RAIS Journal for Social Sciences | Vol. 10, No. 1, 2026
ISSN 2574-0245 (Print) | ISSN 2574-1179 (Online) | DOI:10.5281/zenodo.19661381

Epistemological Gatekeeping in Technology Startup AI Adoption: A Narrative Review of Epistemic Injustice, Concentrated Leadership, and Algorithmic Bias

Tonia R. Ruck

Marymount University, Arlington, VA, USA
trr94454@marymount.edu

Abstract: Technology startups are adopting artificial intelligence at an increasing pace, yet smaller firms are more than twice as likely as larger organizations to lack governance roadmaps or dedicated teams for overseeing AI adoption. In startups with 10 to 100 employees, a small number of founders and C-suite leaders hold concentrated authority over AI decisions, functioning as epistemological gatekeepers who determine whose knowledge shapes adoption choices and whose perspectives are excluded. This narrative literature review integrates scholarship across three fields that have developed largely in isolation: Miranda Fricker's epistemic injustice theory, entrepreneurship research on startup decision-making and governance, and critical AI studies examining algorithmic bias. The review identifies particular organizational mechanisms through which epistemic injustice operates in startup AI governance. These include inferential inertia, temporal shifts in credibility, hermeneutical closure, and the active promotion of dominant epistemological assumptions by AI systems themselves. Results show that epistemological gatekeeping in startups is not a matter of individual bias but rather a structural phenomenon arising from the interaction of concentrated authority, informal governance, speed culture, resource scarcity, and demographic homogeneity. The review concludes that bias mitigation is fundamentally an epistemological problem requiring attention to whose knowledge informs AI decisions, and recommends phenomenological research to investigate how startup leaders experience and enact their gatekeeping role.

Keywords: Epistemic Injustice, Epistemological Gatekeeping, Artificial Intelligence Governance, Startup Leadership, Algorithmic Bias, Testimonial Injustice, Hermeneutical Injustice

Introduction

AI development has shifted from academia into the private tech industry. In 2024, nearly 90% of notable AI models originated from industry, up from 60% the prior year (Stanford AI Index Report, 2025). This concentration means AI governance increasingly depends on organizational practices within companies, rather than on external regulatory systems. Technology startups are among the most aggressive adopters of AI. In these firms, founders often have more education and experience and are motivated to present new ideas. They drive adoption decisions, with a strong focus on growth (McElheran et al., 2024). Yet, smaller firms are more than twice as likely as larger organizations to lack defined AI adoption roadmaps or dedicated governance teams (Stanford AI Index Report, 2025). A significant gap exists between how quickly startups integrate AI and the structures they have for evaluating what systems do, who they affect, and whose concerns shape their design. This gap matters because startup AI decisions tend to scale rapidly. Venture-backed companies expand fast. Evaluation systems, credibility hierarchies, and norms set with 20 employees persist as firms grow. Leaders who adopt AI early embed their assumptions into systems. As companies expand, these assumptions spread, affecting more users. A local technology choice becomes an infrastructure-level commitment with broad social effects.

This review addresses not just whether startups adopt AI responsibly, but who defines "responsible," whose knowledge shapes that definition, and which perspectives are excluded. In startups with 10 to 100 employees, a handful of leaders hold concentrated authority over AI and act as epistemological gatekeepers. They choose which AI risks to prioritize, which concerns are credible, and whose voices shape governance, often informally.

This review synthesizes scholarship from three largely separate fields: epistemic injustice theory, entrepreneurship research on decision making and governance, and critical AI studies on algorithmic bias. Each field offers distinct analytical tools. Epistemic injustice theory provides a language for understanding the assignment and denial of credibility. Startup research shows that these assignments concentrate among a few individuals. Critical AI studies reveal outcomes when these choices become built into technical systems. Few studies combine these perspectives to examine epistemological gatekeeping in resource-limited startups adopting AI. This review fills that gap.

Problem Statement

Founder and C-suite leaders in small technology startups have concentrated authority over AI adoption decisions (the process of implementing artificial intelligence systems). They operate in conditions that limit the range of knowledge informing those decisions. Informal governance (decision-making outside formal rules or policies), pressure to move quickly, limited resources for thorough evaluation, and homogeneous leadership (teams with similar backgrounds or perspectives) interact to create an environment in which some perspectives on AI risk and bias cannot gain traction. To better understand these dynamics, epistemic injustice theory (the study of unfair distribution of knowledge and recognition in society; Fricker, 2007) provides precise analytical methods. However, scholars have not applied these frameworks to AI governance in startups. Research on startup decision-making has examined founder authority concentration and entrepreneurial cognitive patterns, but not how these factors determine which knowledge shapes technology decisions. Critical AI studies have shown that algorithmic systems encode social bias at scale. Yet much of this work focuses primarily on large enterprises with formal structures, excluding resource-constrained organizations where much AI adoption occurs (Hopkins & Booth, 2021).

