American Sociological Association

Social Psychology Quarterly
1–28

DOI: 10.1177/01902725261425997
journals.sagepub.com/home/spq
Sage

How Does AI Disclosure Shape Trust? Unpacking the Role of Legitimacy

Oliver Schilke[1], **and Martin Reimann**[1]

Abstract

As generative artificial intelligence (AI) is increasingly adopted, understanding how its usage is perceived has become crucial for theory and practice. Our investigation highlights how disclosing AI usage reduces trust by triggering legitimacy concerns arising from deviations from taken-for-granted human-centered norms. Drawing on a micro-institutional perspective, we unpack legitimacy into its dimensions and propose that they operate via three context-specific processes—perceived typicality, commitment, and authenticity—that jointly account for the erosion of trust resulting from AI disclosure. An initial structured content-analytic study of directed written interviews reveals that people indeed voice these legitimacy concerns when scrutinizing AI usage and addresses research questions about how such concerns manifest across facets. A subsequent vignette experiment shows that disclosing AI usage sequentially diminishes perceptions of typicality, commitment, and authenticity, ultimately lowering trust. A supplementary replication experiment confirms this pattern. Altogether, our investigation clarifies the paradoxical nature of transparency, advances empirical testing of legitimacy theory, and helps bridge the literatures on trust and institutional theory.

Keywords

disclosure, generative artificial intelligence, legitimacy, trust, vignette experiment

Generative artificial intelligence (AI) is fundamentally reshaping social and economic relationships (Capraro et al. 2024; Lei and Kim 2024; Schenk, Müller, and Keiser 2024). Even though the technology has become mainstream only within the past three years, recent surveys show that 20 percent to 40 percent of Americans are regularly using AI in their work (Crane, Green, and Soto 2025)—a number that will likely continue to grow (McKinsey and Company 2025). Although many people are employing AI without revealing its involvement, others are openly disclosing their use of AI to maintain transparency and ethical integrity (Ali et al. 2024; Fishbowl 2023). Such AI disclosure can backfire, however: recent research has suggested that, rather ironically, it may reduce the trust in the disclosing actor (Schilke and Reimann 2025).

[1]University of Arizona, Tucson, AZ, USA

Corresponding Author:
Oliver Schilke, The University of Arizona, McClelland Hall 405GG, 1130 E. Helen St, Tucson, AZ 85721-0108, USA.
Email: oschilke@arizona.edu

But why? Understanding the social implications of AI disclosure demands closer study of how it conflicts with widely shared expectations of human-centered work. Although prior research has hinted at the role of legitimacy perceptions (Schilke and Reimann 2025), how these perceptions manifest in the context of AI disclosure remains underexplored. To that end, our research uses a micro-institutional approach (Powell 2019; Schilke 2018; Zucker 1977) to unpack the AI disclosure effect in order to shed greater light on the precise mechanisms through which transparency can paradoxically undermine trust. We dissect the legitimacy process into its cognitive, pragmatic, and moral dimensions (Suchman 1995). This examination enables us to uncover how distinct facets of legitimacy jointly shape trust erosion. An initial structured content analysis (Study 1) identifies perceptions of typicality, commitment, and authenticity as context-specific instantiations of legitimacy concerns associated with AI disclosure and quantifies their salience in written interview responses. A subsequent vignette experiment (Study 2) tests our serial mediation model, in which these three processes jointly explain the effect of AI disclosure on trust.

This article makes several contributions. First, our work contributes to the sociological discourse on transparency in technology use (Pasquale 2015; Schudson 2015) by helping clarify the complex dynamics of transparency in AI use, adding greater depth to understanding a paradox where disclosure intended to increase trustworthiness instead reduces it (Schilke and Reimann 2025). This pattern appears puzzling in light of long-standing assumptions that increased openness should foster trust (Schnackenberg and Tomlinson 2016). But rather than bolstering confidence in one's integrity, our study shows how AI disclosure introduces new grounds for skepticism. Unpacking this paradox not only clarifies how transparency can fail to achieve its presumed benefits but also challenges received wisdom in the field about the uniformly positive effects of disclosure.

Second, our investigation heeds calls to enhance the empirical testability of legitimacy theory (Deephouse and Suchman 2008; Haack, Schilke, and Zucker 2021; Schilke, Xue, and Haack 2025). It not only extends legitimacy into the realm of AI technology use and dissects the legitimacy process into cognitive, pragmatic, and moral dimensions but also specifies these concepts by delineating typicality, commitment, and authenticity. These are presented as measurable, context-specific elements that mediate the relationship between AI disclosure and trust erosion. This approach enriches both theoretical perspectives and empirical applicability, making it possible to assess and test legitimacy concerns in concrete ways.

Third, this investigation responds to calls for better integration of two important research streams—trust and institutional theory—that scholars have repeatedly noted as being insufficiently connected (Möllering 2006; Schilke and Cook 2013; Zucker and Schilke 2019). This theoretical synthesis offers a novel framework for examining how institutional norms shape people's trust judgments, particularly in emerging technological contexts.

We begin by briefly reviewing relevant concepts of trust and AI and then develop a theoretical argument for how AI disclosure influences trust by shaping legitimacy perceptions. Next, we present a content-analytic study before turning to experimental results. Finally, we discuss our findings in relation to literatures on transparency, trust, and legitimacy. We then acknowledge study limitations and outline broader implications for sociological research on AI.

TRUST AND AI

Across the social sciences, researchers now widely consider trust—the willingness to make oneself vulnerable to the actions of another party (Mayer, Davis, and Schoorman 1995)—to be a critical ingredient to social exchange (Fehr 2009; Kramer 1999; Robbins 2016; Schilke, Reimann, and Cook 2021). Trust serves as a fundamental mechanism facilitating cooperation, coordination, and the navigation of uncertainty in complex interactions (Arrow 1974; Harkness et al. 2022; Luhmann 1979; Schilke et al. 2015). As a result, trust has drawn substantial attention in organizational studies (Dirks and de Jong 2022; Schilke et al. 2026), and especially over the past two decades, a significant body of scholarship has emerged to examine how trust is shaped by and interacts with modern technologies (Cook et al. 2009; Lumineau, Schilke, and Wang 2023; Schor and Vallas 2021), with AI becoming a prominent focus in the most recent literature (Glikson and Woolley 2020; Tsvetkova et al. 2024). Because workplaces have become digital and many have even recently shifted to being AI-centric, scholars are increasingly focused on how advanced technologies might reshape interpersonal processes at the workplace—a core issue in contemporary trust research (Gkinko and Elbanna 2023; Glikson and Woolley 2020).

According to the widely used definition by the Organisation for Economic Cooperation and Development (2024), AI is "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." As AI becomes increasingly embedded across numerous areas of human activity (Burrell and Fourcade 2021), grasping how trust functions in these interactions grows ever more critical. Recent research has begun to investigate factors influencing whether people develop greater or lesser trust in AI technologies (Lockey and Gillespie 2024). Most existing research focuses on trust in AI itself, but it is equally important to consider how AI matters for trust in the people who deploy it (Elish 2019), the perspective employed in the current research. Here, our focus is on understanding how an individual's decision to reveal their use of AI influences others' trust in them.

Why does this matter, and what are its implications? Trust is a critical cooperation, coordination, and uncertainty-navigation resource (Schilke et al. 2017): when it falls, people share less information, rely more on costly monitoring, discount others' guidance, and hesitate to collaborate more broadly (Dirks and de Jong 2022; Schilke et al. 2021)—this undercuts performance and slows organizational learning and success. Because organizations have started to encourage AI disclosure, with some even mandating it, such as in academic research (Resnik and Hosseini 2026), a systematic disclosure-triggered drop in interpersonal trust also has policy and equity implications: it can discourage transparency in future interactions, incentivize concealment, and selectively penalize compliant actors even when work quality is held constant. As such, as AI becomes ubiquitous, a disclosure-induced trust deficit can blunt core organizational functions (Schilke and Reimann 2025), so the very productivity gains AI promises may be offset by relational frictions. These are reasons underscoring the need to understand and mitigate how AI disclosure lowers trust.

HOW AI DISCLOSURE ERODES TRUST: THE ROLE OF LEGITIMACY

As a baseline, we expect that disclosing AI usage will decrease the level of trust

placed in the actor, consistent with recent empirical findings (Schilke and Reimann 2025). This trust erosion occurs because disclosing AI involvement signals to observers that the actor's work diverges from traditional expectations of human agency (Hancock, Naaman, and Levy 2020). According to micro-institutional theory (Powell 2019; Schilke 2018; Zucker 1977), such deviation from taken-for-granted assumptions prompts concerns about legitimacy—that is, the perception that an entity's actions or decisions are desirable, proper, or appropriate in the given setting (Hawks et al. 2024; Suchman 1995). Put differently, publicly acknowledging the use of AI can appear misaligned with socially recognized norms of how things are done, thus undercutting perceived legitimacy. When someone's behavior casts doubt on their alignment with social standards and upends the expected order, it sets off mental alarms (Harmon 2019) and undermines trust (Kramer 1999). This is because legitimacy acts as a shortcut for assessing whether reliance on this person is safe. When a trustee appears legitimate, trustors anticipate predictable, role-congruent behavior, which lowers perceived relational uncertainty. When a trustee is seen as illegitimate, however, norm deviance increases uncertainty about their motives and behavior and triggers caution—conditions that, in turn, erode trust. In short, perceived (il)legitimacy provides the scaffolding on which trust judgments rest; when AI disclosure shifts legitimacy downward, trust falls.

To provide an illustration, imagine a human resource manager disclosing that she used ChatGPT to evaluate and shortlist job candidates. Although the tool may have helped her work quickly and objectively, some team members might worry that her reliance on AI means she is sidestepping direct human judgment, thereby breaking with familiar hiring norms. As a result, they may perceive her actions as less legitimate and reduce their trust in her decision-making.

Deepening this general argument, we embrace recent recommendations by micro-institutional scholars to provide greater precision by decomposing the broad legitimacy construct into its underlying dimensions (Suddaby, Bitektine, and Haack 2017; Tost 2011), and we use these dimensions to derive context-specific micro-processes that help explain relevant intricacies of the AI disclosure–trust effect. In particular, we leverage Suchman's (1995) foundational typology—distinguishing among cognitive, pragmatic, and moral legitimacy—as a starting point and adapt it to the specific context of AI disclosure, which leads us to pinpoint typicality (cognitive), commitment (pragmatic), and authenticity (moral) as relevant manifestations.[1] Importantly, we certainly do not claim that these mappings are universal, and different representations may be salient in other contexts. Theoretical adaptation (Cohen 1989) to an investigation's substantive focus is crucial given that legitimacy processes are highly context-specific (Suddaby et al. 2017). This approach is consistent with prior research that posits that legitimacy

[1]Beyond Suchman's (1995) cognitive, pragmatic, and moral dimensions that dominate in institutional research, some social-psychological work (e.g., Tyler 1997; Tyler and Lind 1992) specifies a relational basis of legitimacy that centers on whether a target accords the evaluator respect, dignity, and identity affirmation (for a detailed discussion, see Tost 2011). In our application, modeling a separate relational mediator would be conceptually overlapping with the trustworthiness facets examined in our supplemental study (see the Online Supplement), which capture audiences' relational evaluations of the actor. To preserve parsimony and avoid overlap, we therefore bracketed relational legitimacy in the causal chain we test and return to it in the "Discussion" as an important extension.

and its dimensions are inherently latent, serving as a metaframework that allows for pinpointing more context-specific constructs (Hamm, Trinkner, and Carr 2017). As such, rather than applying (overly) generic scales for "cognitive/pragmatic/moral legitimacy," our approach specifies the concrete constructs that are actually made salient along those legitimacy dimensions (Study 1 was designed precisely to assess which specific concerns people have under each dimension in this setting). In what follows, we discuss each type of legitimacy in general terms; develop an argument that typicality, commitment, and authenticity represent key elements of these dimensions that are salient in the context of AI disclosure; and link each to trust. Throughout this discussion, we treat cognitive, pragmatic, and moral legitimacy as analytical lenses rather than mutually exclusive bins; real-world judgments often engage multiple lenses, so some overlap is inevitable. Our mapping therefore identifies the dominant diagnostic for each lens in this setting (without claiming one-to-one exclusivity):

Typicality → Cognitive
Commitment → Pragmatic
Authenticity → Moral.

