-cognitive responses (e.g., awareness). Meanwhile, for others, the evidence presented is either still very limited (as in the case of perceived control and social presence), or even contradictory and fraught with a number of inconsistencies (as in the case of perceived advice value and perceived fairness).

An additional nuance is added through a series of factors indirectly shaping (or "moderating"[3]) the rational-cognitive responses to AT as perception (see Table 7 below). First, AT as perception is found to lower

[3] Although the shaping factors described here may be considered *moderators* in the broader sense, we limit our use of this term. Given that some of the reported studies are qualitative, not all of the factors described here have been found to be "moderating variables" through the respective statistical tests. Accordingly, we use the term "factors shaping the responses" to describe these relationships in general terms. Nevertheless, future quantitative researchers may find the list of these factors a helpful starting point for statistical tests of moderating relationships (also see research opportunity C below). Also, there are likely other moderators influencing the mixed outcomes described here, which have, however, not yet been identified by extant research and require further study.

perceived social presence only for recipients with low **anthropomorphism tendency** (B. Liu & Wei, 2021). Second, where an algorithm has lower **certainty**, the effect of increased AT as perception on the user experience is negative, whereas for higher certainty it is positive (Lim & Dey, 2011). Third, when **recipients benefit** from the decisions of an algorithm-enabled system (e.g., in credit approval), this appears to increase the positive effect of AT as perception on perceived fairness (Lu et al., 2020) and on recipients' stance towards the algorithm (Eslami et al., 2019). However, when recipients cannot **control the outcome** of a decision or when their **expectations are violated**, the effect of AT as perception on perceived fairness can also be negative (Lee et al., 2019). In addition, for algorithm-enabled advertising, recipients' **general liking of advertising** influences the effect of AT as perception on the perceived value of ads. For recipients with either low or high liking, the influence of AT as perception on perceived value is inexistent. These recipients tend to always ignore (low liking) or always appreciate (high liking) ads, irrespective of the perceived AT. Meanwhile, for recipients with moderate levels of general liking, the effect of AT as perception on perceived value follows the aforementioned positive direction (Eslami et al., 2018). Also, when users have greater **engagement** with an algorithm, the effects of low AT as perception on their appreciation of an algorithm are found to be attenuated (Eslami et al., 2019). Lastly, the positive effect of AT as perception on recipients' willingness to be profiled can be weakened by greater **general privacy concerns** (Awad & Krishnan, 2006).

Table 7. "Moderating" Factors Shaping the Rational-Cognitive Responses to AT as Perception

Factor	Short description (References)
Anthropomorphism tendency	Degree to which recipients attribute human capacities/characteristics to a given technology (B. Liu & Wei, 2021).
Certainty	Degree to which a given algorithm-enabled technology shows confidence that its outputs are correct (manifested e.g., through high confidence scores) (Lim & Dey, 2011).
Recipient benefit	Extent to which recipients are positively affected by the decisions of an algorithm-enabled system (Eslami et al., 2019; Lu et al., 2020).
Outcome control	Recipients' ability to influence/change the algorithm's results (Lee et al., 2019).
Expectation violation	Degree to which recipients have anticipated a given result provided by an algorithm-enabled system (Lee et al., 2019; Springer & Whittaker, 2019, 2020).
General liking of advertising	Degree to which recipients generally appreciate the information provided through advertising (Eslami et al., 2018).
Engagement	Degree to which recipients actively interact with an algorithm-enabled system; e.g., by making use of the system features and outputs (Eslami et al., 2019).
General privacy concern	Degree to which recipients generally fear that their privacy is at risk when interacting with algorithm-enabled systems or related digital technologies (Awad & Krishnan, 2006).

4.3.2 Relationship 6: Affective-Emotional Responses to AT as Perception

Pertaining to recipients' emotions and feelings, affective-emotional responses describe a second category of direct consequences of AT as perception (see Table 8 for an overview). Most studies on affective-emotional responses focus on **trust**. The majority of these studies generally point to a positive relationship between AT as perception and trust (Ananthakrishnan et al., 2020; de Oliveira Cesar de Moraes et al., 2019; Diakopoulos & Koliska, 2017; Hofeditz et al., 2021; Lai & Tan, 2019; Ochmann et al., 2021; Shin, 2020b; Shin et al., 2020). Two studies by Wang and Benbasat provide a more nuanced perspective. The first one (2007) shows that different sub-types of AT as perception (i.e., explanations on how vs. why) link to different types of trust (competence, benevolence, and integrity). The second one (2016) re-confirms the high influence of AT as perception by showing that it affects all three types of trust, whereas perceived cognitive effort and advice quality only affect recipients' competence beliefs. Somewhat contrary to these findings, two studies find a negative effect of AT as perception on the trustworthiness and credibility of algorithmic output (Salminen et al., 2020; Seo et al., 2018). In the case of the first study (Salminen et al., 2020), this might be due to the use of algorithms to generate human-like *personas*, which may be at odds with the technical nature of algorithms and cause recipients to feel lower trust. In the case of the second study (Seo et al., 2018), recipients' trust might have been lowered by a lack of understanding of the information disclosed. Another study by Kizilcec (2016) finds a bell-shaped relationship between AT as perception and trust, where trust only increases up to a medium level of AT as perception and then decreases again. The author argues that providing too much information may have confused recipients and thus lead to a decline in trust. Furthermore, the study finds the bell-shaped relationship to only apply for recipients whose expectations have been violated by the results of an algorithm, whereas for others there appears to be no effect. Generally, the above-presented findings in relation to trust are corroborated by studies on **distrust**, which is found to decrease with growing AT as perception (de Oliveira Cesar de Moraes et al., 2019; Kim & Moon, 2021).

Table 8. Affective-Emotional Responses to AT as Perception

Response	Short description (References)
Trust	Degree to which recipients consider an algorithm-enabled system to be competent, benevolent, credible, and honest (Ananthakrishnan et al., 2020; de Oliveira Cesar de Moraes et al., 2019; Diakopoulos & Koliska, 2017; Hofeditz et al., 2021; Kizilcec, 2016; Lai & Tan, 2019; Ochmann et al., 2021; Salminen et al., 2020; Seo et al., 2018; Shin, 2020b; Shin et al., 2020; Wang & Benbasat, 2007, 2016).
Distrust[4]	Degree to which recipients believe that an algorithm-enabled system will not act in their best interest or act injuriously, that an algorithm-enabled system has insufficient capabilities and negative motives, and that an algorithm-enabled system is dishonest, false, or misleading (de Oliveira Cesar de Moraes et al., 2019; Kim & Moon, 2021).
Information overload	Recipients' feeling of being overwhelmed by too much information on an algorithm-enabled system (Diakopoulos & Koliska, 2017; Schaffer et al., 2015).
Algorithmic anxiety	Feelings of uncertainty, loss, frustration, and lack of control by recipients in relation to a given algorithm-enabled system (Jhaver et al., 2018).
Confidence/confirmation vs. algorithm disillusionment	Degree to which recipients feel that a given algorithm-based system meets their expectations, is powerful and performative to help them successfully complete a given task, and/or provides them with the right results (Eslami et al., 2018; Schaffer et al., 2015; Shin, 2020a; Sinha & Swearingen, 2002).
Satisfaction	Degree to which recipients are pleased with an algorithm-enabled system, the interaction with this system, and the results it produces (Eslami et al., 2018; Lu et al., 2020; Schaffer et al., 2015; Shin & Park, 2019; Sinha & Swearingen, 2002).
Empathy	Perception of (algorithm-/human-produced) output as human-like (Salminen et al., 2020).

As well, studies indicate that AT as perception can lead to **information overload** by increasing recipients' sense of being overwhelmed due to information that is too complex or plentiful (Diakopoulos & Koliska, 2017), for example, a "daunting Twitter dataset" (Schaffer et al., 2015, p. 354). Akin to this, one study finds that AT as perception is associated with **algorithmic anxiety**, including feelings of loss, frustration, and uncertainty towards algorithms. These feelings are reduced when users affected by the outcomes of an algorithm (e.g., algorithmic ratings of individuals) receive additional information (Jhaver et al., 2018). Interestingly, this study also suggests that moderate levels of perceived AT might still exacerbate the problem of anxiety, since recipients will develop their own understanding of how the algorithm works, and

[4] After careful consideration, and in line with the literature reviewed (e.g., De Oliveira Cesar de Moraes et al., 2019), we chose to retain two separate concepts for trust and distrust because distrust does not merely refer to the opposite of trust but to an own, strongly negative affect that extends well beyond low trust (i.e., low trust is a necessary but not a sufficient condition for distrust).

these "folk theories" (p. 7) can make them even less certain and more anxious about an algorithm. Similarly, mixed study findings can also be observed for recipients' **confidence/confirmation** in an algorithm-based system. While some studies find positive effects on recipients' confidence (Sinha & Swearingen, 2002) and on the confirmation of their beliefs (Shin, 2020a), others find that AT as perception decreases confidence and even leads to algorithm disillusionment by exposing algorithmic errors or simplicity (Eslami et al., 2018; Schaffer et al., 2015). Essentially, the same (mixed findings) applies to the effect of AT as perception on recipients' **satisfaction**. Here, AT as perception has been found to increase satisfaction (Lu et al., 2020; Shin & Park, 2019; Sinha & Swearingen, 2002), but also to increase dissatisfaction when recipients' expectations are not met (Eslami et al., 2018), or when they are overburdened with information (Jhaver et al., 2018; Schaffer et al., 2015). Finally, one study finds that for some outputs (in this case *female* algorithm-generated personas) AT as perception can also increase (perceived) **empathy** (Salminen et al., 2020).