This literature gap leaves three fields with tools to analyze how startup AI governance privileges or marginalizes knowledge, but without integration among them. This review unites them and identifies specific mechanisms by which epistemological gatekeeping works in startups.

Significance Statement

Concentrated startup leadership shapes AI adoption beyond single organizations. Early adoption establishes patterns that become embedded into products, services, and company cultures as firms grow. If testimonial and hermeneutical injustices arise in a 30-person company, leaders normalize them as the company expands. Intervening early is easier and more effective than correcting later.

This review addresses which organizations researchers study and which organizations they overlook. Most AI governance research focuses on large enterprises, assuming these firms make key decisions (Hopkins & Booth, 2021). But in startups, where governance stays informal and leadership is concentrated, epistemological gatekeeping is most visible. These conditions also ease investigation. Startup leaders, investors, and regulators all need to understand how knowledge is validated in organizations that make high-stakes AI decisions with minimal oversight.

Nature of the Study

This study is a narrative literature review. Narrative reviews synthesize and interpret existing research to identify patterns, themes, and gaps across bodies of scholarship (Baumeister & Leary, 1997). This approach is appropriate for the present study because the research question spans three distinct disciplinary traditions that require interpretive integration rather than statistical aggregation.

The review drew on multiple research sources used to ensure coverage across the relevant fields. ABI/Inform Complete (ProQuest), Academic Search Premier (EBSCO), Business Source Premier (EBSCO), and Web of Science provided scholarship in management, entrepreneurship, and organizational behavior. Sage Journals and Taylor & Francis Online provided access to journals in philosophy, sociology, and feminist studies. The ACM Digital Library contributed technology-focused scholarship on AI ethics. Google Scholar supplemented these databases and enabled citation tracking from pivotal works.

Search terms were developed iteratively as the research focus narrowed. Initial searches combined broad terms: *artificial intelligence, AI bias, algorithmic fairness,* and *AI governance*, with *leadership, decision-making, and organizational culture*. Targeted searches added *epistemic injustice*, *testimonial injustice*, *hermeneutical injustice*, and *standpoint epistemology,* combined with *organization*, *workplace*, and *technology*. Startup-specific searches used *startup* OR *entrepreneurial firm* OR *SME,* paired with *AI adoption*, *technology governance*, *founder authority*, and decision-making. Boolean operators combined terms strategically. Citation tracking identified additional relevant literature.

Sources were included based on relevance to the research question, scholarly strictness (prioritizing peer-reviewed publications), currency (emphasizing 2020-2025 while retaining pivotal works), and methodological quality. Non-peer-reviewed publications, opinion pieces without empirical grounding, and generic small-business literature lacking a technology focus were excluded. The final literature base comprises 125 references, of which 118 (94.4%) are peer-reviewed, and 97 (77.6%) were published within the past five years. The review organized sources into three thematic clusters: epistemic injustice theory and feminist epistemology, startup decision-making and governance, and critical AI studies and bias mitigation.

Concepts and Studies in the Literature

Epistemic Injustice and Its Organizational Mechanisms

Miranda Fricker's landmark work (2007) identified two forms of epistemic injustice that target individuals as knowers. Epistemic injustice refers to wrongs done to someone in their capacity as a knower. Testimonial injustice occurs when prejudice leads a listener to assign reduced credibility to a speaker's claims, often because of identity-based bias. Hermeneutical injustice arises when gaps in collective interpretive resources prevent people from articulating their experiences in terms others can understand, limiting recognition of those experiences. Both forms operate through unequal participation in meaning-making practices. Fricker argued that testimonial injustice degrades the speaker not only as a source of information but also as a person, and that hermeneutical injustice prevents individuals from contesting the conditions that affect them because they lack the conceptual vocabulary to describe their experiences (Fricker, 2007).

Fricker's later scholarship extended this system from interpersonal exchanges to institutional settings. Her analysis of organizational epistemic vices (Fricker, 2020) introduced the concept of inferential inertia: a pattern where organizations hold relevant information without making the cognitive effort to draw available inferences. Studying the BBC's institutional failures, Fricker (2020) demonstrated that informational compartmentalization kept scattered knowledge from aggregating into operable evidence. Individual employees possessed pieces of the picture. No organizational mechanism assembled them. This concept translates directly to startup AI governance. A customer-facing employee notices an AI system performing poorly for certain demographics. An engineer observes unexpected output patterns. A sales representative hears complaints from a specific segment. Each observation remains isolated. Without formal channels for aggregation, the information exists, while the organizational inference does not.