Typicality

Following Suchman (1995), cognitive legitimacy is the extent to which observers see an actor's activities as understandable, normal, and taken-for-granted. Practices that are familiar and standard are viewed as the natural way of doing things (Hawks 2025; Schilke and Rossman 2018; Zucker 1983). Typicality—defined as the extent to which a behavior is characteristic of a particular category—is at the core of the cognitive legitimacy dimension. The more typical an activity is perceived to be, the higher is its cognitive legitimacy (Tolbert and Zucker 1983). A typical activity matches established cognitive patterns and is thus less likely to challenge the mental schemas of evaluators (DiMaggio 1997; Garfinkel 1963), thereby enhancing cognitive legitimacy. Conversely, actions that deviate from typical behavior violate cultural or legal expectations and elicit cognitive legitimacy challenges (Deephouse and Carter 2005).

Disclosing the usage of AI can alter perceptions of how typical a person's work practices are because reliance on AI might be seen as a departure from the norm in which the predominant practice is based on human expertise. Such disclosure may thus reduce perceptions of typicality because AI represents a nontraditional approach that challenges familiar social and organizational norms. The extent to which someone's practices fail to align with what is typical will inform observers' trust judgments. Actions deviating from convention are more likely to be perceived as unpredictable, thereby undermining trust (Nowak et al. 2023). Conversely, actions viewed as typical are more predictable and thus gain trust (Zucker 1986). These arguments lead us to believe that typicality mediates the negative effect of AI disclosure on trust.

Commitment

Next, we adapt Suchman's (1995) idea of pragmatic legitimacy to our context. Pragmatic legitimacy involves an instrumental perspective in which legitimacy is granted because the evaluator sees the focal actor as useful. It is grounded in the perception that an individual is responsive to the constituency's larger interests and thus valuable (Tost 2011). Pragmatic legitimacy is thus exchange-based: audiences ask whether the actor will advance their interests—that is, do the job for us. Commitment—defined as

one's psychological attachment, dedication, and sense of responsibility toward one's job (Eisenberger, Fasolo, and Davis-LaMastro 1990)—is a core element signaling an individual's motivation to respond to and align with stakeholders' interests (Kim 2025; Meyer and Herscovitch 2001). When an evaluator depends on an actor's work outputs, the degree to which the evaluator perceives this actor to be committed to their work is a key signal of the actor's ability to deliver ongoing benefits and thus an important component of pragmatic legitimacy (Cooper-Hakim and Viswesvaran 2005; Katz and Kahn 1978). Effort and responsibility—key ingredients to commitment—are especially diagnostic here because they are costly, observable investments of time, attention, and diligence to stakeholders' ends, thus credible signals of responsiveness.[2]

In line with this reasoning, AI disclosure is likely to diminish perceptions of commitment by signaling a shift away from human effort, accountability, and dedication in key organizational tasks. People delegating significant portions of their tasks to a machine can be interpreted as "not doing their work" and attempting to "take the easy way out" rather than putting in effort themselves (Fügener et al. 2022). Commitment to work performance, in turn, acts as a crucial trust ingredient, assuring the evaluator of the actor's dedication in delivering consistent results over time (Sitkin and Roth 1993). This perceived reliability fosters confidence that the trustee will not only meet immediate expectations but also continue to prioritize the evaluator's long-term interests, thereby enhancing the stability of the relationship (Kramer 1999). Conversely, if an individual is perceived as lacking commitment, an evaluator may view them as unaligned with their goals, leading to diminished trust. Commitment should thus act as another mediator in the AI disclosure–trust effect.

Authenticity

Finally, Suchman's (1995) framework offers a third dimension—moral legitimacy—that we apply to our focal context. At a general level, moral legitimacy concerns the normative approval of an actor or activity based on a judgment that certain actions adhere to widely accepted values, are the right things to do, and/or are perceived as sincere (Suchman 1995; Tost 2011). We argue that in the specific context of AI disclosure, authenticity serves as an important manifestation of moral legitimacy. Compared to other aspects of legitimacy, "it is the moral dimension that distinguishes authenticity" (Reilly 2018:936). Because authenticity reflects sincerity, honesty, and value-congruent conduct in role—that is, being genuine rather than performative (Barasch et al. 2014; Gershon and Smith 2020)—it is a critical basis for moral approval (Brown and Toyoki 2013; Lim and Zhang 2025; Sidani and Rowe 2018).[3] From a moral perspective, observers often treat AI involvement primarily as an authenticity problem: algorithmic work is

[2]By contrast, competence/capability under AI assistance is ambiguous: disclosure can be read as augmentation (capability up) or as overreliance/deskilling (capability down), and disclosure introduces credit assignment ambiguity about whether performance stems from the human or the tool. In this setting, commitment therefore provides a cleaner, less noisy mapping to the instrumental question of pragmatic legitimacy than capability does.

[3]In our study, we treat the precise content of the actor's true values as exogenous. That is, such values may in principle not align with those of the evaluator. Nonetheless, inauthenticity undermines moral approval independent from whether or not the actor's values themselves are perceived as objectionable (Simons 2002). Some research even suggests that perceptions of inauthenticity can weigh more heavily in moral evaluations than disagreement with values (Jordan et al. 2017).

judged less morally authentic—less sincere and person-grounded—than equivalent human work such that AI disclosure depresses perceived authenticity with downstream negative reactions (Brüns and Meißner 2024; Jago 2019).

In the AI-usage setting in particular, a salient focus of evaluators' moral assessments lies in whether outputs still reflect the actor's own judgment and accountability (Jago 2019; Kirk and Givi 2025). When people disclose their usage of AI, such disclosure calls into question whether their work outcomes are truly representative of their human judgment (Jago 2019; Reimann and Kronrod 2026). The use of AI can suggest a detachment from the personal, subjective decision-making processes central to authenticity perceptions, making the work appear more mechanical and less reflective of the individual's unique perspective and values (Kirk and Givi 2025). This doubt may lead audiences to view the disclosing individual and their work as lacking authenticity. Perceived authenticity, in turn, is an influential component in trust judgments because it suggests that an individual is acting in alignment with their true self and values, which fosters confidence in their trustworthiness. When people are seen as authentic, they are more likely to be trusted because authenticity reduces uncertainty and signals consistency in behavior (Nguyen et al. 2022). Conversely, a lack of perceived authenticity, particularly in contexts such as AI disclosure, can raise doubts about personal engagement and sincerity, thereby undermining trust (Kim et al. 2022). Thus, we expect authenticity to mediate the AI disclosure–trust effect.

Serial Mediation

So far, we have treated each of the three facets of legitimacy in isolation. Recent micro-institutional theorizing suggests, however, a staged model of legitimacy construal whereby more intuitive cognitive aspects are processed first, followed by the more deliberate pragmatic and moral dimensions (Bitektine 2011; Bitektine and Haack 2015; Haack, Pfarrer, and Scherer 2014). This idea of serial legitimacy construction from micro-institutional theory aligns with dual-process theories, which propose that System 1 (intuition) initiates processing before System 2 (deliberation) takes over (Diederich and Trueblood 2018; Kahneman 2011; Tost 2011). Building on this logic, we anticipate that typicality, as our focal manifestation of cognitive legitimacy, will be the mediator most immediately engaged by AI disclosure, in turn shaping perceptions of commitment and authenticity (representing pragmatic and moral legitimacy) and ultimately, trust. That is, AI disclosure will trigger intuitions that the focal actor's usage of AI is atypical. Such typicality judgments involve a swift comparison of an actor to a preexisting mental prototype (Rosch 1975; Voorspoels, Vanpaelmel, and Storms 2008), a process that can occur relatively quickly and without much conscious reflection (Devine 1989). In turn, these more immediate perceptions of atypicality can further provoke deliberate evaluations concerning the commitment and authenticity of the actor. Evaluating someone's commitment to their work requires active, multifaceted sensemaking and entails cognitive effort and time to interpret, weigh, and reconcile the various signals that may indicate someone's devotion (Shore, Barksdale, and Shore 1995). Similarly, authenticity judgments involve a comparatively complex and deliberative evaluation, requiring the assessment of whether actions genuinely reflect a person's beliefs by examining unobservable internal states and motives and integrating multiple cues (Beverland

and Farrelly 2009). This is why we propose that swift typicality assessments will causally precede the more deliberate evaluations of commitment and authenticity.

In further support of our proposed sequence, Tost's (2011) dual-process account of legitimacy judgments suggests that audiences typically begin in a low-effort "passive" mode that relies on taken-for-grantedness cues and conserves cognitive resources rather than actively appraising pragmatic or moral qualities. When a practice violates cultural expectations—that is, when cognitive legitimacy (typicality) is low—this discrepancy sounds the alarm and shifts observers from the use stage into a judgment reassessment stage in which effortful evaluation replaces heuristics. In that active evaluative mode, observers invest time and effort to assess the entity along pragmatic and moral dimensions to form a judgment. This sequence maps directly onto our serial mediation: AI disclosure first depresses typicality (cognitive), which then prompts more effortful scrutiny of commitment (pragmatic) and authenticity (moral), yielding downstream effects on trust. In combination, these arguments lead us to expect the following pattern of serial mediation:

Hypothesis 1: The negative effect of AI disclosure on trust is mediated by perceived (1) typicality, followed concurrently by (2) commitment and (3) authenticity.

Figure 1 illustrates this mediation model.

DOES LEGITIMACY, OR TRUSTWORTHINESS, OR BOTH EXPLAIN THE EFFECT OF AI DISCLOSURE ON TRUST?

Because of the centrality of the construct of perceived trustworthiness in the trust literature (Colquitt, Scott, and LePine 2007; Mayer et al. 1995; Schilke and Cook 2015), it is imperative to clarify that legitimacy and trustworthiness are related yet conceptually distinct constructs, as perhaps most explicitly discussed by Bitektine, Gillespie, and Lange (2026). Whereas legitimacy addresses the question of whether I find a counterpart socially acceptable, trustworthiness pertains to whether I think that counterpart will not take advantage of me or let me down. In terms of points of comparison, legitimacy involves comparing an actor with a set of social norms (Tost 2011), whereas trustworthiness involves comparing an actor with characteristics that may be predictive of their reliable behavior—notably, ability, benevolence, and integrity (Mayer et al. 1995). Therefore, legitimacy is a generalized judgment about contextual fit—entities are judged to be legitimate when they are seen as appropriate for their social context (Suchman 1995; Tost 2011)—rather than a trait-based appraisal of a person's qualities.