Like rational-cognitive responses, the studies in our review sample also point to a few factors that shape (or "moderate"[5]) the affective-emotional responses to AT as perception (see Table 9). For example, one of these studies finds that only in cases of **expectation violation**, AT as perception affects trust (along a bell-shaped relationship). Meanwhile, there is no effect when expectations have not been violated – recipients simply seem not to care about transparency when their expectations are met (Kizilcec, 2016). Another study also finds a positive effect of AT as perception on satisfaction for negative outcomes (i.e., expectation violations), and even finds a *negative* effect of AT as perception on satisfaction when the algorithm's decision is positive for recipients (Lu et al., 2020). The authors explain this counter-intuitive finding with a mismatch of actual (low) and anticipated (high) complexity of the algorithm, which becomes evident through greater AT as perception, and lets recipients lose trust in the overly simplistic algorithm. Further, the positive relationship between AT as perception and algorithm aversion is found to be more pronounced for recipients with lower levels of **general trust in AI** (Ochmann et al., 2020). In contrast, the positive impact of AT as perception on satisfaction is more pronounced for recipients with greater levels of trust (Shin & Park, 2019). Finally, two studies point to the fact that the **type of output** that is produced by an algorithm may also influence the effect of AT as perception on recipients: First, Hofeditz et al. (2021), find a stronger positive influence of AT as perception when the output of an algorithm is a recommendation in a social network rather than on a traditional news website. Second, Salminen et al. (2020) find that AT increases the perceived empathy of algorithm-generated personas (i.e., fictitious people representing product users) only when the personas are **female**.

In sum, existing research on affective-emotional responses to AT as perception is less comprehensive in number and more contradictory in terms of its findings than extant research on rational-cognitive responses. For example, corresponding studies find opposing effects of AT as perception on constructs such as trust, anxiety, and confidence. These contradictions can only partly be explained by the above-introduced factors shaping affective-emotional responses.

Table 9. "Moderating" Factors Shaping the Affective-Emotional Responses to AT as Perception

Factor	Short description (References)
Expectation violation	Degree to which recipients have anticipated a given result provided by an algorithm-enabled system (Kizilcec, 2016; Lu et al., 2020).
General trust in AI	Degree to which recipients believe that the opinion of AI is worthy of consideration and in their best personal interest (Ochmann et al., 2020; Shin & Park, 2019).
Type of output	Context-specific nature of the result generated by an algorithm, e.g., social network vs. news website recommendation (Hofeditz et al., 2021), or female vs. male persona (Salminen et al., 2020).

4.3.3 Relationship 7: Behavioral Consequences of Responses to AT as Perception

Besides rational-cognitive and affective-emotional responses, AT as perception is also, at least indirectly, related to recipient behaviors. These behavioral effects include both *intended* effects that are considered desirable by the disclosing party (e.g., adherence to algorithmic recommendations or usage intention) and *unintended* (side) effects (e.g., lower recommendation acceptance or discontinuation intention). Here, some studies make it explicit that behavioral effects in general are a consequence of AT as perception *through* rational-cognitive or affective-emotional responses, such as understanding (Criado et al., 2020) or

[5] For a discussion of the term "moderating variable" please refer to footnote 3 above.

trust (Lai & Tan, 2019). Conversely, other studies consider the direct link between recipients' perceptions of AT and their resulting behaviors, such as the acceptance of algorithmic recommendations (Sinha & Swearingen, 2002). In this section, we report the findings of both 'types' of studies (see Table 10 for a summary).

As indicated above, one intended behavioral effect highlighted in the reviewed studies is **recommendation acceptance** by recipients. More specifically, several studies offer evidence that increasing AT as perception positively influences acceptance. This holds for private contexts, such as recommendations for art (Cramer et al., 2008) or detection of fake news (Seo et al., 2018), as well as for professional contexts, including team recruitment, IT investments (Fuchs et al., 2016; Ochmann et al., 2021), and retirement planning (Hardin et al., 2017). Further, regardless of the specific context, several studies find that the relationship between recipients' perceptions of AT and their recommendation acceptance is dependent on 'shaping' factors such as **recipient type** (students vs. retirement planners) (Hardin et al., 2017), **general trust in AI** (Ochmann et al., 2020), or the level of **outcome ambiguity** of a given decision (selecting an IT investment vs. recruiting a new team member) (Fuchs et al., 2016). Contrasting these findings, two studies on demand forecasting (Lehmann et al., 2020) and job recommendations (Ochmann et al., 2020) find a negative link between AT as perception and the 'behavioral' acceptance of algorithmic recommendations. To explain their counterintuitive finding, the authors refer to recipients' disappointment with a relatively simple, "underwhelm[ing]" (Lehmann et al., 2020, p. 4) algorithm or to algorithm aversion (Ochmann et al., 2020). Also, they suspect that the observed negative link between AT as perception and the acceptance of algorithmic recommendations may depend on a system's level of **voluntariness** (Ochmann et al., 2021) and be more pronounced in mandatory than in voluntary use contexts. A third study (Bader & Kaiser, 2019) finds mixed effects and also identifies human shaping factors, such as prior domain experience, as key shaping variables. Please refer to Table 11 for a summary of factors that have been found to affect, or shape, the *relationship* between AT as perception and its behavioral consequences.

Another behavioral consequence examined in the reviewed studies is **usage** (intention)[6] (or continuance intention, respectively)**,** which is found to be increased by greater levels of AT as perception and – similar to recommendation acceptance – also exhibits moderation effects of trust (Shin et al., 2020; Shin, 2020a). Related studies find that recipients are even willing to forego control over an algorithm-enabled technology (e.g., an investment robo-advisor) when perceiving high levels of AT (Rühr, 2020). On the other hand, one study finds a somewhat mixed picture, where about half of the recipients expressed their intention to leave an algorithm-enabled technology when they perceived greater AT (Eslami et al., 2019). Further, extant literature also links AT as perception to recipients' **platform engagement.** One study finds that recipients who perceive greater AT about a platform's use of an algorithm for fraud detection (as opposed to the deployment of the algorithm without informing recipients) tend to be more cautious and spend more time on that platform, reviewing and calibrating the information received (Ananthakrishnan et al., 2020). This study suggests that perceptions of high AT trigger more in-depth thinking processes, prompting recipients to extend the choice set considered and the decision time invested. Similarly, one study suggests that when they perceive greater AT, some recipients are inclined to spend *more* time and effort on a given platform to make their inputs (e.g., reviews they write) suitable for a given algorithm (Eslami et al., 2019).

Table 10. Behavioral Consequences of AT as Perception

Consequence	**Short description (References)**
Recommendation acceptance	Recipients' incorporation of guidance provided by an algorithm-enabled system into their decisions (Bader & Kaiser, 2019; Cramer et al., 2008; Fuchs et al., 2016; Hardin et al., 2017; Lehmann et al., 2020; Ochmann et al., 2020; Ochmann et al., 2021; Seo et al., 2018; Yeomans et al., 2019).
Usage (intention)	Recipients' choice to use (or leave) a given algorithm-enabled system (Eslami et al., 2019; Rühr, 2020; Shin, 2020a; Shin et al., 2020).
Platform engagement	Amount of time, effort, and rigor that recipients of AT as perception spend on a given algorithm-enabled platform to provide inputs (e.g., writing reviews), or collect outputs (e.g., reading and judging reviews) (Ananthakrishnan et al., 2020; Eslami et al., 2019).

[6] In line with prior IS research, we view usage intention as a proxy for actual usage, and thus classify it as a *behavioral* consequence; yet, in this regard, it should be acknowledged that there might be an intention-behavior gap.

On balance, most studies focus exclusively on the intended, positive effects of AT as perception (e.g., Cramer et al., 2008; Fuchs et al., 2016; Hardin et al., 2017; Yeomans et al., 2019), whereas only a few also indicate unintended side effects such as lower acceptance (Bader & Kaiser, 2019; Lehmann et al., 2020; Ochmann et al., 2020), or intention to leave an algorithm-enabled platform (Eslami et al., 2019). We consider this lack of studies on unintended behavioral (side) effects of AT as perception an important gap in extant literature.

Table 11. "Moderating" Factors Shaping the Behavioral Effects of AT as Perception

Factor	Short description (References)
Recipient type	Recipients' domain experience (Bader & Kaiser, 2019) and background (student vs. retirement planner) (Hardin et al., 2017).
Trust in AI	Degree to which recipients believe that an algorithm-enabled technology acts in their best interest and is capable of providing them with suitable recommendations (Ochmann et al., 2020; Shin et al., 2020; Shin, 2020a).
Outcome ambiguity	Degree to which "the […] results of a decision are not clear" (Fuchs et al., 2016, p. 5).
Voluntariness	Degree to which recipients are forced to accept the recommendations of a given algorithm-enabled system or can freely opt for/ignore them (Ochmann et al., 2021).