Whether these observations surface at all depends on leadership conditions. Brown et al. (2005) developed the core construct of ethical leadership, defining it as the demonstration of normatively appropriate conduct through personal actions and interpersonal relationships, combined with the promotion of such conduct through two-way communication, reinforcement, and decision-making. Brown et al. (2005) found that ethical leadership predicts follower willingness to report problems to management. This finding is directly relevant to AI governance: employees must be willing to surface concerns about AI systems for those concerns to be addressed. Leadership that models openness to critical feedback creates conditions in which testimonial injustice is less likely to occur. Leadership that signals impatience with dissent or penalizes questioning ensures gatekeeping occurs before any formal credibility assessment. The concern never enters the room.

Fricker (2013) also connected hermeneutical injustice to political contestation, arguing that people who lack interpretive frameworks for describing their experience are disabled as contesters. They cannot challenge decisions through available procedures because their concerns resist articulation in terms that those procedures accept. In startup settings where governance is relaxed and no established process exists for raising AI ethics concerns, the hermeneutical barrier compounds the procedural one. An employee who senses that an AI system produces unfair results but lacks the vocabulary to articulate concepts such as disparate impact, representational harm, or algorithmic bias cannot translate that intuition into a claim the organization can process.

Credibility assessments also shift over time. Drage et al. (2024) documented that in technology organizations, responsibility for AI harms often surfaces only in crises, provoking blame rather than systematic attention to earlier concerns. Warnings issued before harm becomes visible are given less credibility. The same information attracts attention after the fact. This temporal-based dimension offers specificity to the phenomenon of testimonial injustice in startup contexts: an employee who raises concerns about AI bias before deployment faces credibility assessments shaped by speed culture, in which the concerns are deemed speculative and lack quantitative evidence. After deployment, if harm is measurable, the same concern becomes urgent. The knowledge was available. Its credibility was time-dependent.

The application of Fricker's framework to organizational AI governance requires extending it beyond her original interpersonal scope. Three lines of evidence support this move. First, Fricker herself moved in this direction through her institutional analysis (Fricker, 2020) and her connection of epistemic injustice to structural political freedom (Fricker, 2013). Second, expertise-category privileging in technology companies carries identity-based prejudice. Engineering and technical roles are disproportionately held by white and Asian men, while non-technical roles are disproportionately held by women and underrepresented minorities (Nadeem et al., 2022). When leaders privilege technical knowledge over experiential knowledge, the resulting credibility deficit is not identity neutral. Drage et al. (2024) documented this directly: maintenance and care work essential for responsible AI is undervalued, gendered, racialized, and frequently assigned to lower-status workers. Third, the cognitive mechanisms that produce credibility deficits in entrepreneurial settings generate outcomes functionally equivalent to testimonial injustice. Busenitz and Barney (1997) found that entrepreneurs are significantly more prone to overconfidence and representativeness heuristics than managers in large organizations. These cognitive shortcuts systematically inflate the credibility of the founder's own knowledge while deflating the credibility of incoming claims. Kraft et al. (2022) identified overprecision, excessive certainty in one's own beliefs, as the specific overconfidence type most consistently harmful for entrepreneurial decision-making. A founder who is overprecise discounts alternative viewpoints not out of explicit prejudice but because of the certainty that their established structure is sufficient. The effect is equivalent: certain voices are dismissed before their claims obtain genuine evaluation.

Standpoint epistemology deepens the analysis by explaining what is lost when these dismissals occur. Sandra Harding (1993) argued that conventional claims to objectivity actually encode dominant-group perspectives while presenting them as universal truths. Genuine objectivity necessitates examining how social location shapes inquiry, extending critical scrutiny to the background assumptions that conventional methodology leaves unexamined. Harding distinguished a standpoint from a mere perspective: a standpoint is an achievement requiring critical engagement with how social systems operate from structurally marginalized positions (Harding, 1991). Patricia Hill Collins (2000) developed the "outsider within" concept to describe how individuals working within institutions without full membership develop distinctive insights about contradictions between stated values and actual practices. Collins also articulated an alternative epistemological framework emphasizing lived experience as a criterion for meaning, dialogue in knowledge assessment, an ethics of caring, and personal accountability that connects character to credibility (Collins, 2000). These criteria are precisely what technical meritocracy in startup cultures tends to reject as insufficiently objective. When leaders privilege quantitative metrics over qualitative concerns and detached analysis over engaged dialogue, they systematically exclude epistemological approaches that standpoint theory argues reveal what dominant frameworks cannot see.