In terms of causal order, recent micro-institutional research indicates that legitimacy precedes trustworthiness judgments (Chen et al. 2022; Lamertz and Bhave 2017) such that positive assessments of legitimacy suggest that a counterpart adheres to relevant social norms and is thus unlikely to act in a nontrustworthy fashion. From this perspective, legitimacy signals the partner's appropriateness for interactions involving risk. To further examine our position that both legitimacy and trustworthiness, along with their subfacets, play a role in the effect of AI disclosure on trust, we ran another supplementary experiment, briefly discussed under Study 2.[4]

[4]Reported as Supplementary Study in the Online Supplement.

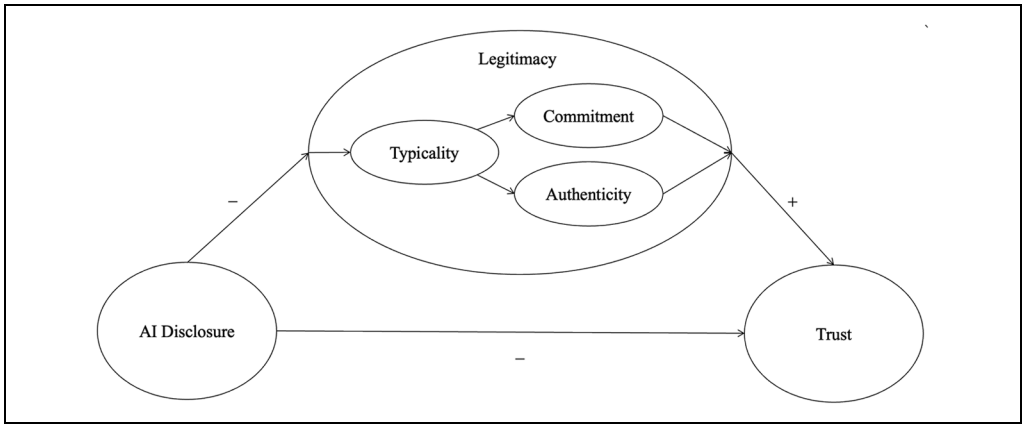

Figure 1. Mediation Model: Legitimacy Mediates the Negative Effect of Artificial Intelligence (AI) Disclosure on Trust Such That Disclosing (vs. Not Disclosing) the Usage of AI for Work Tasks Reduces Perceptions of Legitimacy, Thereby Eroding Trust

METHODS

Study Overview

This article offers results from a structured content-analytic study (Study 1) and a vignette experiment (Study 2). Study 1's purpose is to ground the theorized mapping by analyzing written interview responses using a theory-guided, directed content analysis with dictionary-assisted coding and quantitative comparisons. Consistent with our context-first approach, Study 1 aims to empirically ground the mapping from broad legitimacy dimensions to concrete concerns in this setting rather than assume that generic legitimacy scales would suffice. A content-analytic approach is particularly well suited to establishing construct specification (Elo and Kyngäs 2008; Schreier 2012)—that is, verifying that one focuses on constructs that appropriately represent broader categories prior to testing relations among them. In our content analysis, we examine how people assess the legitimacy of AI disclosure in work tasks and validate the three-dimensional structure suggested by our hypothesis. For this purpose, we analyze written interview responses using directed content analysis with dictionary-assisted coding, manual verification, and quantitative comparisons.

Study 2 then provides a test of the relationships proposed in this serial mediation hypothesis. We employ the experimental method, which can be particularly useful for examining mediators explaining a main effect (e.g., Levine et al. 2023) and has been frequently employed to study antecedents of trust (Schilke, Powell, and Schweitzer 2023). In our experiment, participants took on the role of a trustor charged with assessing another actor who either disclosed or did not disclose AI usage. Together, these two studies complement each other by first contextualizing and validating the mapping of constructs and then testing relationships among them. This sequencing combines rich insight into legitimacy construal with causal evidence, offering a holistic view of the phenomenon (Small 2011).

STUDY 1: STRUCTURED CONTENT ANALYSIS OF WRITTEN INTERVIEWS

The aim of Study 1, a structured (directed) content analysis of written interviews,

was to contextualize and validate how legitimacy concerns manifest in the AI disclosure setting. More precisely, Study 1 addresses three research questions:

Research Question 1a (cognitive): Do participants' elaborations of cognitive legitimacy invoke typicality?
Research Question 1b (pragmatic): Do elaborations of pragmatic legitimacy invoke commitment?
Research Question 1c (moral): Do elaborations of moral legitimacy invoke authenticity?

These questions guided our coding and summary comparisons.

In line with the deliberation-design approach advocated by Haack et al. (2021) for investigating the intricacies of legitimacy, we sought to capture detailed personal accounts of how people think about the appropriateness of disclosed AI usage. We were particularly interested in the extent to which participants would express concerns about typicality, commitment, and authenticity when queried about cognitive, pragmatic, and moral legitimacy. For example, if participants brought up concerns about authenticity when elaborating moral legitimacy—defined to them as the extent to which someone's actions are the right things to do and sincere—it would support our argument that in the context of AI disclosure, moral legitimacy can be effectively represented through perceptions of authenticity. This phrasing ("right things to do and sincere") operationalizes moral legitimacy in terms of sincerity/integrity, consistent with our use of authenticity as its context-specific indicator.

Participants and Design

One hundred thirty individuals were recruited to participate in structured written interviews (e.g., Raclaw, Barchas-Lichtenstein, and Bajuniemi 2020). Written interviews allow participants time to reflect and craft in-depth responses while also minimizing interviewer presence or reaction cues, reducing social-desirability biases and providing a sense of privacy that can foster honesty and candor (James 2016; Salmons 2014). We recruited participants from the United States using the CloudResearch Connect platform (Litman, Robinson, and Abberbock 2017) in exchange for monetary compensation. The CloudResearch platform facilitates the recruitment of panelists who have shown prior evidence of attention and engagement and have not been associated with suspicious locations or duplicate IP addresses, thus addressing potential concerns regarding Amazon Mechanical Turk data quality (Eyal et al. 2022; Hauser et al. 2023). Although collecting data online may sacrifice a certain degree of closeness to the participants, it makes data collection significantly more affordable, fast, and efficient (Bitektine, Lucas, and Schilke 2018) and allows for involving a population with higher demographic diversity than a student sample (Weinberg, Freese, and McElhattan 2014). In particular, Weinberg et al.'s (2014) analyses suggest that crowdsource-recruited samples produce results that are substantively similar to representative samples. Moreover, the primary focus of our research is on isolating causal mechanisms in a controlled scenario (rather than making population-level inferences), a context in which a nonrepresentative sample is commonly considered acceptable for theory-testing purposes (Auspurg and Hinz 2014). Table 1 summarizes the sociodemographic characteristics for both studies. We employed a directed content-analytic approach anchored in our theory, eliciting open-ended text and then coding responses with dictionary-

Table 1. Descriptive Statistics of Sociodemographic Variables across Both Studies

Study	Sex assigned at birth (percent female)	Age (average years)	Race (percent White)	Education (median on scale from 1 = less than high school to 5 = graduate college degree)	Income (median on scale from 1 = less than $10,000 to 11 = $100,000 or more)	Work experience (average years)
1	59.21%	36.04	42.31%	Undergraduate college degree	$60,000–$69,999	14.55
2	60.64%	36.19	73.40%	Undergraduate college degree	$70,000–$79,999	16.04

Note: In Study 1, sociodemographic information was recorded only for those 76 participants who found their acquaintance's behavior inappropriate and thus qualified for the rest of the study.

assisted procedures followed by quantitative scoring for each focal construct.

Interviewees were first provided with a consent form and then asked to imagine an acquaintance of theirs who works as a human resource manager at a large company.[5] They were told that she had disclosed to them that she recently started using ChatGPT to screen job applicants and shortlist candidates. Consistent with our directed interview approach, all respondents then received the same set of questions in a structured survey with open-ended response fields to fill in their thoughts (Patton 2015; Turner 2010). They were asked whether they would have any concerns about the social appropriateness of her disclosure and if so, what concerns they would have. Next, interviewees were queried on three facets of legitimacy—cognitive, pragmatic, and moral. Each facet was defined, and participants were then asked to explain how their acquaintance's disclosure might violate their perceptions of each.

Our primary aim in Study 1 was not to test whether legitimacy arises spontaneously but to map the context-specific content of each legitimacy dimension via directed free-text questions (Hsieh et al. 2018; Merton, Fiske, and Kendall 1956; Turner 2010). Accordingly, our prompts featured brief, standard definitions of cognitive, pragmatic, and moral legitimacy and invited elaboration for each facet. This design balances (a) capturing top-of-mind reactions and (b) enabling within-person comparisons across the three theoretically specified facets.

[5]Because all study data were recorded in completely anonymized form, we did not need to change or exclude any evidence to protect subjects from identification or distress. We also note that participants were not given the opportunity to review or comment on drafts of the article.

Question Prompts

We used the following prompts to query participants: "Imagine an acquaintance of yours, who works as a human resource manager at a large company, tells you that she has recently started using ChatGPT to screen job applicants and shortlist candidates. Would you have any concerns about the social appropriateness of her disclosure?" (answer choices: yes/no). In case participants chose "yes," they were asked: "Please elaborate. In what ways do you find her disclosure socially inappropriate?" (participants entered responses into an open text field). We then stated:

> Thanks for sharing your assessment! According to academic research, there are three different facets of social appropriateness. One is called "cognitive legitimacy," which denotes the extent to which someone's activities are simply accepted as a given and seen as common practice. In what ways may your acquaintance's disclosure of using ChatGPT for hiring decisions violate your perception of cognitive legitimacy? (open text field entry).

Then, we asked: "A second facet of social appropriateness is 'pragmatic legitimacy'—the extent to which someone is seen as useful and doing their job properly. In what ways may your acquaintance's disclosure of using ChatGPT for hiring decisions violate your perception of pragmatic legitimacy?" (open text field entry). This was followed by:

> Finally, the third facet of social appropriateness is "moral legitimacy"—the extent to which someone's actions are the right things to do and sincere. In what ways may your acquaintance's disclosure of using ChatGPT for hiring decisions violate your perception of moral legitimacy? Thank you for helping us breaking down your appropriateness perceptions! (open text field entry).[6]

The survey concluded with demographic items, questions about the study purpose and other background variables, and a debriefing.