5 Discussion of Findings and Avenues for Future Research

Drawing on a nuanced conceptualization of AT, including the distinction between *AT as action* and *AT as perception*, our study reviewed and synthesized existing literature on the antecedents and consequences of AT. In addition, we integrate the review results into a research framework consisting of seven central relationships. Using the link between AT as action and AT as perception as its nucleus, the derived framework offers a coherent overview of relevant relationships: from factors triggering and shaping AT as action, to factors shaping AT as perception, to the latter leading to rational-cognitive and affective-emotional responses, and ultimately to behavioral effects. Our review also indicates considerable variation in the depth of research on the different relationships, as well as in the strength of support for the identified factors and their effects. This brings to light some noteworthy research gaps and inconsistencies. In the following section, we elaborate on these gaps and inconsistencies and discuss five opportunities for future research resulting from them (please see Figure 3 for an overview). In doing so, we complement our review results with insights gained from extant conceptual research on transparency and algorithms. At this point, it should be noted that, beyond the research opportunities highlighted below, there may be other promising areas for future research. Furthermore, the rapid technological and societal change in the field will also lead to the emergence of additional interesting questions.

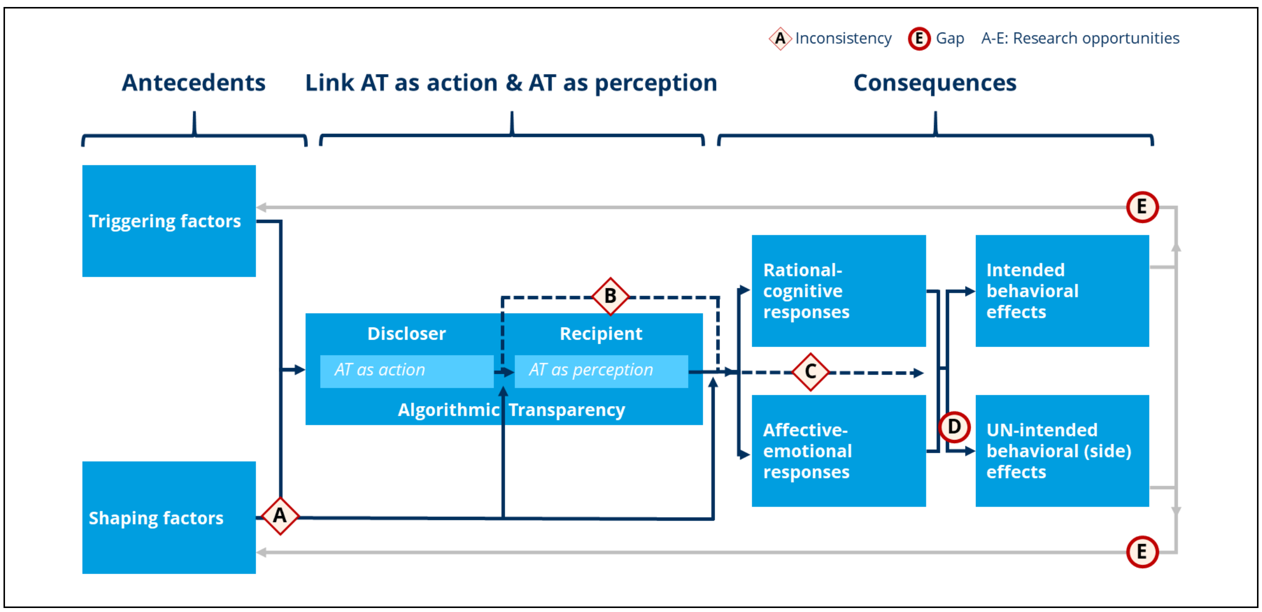

Figure 3. Inconsistencies/gaps Derived from the Research Framework and Resulting Research

5.1 Discussion of Inconsistencies Observed and Resulting Research Opportunities

Based on our analyses, we identify three particularly noteworthy research inconsistencies. These refer to (A) the factors shaping AT as action (relationship 3), (B) the process leading to both rational-cognitive and affective-emotional responses (relationships 5 and 6), and (C) the behavioral effects of AT as perception (relationship 7). First, regarding relationship 3, we find a broad set of factors — related to legal, human/social, and technology categories — that shape AT as action (see Table 4). Though numerous, the findings on these factors offer inconsistent answers to the question of what shapes AT as action. Most importantly, the factors suggested by prior studies often appear to counteract each other. For example, the cost of disclosing information and concerns about intellectual property protection suggest *less* AT as action. On the other hand, high recipient knowledge, ethical obligations, and/or legal requirements suggest *more* AT as action. At present, studies investigating such trade-off decisions appear to be largely wanting.

One potential way to resolve this inconsistency would be to examine closely the three different stages of the algorithmic process (input, transformation, output) suggested by extant conceptual work (Gillespie, 2014). Although there have been some initial proposals for a conceptual distinction among input, process, and output AT (Crawford & Schultz, 2014; Zhao et al., 2019a), to the best of our knowledge these have not yet been linked empirically to the shaping factors of AT as action. By studying what shaping factors influence AT as action in terms of the **input, transformation, and output of an algorithm** (research opportunity A), researchers are likely to find differences across the three subtypes of AT as action. Such research could thus lead to a deeper understanding of different shaping factors. For example, (high) costs of disclosing information and risks for intellectual property may lead to *low* levels of AT as action regarding the algorithmic transformation since the latter is at the 'heart' of an algorithm. Meanwhile, both legal and ethical obligations could potentially lead to *high* levels of input AT as action, as this will allow the detection of protected class data and satisfy privacy law demands for the disclosure of personal data use. Similarly, with regard to recipients' knowledge, one could imagine that disclosing parties include information on the transformation step of an algorithm for experts or auditors, whereas for lay users, information on algorithmic inputs and outputs could suffice (Hosanagar & Jair, 2018). These are but a number of initial proposals for relationships to be confirmed using quantitative surveys or experiments. In addition, qualitative studies could help explore additional links and shed further light on the question of how and why disclosing parties disclose information.

A second inconsistency emerges with regard to the process leading from AT as action to AT as perception, and to the rational-cognitive and affective-emotional responses (relationships 5 and 6). Many studies do not distinguish between AT as action and AT as perception. However, in our view, such a distinction is of critical importance, and we argue that a direct link from AT as action to rational-cognitive and affective-emotional responses is logically flawed. In order for any response to emerge from *recipients* of transparency, the extent to which these recipients *perceive* transparency (AT as perception) would need to be studied first. Otherwise, there is a risk of finding inconclusive consequences of AT as perception. For example, extant studies find that an increase in AT as perception can lead to both greater (Jarrahi & Sutherland, 2019) or lower (Rader et al., 2018) levels of perceived control, more (Lee et al., 2019; Lu et al., 2020), or less (Rader et al., 2018) perceived fairness, and higher (Shin & Park, 2019; Sinha & Swearingen, 2002) or lower (Jhaver et al., 2018; Schaffer et al., 2015) satisfaction, to only name some of the most contested relationships.

A possible explanation to these contradictions may come from conceptual research that identifies a number of **different meanings of transparency** (research opportunity B). As such, transparency may not only refer to the notion of *disclosure*, which guided our literature search, but also have additional meanings of *monitoring*, *process visibility*, or even *surveillance* (Bernstein, 2017). Accordingly, when they perceive AT, recipients may attribute additional different meanings to it, which go beyond disclosure. For example, recipients may perceive AT and realize that they now have the opportunity to understand how an algorithm derives its results from a given set of inputs and can thereby monitor a given algorithm's proper functioning. Conversely, recipients may perceive AT and develop a sense of being observed by an algorithm that they were previously unaware of, which may lead them to perceive AT with an additional meaning of surveillance.

Evidently, the rational-cognitive and affective-emotional responses of recipients would be very different in the two scenarios: In the prior scenario, greater perceived AT would likely lead recipients to perceive *greater* levels of control. Meanwhile, in the latter scenario, the opposite would likely be the case and recipients would perceive *lower* levels of control. Arguably, fairly similar scenarios could be imagined for related consequences of AT as perception, like perceived fairness or advice value. Accordingly, separating the different meanings of AT as perception beyond disclosure could help resolve the resulting inconsistencies that we find in our analysis. Recent empirical results lend support to this argument: *repeated* – as opposed to one-time – explanations along the algorithmic process show a stronger effect of AT on trust (Matt & Schlusche, 2022). Conceptualizing transparency as monitoring may thus indeed help resolve contradictions.

Finally, we note a third set of inconsistencies with regard to the behavioral effects of AT as perception (relationship 7). Here, extant studies find opposite effects of AT as perception on the same behavioral consequence. For example, the majority of studies suggest a positive (indirect) effect of AT as perception on recommendation acceptance (Cramer et al., 2008; Fuchs et al., 2016; Hardin et al., 2017; Shin, 2020a; Shin et al., 2020; Yeomans et al., 2019). Yet, there are also a number of studies that show a negative link (Lehmann et al., 2020; Ochmann et al., 2020). Similarly, the picture of usage (intention) is also mixed. One study shows a positive effect of AT as perception (Rühr, 2020), while another study shows negative consequences (Eslami et al., 2019).

The potential explanations for these inconsistencies are twofold (research opportunity C). First, a number of studies (e.g., Sinha & Swearingen, 2002; Yeomans et al., 2019) investigate the behavioral consequences of AT as perception as *direct* effects of perceived AT. These studies thereby disregard the critical **"mediating" role of rational-cognitive and affective-emotional responses.** These responses may potentially differ due to influences other than AT as perception. For example, recipients may not trust an algorithm-enabled system due to general distrust in the disclosing party. They may thus indicate low recommendation acceptance despite high perceived AT. Second, even where they do acknowledge that AT as perception only has an *indirect* influence on behavior, the vast majority of studies only investigate a very **limited set of mediating and moderating factors**. Often, these studies show a strong emphasis on trust and do not capture **interaction effects between different factors**. Meanwhile, a number of extant theories do suggest additional factors play a critical role in behavioral effects. For example, factors such as *privacy concerns* (Bélanger & Crossler, 2011; Malhotra et al., 2004) and *online trust* (Xu & Chau, 2018) have been suggested as important influences on decision behavior. These can differ depending on recipient or context characteristics. Relatedly, conceptual research on transparency argues that "[i]ncorporating the perspective of the observed — that is, the behavioral consequences of feeling observed and the desire for *privacy* — will benefit future research on transparency" (Bernstein, 2017, p. 222, emphasis added).

