

Auditing AI-Washing in German Startups: A Mixed-Methods Study of Marketing Claims and Perceptions

A Pranav, Luke Hartmann, Anne Lauscher

Trustworthy AI Lab
University of Hamburg, Germany

Abstract

AI-washing is the use of ‘AI’ as a marketing term where products’ claims are exaggerated, likely to capture the hype from investors and consumers. In this paper, we examine how startups market AI and how people perceive those claims. We audited the German startup market, applying an annotation codebook of seven dimensions to 100 startups and conducting 63 expert interviews. Startups exaggerated AI capability in three forms: claims without inspectable evidence, performance numbers without methodology, and claims without acknowledged limits. The capabilities they claimed rarely matched the underlying system, and consumers who had encountered exaggerated claims before withdrew trust from later AI marketing. AI-washing has also been linked to the marketing of unethical products: 38% of the sample were flagged for moderate or serious ethical concern, in domains such as surveillance and automated decisions in regulated fields. Together these findings show how AI-washing happens: founders exaggerate capability to attract investors who reward AI branding they cannot verify, and consumers in unfamiliar domains take the marketing at face value. We recommend that startup AI claims be brought under disclosure requirements, with the codebook offered to regulators, investors, and consumer-protection bodies for routine audits.

Introduction

Media coverage and investor attention have inflated what the term ‘artificial intelligence’ promises, and startups increasingly build their products and pitches around it (Floridi 2024; Garud et al. 2025). In Germany, 45.1% of the 2025 startup cohort name AI as a core product component (Hirschfeld, Gilde, and Walk 2025). Investors drive towards AI branding: founders who present their product as AI-driven raise capital more easily than those who do not, and funding often goes towards AI-based business models (Rady et al. 2025). Many consumers treat AI as a “*near-magical capability*” rather than a bounded engineering system, and consumers with lower AI literacy overestimate what an advertised product can do (Cave and Dihal 2019; Tully, Longoni, and Appel 2025). These conditions reward what we call **AI-washing**: *the deceptive use of AI-related claims, in which a company exaggerates, misrepresents, or obscures*

the actual role of AI in its products and operations (Elhajjar and Itani 2025; Kozarkiewicz 2025).

In practice, deception in marketing happens in three ways: claiming an ability the system does not have, reselling another company’s system as proprietary technology, and exaggerating a capability without saying what the system actually does (Aditya 2001). Often, a customer or an investor cannot check what the advertised AI does behind a product, or whether it exists at all, because the marketing does not disclose the method and no independent test is available (CFA Institute 2025; Barrios et al. 2025). The claim is then believed on trust, and that trust is misplaced when the advertised capability is absent or shallow (University of Melbourne and KPMG International 2025).

Investors often pressure startups to brand their products as *AI-powered* or *smart*. Behind that label, some of these products gather and model large amounts of personal data (Zuboff 2015; Sadowski and Pasquale 2015). Such products often result in tracking, behavioural profiling, and automated decisions the user never knowingly agreed to, even as the marketing calls this personalisation or assistance (Susser, Roessler, and Nissenbaum 2019b; Nissenbaum 2004). The AI label reframes these functions as product features, so users and regulators do not subject them to the scrutiny that surveillance and profiling would otherwise attract (Mathur et al. 2019; Crawford 2021).

In this paper, we ask: *What does AI-washing look like in the marketing of German AI startups, what drives founders and marketers to produce it, and how do consumers and domain experts receive it?* We study supply, motivation, and demand together: a **scoping study** scores the marketing copy of 100 German AI startups on a seven-dimension codebook and a composite index, and 63 expert **interviews** (53 written, ten oral) examine why founders and marketers make these claims and how audiences receive them. We report three findings:

1. **Startups market capabilities current AI cannot deliver.** Startups advertised AI functions that exceed what current systems can reliably do, and applied AI to tasks a simpler approach would handle; most described their AI only in buzzwords without saying what it did (59%), and many gave AI more marketing weight than its role in the product warranted (40%).
2. **Startups exaggerate AI capability, and the exaggera-**

tion costs them consumer trust. Startups overstated capability in three recurring forms: claims with no publicly inspectable evidence (43%), performance numbers with no disclosed methodology (79% of startups making any quantitative claim), and capability claims that acknowledge no limits (76%). Consumers who had encountered exaggerated claims before reported lower trust in “AI-powered” claims.

3. **AI-washing enables ethical harm by marketing high-risk applications as ordinary AI products.** Some of the startups openly marketed surveillance, autonomous decisions in regulated domains, and behavioural tracking and manipulation as ordinary AI products. 38% of the sample were flagged for moderate or serious ethical concern (19% serious), all in categories the EU AI Act classifies as high-risk or prohibits outright.

Related work. Even before AI, startups have long exaggerated what their products can do to convince investors and customers that the venture is ahead of its market (Zott and Huy 2007; Rutherford, Buller, and Stebbins 2009). Venture capital rewards this behaviour, because valuations follow what investors anticipate a technology to become rather than what the product can demonstrably do (Kampmann 2024). Studies of AI-washing document this exaggeration in forms that range from inflated capability claims (Nyilasy and Gangadharbatla 2025; Barrios et al. 2025; Elsayed 2026) to third-party technology resold as proprietary AI (Elhajjar and Itani 2025; Kozarkiewicz 2025). Zhang, Blodgett, and Markl (2026) analysed the capability discourses of large AI companies and argued that such claims persuade by reinstating hegemonic ideas about society and identity.