Content Analysis and Scoring

We conducted a structured content analysis of responses from those 76 participants who reported appropriateness concerns in response to the initial screening question. We read each response to identify representative quotes that illustrate common themes (Roller and Lavrakas 2015). Following the approach of Braun and Clarke (2006), the interview data were then systematically coded, categorized, and condensed. We implemented a directed content-analysis pipeline in which a large language model (LLM; ChatGPT; OpenAI 2023) served as primary coder, consistent with both guidance on AI-assisted qualitative analysis and mixed-methods text coding (Grimmer and Stewart 2013; Hsieh and Shannon 2005; Morgan 2023; Wachinger et al. 2024). Starting from theory-driven seed terms for each construct (typicality, commitment, authenticity), we prompted the model to (a) expand each seed into a semantically proximate keyword list and (b) apply those dictionaries to every free-text response. For each Response × Construct, the model returned a numeric score on a range from 1 to 10. To ensure interpretive validity, a human audit followed: the authors reviewed the keywords, their assigned values, and participant responses, and their review confirmed that the model's scores and

[6]We subsequently also asked several other open-ended questions, which we ultimately did not use for this article but are included in the Qualtrics survey posted on OSF.

keyword matches reflected the intended meaning. Beyond this sanity check, no qualitative recoding was performed by humans; the LLM outputs formed the data set for further analyses, leveraging evidence that LLMs provide high-quality text coding (Gilardi, Alizadeh, and Kubli 2023). Finally, we ran t tests for comparisons to address Research Questions 1a through 1c (i.e., which construct was most prevalent within each legitimacy facet). The study materials, keyword list used for coding, data, and syntax are posted on the Open Science Framework (OSF).[7]

Results

Respondents' elaborations of the legitimacy dimensions mapped onto the three proposed instantiations. For cognitive legitimacy, a dominant theme was lack of typicality and taken-for-grantedness. For example, respondents thought that AI usage is "not yet seen as expected," "stands out as unusual," and "isn't something that is common practice." Pragmatic legitimacy concerns, furthermore, centered on perceived commitment, effort, and duty: participants worried the manager was "not paying attention herself," "outsourcing her job to an unreliable piece of technology," and "taking the easy route," with risks of missed job candidates and errors. Moreover, moral legitimacy comments foregrounded authenticity, such as concerns that the behavior would lead the manager "not to see the real person beneath"; respondents also imagined it was possible that "someone would assume they are talking to a real person when in reality it would not be." Respondents were worried that "relying on AI in such a significant human process could be seen as insincere" and would require "clarity about accountability" by questioning who is really responsible in the hiring decision. To illustrate content-analytic themes of the free-text responses to each of the three legitimacy questions and how they relate to typicality, commitment, and authenticity, respectively, Table 2 lists five representative quotes for each.

We then constructed typicality, commitment, and authenticity scores for each of the three free-text responses participants provided when elaborating on their cognitive, pragmatic, and moral legitimacy concerns. As shown in Table 3, the pattern of scores reveals that concerns about typicality are most prevalent in cognitive legitimacy, commitment in pragmatic legitimacy, and authenticity in moral legitimacy assessments. A series of paired t tests provided further evidence. For example, in the responses to the cognitive legitimacy prompt, the typicality score ($M = 5.63$, $SD = 3.50$) was significantly larger than both the commitment score ($M = 3.36$, $SD = 2.76$; $t[75] = 4.42$, $p < .001$, $d = .51$) and the authenticity score ($M = 3.01$, $SD = 2.31$; $t[75] = 5.31$, $p < .001$, $d = .61$). These results support our proposed alignment of the three concepts with their respective types of legitimacy. That is, the identified patterns are consistent with Research Questions 1a through 1c: concerns about typicality were most prevalent when elaborating cognitive legitimacy, commitment when elaborating pragmatic legitimacy, and authenticity when elaborating moral legitimacy. We interpret this as content-analytic construct validation of our mapping.

Taken together, the written interviews provided useful insights that helped further unpack how the dimensions of legitimacy manifest in the context of AI disclosure. Interviewees' detailed accounts highlighted the centrality of typicality, commitment, and authenticity, suggesting that these constructs are useful

[7] https://osf.io/wbj98/.

Table 2. Representative Quotes Extracted from the Free-Text Responses

In what ways may your acquaintance's disclosure of using ChatGPT for hiring decisions violate your perception of cognitive legitimacy?
"I think it violates cognitive legitimacy because using AI in this manner is not yet seen as expected, so it stands out as unusual and perhaps not an acceptable practice."
"The use of artificial intelligence is still in its infancy. As such it hasn't entered into the domain of being considered common practice."
"It isn't something that is common practice."
"The use of ChatGPT isn't normal or common practice."
"My acquaintance's use of ChatGPT for hiring decisions might violate your sense of cognitive legitimacy because it's not something that is typically seen as normal or accepted in hiring practices."
In what ways may your acquaintance's disclosure of using ChatGPT for hiring decisions violate your perception of pragmatic legitimacy?
"I don't think this counts as doing her job properly, because it means that she is not paying attention herself to all of the candidates but instead letting many get screened out before they even get to her."
"They'd definitely seem like they were doing less work than everyone else. It seems lazy and like it would be prone to errors compared to sorting them normally."
"She's outsourcing her job to an unreliable piece of technology."
"Then they would not be doing their job properly and the way it is stated in their job description."
"They wouldn't be doing their job correctly and could be seen as taking the easy route. They aren't doing their due diligence."
In what ways may your acquaintance's disclosure of using ChatGPT for hiring decisions violate your perception of moral legitimacy?
"Additionally, relying on AI in such a significant human process could be seen as insincere."
"I definitely think that moral legitimacy is the principle most challenged by my acquaintance's disclosure, because to me, the process of recruitment is a personal one, in which sincerity plays a major role. To me, using AI denies and negates that sincerity, instead replacing it with coldness and indifference."
"It is against morals that a persons [*sic*] job fate is determined by AI and looking at numbers. Not to see the real person beneath."
"Because someone would assume they are talking to a real person when in reality it would not be."
"Moral legitimacy requires clarity about accountability."

representations of the broader concepts of cognitive, pragmatic, and moral legitimacy. Study 1 thus builds additional theory around our premise that legitimacy concerns are at the very core of the AI disclosure effect. Its findings validate the context-specific instantiations used to represent legitimacy and thereby justify proceeding to the experimental test of the serial process in Study 2.

To further clarify, Study 1 was designed—per the deliberation-design approach (Haack et al. 2021)—to illuminate how legitimacy concerns are articulated when they arise rather than to estimate their prevalence. The initial yes/no screen of respondents who were (vs. were not) concerned about AI usage at work, therefore, served to identify information-rich cases for thematic analysis. Focusing on the 76 respondents who reported concerns allowed us to unpack the context-specific meanings participants attached to cognitive, pragmatic, and moral legitimacy. Nonetheless, future work should further probe

Table 3. Means of Content-Analytic Indices by Legitimacy Prompt

Coded concept	Cognitive legitimacy prompt	Pragmatic legitimacy prompt	Moral legitimacy prompt
Typicality	5.63(3.50)	2.28(1.37)	2.13(.87)
Commitment	3.36(2.76)	6.80(3.21)	3.33(2.71)
Authenticity	3.01(2.31)	2.30(1.25)	7.89(2.57)

Note: $N = 76$. Standard deviations are in parentheses.

those respondents who have no concerns about AI disclosure and compare their reasoning with the reasoning of those who do.

Moreover, we acknowledge that Study 1's scenario about the hiring manager's AI use features a bounded, one-off managerial communication instance rather than a behavior that repeatedly shapes current employees' day-to-day experience. If participants perceived this as lower in ongoing consequences, our study design likely constitutes a conservative lower-bound estimate on trust penalties stemming from AI disclosure: as stakes or repetition increase, we anticipate that atypicality becomes more diagnostic and expectations of commitment and authenticity intensify, which should reduce trust. Future research should orthogonally manipulate stakes (low vs. high) and repetition (single vs. repeated) and test whether the effects change as a result.

STUDY 2: VIGNETTE EXPERIMENT

Building on Study 1's mapping that established the context-specific salience of typicality, commitment, and authenticity in their respective legitimacy facets, Study 2 provides a pre-registered causal test of Hypothesis 1, according to which AI disclosure lowers typicality, which in turn shapes commitment and authenticity, ultimately affecting trust. The vignette methodology is particularly suitable for this purpose because it allows controlled investigation of social perceptions in realistic social contexts (Finch 1987), making it one of the most frequently employed approaches in the social sciences (Wallander 2009). Experiments enable us to separate AI disclosure from AI use (Schaap, Bosse, and Hendriks Vettehen 2024) by keeping the actual deployment of AI constant while systematically varying whether its use is communicated, allowing us to clearly examine disclosure effects on trust. The anonymized preregistration, study materials, data, syntax, and manipulation check results are posted on OSF,[8] and all administered measures, conditions, and data exclusions and the determination of our sample sizes have been reported (Nosek et al. 2013).

Participants and Design

Ninety-four individuals were recruited to participate via CloudResearch Connect and randomly assigned to one of two conditions in a one-factor between-subjects design with two levels (AI disclosure, no disclosure control). Descriptive sociodemographic information on the sample is provided in Table 1.

Experimental Procedure

In the study's vignette, participants were asked to picture themselves employed at a warehouse company called Dock

[8]https://osf.io/89xgu/.

Logistics Storage & Fulfillment, Inc. They were told that as warehouse team members, their role was important to maintaining smooth operations—from coordinating incoming shipments to dispatching orders. After reviewing a short job description, participants were shown a notice of termination from the warehouse's managing director, Alexander Vanderberg, to a coworker named Zac Mayers. Subsequently, participants either saw that the message was drafted using ChatGPT (AI disclosure condition) or observed a simple "loading wheel"—that is, a standard animated spinning on-screen indicator, commonly used in computer interfaces to signal that a system is processing something—with no mention of AI (no disclosure control condition).

Next, participants responded to a five-item measure of the managing director's typicality (e.g., "How typical are the managing director's work practices for this kind of setting?"; $\alpha = .95$; Johnston and Hewstone 1992), a three-item measure of the managing director's authenticity (e.g., "How authentic is the managing director?"; $\alpha = .96$; Gershon and Smith 2020), and a five-item measure of the managing director's commitment (e.g., "What level of commitment does the managing director have toward their job?"; $\alpha = .96$; Perrewe, Fernandez, and Morton 1993). Participants were then asked to respond to four items about the degree to which they trusted the managing director (e.g., "I would be comfortable giving the managing director a task or problem, which was critical to me, even if I could not monitor his actions"; $\alpha = .86$; Mayer and Davis 1999). When the items for all four variables (i.e., trust, typicality, commitment, and authenticity) were submitted to an exploratory factor analysis with iterated principal factors and orthogonal varimax rotation, we found four distinct factors along the items that are specific to each variable. This structure supports the conceptual separation among typicality, commitment, and authenticity.