For the specific case of AT as perception, differences in behavioral responses may arguably be due to differing levels of recipients' privacy concerns or context-dependent online trust. For example, in professional contexts like demand planning or recruitment extant research has found a negative effect of AT as perception on decision acceptance (Lehmann et al., 2020; Ochmann et al., 2020). In these contexts, recipients' sense of violated privacy because they are 'being looked at by the algorithm' may lead to negative consequences, whereas the same may not equally be the case for non-professional contexts. Similarly, the declining usage intention for algorithm-enabled online platforms when AT as perception increases (Eslami et al., 2019) could potentially be related to a drop in online trust in the specific context of online platforms. Only very recently researchers have begun to propose more comprehensive models that intend to link together AT, trust, and behavioral outcomes (Vorm & Combs, 2022). However, to our knowledge, there has not yet been an incorporation of factors like privacy or context-dependent online trust into research on AT as perception. These factors – in combination with a focus on interaction effects – could offer a promising explanation for the aforementioned inconsistencies.

5.2 Gaps Observed

In addition to the inconsistencies highlighted above, the results of our review also revealed two noteworthy research gaps in extant literature. In particular, our results point to (D) a lack of research on the unintended (side) effects of AT as perception (relationship 7), as well as to (E) a lack of process-like understanding in terms of how the consequences of AT as perception loop back into (future) decisions around AT as action.

First, while we identified a total of 15 studies that analyzed the intended effects of AT as perception (relationship 7, see Table 10 for an overview), we found only four studies that pointed to unintended (side) effects (Bader & Kaiser, 2019; Eslami et al., 2019; Lehmann et al., 2020; Ochmann et al., 2020). This gap stands in stark contrast to the variety of negative rational-cognitive and affective-emotional responses to AT as perception that have been identified by extant studies. Based on this, one would also expect a more extensive set of negative behavioral consequences. The relatively positive picture of the influence of AT as perception on recipients' behavior may thus be incomplete and skewed. This potential issue has already been foreseen by earlier expert statements that tended to find numerous concrete drawbacks but few concrete benefits of AT as perception (Diakopoulos & Koliska, 2017).

One potential way to address the gap in present-day research on negative behavioral consequences of AT as perception, could be through **new sources of data** (research opportunity D). For example, negative behavioral effects of AT as perception could be particularly prevalent when individuals are subjected to algorithm-enabled decision-making by public institutions. The benefits of deploying algorithmic instead of human decision-makers for public tasks like law enforcement can be material (Xu & Chen, 2004). However, for technologies like preventive policing or automated prosecution, extant conceptual research identifies a high risk that individuals respond negatively when they perceive algorithms. For example, this might be due to perceived discrimination, perceived loss of control, or perceived privacy violations (Citron, 2007; Giest & Grimmelikhuijsen, 2020; Levmore & Fagan, 2021). Accordingly, this context could offer particularly rich data on the negative behavioral consequences of AT as perception. Future researchers could collaborate with public institutions to design (semi-)controlled experiments, where algorithmic decisions are introduced to recipients. In these experiments, one group might be subjected to high levels of AT as perception, whereas another group receives no information on the algorithm making the decision[7]. Such an experiment would potentially point to negative influences of AT as perception with recipients and lead to unique new insights. An initial attempt to study the effects of AT as perception in a similarly delicate question – the limitation of free speech through algorithmic comment moderation – a study by Müller, Koelmann, Niemann, Plattfaut, and Becker (2022) may be partially addressing this, yet at the time of concluding this review, the research was still underway.

Table 12. Inconsistencies and Gaps Observed and Resulting Research Opportunities

Research opportunity	Relation-ship(s)	Inconsistency / gap observed	Possible avenue for future research
A	3	Extant research suggests a broad set of factors shaping AT as action, yet the factors found by prior studies often appear to counteract each other. For example, risk of losing intellectual property suggests less AT as action, whereas ethical/legal obligations suggest more AT as action.	The **input, transformation, and output** stage of the algorithmic process may imply different (sub-)types of AT as action, which are in return influenced by different shaping factors. Thus, by examining the three sub-types of AT as action and understanding how they are related to shaping factors, future researchers may be able to reconcile current inconsistencies.
B	5 and 6	Many studies do not distinguish between AT as action and AT as perception, which is a logically flawed argument, because any response *to* AT requires it to first be *perceived.* This in return leads to conflicting findings on the cognitive and affective responses to AT, where AT is found to both increase and decrease consequences such as satisfaction.	There may be **different additional meanings of transparency**, which recipients of AT may perceive beyond the notion of disclosure, for example, monitoring or surveillance. Depending on the additional meaning they associate with AT, recipients may react differently (e.g., greater trust due to the ability to monitor vs. distrust due to a notion of surveillance). By studying these meanings, future research may be able to explain the differing responses to AT as perception.

[7] We acknowledge that there may be delicate ethical questions that need to be resolved prior to conducting such research.

C	7	Extant research finds opposite effects of AT as perception on the same behavioral consequence. In particular, for recommendation acceptance and usage (intention) some studies show a positive effect of AT as perception while others show a negative effect.	Rational-cognitive and affective-emotional responses to AT as perception may play a critical mediating or moderating role that can explain different behavioral responses. Future research could try to capture this role by investigating behavioral consequences as **indirect (mediated) effect**, expanding to a **greater set of mediating/moderating factors** studied (e.g., to privacy), and investigating the interaction **effects between multiple factors**.
D	7	Research on unintended behavioral (side) effects of AT as perception is very limited and not comparable in breadth and depth to the research conducted on intended behavioral effects. Meanwhile, findings on negative rational and affective responses suggest that negative behavioral effects may likely exist.	**New sources of data**, for example from algorithmic decision-making in public institutions, and **extant findings on negative rational-cognitive and affective-emotional responses** to AT as perception, could help future research to uncover additional insights on unintended behavioral (side) effects.
E	N/A	Extant studies of AT almost exclusively adopt a static variance logic, a dynamic perspective on AT is largely wanting. Such a perspective could shed light on potential virtuous and vicious circles, for example, based on recipients' behavioral responses and the resulting reactions of disclosing parties.	**Longitudinal studies with a process view on AT** that observe the full set of relationships and potential feedback loops in the research model – from antecedents of AT as action to behavioral consequences of AT as perception – could be conducted for an extended period of time; for example, in the application domain of recommendation agents.

Another opportunity to unveil and better understand the negative consequences of AT as perception may be to **build on the negative rational-cognitive and affective-emotional responses** that have already been identified in extant studies, such as lower satisfaction due to information overload (Schaffer et al., 2015), or algorithmic anxiety (Jhaver et al., 2018). By adapting the underlying research designs to focus (more[8]) on the presumed negative behavioral effects of AT as perception that are resulting from these negative responses, researchers may be able to shed additional light on this question.

Finally, our second research gap refers to a lack of a process-like understanding on how the consequences of AT as perception loop back into (future) decisions around AT as action. Based on our review, we find that present-day research largely applies a static, unidirectional view of AT. Thereby, AT as action is a consequence of shaping and triggering factors and leads to AT as perception. This, in return, has direct cognitive and affective responses, as well as indirect behavioral consequences. Meanwhile, there have – to our knowledge – not yet been studies applying a process view to AT. Such a view would extend the current research from a static to a *dynamic* perspective and could help reveal the formation of virtuous or vicious circles around AT. For example, based on recipients' favorable behavioral responses to AT as perception, such as greater decision acceptance, disclosing parties may be motivated to further increase AT. This could cause further positive responses, leading to the emergence of a virtuous cycle of AT. Conversely, when recipients react negatively to AT, they may perceive and show lower decision acceptance, and disclosing parties may conclude that they should provide less AT. This could lead to further distrust and lower decision acceptance, ultimately leading to the emergence of a vicious cycle of AT.

Arguably, the most suitable way for addressing this gap is through **longitudinal studies on AT.** Such studies observe the entire set of relationships described in our research model from antecedents of AT as action to behavioral consequences of AT as perception (research opportunity E). Given that the provisioning and perception of AT takes time, changes in cognitive and affective responses, as well as changed behaviors and feedback loops, will take time to emerge. One potential application domain for such studies could be relatively simple recommendation agents, which were among the first to provide algorithms (and disclose information on them) with (lay) end users (e.g., Cramer et al., 2008; Sinha & Swearingen, 2002). Based on these comparatively long-standing experiences with disclosing and perceiving AT, one might expect feedback loops to be most advanced in this domain. In addition, such

[8] The underlying studies also consider behavioral consequences, yet these are usually the behavioral responses to the algorithm itself, rather than the responses to the perceived *transparency on* the algorithm.

tools for individuals can be adapted faster than full-scale algorithmic technologies in organizations and thus allow for earlier findings.