Recent empirical work confirms these dynamics in technology organizations. Ali et al. (2023) studied technology workers responsible for integrating AI ethics and found that ethics practitioners rely on personal influence rather than formal authority within environments that value product metrics and launch timelines. Ethical issues are systematically subordinated to shipping deadlines. The researchers noted that these conditions are particularly acute for practitioners from disadvantaged backgrounds. Browne et al. (2024) interviewed 63 employees at a major AI firm and documented confusion about the concept of "bias" alongside weak connections between diversity initiatives and AI development practice. Diversity programs were siloed as HR concerns, disconnected from the technical processes they were meant to inform. Stinson and Vlaad (2024) developed the concept of "emergent expertise" to theorize why diverse teams produce better AI outcomes, arguing that networks combining technical ability with standpoint expertise are necessary for technologies that serve differently situated communities.

Startup Conditions That Boost Gatekeeping

The organizational characteristics of startups do not simply host epistemological gatekeeping. They intensify it. Concentrated founder authority, informal governance, speed culture, resource restrictions, and demographic homogeneity all converge into a system that determines how knowledge claims about AI will be received before any particular claim is made.

Fan (2022) conducted empirical research on startup corporate governance, finding that a founder-centric model predominates and that governance measures are emphasized mainly through economic slumps and immediately prior to exits, such as IPOs or acquisitions. AI adoption decisions occur during expansion stages, precisely when governance attention is lowest. Colombo et al. (2024) documented that, even when venture capital introduces formal delegation structures, a gap emerges between the established and the actual organization. Nominal authority may be distributed. Actual decision-making power remains concentrated. Employees may have theoretical standing to convey concerns about AI systems, even if they have no practical influence over adoption decisions. The established system creates an appearance of inclusive governance that the actual power relations do not support. Salgado-Criado et al. (2024) confirmed this pattern on the investor side, documenting that venture capital's attention to governance varies across investment phases. Early-stage investments receive less governance scrutiny than later rounds. AI adoption decisions in startups occur during a period when neither investors nor researchers pay close attention.

Leadership structure shapes organizational epistemology independently of who occupies leadership positions. Sandberg (2025) found that vertical leadership correlates with higher levels of masculine contest culture across all measured dimensions, while the proportion of female founders does not significantly influence those scores. Adding diverse leaders to a concentrated authority structure does not necessarily change who has their knowledge validated. Shared leadership creates conditions where alternative perspectives can gain influence. The structure of authority must shift, not just the demographics of who holds it.

McElheran et al. (2024) found that AI-adopting startup founders are better educated, more experienced, and more motivated to bring new ideas to market than non-adopting founders. These characteristics signal innovation orientation but do not guarantee attention to AI ethics or social impact. The combination of higher education, a healthy market orientation, and resource constraints may produce leaders who are technically sophisticated about how machine learning models work. Yet those same leaders lack an understanding of how those models affect communities not represented in the training data. Technical knowledge and social-impact awareness are different competencies. Startup leadership selection processes reward the first. Nothing within the startup ecosystem systematically develops the second.

This pattern becomes most visible when startups use AI explicitly to replace human reasoning and labor. Companies that position themselves as "lean because of AI," operating with 30 employees while claiming the productivity of 100, are making epistemological choices about whose knowledge matters. The decision to replace a human role with an algorithmic system is a decision about whose judgment is dispensable. When those choices flow through concentrated founder authority without input from the people whose roles are being eliminated or the communities those roles served, they constitute testimonial injustice operating through organizational structure. The knowledge embedded in displaced human work is viewed as less credible than the algorithmic output replacing it.

The lean startup methodology, often presented as a systematic evaluation framework, introduces its own epistemological limitations. Felin et al. (2020) critiqued the method's emphasis on rapid experimentation and customer validation, arguing that it creates a streetlight effect. Search occurs only where observation is already easy. Concerns that require time to articulate, evidence to assemble, or perspectives that are missing become structurally invisible among current stakeholders. The method itself produces hermeneutical gaps by design. It is not that leaders choose to ignore concerns about bias. The evaluation framework they use cannot detect them.