Results

In support of the baseline main effect of AI disclosure on trust, an independent samples t test revealed that participants trusted the managing director less when he disclosed preparing the letter using generative AI ($M = 2.38$, $SD = .96$) compared to when he made no such disclosure ($M = 2.81$, $SD = .93$; $t[92] = 2.21$, $p = .03$, $d = .46$). In support of our serial mediation model, when entering all three mediators in one serial mediation model (PROCESS Model 81 with 5,000 bootstrapped samples), we found that the negative effect of disclosure on trust is mediated by typicality, followed by commitment and authenticity (Figure 2).[9] Consistent with our preregistered analyses and with pertinent guidance cautioning against indiscriminate use of controls in randomized experiments (Atinc, Simmering, and Kroll 2012; Bernerth and Aguinis 2016; Spector and Brannick 2011), we did not include demographics in our default specification of the mediation model. A covariate-adjusted robustness analysis—documented in the SPSS

[9]Three separately run mediation analyses using the PROCESS macro (Model 4) with 5,000 bootstrapped samples (Hayes 2022) revealed that perceptions of the managing director's typicality ($ab = -.37$, SE = .12, 95% confidence interval [CI] = [−.61, −.16]), authenticity ($ab = -.52$, SE = .13, 95% CI = [−.77, −.27]), and commitment ($ab = -.63$, SE = .13, 95% CI = [−.89, −.37]) each mediated the effect of AI disclosure on trust. In an exploratory post hoc analysis, furthermore, we modeled the three mediators in parallel rather than in the hypothesized sequence. In this model, the second-stage effect of typicality on trust was notably weak ($b = .02$, SE = .05, $p = .71$), suggesting that this effect is mediated by commitment and authenticity, consistent with our proposed model.

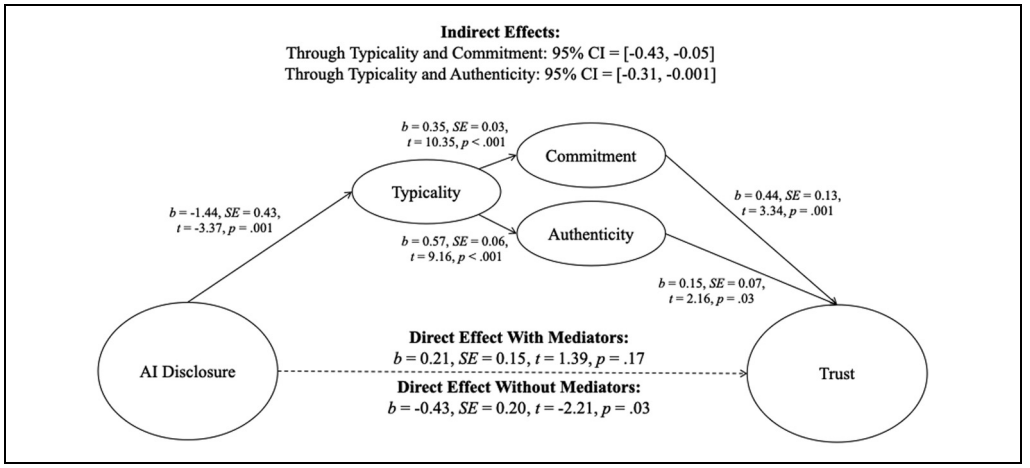

Figure 2. The Negative Effect of Artificial Intelligence (AI) Disclosure on Trust Is Sequentially Mediated by Typicality, Followed by Commitment and Authenticity, Such That Disclosing the Usage of AI for Work Tasks Diminishes Typicality Perception, Which in Turn Informs Both Commitment and Authenticity Perceptions, Ultimately Eroding Trust

syntax file on OSF—revealed, however, that the regression results remained virtually unchanged.

We also conducted a replication study that reproduces the design of Study 2 while adding a perceived trustworthiness measure. This additional study provides convergent evidence for the proposed mediation structure.[10] The supplementary study further shows that the three legitimacy subfacets adopted in this research are empirically distinct from trustworthiness dimensions (ability, benevolence, integrity).

DISCUSSION

The starting problem of this investigation is a practical and theoretical puzzle: transparency about AI use can paradoxically reduce trust in the disclosing actor. We argue that AI disclosure violates taken-for-granted expectations of human-centered work, thereby inviting legitimacy concerns. Our argument specifies how this unfolds: legitimacy concerns become concrete through three context-specific processes—typicality (cognitive), commitment (pragmatic), and authenticity (moral)—that operate in a serial fashion as people move from fast pattern matching to more deliberative evaluation.

Study 1 employed a structured content analysis of written interviews to examine three related research questions regarding the legitimacy dimensions and their representations in the study context: When assessing AI disclosure, do evaluators (a) chiefly invoke typicality when elaborating cognitive legitimacy, (b) commitment when elaborating pragmatic legitimacy, and (c) authenticity when elaborating moral legitimacy? Results are consistent with these research questions: typicality is most salient under cognitive legitimacy prompts, commitment under pragmatic prompts, and authenticity under moral prompts. This provides construct validation that the three processes we theorized are indeed the ones people mobilize when assessing AI disclosure.

Study 2 then tested Hypothesis 1, a preregistered serial mediation model, in which AI disclosure reduces perceived

[10]The Supplementary Study in the Online Supplement provides convergent evidence for the proposed mediation structure.

typicality, which in turn lowers commitment and authenticity and ultimately decreases trust. Findings of Study 2 support this hypothesized mediation structure. Taken together, the two studies translate our theoretical logic into a coherent empirical sequence: Study 1 provides content-analytic construct validation of the legitimacy concerns salient in our context, and Study 2 demonstrates their serial pathway from AI disclosure to lower trust. Stated most directly, Study 1 validates the constructs; Study 2 tests the theorized process among them.

Trust serves as our focal outcome, conceptually defined as a willingness to be vulnerable to another's actions. In our design, it is the result of legitimacy-driven social evaluations about a human actor who discloses AI use. The findings in the Online Supplement help situate trust within a broader evaluation sequence by showing that trustworthiness (ability, benevolence, integrity) can be modeled as a downstream mediator following commitment and authenticity and preceding trust; in other words, legitimacy perceptions shape expectations about the actor's degree of trustworthy conduct, which then inform trust.[11]

Theoretical Contributions

AI has swiftly moved from a novel technological gimmick to a widespread force reshaping a wide range of interactions, making it imperative to understand the social consequences of its use. Whereas much existing research focuses on the question of whether people trust AI itself, our study redirects the conversation toward an examination of its impact on the actors who deploy it, examining how disclosing AI usage influences social evaluations. Adopting a micro-institutional lens, our investigation highlights that legitimacy perceptions play a central role in eroding trust. In doing so, we expand prior scholarship beyond the technical attributes of AI toward the broader institutional and social frameworks that drive trust dynamics in AI-enhanced work settings.

Going beyond prior research (e.g., Schilke and Reimann 2025), we dissect the legitimacy construct into cognitive, moral, and pragmatic dimensions, thus providing greater insight into how trust erosion unfolds. The pathways by which AI disclosure impacts trust—by lowering perceptions of typicality, which then decrease perceptions of both commitment and authenticity, ultimately reducing trustworthiness perceptions and trust—are particularly revealing. This sequence of effects highlights how socially embedded expectations and norms profoundly influence trust dynamics. Of note, knowing that typicality comes first matters because it functions as a gatekeeper appraisal that determines whether people ever engage the more effortful, deliberative evaluations of pragmatic and moral legitimacy. When AI usage deviates from taken-for-granted scripts, it triggers a System 1 alarm that moves observers into System 2 processing (Tost 2011), which prompts closer scrutiny of pragmatic and moral legitimacy. As AI usage becomes routine over time and thus cognitively more legitimate, however, this alarm may be less likely to sound; even if observers harbor latent instrumental or moral qualms, they may never enter that second stage, effectively short-circuiting the chain and attenuating any downstream trust effect of AI disclosure. Conversely, in contexts where AI remains atypical, cognitive flags are more likely to be raised as the result of disclosure, activating those subsequent evaluations and preserving the trust penalty.

Our research makes contributions to several scholarly debates. First, it sheds

[11]See the Online Supplement.

light on the complicated interplay between transparency and trust (Bernstein 2017): although disclosing AI usage may signal openness, it simultaneously draws attention to departures from human-centered norms, undermining legitimacy. Our investigation thus highlights a striking contradiction: transparency about AI usage triggers legitimacy concerns and ultimately diminishes rather than bolsters trust. This finding challenges established views that generally link transparency with greater levels of trust (Schnackenberg and Tomlinson 2016). In other words, what might seem like a path to honesty can paradoxically erode confidence in the disclosing actor. Our micro-institutional lens helps clarify why this paradox emerges. In particular, disclosure renders institutional scripts and taken-for-granted norms salient, moving the practice from the realm of the routine to explicit evaluation against role expectations. Once this shift occurs, observers construe AI usage as atypical (cognitive), less committed and dutiful (pragmatic), and less authentic and sincere (moral), so transparency itself precipitates legitimacy penalties that undermine trust. Of note, the effect we document is not simply reflective of transparency about a negative act reducing trust—it is that AI usage disclosure per se triggers legitimacy screening and a trust penalty. This contributes to the transparency–trust literature (Bernstein 2017; Schnackenberg and Tomlinson 2016) by specifying why and when transparency backfires and by offering a mechanism grounded in legitimacy. Similarly, it adds to research on trust signaling by providing further evidence that communication meant to bolster trustworthiness perceptions may instead reduce it (Reimann et al. 2022). Decomposing legitimacy into the context-specific instantiations of typicality, commitment, and authenticity is helpful because it turns a broad concept into diagnostic levers: typicality explains the initial "alarm" when practices feel nonstandard, commitment captures concerns that the actor may be economizing on effort or care, and authenticity captures worries about sincerity and human agency. Seeing these facets—and their sequence—clarifies trust dynamics: norm violation (low typicality) invites more deliberative evaluations (commitment, authenticity), which then depress trust. This finer-grained account not only advances theory but also indicates possible interventions—for example, normalize the practice, signal effort, and demonstrate human ownership—to mitigate the trust costs of AI usage disclosure.

Second, by decomposing legitimacy into three conceptually distinct and measurable dimensions—typicality, commitment, and authenticity—we enhance the empirical testability of legitimacy theory in specific contexts. Each dimension is not only coherent but also readily operationalizable, allowing scholars to capture the subtle ways AI disclosure reshapes legitimacy perceptions. By specifying these context-specific elements, our study bridges the gap between broad, theoretical discussions of legitimacy and concrete empirical inquiry (Deephouse and Suchman 2008; Haack et al. 2021).

Third, our study helps bring trust and institutional theory closer together by showing how institutional norms and expectations fundamentally shape trust judgments. Recognizing that trust does not operate in a vacuum (Buskens and Raub 2002) but is deeply embedded in social and cultural frameworks underscores the synergy between these two literatures (Möllering 2006; Schilke and Cook 2013; Zucker and Schilke 2019). By foregrounding how cognitive, pragmatic, and moral legitimacy concerns influence trust, we demonstrate the rich

potential of integrating institutional insights into trust research.

Finally, although our research does not explicitly test disclosure policies, our findings inform broader discussions in that literature (e.g., King and Bearman 2017; Sah and Fugh-Berman 2013). Although mandatory disclosure could enhance accountability, our results indicate that those who comply may inadvertently incur legitimacy penalties—a cost that rational actors may weigh against the potential repercussions of nondisclosure. This consideration suggests that if mandated disclosure is adopted, it should be coupled with robust enforcement to prevent an adverse selection effect in which complying actors bear an undue burden. Another implication of our research is that mandated disclosure may be less disruptive when paired with a communication campaign that frames AI use as typical or role-consistent, thereby reducing the cognitive legitimacy and trust penalties that might otherwise accompany transparency. In the context of AI, furthermore, it is unclear whether it is in society's best interest to curb rapid adoption of the technology by introducing barriers, such as mandatory disclosure, or to encourage diffusion by keeping its use unregulated (Cuéllar et al. 2024). Policymakers must carefully weigh transparency goals against potential unintended consequences, recognizing that these trade-offs can vary widely across different disclosure contexts—be it conflict of interest or AI usage.