6 Conclusion

The overarching goal of this research was to facilitate and inspire future research on the increasingly important topic of AT. We meet this goal in three ways. First, the study strengthens the conceptual foundations of AT by offering a more nuanced conceptualization, including the explicit distinction between *AT as action* and *AT as perception,* as well as by delineating AT from related concepts, such as xAI and algorithmic accountability. Second, following established guidelines and drawing on a sample of 50 studies (published between 2002 and 2021), our study provides a systematic and comprehensive literature review of current research findings regarding the antecedents and consequences of AT. Third, our results integrate the review insights into a coherent research framework. This framework is then used to highlight a set of research inconsistencies and gaps, along with the identification of promising research opportunities.

In conclusion, our study illustrates the relevance of, and increasing research interest in, AT in the field of IS and beyond. It sets the stage for future research on the topic by synthesizing extant work and by highlighting exciting opportunities for future study. With the ongoing digital transformation of businesses and society and the associated emergence of increasingly intelligent and powerful algorithms, it can be expected that AT will continue to gain momentum and grow in both academic and practical relevance. As such, we hope that our study can provide guidance and serve as inspiration for future research on the topic.

References

ACM US Public Policy Council (2017, January 12). *Statement on algorithmic transparency and accountability*. Retrieved from https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Ahonen, P., & Erkkilä, T. (2020). Transparency in algorithmic decision-making: Ideational tensions and conceptual shifts in Finland. *Information Polity*, *25*(4), 419–432.

Allen, G. N., Burk, D. L., & Davis, G. B. (2006). Academic data collection in electronic environments: Defining acceptable use of internet resources. *MIS Quarterly*, *30*(3), 599–610.

Ammara (2020, August 17). *“F*ck the algorithm”: A rallying cry for the future*. Digital Diplomacy. Retrieved from https://medium.com/digital-diplomacy/fuck-the-algorithm-the-rallying-cry-of-our-youth-dd2677e190c

Amoore, L. (2020, August 19). *Why 'Ditch the algorithm' is the future of political protest*. The Guardian. Retrieved from https://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.

Ananthakrishnan, U. M., Li, B., & Smith, M. D. (2020). A tangled web: Should online review portals display fraudulent reviews? *Information Systems Research*, *31*(3), 950–971.

Asatiani, A., Malo, P., Nagbol, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2020). Challenges of explaining the behavior of black-box AI systems. *MIS Quarterly Executive*, *19*(4), 259–278.

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, *30*(1), 13–28.

Bader, V., & Kaiser, S. (2019). Algorithmic decision-making? The user interface and its role for human involvement in decisions supported by artificial intelligence. *Organization*, *26*(5), 655–672.

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., & Beekhuyzen, J. (2015). Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support. *Communications of the Association for Information Systems*, *37*, 154–204.

Bélanger, & Crossler (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, *35*(4), 1017–1041.

Bernstein, E. S. (2017). Making transparency transparent: The evolution of observation in management theory. *Academy of Management Annals*, *11*(1), 217–266.

Boell, S. K., & Cecez-Kecmanovic, D. (2015a). Debating systematic literature reviews (SLR) and their ramifications for IS: A rejoinder to Mike Chiasson, Briony Oates, Ulrike Schultze, and Richard Watson. *Journal of Information Technology*, *30*(2), 188–193.

Boell, S. K., & Cecez-Kecmanovic, D. (2015b). On being ‘systematic’ in literature reviews. *Journal of Information Technology*, *30*, 161–173.

Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale Journal of Law & Technology*, *20*, 103–176.

Brill, J. (2015). *Scalable approaches to transparency and accountability in decisionmaking algorithms: Remarks at the NYU conference on algorithms and accountability* (No. 28). Federal Trade Commission. Retrieved from https://loadtest.ftc.gov/system/files/documents/public_statements/629681/150228nyualgorithms.pdf

Bunt, A., Lount, M., & Lauzon, C. (2012). Are explanations always important? A study of deployed, low-cost intelligent interactive systems. In Association for Computing Machinery (Chair), *Proceedings of the 2012 ACM international conference on intelligent user interfaces*.

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1-12.

Caplan, R., Donovan, J., Matthews, J., & Hanson, L. (2018). *Algorithmic accountability: A primer*. Data & Society Research Institute. Retrieved from https://datasociety.net/library/algorithmic-accountability-a-primer/

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23.

Cech, F. (2020). Tackling algorithmic transparency in communal energy accounting through participatory design. In F. Cech & S. Farnham (Eds.), *C&T '21: Proceedings of the 10th international conference on communities & technologies* (pp. 258–268). ACM.

Chiasson, M. W. (2015). Avoiding methodological overdose: A declaration for independent ends. *Journal of Information Technology*, *30*(2), 174–176.

Citron, D. K. (2007). Technological due process. *Washington University Law Review*, *85*, 1249–1314.

Cram, W. A., Templier, M., & Pare, G. (2020). (Re)considering the concept of literature review reproducibility. *Journal of the Association for Information Systems*, *21*(5), 1103–1114.

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, *18*(5), 455–496.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, *55*(1), 93–128.

Criado, J. I., Valero, J., & Villodre, J. (2020). Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system. *Information Polity*, *25*(4), 449–470.

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, *29*(3), 245–268.

De Oliveira Cesar de Moraes, H. R., Sanchez, O., Brown, S., & Zhang, B. (2019). Trust and distrust in big data recommendation agents. In *ICIS 2019 Proceedings*.

De-Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology*, *31*(4), 525–541.

Desai, D. R., & Kroll, J. A. (2017). Trust but verify: A guide to algorithms and the law. *Harvard Journal of Law & Technology*, *31*(1), 1–64. Retrieved from https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/hjlt31§ion=4

Diakopoulos, N., Cass, S., & Romero, J. (2014). Data-driven rankings: The design and development of the IEEE Top Programming Languages news app. In *Proceedings of the Symposium on Computation+ Journalism*.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–828.

Dimensions.ai (2021). *Free text search on "Algorithmic Transparency" in full data*. Dimensions.ai Retrieved from https://app.dimensions.ai/discover/publication?search_mode=content&search_text=%22Algorithmic%20Transparency%22&search_type=kws&search_field=full_search

Domingo-Ferrer, J., Pérez-Solà, C., & Blanco-Justicia, A. (2019). Collaborative explanation of deep models with limited interaction for trade secret and privacy preservation. In L. Liu (Ed.), *ACM digital library, companion proceedings of The 2019 World Wide Web Conference* (pp. 501–507). Association for Computing Machinery.

Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism*, *5*(4), 404–419.

Durcikova, A., & Gray, P. (2009). How knowledge validation processes affect knowledge contribution. *Journal of Management Information Systems*, *25*(4), 81–108.

Eslami, M., Krishna Kumaran, S. R., Sandvig, C., & Karahalios, K. (2018). Communicating algorithmic process in online behavioral advertising. In R. Mandryk, M. Hancock, M. Perry, & A. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery.

Eslami, M., Vaccaro, K., Lee, M. K., Elazari Bar On, A., Gilbert, E., & Karahalios, K. (2019). User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *CHI 2019: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery.

Fink, K. (2018). Opening the government's black boxes: Freedom of information and algorithmic accountability. *Information, Communication & Society*, *21*(10), 1453–1471.

Flyverbom, M. (2016). Transparency: Mediation and the management of visibilities. *International Journal of Communication*, *10*(1), 110–122.

Fuchs, C., Matt, C., Hess, T., & Hoerndlein, C. (2016). Human vs. algorithmic recommendations in big data and the role of ambiguity. In *AMCIS 2016 Proceedings*.

Galliers, R. D., Newell, S. [S.], Shanks, G., & Topi, H. [H.] (2017). Datification and its human, organizational and societal effects: The strategic opportunities and challenges of algorithmic decision-making. *The Journal of Strategic Information Systems*, *26*(3), 185–190.

Gerlings, J., Shollo, A., & Constantiou, I. (2021). Reviewing the need for explainable artificial intelligence (xAI). In T. Bui (Chair), *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Giest, S., & Grimmelikhuijsen, S. (2020). Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective. *Information Polity*, *25*(4), 409–417.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167–193). MIT Press.

Goad, D., & Gal, U. (2018). Understanding the impact of transparency on algorithmic decision making legitimacy. In U. Schultze, M. Aanestad, M. Mähring, C. Østerlund, & K. Riemer (Eds.), *Living with monsters? Social implications of algorithmic phenomena, hybrid agency, and the performativity of technology* (pp. 64–79). Springer.

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., Holzinger, A. (2018). Explainable AI: The new 42? In A. Holzinger, P. Kieseberg, A. M. Tjoa, & E. Weippl (Eds.), *LNCS Sub: Library: SL3 - Information Systems and Applications, incl. Internet/Web, and HCI, Machine Learning and Knowledge Extraction* (pp. 295–303). Springer.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 1–42.

Hansen, H. K., & Flyverbom, M. (2015). The politics of transparency and the calibration of knowledge in the digital age. *Organization*, *22*(6), 872–889.

Hardin, A., Looney, C. A., & Moody, G. D. (2017). Assessing the credibility of decisional guidance delivered by information systems. *Journal of Management Information Systems*, *34*(4), 1143–1168.

Hepenstal, S., Zhang, L., Kodagoda, N., & Wong, B. L. W. (2020). Pan. In *IUI 2020: Proceedings of the 25th International Conference on Intelligent User Interfaces companion* (pp. 134–135). Association for Computing Machinery.

Hofeditz, L., Mirbabaie, M., Holstein, J., & Stieglitz, S. (2021). Do you trust an AI-journalist? A credibility analysis of news content with AI-authorship. In *ECIS 2021 Proceedings*.