Speed culture compounds the streetlight effect by treating deliberation itself as an obstacle. Chesterman (2021) tracked the phrase “move fast and break things” from its origins as an internal Facebook motto, featured on office posters and in Mark Zuckerberg’s 2012 letter to investors. Then, it was adopted as a general principle of technological disruption, embraced by countless Silicon Valley imitators. What began as one company’s internal culture became the ideological foundation for an entire industry’s approach to governance, evaluation, and deliberation. Quandt and Klapproth (2025) extended this analysis, documenting how Silicon Valley ideology produces what they term “hermeneutical closure.” This is a coherent system of beliefs in which disruption functions as a moral necessity, meritocracy serves as a legitimating mythology, and speed operates as a cardinal virtue. Within this system, governance concerns are not simply deprioritized. They are structurally inarticulate. The concepts needed to question AI neutrality are unavailable within the dominant interpretive system. That absence is itself a form of hermeneutical injustice operating at the cultural level.

Resource boundaries operate as epistemic constraints, shaping not only what organizations can implement but also what knowledge they can produce. Hopkins and Booth (2021) found that leadership in smaller firms often combined strong intuition with low data literacy, creating incompatibility between what leaders demand and what AI systems can deliver. Practitioners in resource-constrained settings accepted opaque models and relied on informal testing approaches because comprehensive interpretability tools were too costly to operationalize. When interpretability is unaffordable, the knowledge that interpretable systems would generate remains permanently unavailable. The organization does not reject that knowledge. It never produces it. The gap is invisible because no one learns what they would have discovered.

Demographic homogeneity magnifies these organizational factors. Wise et al. (2022) found a significant positive association between startup team ethnic diversity and investment capital raised, suggesting that diverse teams have access to resources unavailable to homogeneous teams. Yet, founders consistently prefer homogeneous teams. The gap between what homogeneous teams prefer and what diverse teams achieve suggests that founders may lack interpretive frameworks to recognize how team composition constrains their organizations. They choose sameness not from evaluated tradeoffs but because the advantages of difference are invisible from inside a uniform perspective. Chowdhury (2005) directly examined demographic diversity in entrepreneurial teams and found that it did not predict overall team effectiveness. The most revealing finding, however, was methodological: ethnic diversity data could not be analyzed because there was negligible variability within teams. High-tech entrepreneurial teams were so demographically uniform that diversity effects could not even be measured. The absence of measurable diversity is itself a finding. It reveals the depth of homogeneity in startup leadership and raises the question of what perspectives are structurally unavailable to teams that lack demographic variation.

AI Systems as Sites of Encoded Injustice

Critical AI scholarship offers empirical evidence that the epistemological dynamics described above have concrete consequences when embedded in technical systems. The assumption that algorithmic systems transcend human subjectivity has deep roots in the development of computing and statistics, where mathematical formalization was equated with objectivity. Joyce et al. (2021) observed that contemporary AI builds on a longer historical fascination with intelligent machines, and that sociologists recognize that what counts as data is socialized, politicized, and multilayered, because data about humans is also data about structural inequalities related to gender, race, and class. Yet, the technical complexity of these systems, combined with their mathematical foundations, creates what scholars have termed algorithmic authority: a presumption that computational outputs are more reliable than human assessment precisely because they appear to transcend human subjectivity.

Safiya Noble (2018) documented how commercial search engines systematically reinforce racism and sexism through their algorithmic design, arguing that the mathematical formulations underlying automated decisions reflect human choices, despite common perceptions of neutrality. Noble introduced the concept of technological redlining to describe how digital systems replicate patterns of historical discrimination. Virginia Eubanks (2018) extended this analysis to automated public service systems, demonstrating how algorithmic decision-making in welfare, housing, and child welfare disproportionately influences poor and working-class communities. Eubanks identified these systems as empathy overrides that allow decision-makers to avoid direct engagement with difficult social realities by delegating moral choices to machines. Catherine Stinson (2022) provided theoretical grounding for the claim that algorithmic neutrality is impossible in principle, demonstrating that joint filtering algorithms introduce bias through their mathematical design, independent of training data. Those who already encounter diminished credibility in human interactions face additional disadvantage from systems that treat their data as statistical outliers.

Russo et al. (2024) proposed a further epistemological adjustment to how AI systems are understood. AI systems are not merely value-laden, passively containing the biases of their designers. They are value-promoting, actively shaping what counts as relevant knowledge and whose experiences the system can recognize and respond to. This distinction matters to understand epistemological gatekeeping. When a startup adopts an AI system, it does not simply inherit static biases. It deploys a system that actively promotes certain ways of knowing while discounting others. The system continues the gatekeeping work after the adoption decision is made, encoding the founder's epistemological assumptions into ongoing organizational practice.