Limitations and Avenues for Future Research

We would like to point to some limitations of our work, which offer avenues for future research. First, our research examines trust formed in initial encounters with largely unknown actors—often referred to as "swift trust" (Blomqvist and Cook 2018; Schilke and Huang 2018). Whether our results generalize to contexts involving more established relationships (for a discussion, see Schilke et al. 2013) is an important question for future research to explore.

Second, our empirical work centers on a relatively narrow set of tasks situated in the context of organizations' human resource management, and the generalizability of our findings to other settings has yet to be evaluated. Future research might explore how AI disclosure operates in different areas, such as legal case review, health care diagnostics, creative content production, and other contexts where AI is commonly employed.

Third, we acknowledge the trade-off inherent in our Study 1 design. By asking respondents first whether they had concerns and then inviting them to elaborate with reference to cognitive, pragmatic, and moral legitimacy—consistent with the deliberation design approach (Haack et al. 2021)—we sacrificed some spontaneity in order to focus attention on the theorized mechanism and obtain comparable, information-rich accounts. Our primary aim in the directed content-analytic study was not to gauge prevalence or to elicit unconstrained reactions but to unpack how concerns are construed when they arise; importantly, responses were fully open-ended and in participants' own language, and they repeatedly invoked themes of typicality, commitment, and authenticity (illustrated in Table 2). Moreover, Study 2 then tests the same constructs experimentally and shows that typicality, followed by commitment and authenticity, statistically mediate the disclosure–trust link, which helps reduce concerns that the Study 1 themes are mere artifacts of prompting. Still, we acknowledge that more naturalistic elicitation could potentially be valuable: future work could first capture unprompted reactions to the scenario

and compare the prevalence and content of emergent themes across prompted versus unprompted conditions. This would quantify any priming and further establish whether legitimacy concerns surface spontaneously.

Fourth, the serial mediation in Study 2 supports the theorized sequence—AI disclosure is negatively related to perceived typicality, which is associated with commitment and authenticity and, in turn, with lower trust—but we caution that the exact ordering among the mediators is not causally identified in this design. Because typicality, commitment, and authenticity were measured and not experimentally manipulated, the paths from typicality to commitment and authenticity need to be interpreted as correlational rather than causal. The experiment establishes a causal effect of disclosure on each mediator and on trust (via random assignment), but adjudicating whether typicality causes shifts in commitment and authenticity will require future research that manipulates the mediators independently.

Fifth, we followed the majority of institutional research and modeled our theory on the three-dimensional conceptualization of legitimacy that goes back to Suchman (1995). Other scholars (e.g., Tost 2011; Tyler and Lind 1992), however, have suggested a fourth dimension, relational legitimacy, that centers on whether a target accords audiences respect, dignity, and identity affirmation, and that should be considered in future research on the legitimacy–trust nexus.

Finally, as AI becomes more commonplace across professional settings, the implications of disclosing its use may evolve. Over time, as generative AI transitions from relatively novel to routine, reactions toward disclosures of AI involvement could change. Therefore, future research should explore how ongoing AI diffusion influences the relationship between AI disclosure and trust and aim to develop dynamic theoretical accounts that capture how AI may achieve legitimacy as practices evolve.

CONCLUSION

In closing, we would like to underscore the importance of further exploring the social implications of AI. Sociologists are increasingly turning their attention toward artificial intelligence as a critical area of inquiry (Joyce and Cruz 2024)—and rightly so. A sociological approach provides unique value because it views algorithms not only as purely technical objects but also as sites of social struggle. Sociologists are well equipped to explore the boundary work that emerges as AI challenges our taken-for-granted expectations of human agency, focusing on contexts where machines take on roles previously viewed as distinctly human (Airoldi 2021). By advancing such lines of inquiry, sociologists can integrate the micro-institutional processes identified here with broader structural analyses—from the "coding elite" and the "cybertariat" to shifting discourses and inequalities (Burrell and Fourcade 2021)—ultimately helping to forge an AI landscape that is not only technologically advanced but also socially equitable, accountable, and genuinely trusted.

AUTHORS' NOTE

The authors contributed equally to this article. During the preparation of this work, the authors used ChatGPT-4 and Dall-E to generate experimental stimuli (which are posted in OSF repositories linked in the article), code free-text responses of the content-analytic study, and preliminarily copyedit the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

ACKNOWLEDGMENTS

The authors are thankful for capable research assistance by Stephanie Clauson. The authors thank Alex Bitektine, Patrick Haack, Bart Vanneste, and the audience of the 2025 MSI AI Summit for helpful comments on earlier versions of this research.

FUNDING

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research support was provided by a National Endowment for the Arts research grant (No. 1925643-38-24) and a National Security Systems (TRIF NSS) research grant to the second author and by a National Science Foundation CAREER Award (No. 1943688) to the first author. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ORCID iD

Oliver Schilke https://orcid.org/0000-0001-6832-1677

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

Airoldi, Massimo. 2021. *Machine Habitus: Toward a Sociology of Algorithms*. Cambridge, UK: Polity.

Ali, Abdallah El, Karthikeya Puttur Venkatraj, Sophie Morosoli, Laurens Naudts, Natali Helberger, and Pablo Cesar. 2024. "Transparent AI Disclosure Obligations: Who, What, When, Where, Why, How." Pp. 1–11 in *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery. doi:10.1145/3613905.3650750.

Arrow, Kenneth J. 1974. *The Limits of Organization*. New York, NY: W. W. Norton.

Atinc, Guclu, Marcia J. Simmering, and Mark J. Kroll. 2012. "Control Variable Use and Reporting in Macro and Micro Management Research." *Organizational Research Methods* 15(1):57–74.

Auspurg, Katrin, and Thomas Hinz. 2014. *Factorial Survey Experiments*. Los Angeles, CA: Sage.

Barasch, Alixandra, Emma E. Levine, Johnathan Z. Berman, and Deborah A. Small. 2014. "Selfish or Selfless? On the Signal Value of Emotion in Altruistic Behavior." *Journal of Personality and Social Psychology* 107(3):393–413.

Bernerth, Jeremy B., and Herman Aguinis. 2016. "A Critical Review and Best-Practice Recommendations for Control Variable Usage." *Personnel Psychology* 69(1):229–83.

Bernstein, Ethan S. 2017. "Making Transparency Transparent: The Evolution of Observation in Management Theory." *Academy of Management Annals* 11(1):217–66.

Beverland, Michael B., and Francis J. Farrelly. 2009. "The Quest for Authenticity in Consumption: Consumers' Purposive Choice of Authentic Cues to Shape Experienced Outcomes." *Journal of Consumer Research* 36(5):838–56.

Bitektine, Alex. 2011. "Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status." *Academy of Management Review* 36(1):151–79.

Bitektine, Alex, Nicole Gillespie, and Donald Lange. 2026. "From the Evaluator's Perspective: A Functional Approach to Social Judgments." *Academy of Management Review*. doi:10.5465/amr.2020.0405.

Bitektine, Alex, and Patrick Haack. 2015. "The 'Macro' and the 'Micro' of Legitimacy: Toward a Multilevel Theory of the Legitimacy Process." *Academy of Management Review* 40(1):49–75.

Bitektine, Alex, Jeff Lucas, and Oliver Schilke. 2018. "Institutions under a Microscope: Experimental Methods in Institutional Theory." Pp. 147–67 in *Unconventional Methodology in Organization and Management Research*, edited by A. Bryman and D. A. Buchanan. Oxford, UK: Oxford University Press.

Blomqvist, Kirsimarja, and Karen S. Cook. 2018. "Swift Trust: State-of-the-Art and Future Research Directions." Pp. 29–49 in *The Routledge Companion to Trust*, edited by R. H. Searle, A.-M. I. Nienaber, and S. B. Sitkin. New York, NY: Routledge.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology."

Qualitative Research in Psychology 3(2):77–101.

Brown, Andrew D., and Sammy Toyoki. 2013. "Identity Work and Legitimacy." *Organization Studies* 34(7):875–96.

Brüns, Jasper David, and Martin Meißner. 2024. "Do You Create Your Content Yourself? Using Generative Artificial Intelligence for Social Media Content Creation Diminishes Perceived Brand Authenticity." *Journal of Retailing and Consumer Services* 79:103790. doi:10.1016/j.jretconser.2024.103790.

Burrell, Jenna, and Marion Fourcade. 2021. "The Society of Algorithms." *Annual Review of Sociology* 47:213–37.

Buskens, Vincent, and Werner Raub. 2002. "Embedded Trust: Control and Learning." *Advances in Group Processes* 19:167–202.

Capraro, Valerio, Austin Lentsch, Daron Acemoglu, Selin Akgun, Aisel Akhmedova, Ennio Bilancini, and Jean-François Bonnefon, et al. 2024. "The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making." *PNAS Nexus* 3(6):pgae191. doi:10.1093/pnasnexus/pgae191.

Chen, Shijiao, Jing A. Zhang, Hongzhi Gao, Zhilin Yang, and Damien Mather. 2022. "Trust Erosion during Industry-Wide Crises: The Central Role of Consumer Legitimacy Judgement." *Journal of Business Ethics* 175(1):95–116.

Cohen, Bernard P. 1989. *Developing Sociological Knowledge: Theory and Method*. Chicago, IL: Nelson-Hall.

Colquitt, Jason A., Brent A. Scott, and Jeffery A. LePine. 2007. "Trust, Trustworthiness, and Trust Propensity: A Meta-analytic Test of Their Unique Relationships with Risk Taking and Job Performance." *Journal of Applied Psychology* 92(4):909–27.

Cook, Karen S., Chris Snijders, Vincent Buskens, and Coye Cheshire. 2009. *eTrust: Forming Relationships in the Online World*. New York, NY: Russell Sage Foundation.

Cooper-Hakim, Amy, and Chockalingam Viswesvaran. 2005. "The Construct of Work Commitment: Testing an Integrative Framework." *Psychological Bulletin* 131(2):241–59.

Crane, Leland, Michael Green, and Paul Soto. 2025. "Measuring AI Uptake in the Workplace." https://www.federalreserve.gov/econres/notes/feds-notes/measuring-ai-uptake-in-the-workplace-20240205.html#:~:text=Surveys%20of%20firms%20show%20a,suggest%20rapid%20growth%20in%20adoption.

Cuéllar, Mariano-Florentino, Benjamin Larsen, Yong Suk Lee, and Michael Webb. 2024. "Does Information about AI Regulation Change Manager Evaluation of Ethical Concerns and Intent to Adopt AI?" *Journal of Law, Economics, and Organization* 40(1):34–75.

Deephouse, David L., and Suzanne M. Carter. 2005. "An Examination of Differences between Organizational Legitimacy and Organizational Reputation." *Journal of Management Studies* 42(2):329–60.

Deephouse, David L., and Mark C. Suchman. 2008. "Legitimacy in Organizational Institutionalism." Pp. 49–77 in *The Sage Handbook of Organizational Institutionalism*, edited by R. Greenwood, C. Oliver, K. Sahlin and R. Suddaby. Los Angeles, CA: Sage.