Hosanagar, K., & Jair, V. (2018). *We need transparency in algorithms, but too much can backfire*. Harvard Business Review. Retrieved from https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire

Jarrahi, M. H., & Sutherland, W. (2019). Algorithmic management and algorithmic competencies: Understanding and appropriating algorithms in gig work. In N. Greene Taylor, C. Christian-Lamb, M. H. Martin, & B. Nardi (Chairs), *Information in contemporary society: 14th International Conference, iConference 2019.* Washington, DC.

Jhaver, S., Karpfen, Y., & Antin, J. (2018). Algorithmic anxiety and coping strategies of Airbnb hosts. In R. Mandryk, M. Hancock, M. Perry, & A. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery.

Karwatzki, S., Dytynko, O., Trenz, M., & Veit, D. (2017). Beyond the personalization–privacy paradox: Privacy valuation, transparency features, and service personalization. *Journal of Management Information Systems*, *34*(2), 369–400.

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081–2096.

Khovanskaya, V., Bezaitis, M., & Sengers, P. (2016). The case of the strangerationist. In M. Foth, W. Ju, R. Schroeter, & S. Viller (Eds.), *DIS 2016: Proceedings of the 2016 ACM conference on Designing Interactive Systems* (pp. 134–145). Association for Computing Machinery.

Kim, K., & Moon, S.-I. (2021). When algorithmic transparency failed: Controversies over algorithm-driven content curation in the South Korean digital environment. *American Behavioral Scientist*, *65*(6), 847–862.

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In J. Kaye, A. Druin, C. Lampe, D. Morris, Hourcade, & Juan P. (Eds.), *CHI 2016: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390–2395). Association for Computing Machinery.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*(3), 633–705.

Lai, V., & Tan, C. (2019). On human predictions with explanations and predictions of machine learning models. In D. Boyd & J. Morgenstern (Eds.), *FAT* '19: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency* (pp. 29–38). Association for Computing Machinery.

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction*, *3*, 1–26.

Lehmann, C. A., Haubitz, C., Fuegener, A., & Thonemann, U. (2020). Keep it mystic? – The effects of algorithm transparency on the use of advice. In *ICIS 2020 Proceedings*.

Levmore, S., & Fagan, F. (2021). Competing algorithms for law. *The University of Chicago Law Review*, *88*(2), 367–412. Retrieved from https://www.jstor.org/stable/26986410

Lim, B. Y., & Dey, A. K. (2011). Investigating intelligibility for uncertain context-aware applications. In J. Landay & Y. Shi (Eds.), *UbiComp '11: Proceedings of the 13th International Conference on Ubiquitous Computing*. Association for Computing Machinery.

Liu, B., & Wei, L. (2021). Machine gaze in online behavioral targeting: The effects of algorithmic “humanlikeness” on social presence and social influence. *Computers in Human Behavior*, *124*(106926).

Lu, J., Lee, D. [Dokyun], Kim, T. W., & Danks, D. (2020). Good explanation for algorithmic transparency. In A. Markham (Ed.), *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (p. 93). Association for Computing Machinery.

Maak, N. (2021, January 24). *Ein gutes zeichen: Bewegung "Fuck the Algorithm".* Frankfurter Allgemeine Zeitung. Retrieved from https://www.faz.net/aktuell/feuilleton/fuck-the-algorithm-bewegung-in-grossbritannien-17160716.html

Malhotra, N. K., Sung, S. K., & Agarwal, J. (2004). Internet users' information privacy concerns: The construct, the scale, and a causal model. *Information Systems Research*, *15*(4), 336–355.

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2018). Algorithmic pollution: Understanding and responding to negative consequences of algorithmic decision-making. In U. Schultze, M. Aanestad, M. Mähring, C. Østerlund, & K. Riemer (Eds.), *Living with monsters? Social implications of algorithmic phenomena, hybrid agency, and the performativity of technology* (pp. 31–47). Springer.

Mark, R., & Anya, G. (2019). Ethics of using smart city AI and big data: The case of four large European cities. *The ORBIT Journal*, *2*(2), 1–36.

Markus, M. L. (2017). Datification, organizational strategy, and IS research: What’s the score? *The Journal of Strategic Information Systems*, *26*(3), 233–241.

Markus, M. L., & Topi, H. [Heikki] (2015). *Big data, big decisions for science, society, and business: Report on a research agenda setting workshop*. Retrieved from https://www.cloud-finder.ch/fileadmin/dateien/pdf/themenkategorien/bigdata/bigdata_big_decisions.pdf

Matt, C., & Schlusche, C. (2022). How transparency measures can attenuate initial failures of intelligent decision support systems. In *ECIS 2022 Proceedings*. Retrieved from https://aisel.aisnet.org/ecis2022_rp/10/

Miltgen, C. L., & Peyrat-Guillard, D. (2014). Cultural and generational influences on privacy concerns: A qualitative study in seven European countries. *European Journal of Information Systems*, *23*(2), 103–125.

Müller, K., Koelmann, H., Niemann, M., Plattfaut, R., & Becker, J. (2022). Exploring audience's attitudes towards machine learning-based automation in comment moderation. In *Wirtschaftsinformatik 2022 Proceedings*. Retrieved from https://aisel.aisnet.org/wi2022/human_rights/human_rights/1/

Newell, S. [Sue], & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems*, *24*(1), 3–14.

Oates, B. J. (2015). On systematic reviews for evidence-based practice. *Journal of Information Technology*, *30*(2), 177–179.

Ochmann, J., Michels, L., Zilker, S., Tiefenbeck, V., & Laumer, S. (2020). The influence of algorithm aversion and anthropomorphic agent design on the acceptance of AI-based job recommendations. In *ICIS 2020 Proceedings*.

Ochmann, J., Zilker, S., & Laumer, S. (2021). The evaluation of the black box problem for AI-based recommendations: An interview-based study. In *Wirtschaftsinformatik 2021 Proceedings*.

Okoli, C., & Schabram, K. (2010). *A guide to conducting a systematic literature review of information systems research*. SSRN. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1954824

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishers.

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems*, *25*(6), 493–508.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful explanations of black box AI decision systems. In Association for the Advancement of Artificial Intelligence (Ed.), *AAAI-19: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence* (pp. 9780–9784). AAAI Press.

Preece, A. (2018). Asking 'Why' in AI: Explainability of intelligent systems - perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, *25*(2), 63–72.

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. In R. Mandryk, M. Hancock, M. Perry, & A. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141.

Rosenfeld, A., & Richardson, A. (2019). Explainability in human–agent systems. *Autonomous Agents and Multi-Agent Systems*, *33*(6), 673–705.

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems*, *23*(3), 241–255.

Rühr, A. (2020). Robo-advisor configuration: An investigation of user preferences and the performance-control dilemma. In *ECIS 2020 Proceedings*.

Salminen, J., Santos, J. M., Jung, S.-G., Eslami, M., & Jansen, B. J. (2020). Persona transparency: Analyzing the impact of explanations on perceptions of data-driven personas. *International Journal of Human-Computer Interaction*, *36*(8), 788–800.

Schaffer, J., Giridhar, P., Jones, D., Höllerer, T., Abdelzaher, T., & O'Donovan, J. (2015). Getting the message? A study of explanation interfaces for microblog data analysis. In O. Brdiczka & P. Chau (Eds.), *IUI '15: Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 345–356). Association for Computing Machinery.

Schultze, U. (2015). Skirting SLR's language trap: Reframing the 'systematic' vs 'traditional' literature review opposition as a continuum. *Journal of Information Technology*, *30*(2), 180–184.

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, *7*(4), 233–242.

Seo, H., Xiong, A., & Lee, D. [Dongwon] (2018). Trust it or not: Effects of machine-learning warnings in helping individuals mitigate misinformation. In H. Akkermans, K. Fontaine, & I. Vermeulen (Eds.), *WebSci '18: Proceedings of the 10th ACM conference on web science* (pp. 265–274). Association for Computing Machinery.

Shah, H. (2018). Algorithmic accountability. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *376*(2128).

Shaikh, M., & Vaast, E. (2016). Folding and unfolding: Balancing openness and transparency in open source communities. *Information Systems Research*, *27*(4), 813–833.

Shi, J., & Hurdle, J. F. (2018). Trie-based rule processing for clinical NLP: A use-case study of n-trie, making the ConText algorithm more efficient and scalable. *Journal of Biomedical Informatics*, *85*, 106–113.

Shin, D. (2020a). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, *109*.

Shin, D. (2020b). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, *64*(4), 541–565.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284.

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management*, *52*.

Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. In L. Terveen (Ed.), *CHI 2002: Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery.

Springer, A., & Whittaker, S. (2019). Making transparency clear: The dual importance of explainability and auditability. In C. Trattner, D. Parra, & N. Riche (Eds.), *Joint Proceedings of the ACM IUI 2019 Workshops*. Association for Computing Machinery.

Springer, A., & Whittaker, S. (2020). Progressive disclosure: When, why, and how do users want algorithmic transparency information? *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1–32.

Stohl, C., Stohl, M., & Leonardi, P. M. (2016). Managing opacity: Information visibility and the paradox of transparency in the digital age. *International Journal of Communication*, *10*, 123–137.

Strich, F., Mayer, A.-S., & Fiedler, M. (2021). What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. *Journal of the Association for Information Systems*, *22*(2), 304–324.