Generative AI systems introduce epistemological challenges that go beyond those of earlier predictive models. Newstead et al. (2023) examined AI-generated leadership reports and found that nearly half of the women leaders mentioned were characterized negatively, compared to only a quarter of men. Gender-neutral prompts generated content with no examples of women leaders. Gorska and Jemielniak (2023) documented similar patterns in AI-generated professional images: across nine text-to-image generators, men accounted for 76% of images, while women accounted for only 8%. These results show that generative AI does not simply reproduce existing bias through training data and optimization choices. It actively produces new content that reinforces credibility patterns, presenting skewed representations as authoritative while offering no visible mechanism for interrogating the assumptions embedded in the output. The interpretive gap is wider in generative systems because the connection between input assumptions and output content is less transparent than in traditional predictive systems. The Stanford AI Index Report (2025) confirmed that even language models trained to be explicitly unbiased continue to demonstrate implicit bias, suggesting that deliberate debiasing efforts within well-resourced organizations have fallen short. If large companies with dedicated research teams cannot eliminate bias through intentional effort, the prospect of resource-constrained startups achieving better results without formal governance structures is remote.

The field of technical bias mitigation has developed interventions at every stage of the development pipeline. Pre-processing approaches modify training data. In-processing approaches incorporate fairness limitations during model training. Post-processing approaches adjust outputs after training is complete. Yet different fairness metrics are mathematically incompatible, and achieving one form of fairness frequently requires sacrificing another (Mehrabi et al., 2021). Someone must decide which definition of fairness to apply, which harms to focus on, and whose interests to weigh more heavily. Those choices are value-laden, not technical. When startup leaders serve as epistemological gatekeepers without input from affected communities, they embed their own assumptions concerning fairness into systems designed to address bias. The apparatus of bias mitigation becomes a vehicle for the concentrated decision-making it was intended to address.

Resource-constrained startups face an additional layer of inherited gatekeeping. Hopkins and Booth (2021) documented that smaller organizations rely on opaque third-party AI systems because they lack the data and talent to build their own. Adopting a third-party system means inheriting the knowledge-validation practices that guided its development. Whatever communities were excluded from evaluation, whatever fairness definitions were selected, and by whom: those decisions transfer invisibly to the adopting startup. The startup leader makes governance choices about a system whose epistemological foundations are opaque to them.

Ethics frameworks intended to address these challenges may create additional complications. The AI ethics field has produced numerous frameworks, principles, and guidelines since 2016, yet Schiff et al. (2020) found that many of these documents treat civic values as discrete and rarely explain how they connect. This disaggregation allows organizations to claim commitment to specific principles without developing integrated approaches, a pattern critics have termed ethics washing. Drage et al. (2024) distinguished responsibility from response-ability, finding that current practices encourage procedural compliance rather than honest engagement with whose knowledge informs AI decisions. Camacho Ibanez and Villas Olmeda (2022) found that regulation drives compliance rather than governance, resulting in minimal adherence to requirements rather than substantive organizational transformation. When startups adopt lightweight ethics checklists in response to external pressure, these mechanisms may satisfy formal requirements without changing whose viewpoints shape adoption decisions.

Conclusion and Recommendations

This review integrated three bodies of scholarship to examine how founder and C-suite leaders in technology startups function as epistemological gatekeepers in AI adoption decisions. The central finding is that epistemological gatekeeping in startups is not a matter of individual bias or isolated poor judgment. It is a fundamentally structural phenomenon. Concentrated founder authority, informal governance, speed culture, entrepreneurial cognitive patterns, resource restrictions, and demographic homogeneity operate as an interconnected system. That system determines whose knowledge shapes AI decisions before any distinct claim is evaluated on its merits.

Several important insights emerge from this summary. First, the mechanisms of epistemic injustice in startup AI governance are organizational, not only interpersonal. Inferential inertia prevents scattered observations from aggregating into evidence. Temporal credibility shifts ensure that warnings receive less weight before harm is visible than after. Hermeneutical gaps disable employees from articulating concerns in terms that organizational processes can recognize. These mechanisms interact: information fails to aggregate partly because early warnings lack credibility, partly because employees lack the vocabulary to express their concerns, and partly because whatever procedural mechanisms exist satisfy formal requirements without changing whose perspectives inform decisions.

Second, the assumption of AI neutrality in startup cultures is not a belief that founders arrive at through examination. It is a default produced by an ideological system that lacks the interpretive resources to question it. Speed culture, meritocracy mythology, and founder authority combine to create conditions of hermeneutical closure in which governance concerns cannot be articulated in terms recognized as legitimate by the prevailing framework. Addressing this requires more than individual awareness. It requires changing the interpretive resources within the organization's culture.