Devine, Patricia G. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components." *Journal of Personality and Social Psychology* 56(1):5–18.

Diederich, Adele, and Jennifer S. Trueblood. 2018. "A Dynamic Dual Process Model of Risky Decision Making." *Psychological Review* 125(2):270–92.

DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263–87.

Dirks, Kurt T., and Bart de Jong. 2022. "Trust within the Workplace: A Review of Two Waves of Research and a Glimpse of the Third." *Annual Review of Organizational Psychology and Organizational Behavior* 9(1):247–76.

Eisenberger, Robert, Peter Fasolo, and Valerie Davis-LaMastro. 1990. "Perceived Organizational Support and Employee Diligence, Commitment, and Innovation." *Journal of Applied Psychology* 75(1):51–59.

Elish, Madeleine Clare. 2019. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction." *Engaging Science, Technology, and Society* 5:40–60.

Elo, Satu, and Helvi Kyngäs. 2008. "The Qualitative Content Analysis Process." *Journal of Advanced Nursing* 62(1):107–15.

Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina. 2022. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods* 54(4):1643–62.

Fehr, Ernst. 2009. "On the Economics and Biology of Trust." *Journal of the European Economic Association* 7(2–3):235–66.

Finch, Janet. 1987. "The Vignette Technique in Survey Research." *Sociology* 21(1):105–14.

Fishbowl. 2023. "70% of Workers Using ChatGPT at Work Are Not Telling Their Boss; Overall Usage among Professionals Jumps to 43%." https://www.fishbowlapp.com/insights/70-percent-of-workers-using-chatgpt-at-work-are-not-telling-their-boss/.

Fügener, Andreas, Jörn Grahl, Alok Gupta, and Wolfgang Ketter. 2022. "Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path toward Productive Delegation." *Information Systems Research* 33(2):678–96.

Garfinkel, Harold. 1963. "A Conception of, and Experiments with, 'Trust' as a Condition of Stable Concerted Actions." Pp. 187–238 in *Motivation and Social Interaction: Cognitive Determinants*, edited by O. J. Harvey. New York, NY: Ronald Press.

Gershon, Rachel, and Rosanna K. Smith. 2020. "Twice-Told Tales: Self-Repetition Decreases Observer Assessments of Performer Authenticity." *Journal of Personality and Social Psychology* 118(2):307–24.

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences* 120(30):e2305016120. doi:10.1073/pnas.2305016120.

Gkinko, Lorentsa, and Amany Elbanna. 2023. "Designing Trust: The Formation of Employees' Trust in Conversational AI in the Digital Workplace." *Journal of Business Research* 158:113707. doi:10.1016/j.jbusres.2023.113707.

Glikson, Ella, and Anita Williams Woolley. 2020. "Human Trust in Artificial Intelligence: Review of Empirical Research." *Academy of Management Annals* 14(2): 627–60.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–97.

Haack, Patrick, Michael D. Pfarrer, and Andreas Georg Scherer. 2014. "Legitimacy-as-Feeling: How Affect Leads to Vertical Legitimacy Spillovers in Transnational Governance." *Journal of Management Studies* 51(4):634–66.

Haack, Patrick, Oliver Schilke, and Lynne G. Zucker. 2021. "Legitimacy Revisited: Disentangling Propriety, Validity, and Consensus." *Journal of Management Studies* 58(3):749–81.

Hamm, Joseph A., Rick Trinkner, and James D. Carr. 2017. "Fair Process, Trust, and Cooperation: Moving toward an Integrated Framework of Police Legitimacy." *Criminal Justice and Behavior* 44(9):1183–212.

Hancock, Jeffrey T., Mor Naaman, and Karen Levy. 2020. "AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations." *Journal of Computer-Mediated Communication* 25(1): 89–100.

Harkness, Sarah K., Coye Cheshire, Karen S. Cook, Cătălin Stoica, and Bogdan State. 2022. "Exchange and the Creation of Trust and Solidarity across Cultures." *Social Psychology Quarterly* 85(4):351–73.

Harmon, Derek J. 2019. "When the Fed Speaks: Arguments, Emotions, and the Microfoundations of Institutions." *Administrative Science Quarterly* 64(3):542–75.

Hauser, David J., Aaron J. Moss, Cheskie Rosenzweig, Shalom N. Jaffe, Jonathan Robinson, and Leib Litman. 2023. "Evaluating Cloudresearch's Approved Group as a Solution for Problematic Data Quality on MTurk." *Behavior Research Methods* 55(8):3953–64.

Hawks, Kate. 2025. "The Role of Personal Values in Shaping Perceptions of the Legitimacy of Public Health Officials during a Global Pandemic." *Social Psychology Quarterly* 88(1):89–110.

Hawks, Kate, Karen A. Hegtvedt, Ryan Gibson, Cathryn Johnson, and Jamica Zion. 2024. "Pathways to Legitimacy for Black and White Authorities: Impressions of Competence and Warmth." *Social Psychology Quarterly* 87(1):84–105.

Hayes, Andrew F. 2022. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Hsieh, Hsiu-Fang, and Sarah E. Shannon. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research* 15(9):1277–88.

Hsieh, Ying-Ying, Jean-Philippe Vergne, Philip Anderson, Karim Lakhani, and Markus Reitzig. 2018. "Bitcoin and the Rise of Decentralized Autonomous Organizations." *Journal of Organization Design* 7(1):14. doi:10.1186/s41469-018-0038-1.

Jago, Arthur S. 2019. "Algorithms and Authenticity." *Academy of Management Discoveries* 5(1):38–56.

James, Nalita. 2016. "Using Email Interviews in Qualitative Educational Research:

Creating Space to Think and Time to Talk." *International Journal of Qualitative Studies in Education* 29(2):150–63.

Johnston, Lucy, and Miles Hewstone. 1992. "Cognitive Models of Stereotype Change: Subtyping and the Perceived Typicality of Disconfirming Group Members." *Journal of Experimental Social Psychology* 28(4):360–86.

Jordan, Jillian J., Sommers Roseanna, Paul Bloom, and David G. Rand. 2017. "Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling." *Psychological Science* 28(3):356–68.

Joyce, Kelly, and Taylor M. Cruz. 2024. "A Sociology of Artificial Intelligence: Inequalities, Power, and Data Justice." *Socius* 10. doi:10.1177/23780231241275393.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Katz, Daniel, and Robert L. Kahn. 1978. *Social Psychology of Organizations*. New York, NY: Wiley.

Kim, Minjae. 2025. "Signaling Commitment via Insincere Conformity: A New Take on the Persistence of Unpopular Norms." *Social Psychology Quarterly* 88(1):111–34.

Kim, Tae Woo, Li Jiang, Adam Duhachek, Hyejin Lee, and Aaron Garvey. 2022. "Do You Mind If I Ask You a Personal Question? How AI Service Agents Alter Consumer Self-Disclosure." *Journal of Service Research* 25(4):649–66.

King, Marissa, and Peter S. Bearman. 2017. "Gifts and Influence: Conflict of Interest Policies and Prescribing of Psychotropic Medications in the United States." *Social Science and Medicine* 172:153–62.

Kirk, Colleen P., and Julian Givi. 2025. "The AI-Authorship Effect: Understanding Authenticity, Moral Disgust, and Consumer Responses to AI-Generated Marketing Communications." *Journal of Business Research* 186:114984. doi:10.1016/j.jbusres.2024.114984.

Kramer, Roderick M. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology* 50:569–96.

Lamertz, Kai, and Devasheesh P. Bhave. 2017. "Employee Perceptions of Organisational Legitimacy as Impersonal Bases of Organisational Trustworthiness and Trust." *Journal of Trust Research* 7(2):129–49.

Lei, Ya-Wen, and Rachel Kim. 2024. "Automation and Augmentation: Artificial Intelligence, Robots, and Work." *Annual Review of Sociology* 50:251–72.

Levine, Sheen S., Oliver Schilke, Olenka Kacperczyk, and Lynne G. Zucker. 2023. "Primer for Experimental Methods in Organization Theory." *Organization Science* 34(6):1997–2025.

Lim, Joon Soo, and Jun Zhang. 2025. "Stakeholder Engagement and Authenticity in Corporate Social Advocacy: Pathways to Corporate Reputation via Perceived Legitimacy." *Journal of Public Relations Research* 37(5):470–97.

Litman, Leib, Jonathan Robinson, and Tzvi Abberbock. 2017. "TurkPrime.com: A Versatile Crowdsourcing Data Acquisition Platform for the Behavioral Sciences." *Behavior Research Methods* 49(2):433–42.

Lockey, Steve, and Nicole Gillespie. 2024. "Understanding Trust in Artificial Intelligence: A Research Agenda." Pp. 11–24 in *A Research Agenda for Trust: Interdisciplinary Perspectives*, edited by R. C. Mayer and B. M. Mayer. Cheltenham, UK: Edward Elgar.

Luhmann, Niklas. 1979. *Trust and Power*. Chichester, UK: Wiley.

Lumineau, Fabrice, Oliver Schilke, and Wenqian Wang. 2023. "Organizational Trust in the Age of the Fourth Industrial Revolution: Shifts in the Nature, Production, and Targets of Trust." *Journal of Management Inquiry* 32(1):21–34.

Mayer, Roger C., and James H. Davis. 1999. "The Effect of the Performance Appraisal System on Trust for Management: A Field Quasi-experiment." *Journal of Applied Psychology* 84(1):123–36.

Mayer, Roger C., James H. Davis, F., and Schoorman, David. 1995. "An Integrative Model of Organizational Trust." *Academy of Management Review* 20(3):709–34.

McKinsey and Company. 2025. "The State of AI: How Organizations Are Rewiring to Capture Value." https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#.

Merton, Robert K., Marjorie Fiske, and Patricia A. Kendall. 1956. *The Focused Interview: A Manual of Problems and Procedures*. New York, NY: Free Press.

Meyer, John P., and Lynne Herscovitch. 2001. "Commitment in the Workplace: Toward a General Model." *Human Resource Management Review* 11(3):299–326.

Möllering, Guido. 2006. "Trust, Institutions, Agency: Towards a Neoinstitutional Theory

of Trust." Pp. 355–76 in *Handbook of Trust Research*, edited by R. Bachmann and A. Zaheer. Cheltenham, UK: Edward Elgar.

Morgan, David L. 2023. "Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT." *International Journal of Qualitative Methods* 22. doi:10.1177/16094069231211248.

Nguyen, Brenda, Hannes Leroy, Carol Gill, and Tony Simons. 2022. "Be Yourself or Adapt Yourself? Authenticity, Self-Monitoring, Behavioural Integrity, and Trust." *Journal of Trust Research* 12(1):24–42.

Nosek, Brian A., Uri Simonsohn, Don A. Moore, Leif D. Nelson, Joseph P. Simmons, Andrew Sallans, and Etienne P. LeBel. 2013. "Standard Reviewer Statement for Disclosure of Sample, Conditions, Measures, and Exclusions." http://osf.io/hadz3.

Nowak, Andrzej, Mikolaj Biesaga, Karolina Ziembowicz, Tomasz Baran, and Piotr Winkielman. 2023. "Subjective Consistency Increases Trust." *Scientific Reports* 13(1):5657. doi:10.1038/s41598-023-32034-4.