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, *76*, 89–106.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, *37*, 205–224.

Vorm, E. S., & Combs, D. J. Y. (2022). Integrating transparency, trust, and acceptance: The intelligent systems technology acceptance model (ISTAM). *International Journal of Human-Computer Interaction*, 1–18.

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, *23*(4), 217–246.

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems*, *33*(3), 744–775.

Watson, H. J., & Nations, C. (2019). Addressing the growing need for algorithmic transparency. *Communications of the Association for Information Systems*, *45*(1), 488–510.

Watson, R. T. (2015). Beyond being systematic in literature reviews in IS. *Journal of Information Technology*, *30*(2), 185–187.

Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*, *26*(2), xiii–xxiii.

Weller, A. (2019). Transparency: Motivations and challenges. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, explaining and visualizing deep learning* (pp. 23–40). Springer.

Wiener, M., Mähring, M., Remus, U., & Saunders, C. (2016). Control configuration and control enactment in information systems projects: Review and expanded theoretical framework. *Management Information Systems Quarterly*, *40*(3), 741–774.

Wieringa, M. (2020). What to account for when accounting for algorithms. In M. Hildebrandt, C. Castillo, E. Celis, S. Ruggieri, L. Taylor, & G. Zanfir-Fortuna (Eds.), *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency: January 27-30, 2020, Barcelona, Spain* (pp. 1–18). Association for Computing Machinery.

Xu, J. J., & Chau, M. (2018). Cheap talk? The impact of lender-borrower communication on peer-to-peer lending outcomes. *Journal of Management Information Systems*, *35*(1), 53–85.

Xu, J. J., & Chen, H. (2004). Fighting organized crimes: Using shortest-path algorithms to identify associations in criminal networks. *Decision Support Systems*, *38*(3), 473–487.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32*(4), 403–414.

Zarsky, T. (2016). The trouble with algorithmic decisions. *Science, Technology, & Human Values*, *41*(1), 118–132.

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019a). Do users always want to know more? Investigating the relationship between system transparency and users' trust in advice-giving systems. In *ECIS 2019 Proceedings*.

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019b). Transparency in advice-giving systems: A framework and a research model for transparency provision. In C. Trattner, D. Parra, & N. Riche (Eds.), *Joint Proceedings of the ACM IUI 2019 Workshops*. Association for Computing Machinery.

Zouave, E. T., & Marquenie, T. (2017). An inconvenient truth: Algorithmic transparency & accountability in criminal intelligence profiling. In J. Brynielsson (Ed.), *2017 European Intelligence and Security Informatics Conference (EISIC)* (pp. 17–23). IEEE.

Appendix A: Papers Included into the Review

Table A1. Overview of Articles Included by Journal and Research Method

Article	Journal	Research method
Ahonen & Erkkilä (2020)	Information Polity	Case study
Ananthakrishnan et al. (2020)	Information Systems Research	Online experiment
Asatiani et al. (2020)	MIS Quarterly Executive	Case study
Awad & Krishnan (2006)	MIS Quarterly	Survey
Bader & Kaiser (2019)	Organization	Case study
Brauneis & Goodman (2018)	Yale Journal of Law & Technology	Document examination
Bunt et al. (2012)	Proceedings of the 2012 ACM Conference on Intelligent User Interfaces (IUI)	In-depth interviews
Cech (2020)	Proceedings of the 10th International Conference on Communities & Technologies	Ethnography
Cramer et al. (2008)	User Modeling and User-Adapted Interaction	Experiment
Criado et al. (2020)	Information Polity	Case study
de Oliveira Cesar de Moraes et al. (2019)	Proceedings of the 2019 International Conference of Information Systems (ICIS)	Online experiment
Diakopoulos et al. (2014)	Proceedings of the Symposium on Computation + Journalism	Mixed
Diakopoulos & Koliska (2017)	Digital Journalism	Focus group
Domingo-Ferrer et al. (2019)	Companion Proceedings of the 2019 World Wide Web Conference	Design proposal & evaluation
Eslami et al. (2018)	Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems	In-depth interviews
Eslami et al. (2019)	Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems	Mixed case study
Fink (2018)	Information, Communication & Society	Document examination
Fuchs et al. (2016)	Proceedings of the 2016 Americas Conference of Information Systems (AMCIS)	Online experiment
Hardin et al. (2017)	Journal of Management Information Systems	Experiment
Hepenstal et al. (2020)	Proceedings of the 25th International Conference on Intelligent User Interfaces	In-depth interviews
Hofeditz et al. (2021)	European Conference of Information Systems 2021 Research Papers	Online experiment
Jarrahi & Sutherl& (2019)	Information in Contemporary Society	Case study
Jhaver et al. (2018)	Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems	In-depth interviews
Khovanskaya et al. (2016)	Proceedings of the 2016 ACM Conference on Designing Interactive Systems	Mixed
Kim & Moon (2021)	American Behavioral Scientist	Case study
Kizilcec (2016)	Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems	Online experiment
Lai & Tan (2019)	Proceedings of the Conference on Fairness, Accountability, and Transparency	Experiment
Lehmann et al. (2020)	Proceedings of the 2020 International Conference of Information Systems (ICIS)	Online experiment
Lee et al. (2019)	Proceedings of the ACM on Human-Computer Interaction	Laboratory experiment
Lim & Dey (2011)	Proceedings of the 13th International	Online experiment

Table A1. Overview of Articles Included by Journal and Research Method

Article	Journal	Research method
	Conference on Ubiquitous Computing	
B. Liu & Wei (2021)	Computers in Human Behavior	Online experiment
Lu et al. (2020)	Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society	Online experiment
Ochmann et al. (2020)	Proceedings of the 2020 International Conference of Information Systems (ICIS)	Online experiment
Ochmann et al. (2021)	Wirtschaftsinformatik 2021 Proceedings	Semi-structured interviews
Rader et al. (2018)	Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems	Online experiment
Rühr (2020)	Proceedings of the 2020 European Conference of Information Systems (ECIS)	Conjoint analysis
Salminen et al. (2020)	International Journal of Human-Computer Interaction	Online experiment
Schaffer et al. (2015)	Proceedings of the 20th International Conference on Intelligent User Interfaces	Online experiment
Seo et al. (2018)	WebSci '18: Proceedings of the 10th ACM Conference on Web Science	Online experiment
Shin (2020a)	Computers in Human Behavior	In-depth interviews; survey
Shin (2020b)	Journal of Broadcasting & Electronic Media	Survey
Shin & Park (2019)	Computers in Human Behavior	In-depth interviews; survey
Shin et al. (2020)	International Journal of Information Management	Survey
Sinha & Swearingen (2002)	CHI 2002 Extended Abstracts on Human Factors in Computing Systems	Survey
Springer & Whittaker (2019)	Joint Proceedings of the ACM IUI 2019 Workshops	Online experiment
Springer & Whittaker (2020)	ACM Transactions on Interactive Intelligent Systems (TiiS)	Online experiment; semi-structured interviews
Wang & Benbasat (2007)	Journal of Management Information Systems	Laboratory experiment
Wang & Benbasat (2016)	Journal of Management Information Systems	Laboratory experiment
H. J. Watson & Nations (2019)	Communications of the Association of Information Systems	Mixed
Yeomans et al. (2019)	Journal of Behavioral Decision Making	Experiment

Appendix B: Descriptive Statistics on the Review Sample

Each paper in our final review sample was classified to determine the publication year (cf. Figure B1) and publication outlet, the research type and methodology, the scientific discipline, the (first) author's country of (professional) origin, the theoretical underpinnings and application domain, as well as the industry context. A detailed view on the descriptive statistics can be found in the subsequent tables.

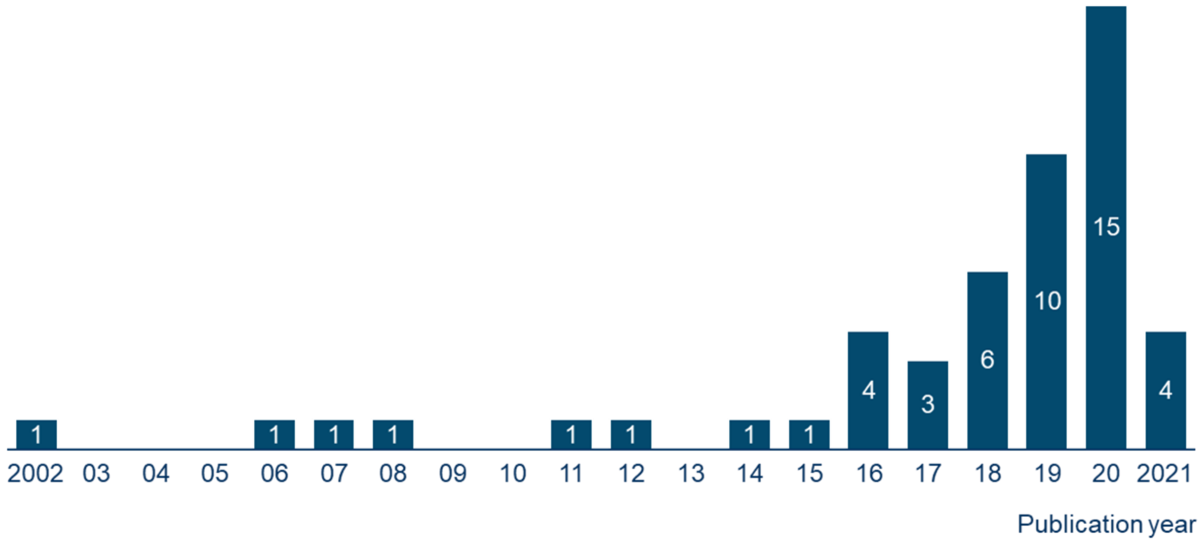

Figure B1. Review Sample by Publication Year

Table B1. Publications by publication outlet

Publication outlet (journal title or conference proceedings)	# of publications
Computers in Human Behavior	3
Journal of Management Information Systems	3
Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems	3
Information Polity	2
Proceedings of the 2020 International Conference of Information Systems (ICIS)	2
ACM Transactions on Interactive Intelligent Systems (TiiS)	1
American Behavioral Scientist	1
CHI 2002 Extended Abstracts on Human Factors in Computing Systems	1
Communications of the Association of Information Systems	1
Companion Proceedings of the 2019 World Wide Web Conference	1
Digital Journalism	1
European Conference of Information Systems 2021 Research Papers	1
Information in Contemporary Society	1
Information Systems Research	1
Information, Communication & Society	1
International Journal of Human-Computer Interaction	1
International Journal of Information Management	1
Joint Proceedings of the ACM IUI 2019 Workshops	1
Journal of Behavioral Decision Making	1
Journal of Broadcasting & Electronic Media	1
MIS Quarterly	1
MIS Quarterly Executive	1