Third, bias mitigation is an epistemological problem that technical solutions alone cannot resolve. Every approach to addressing AI bias depends on a prior decision about what counts as bias, which fairness definitions apply, and whose experiences constitute evidence. When those prior decisions flow through concentrated authority structures without broad input, mitigation efforts reproduce the gatekeeping they are designed to correct.

The literature reviewed here suggests several directions for organizational practice, each grounded in specific findings. Distributed leadership structures reduce the concentration of credibility assessments in a small number of individuals (Sandberg, 2025). Ethical leadership that models openness to dissent creates conditions in which employees are willing to surface concerns about AI systems before they become crises (Brown et al., 2005). Formal channels for aggregating observations about AI system performance address inferential inertia by creating institutional mechanisms that coalesce scattered knowledge (Fricker, 2020). Integrating standpoint expertise alongside technical expertise in AI evaluation provides access to knowledge that homogeneous teams cannot develop on their own (Collins, 2000; Stinson & Vlaad, 2024). Moving beyond procedural compliance toward honest engagement with whose knowledge informs AI decisions requires treating ethics work as a core organizational function rather than a box-ticking exercise (Ali et al., 2023; Drage et al., 2024).

These interventions must be feasible in resource-limited environments, meaning they cannot rely on dedicated ethics teams, large-scale auditing infrastructure, or comprehensive stakeholder consultation processes that startups cannot sustain. The literature suggests that lightweight but structurally significant changes can make a difference: distributing decision-making authority so that AI adoption choices are not concentrated in a single leader (Colombo et al., 2024), creating regular forums where non-technical employees can communicate concerns about AI system performance in their own terms (Fricker, 2013), and deliberately seeking perspectives from employees whose company roles place them closest to the communities AI systems affect (Harding, 1993). None of these interventions requires large budgets. Each requires a willingness to treat epistemic diversity as a governance resource rather than an obstacle to speed.

Recommendations for Future Research

The gap identified in this review calls for empirical investigation into how startup leaders experience and enact their role as epistemological gatekeepers. The three fields synthesized here provide theoretical and contextual foundations, but no identified study has examined how founders and C-suite leaders navigate credibility assessments, knowledge validation, and governance decisions regarding AI from their perspective. Understanding how gatekeeping operates from the inside is essential to developing interventions that are conceptually grounded and practically relevant.

A transcendental phenomenological study following Moustakas (1994) would be well-suited to this investigation. Phenomenological methodology is designed to illuminate the lived experience of a specific phenomenon, in this case, the experience of holding and exercising epistemological authority over AI adoption decisions. The transcendental approach emphasizes bracketing the researcher's assumptions to attend closely to participants' descriptions of their own experience, making it appropriate for examining a phenomenon in which leaders may or may not consciously recognize their gatekeeping role.

This methodologically driven choice aligns with what the review revealed. Quantitative measurement of gatekeeping behaviors would require predefined categories that the existing literature has not established. Survey instruments would impose the researcher's framework on participants, replicating the same top-down epistemological pattern the study seeks to examine. Phenomenological interviews allow participants to describe their experience in their own terms, surfacing the meaning-making processes, tensions, and assumptions that shape how leaders navigate AI governance decisions. The approach provides access to the lived experience of credibility assessment, constraint navigation, and meaning-making that other methods cannot capture.

Semi-structured interviews with 12 to 15 founders and C-suite leaders in U.S.-based technology startups with 10 to 100 employees would provide sufficient depth for phenomenological analysis while focusing on the organizational context where gatekeeping dynamics are most visible and traceable. This methodology would surface whether leaders recognize tensions between speed and scrutiny, how they navigate competing knowledge claims about AI systems, and what organizational conditions either enable or constrain the epistemic justice of their governance practices.

Such a study would contribute both theoretical and practical knowledge. Theoretically, it would test whether epistemic injustice frameworks have explanatory power in organizational AI contexts. Practically, it would identify the specific points in startup decision-making where intervention is both most needed and most feasible. Given that startup AI adoption is accelerating and the systems adopted during formative periods encode assumptions that scale with the organization, this research has urgency that the current pace of scholarship has not yet begun to match in any meaningful way.

References

Ali, S., Constantinides, M., & Quercia, D. (2023). Walking the walk of AI ethics: Organizational challenges and the individualization of risk among ethics workers in technology companies. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 217-227.

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, *1*(3), 311-320.

Brown, M. E., Treviño, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, *97*(2), 117-134.

Browne, K., Bareis, J., Katzenbach, C., Meier, Y., Noscenzo, R., & Henin, L. (2024). Partiality, classification, and the production of bias: Navigating bias in AI within a major AI firm. *Big Data & Society*, *11*(1), 1-14.

Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, *12*(1), 9-30.

Camacho Ibanez, J., & Villas Olmeda, D. (2022). AI regulation on artificial intelligence: An analysis of the EU proposal for an AI act. *Business and Information Systems Engineering*, *64*(6), 1189-1201.

Chesterman, S. (2021). *We, the robots? Regulating artificial intelligence and the limits of the law*. Cambridge University Press.

Chowdhury, S. (2005). Demographic diversity for building an effective entrepreneurial team: Is it important? *Journal of Business Venturing*, *20*(6), 727-746.

Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). Routledge.

Colombo, M. G., Cumming, D. J., Mohammadi, A., & Rossi-Lamastra, C. (2024). Formal delegation and organizational decoupling in startups. *Strategic Entrepreneurship Journal*, *18*(2), 376-406.

Drage, E., Frabetti, F., Catlaw, T. J., & McQuillan, D. (2024). Caring about AI: Ethics of response-ability in the development and deployment of artificial intelligence. *AI & Society*. Advance online publication.

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

Fan, J. S. (2022). Startup governance in the post-Great Recession era. *Berkeley Business Law Journal*, *19*(2), 309-379.

Felin, T., Gambardella, A., Stern, S., & Zenger, T. (2020). Lean startup and the business model: Experimentation revisited. *Long Range Planning*, *53*(4), 101889.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Fricker, M. (2013). Epistemic justice as a condition of political freedom? *Synthese*, *190*(7), 1317-1332.

Fricker, M. (2020). Institutional epistemic vices: The case of inferential inertia. In E. Edenberg & M. Hannon (Eds.), *Political Epistemology* (pp. 89-109). Oxford University Press.

Gorska, A., & Jemielniak, D. (2023). Are AI-generated professional images gender-biased? An analysis of text-to-image generators. *International Journal of Management Reviews*, *25*(4), 1-18.

Harding, S. (1991). *Whose science? Whose knowledge? Thinking from women's lives*. Cornell University Press.

Harding, S. (1993). Rethinking standpoint epistemology: What is "strong objectivity"? In L. Alcoff & E. Potter (Eds.), *Feminist epistemologies* (pp. 49-82). Routledge.

Hopkins, A., & Booth, S. (2021). Machine learning practices outside big tech: How resource constraints challenge responsible development. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 134-145.

Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U., & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, *7*, 1-11.

Kraft, P. S., Günther, C., & Kammerlander, N. (2022). Overconfidence and entrepreneurship: A meta-analysis of different types of overconfidence in the entrepreneurial process. *Journal of Business Venturing*, *37*(4), 106207.

McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L., & Zolas, N. (2024). AI adoption in America: Who, what, and where. *Journal of Economics & Management Strategy*, *33*(2), 375-415.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1-35.

Moustakas, C. (1994). *Phenomenological research methods*. Sage Publications.

Nadeem, A., Abedin, B., & Marjanovic, O. (2022). Gender bias in AI: A review of contributing factors and mitigating strategies. *Journal of Computer Information Systems*, *62*(5), 1-11.

Newstead, A., Gomez, L., & Kinderman, P. (2023). AI-generated leadership: An analysis of artificial intelligence representations of women leaders. *Frontiers in Psychology*, *14*, 1215322.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Quandt, T., & Klapproth, J. (2025). The ideology of AI: How Silicon Valley ideology shapes the discourse on artificial intelligence. *New Media & Society*, *27*(1), 1-22.

Russo, F., Schliesser, E., & Jean-Quartier, C. (2024). Connecting ethics and epistemology of AI. *AI & Society*, *39*, 1485-1499.

Salgado-Criado, J., Manrique de Lara, E., & Tenca, F. (2024). Artificial intelligence and venture capital: Governance challenges and opportunities. *European Business Organization Law Review*, *25*(3), 567-601.

Sandberg, B. (2025). The impact of gender diversity and shared leadership on masculine contest culture in startups. *International Entrepreneurship and Management Journal*, *21*(1), 1-18.

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153-158.

Stanford University. (2025). *Artificial Intelligence Index Report 2025*. Stanford Institute for Human-Centered Artificial Intelligence.

Stinson, C. (2022). Algorithms are not neutral: Bias in collaborative filtering. *AI and Ethics*, *2*(4), 763-770.

Stinson, C., & Vlaad, S. (2024). Emergent expertise: The case for diverse AI development teams. *Philosophy & Technology*, *37*(1), 1-22.

Wise, S., Yeganegi, S., & Laplume, A. (2022). Startup team ethnic diversity and investment capital raised. *Journal of Business Venturing Insights*, *18*, e00318.