OpenAI. 2023. "ChatGPT (Mar 14 Version) [Large Language Model]." https://chat.openai.com/chat.

Organisation for Economic Co-operation and Development. 2024. "Recommendation of the Council on Artificial Intelligence." https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Patton, Michael Quinn. 2015. *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*. Los Angeles, CA: Sage.

Perrewe, Pamela L., Denise Rotondo Fernandez, and Karen S Morton. 1993. "An Experimental Examination of Implicit Stress Theory." *Journal of Organizational Behavior* 14(7):677–86.

Powell, Walter W. 2019. "Institutions on the Ground." *Research in the Sociology of Organizations* 65B:419–28.

Raclaw, Joshua, Jena Barchas-Lichtenstein, and Abby Bajuniemi. 2020. "Online Surveys as Discourse Context: Response Practices and Recipient Design." *Discourse, Context and Media* 38:100441. doi:10.1016/j.dcm.2020.10044.

Reilly, Patrick. 2018. "No Laughter among Thieves: Authenticity and the Enforcement of Community Norms in Stand-Up Comedy." *American Sociological Review* 83(5):933–58.

Reimann, Martin, Christoph Hüller, Oliver Schilke, and Karen S. Cook. 2022. "Impression Management Attenuates the Effect of Ability on Trust in Economic Exchange." *Proceedings of the National Academy of Sciences* 119(30):e2118548119. doi:10.1073/pnas.2118548119.

Reimann, Martin, and Ann Kronrod. 2026. "Language as Means and Ends: How Generative Artificial Intelligence Automates, Amplifies, and Reinvents Language in Marketing." *Journal of the Association for Consumer Research* 12(2). doi:10.1086/740067.

Resnik, David B., and Mohammad Hosseini. 2026. "Disclosing Artificial Intelligence Use in Scientific Research and Publication: When Should Disclosure Be Mandatory, Optional, or Unnecessary?" *Accountability in Research* 33(2):1–13.

Robbins, Blaine G. 2016. "Probing the Links between Trustworthiness, Trust, and Emotion: Evidence from Four Survey Experiments." *Social Psychology Quarterly* 79(3):284–308.

Roller, Margaret R., and Paul J. Lavrakas. 2015. *Applied Qualitative Research Design: A Total Quality Framework Approach*. New York, NY: Guilford Publications.

Rosch, Eleanor. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104(3):192–233.

Sah, Sunita, and Adriane Fugh-Berman. 2013. "Physicians under the Influence: Social Psychology and Industry Marketing Strategies." *Journal of Law, Medicine and Ethics* 41(3):665–72.

Salmons, Janet. 2014. *Qualitative Online Interviews: Strategies, Design, and Skills*. London, UK: Sage.

Schaap, Gabi, Tibor Bosse, and Paul Hendriks Vettehen. 2024. "The ABC of Algorithmic Aversion: Not Agent, but Benefits and Control Determine the Acceptance of Automated Decision-Making." *AI and Society* 39(4):1947–60.

Schenk, Patrick, Vanessa A Müller, and Luca Keiser. 2024. "Social Status and the Moral Acceptance of Artificial Intelligence." *Sociological Science* 11:989–1016.

Schilke, Oliver. 2018. "A Micro-institutional Inquiry into Resistance to Environmental Pressures." *Academy of Management Journal* 61(4):1431–66.

Schilke, Oliver, Reinhard Bachmann, Kirsimarja Blomqvist, Rekha Krishnan, and Jörg Sydow. 2026. "Unpacking the Paradoxes of Trust in Uncertain Times." *Organization Studies* 47(3):359–389.

Schilke, Oliver, and Karen S. Cook. 2013. "A Cross-Level Process Theory of Trust Development in Interorganizational Relationships." *Strategic Organization* 11(3):281–303.

Schilke, Oliver, and Karen S. Cook. 2015. "Sources of Alliance Partner Trustworthiness: Integrating Calculative and Relational Perspectives." *Strategic Management Journal* 36(2):276–97.

Schilke, Oliver, and Laura Huang. 2018. "Worthy of Trust? How Brief Interpersonal Contact Affects Trust Accuracy." *Journal of Applied Psychology* 103(11):1181–97.

Schilke, Oliver, Andy Powell, and Maurice E. Schweitzer. 2023. "A Review of Experimental Research on Organizational Trust." *Journal of Trust Research* 13(2):102–39.

Schilke, Oliver, and Martin Reimann. 2025. "The Transparency Dilemma: How AI Disclosure Erodes Trust." *Organizational Behavior and Human Decision Processes* 188:104405. doi:10.1016/j.obhdp.2025.104405.

Schilke, Oliver, Martin Reimann, and Karen S. Cook. 2013. "Effect of Relationship Experience on Trust Recovery Following a Breach." *Proceedings of the National Academy of Sciences* 110(38):15236–41.

Schilke, Oliver, Martin Reimann, and Karen S. Cook. 2015. "Power Decreases Trust in Social Exchange." *Proceedings of the National Academy of Sciences* 112(42): 12950–55.

Schilke, Oliver, Martin Reimann, and Karen S. Cook. 2021. "Trust in Social Relations." *Annual Review of Sociology* 47:239–59.

Schilke, Oliver, and Gabriel Rossman. 2018. "It's Only Wrong if It's Transactional: Moral Perceptions of Obfuscated Exchange." *American Sociological Review* 83(6):1079–107.

Schilke, Oliver, Gunnar Wiedenfels, Malte Brettel, and Lynne G. Zucker. 2017. "Interorganizational Trust Production Contingent on Product and Performance Uncertainty." *Socio-Economic Review* 15(2):307–30.

Schilke, Oliver, Zeyu Xue, and Patrick Haack. 2025. "Legitimacy Construction in the Presence of Multiple Validity Cues: An Experimental Investigation." Pp. 429–39 in *Handbook of Social Psychology*. Vol. 1, edited by J. E. Stets, K. A. Hegtvedt and L. Doan. Cham, Switzerland: Springer.

Schnackenberg, Andrew K., and Edward C. Tomlinson. 2016. "Organizational Transparency: A New Perspective on Managing Trust in Organization-Stakeholder Relationships." *Journal of Management* 42(7):1784–810.

Schor, Juliet B., and Steven P. Vallas. 2021. "The Sharing Economy: Rhetoric and Reality." *Annual Review of Sociology* 47:369–89.

Schreier, Margrit. 2012. *Qualitative Content Analysis in Practice*. Thousand Oaks, CA: Sage.

Schudson, Michael. 2015. *The Rise of the Right to Know: Politics and the Culture of Transparency, 1945-1975*. Cambridge, MA: Harvard University Press.

Shore, Lynn McFarlane, Kevin Barksdale, and Ted H. Shore. 1995. "Managerial Perceptions of Employee Commitment to the Organization." *Academy of Management Journal* 38(6):1593–615.

Sidani, Yusuf M., and W. Glenn Rowe. 2018. "A Reconceptualization of Authentic Leadership: Leader Legitimation via Follower-Centered Assessment of the Moral Dimension." *The Leadership Quarterly* 29(6): 623–36.

Simons, Tony. 2002. "Behavioral Integrity: The Perceived Alignment between Managers' Words and Deeds as a Research Focus." *Organization Science* 13(1):18–35.

Sitkin, Sim B., and Nancy L. Roth. 1993. "Explaining the Limited Effectiveness of Legalistic 'Remedies' for Trust/Distrust." *Organization Science* 4(3):367–92.

Small, Mario Luis. 2011. "How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature." *Annual Review of Sociology* 37:57–86.

Spector, Paul E., and Michael T. Brannick. 2011. "Methodological Urban Legends: The Misuse of Statistical Control Variables." *Organizational Research Methods* 14(2):287–305.

Suchman, Mark C. 1995. "Managing Legitimacy: Strategic and Institutional Approaches." *Academy of Management Review* 20(3):571–610.

Suddaby, Roy, Alex Bitektine, and Patrick Haack. 2017. "Legitimacy." *Academy of Management Annals* 11(1):451–78.

Tolbert, Pamela S., and Lynne G. Zucker. 1983. "Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform,

1880-1935." *Administrative Science Quarterly* 28(1):22–39.
Tost, Leigh Plunkett. 2011. "An Integrative Model of Legitimacy Judgments." *Academy of Management Review* 36(4):686–710.
Tsvetkova, Milena, Taha Yasseri, Niccolo Pescetelli, and Tobias Werner. 2024. "A New Sociology of Humans and Machines." *Nature Human Behaviour* 8(10):1864–76.
Turner, Daniel W. 2010. "Qualitative Interview Design." *The Qualitative Report* 15(3):754–60.
Tyler, Tom R. 1997. "The Psychology of Legitimacy: A Relational Perspective on Voluntary Deference to Authorities." *Personality and Social Psychology Review* 1(4):323–45.
Tyler, Tom R., and E. Allan Lind. 1992. "A Relational Model of Authority in Groups." *Advances in Experimental Social Psychology* 25:115–91.
Voorspoels, Wouter, Wolf Vanpaelmel, and Gert Storms. 2008. "Modeling Typicality: Extending the Prototype View." *Proceedings of the Annual Meeting of the Cognitive Science Society* 30(30):757–62.
Wachinger, Jonas, Kate Bärnighausen, Louis N. Schäfer, Kerry Scott, and Shannon A. McMahon. 2024. "Prompts, Pearls, Imperfections: Comparing ChatGPT and a Human Researcher in Qualitative Data Analysis." *Qualitative Health Research* 35(9):951–66.
Wallander, Lisa. 2009. "25 Years of Factorial Surveys in Sociology: A Review." *Social Science Research* 38(3):505–20.
Weinberg, Jill D., Jeremy Freese, and David McElhattan. 2014. "Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsource-Recruited Sample." *Sociological Science* 1:292–310.
Zucker, Lynne G. 1977. "The Role of Institutionalization in Cultural Persistence." *American Sociological Review* 42(5):726–43.
Zucker, Lynne G. 1983. "Organizations as Institutions." *Research in the Sociology of Organizations* 2:1–47.
Zucker, Lynne G. 1986. "Production of Trust: Institutional Sources of Economic Structure, 1840-1920." *Research in Organizational Behavior* 8:53–111.
Zucker, Lynne G., and Oliver Schilke. 2019. "Towards a Theory of Micro-institutional Processes: Forgotten Roots, Links to Social-Psychological Research, and New Ideas." *Research in the Sociology of Organizations* 65B:371–89.

BIOS

Oliver Schilke is a professor of management and organizations at the University of Arizona, where he also has a courtesy appointment with the School of Sociology and serves as the director of the Center for Trust Studies. He is also an external faculty affiliate at the Institute for Research in the Social Sciences (IRiSS) at Stanford University. His research interests include collaboration, trust, organizational routines/capabilities, and micro-institutional processes. He received his PhD in sociology from the University of California–Los Angeles.

Martin Reimann is an associate professor of marketing, psychology, veterinary medicine, and cognitive science at the University of Arizona, as well as external faculty affiliate at the Institute for Research in the Social Sciences (IRiSS) at Stanford University. His research interests include trust in AI-mediated contexts, the role of affect as information in decision making, and experience theory. He received his PhD in psychology from the University of Southern California.