Organization	1
Proceedings of the 10th International Conference on Communities & Technologies	1
Proceedings of the 13th International Conference on Ubiquitous Computing	1
Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces	1
Proceedings of the 2016 ACM Conference on Designing Interactive Systems	1
Proceedings of the 2016 Americas Conference of Information Systems (AMCIS)	1
Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems	1
Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems	1
Proceedings of the 2019 International Conference of Information Systems (ICIS)	1
Proceedings of the 2020 European Conference of Information Systems (ECIS)	1
Proceedings of the 20th International Conference on Intelligent User Interfaces	1
Proceedings of the 25th International Conference on Intelligent User Interfaces	1
Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society	1
Proceedings of the ACM on Human-Computer Interaction	1
Proceedings of the Conference on Fairness, Accountability, and Transparency	1
Proceedings of the Symposium on Computation + Journalism	1
User Modeling and User-Adapted Interaction	1
WebSci '18: Proceedings of the 10th ACM Conference on Web Science	1
Wirtschaftsinformatik 2021 Proceedings	1
Yale Journal of Law and Technology	1

Table B2. Categorization of Publications

Article	Application domain	National origin	Industry	Scientific discipline	Theoretical underpinnings
Ahonen & Erkkilä (2020)	Regulation	Finland	Public services	Political Science	Performatives and semantics
Ananthakrishnan et al. (2020)	Website design	USA	Social Media	IS	Theory of reasoned action
Asatiani et al. (2020)	Multiple	Sweden	Public services	IS	None
Awad & Krishnan (2006)	Website design	USA	e-Commerce	IS	Utility maximization th.
Bader & Kaiser (2019)	Decision support systems	Germany	Tele-communication	Organizational Studies	Media theory
Brauneis & Goodman (2018)	Data analysis & prediction	USA	Public services	Law	None
Bunt et al. (2012)	None specified	USA	None specified	Computer Science	None
Cech (2020)	Algorithmic mgmt. & control	Austria	Energy	Computer Science	None
Cramer et al. (2008)	Recommendation agents	Netherlands	Entertainment & Recreation	IS	TAM
Criado et al. (2020)	Algorithmic mgmt. & control	Spain	Public services	Political Science	None
de Oliveira Cesar de Moraes et al. (2019)	Recommendation agents	Brazil	None specified	IS	None
Diakopoulos et al. (2014)	Recommendation agents	USA	General Industry	Journalism	None
Diakopoulos & Koliska	Computer-gen.	USA	Media	Journalism	None

Table B2. Categorization of Publications

Article	Application domain	National origin	Industry	Scientific discipline	Theoretical underpinnings
(2017)	content				
Domingo-Ferrer et al. (2019)	Website design	Spain	Finance & Insurance	Computer Science	None
Eslami et al. (2018)	Behavioral advertising	USA	None specified	Computer Science	None
Eslami et al. (2019)	Website design	USA	Entertainment & Recreation	Computer Science	None
Fink (2018)	Regulation	USA	Public services	Communication Science	None
Fuchs et al. (2016)	Decision support systems	Germany	None specified	IS	Principal-agent perspective
Hardin et al. (2017)	Decision support systems	USA	Finance & Insurance	IS	None
Hepenstal et al. (2020)	Decision support systems	UK	Public services	Computer Science	None
Hofeditz et al. (2021)	Computer-gen. content	Germany	Media	IS	None
Jarrahi & Sutherl& (2019)	Algorithmic mgmt. & control	USA	Gig economy	IS	None
Jhaver et al. (2018)	Algorithmic mgmt. & control	USA	Entertainment & Recreation	Computer Science	None
Khovanskaya et al. (2016)	Recommendation agents	USA	Social Media	IS	Human-comp. interaction
Kim & Moon (2021)	Recommendation agents	South Korea	Social Media	Computer Science	None
Kizilcec (2016)	Decision support systems	USA	Public services	Communication Science	Dual process
Lai & Tan (2019)	Decision support systems	USA	None specified	Computer Science	Decision theory
Lee et al. (2019)	Algorithmic mgmt. & control	USA	Public services	Computer Science	Procedural justice theory
Lehmann et al. (2020)	Decision support systems	Germany	None specified	IS	None
Lim & Dey (2011)	Decision support systems	USA	None specified	Computer Science	None
B. Liu & Wei (2021)	Behavioral advertising	USA	None specified	Communication Science	Human-comp. interaction
Lu et al. (2020)	Website design	USA	None specified	Computer Science	Pragmatic th. of explanation
Ochmann et al. (2020)	Recommendation agents	Germany	None specified	IS	None
Ochmann et al. (2021)	Recommendation agents	Germany	None specified	IS	Theory of planned behavior
Rader et al. (2018)	Website design	USA	Social Media	Computer Science	None
Rühr (2020)	Recommendation agents	Germany	Finance & Insurance	IS	Signaling and agency theory

Table B2. Categorization of Publications

Article	Application domain	National origin	Industry	Scientific discipline	Theoretical underpinnings
Salminen et al. (2020)	Computer-gen. content	Finland	None specified	Computer Science	None
Schaffer et al. (2015)	Recommendation agents	USA	None specified	Computer Science	None
Seo et al. (2018)	Data analysis & prediction	USA	Social Media	Computer Science	None
Shin (2020a)	Recommendation agents	South Korea	Media	Communication Science	Expectation confirmation th.
Shin (2020b)	Recommendation agents	South Korea	Entertainment & Recreation	Communication Science	Elaboration likelihood model
Shin & Park (2019)	None specified	South Korea	None specified	Communication Science	None
Shin et al. (2020)	Recommendation agents	South Korea	None specified	Communication Science	Technology acceptance m.
Sinha & Swearingen (2002)	Recommendation agents	USA	Entertainment & Recreation	IS	None
Springer & Whittaker (2019)	Decision support systems	USA	Entertainment & Recreation	Computer Science	None
Springer & Whittaker (2020)	Data analysis & prediction	USA	None specified	Computer Science	Elaboration likelihood model
Wang & Benbasat (2007)	Recommendation agents	Canada	e-Commerce	IS	Attribution theory
Wang & Benbasat (2016)	Recommendation agents	China	e-Commerce	IS	Agency theory
H. J. Watson & Nations (2019)	None specified	USA	None specified	IS	Privacy theories
Yeomans et al. (2019)	Recommendation agents	USA	None specified	Behavioral Science	None

About the Authors

Tobias Bitzer is a PhD candidate at the Chair of Business Information Systems, especially Business Engineering, TU Dresden, Germany. His research centers around algorithmic transparency, the foundations of the concept, its antecedents, and consequences. His work on algorithmic transparency in applications such as contact-tracing app adoption and time management has been published at the *International Conference of Information Systems* and at the *Americas Conference of Information Systems*.

Martin Wiener is a Chaired Professor of Information Systems and Business Engineering at TU Dresden in Germany, as well as an Affiliated Researcher at the University of Erlangen-Nürnberg (Germany) and the Stockholm School of Economics (SSE) Institute for Research (Sweden). His current research focuses on the control and governance of digital transformation projects, algorithmic management (e.g., on online labor platforms), and data-driven organizations. His work has been published in top-tier IS journals, including *Information Systems Research*, *Journal of Management Information Systems*, and *MIS Quarterly*. Martin currently serves as Associate Editor for *Business & Information Systems Engineering*, *Information Systems Journal*, and *Information Systems Research*, as well as on the Editorial Review Board of the *Journal of the Association for Information Systems* and *Information & Management*. He also serves as Co-Program Chair for the 2022 *European Conference on Information Systems (ECIS)* in Timișoara, Romania.

W. Alec Cram is an Associate Professor and PwC Fellow in the School of Accounting and Finance, University of Waterloo, Ontario, Canada. His research focuses on how information systems control initiatives can contribute to improving the performance of organizational processes, including systems development and cybersecurity management. His work has been published in the *Information Systems Journal, Information & Management, European Journal of Information Systems, MIS Quarterly, Communications of the Association for Information Systems, Information Systems Research, Journal of Management Information Systems,* and *Journal of the Association for Information Systems*. He currently serves as an Associate Editor of the *Information Systems Journal*.