Methods

We designed a mixed-methods study to explore three sides of AI-washing in the German startup market: the supply side of public marketing claims, the motivations behind those claims, and the demand side of audience perception. We combine two methods, each described in turn below: a scoping study of marketing copy, and expert interviews on motivation and audience perception. For the scoping study, we annotated 100 German AI startups against a seven-dimension codebook (D_1 – D_7) that scores how a startup describes its AI in marketing copy and what the underlying product actually does, and aggregated the per-dimension scores into a composite AI-washing index. To explain why these claims are made and how audiences perceive them, we conducted 63 expert interviews with researchers, marketers, investors, and employees of AI startups: 53 in writing and ten orally in semi-structured sessions.

Scoping study

Sample We constructed the sample from two publicly accessible directories of German startups: the Startup-Atlas¹ and the AI Startup Landscape maintained by the appliedAI

Institute². We filtered the Startup-Atlas by the IT and Communication domain and the Artificial Intelligence tag, and drew a random subsample from the AI Startup Landscape to reduce selection bias. Inclusion required a German registered office, an AI claim in public marketing, a functional website with substantive content at the time of annotation, and pre-IPO (Initial Public Offering) startup status. Applying these criteria yielded a final sample of 100 startups.

The first two authors served as annotators. Both have combined backgrounds in business administration, marketing, AI, and AI ethics, matching the codebook’s mix of technical and communication dimensions.

Codebook We developed the codebook through iterative open coding rather than top-down specification, drawing on Birhane et al. (2022) for the dual-annotation protocol and on Bommasani et al. (2024) and Liesenfeld and Dingemanse (2024) for multi-dimensional company scoring. The development workflow ran in five steps:

- Both annotators independently reviewed approximately 30 startups without a predefined scheme and recorded qualitative observations on AI claims, marketing language, technical transparency, evidence of functionality, and ethical implications.
- Comparing the two sets of observations surfaced five recurring patterns: (i) “AI” used as a branding label with no detectable AI functionality, (ii) third-party APIs (most often the OpenAI API) repackaged as proprietary products, (iii) quantitative performance claims made without methodology or source, (iv) headline marketing that promised more than the technical descriptions on product subpages supported, and (v) ethically loaded applications obscured behind positive AI framing.
- We formalised these patterns into seven coding dimensions (D_1 – D_7), each with anchored rating scales and boundary-case decision rules.
- Both annotators dual-coded a 10-startup calibration set and revised anchor wording for cases that produced disagreement.
- Each annotator then independently coded 50 startups, with a separate 10-startup overlap set dual-coded; Cohen’s κ (macro-averaged) on this set was 0.82.

We treated a product as containing AI where its advertised function would use learned models, including large language models, and not where it runs on fixed rules, keyword matching, or scripted workflows. The substance dimension (D_2) then grades where that AI comes from: proprietary models at the top of the scale, adapted or aggregated third-party models in the middle, a single wrapped API near the bottom, and no detectable AI at the floor.

After the seven dimensions were coded, each startup received a composite AI-washing score from 1 to 5 ($\kappa = 0.86$ on the score). The composite is a single overall judgment anchored at 1 (transparent, proportionate, evidenced, and ethically sound) and 5 (misrepresentation across several dimen-

¹<https://www.startup-atlas.de/startups> (Accessed in early 2026)

²<https://www.appliedai-institute.de/ressourcen/ai-startups-deutschland/> (Accessed in early 2026)

Dim.	Scale	One-line definition
D_1	1–5	AI Claim Specificity: how concretely the startup describes what its AI does, from buzzword to fully disclosed model and method.
D_2	1–5	AI Capability Authenticity: technical substance behind the AI label, from no detectable AI to proprietary models with disclosed architecture.
D_3	1–5	Evidence of AI Functionality: what an external observer can verify, from no evidence to independently validated benchmarks.
D_4	0–3	Quantitative Claim Substantiation: whether numeric performance claims are backed by source or methodology; 0 indicates no quantitative claim was made.
D_5	1–5	Marketing Proportionality: match between the prominence of AI in marketing and its role in the product.
D_6	0–3	Ethical Concern Level: ethical risk of the AI application itself, coded independently of marketing behaviour.
D_7	1–4	Transparency About Limitations: whether the startup discloses what its AI cannot do, including accuracy bounds and human-in-the-loop dependencies.

Table 1: Codebook summary.

sions, in some cases with ethical obfuscation). Ethical concern (D_6) raises the composite where AI-washing language normalises a problematic application, and strong functionality evidence (D_3) or disclosed limitations (D_7) lower it for an otherwise concerning profile; annotators verified the provisional composite against these adjustments before finalising.

Table 1 summarises the seven dimensions. We additionally recorded, by manual inspection, whether each website carried text or imagery showing signs of AI generation. We did not use automated AI-content detectors, because current detectors are not reliable enough for evidentiary use (Sadashivan et al. 2023). This observation supports later qualitative discussion but does not feed the composite index.

For the Ethical Concern dimension (D_6), annotators scored from 0 (no apparent concern) through 1 (minor: potential for misuse or minor privacy implications), 2 (moderate: personal-data processing or behavioural tracking), to 3 (serious: surveillance, emotion detection, discriminatory potential, or an application that meets a high-risk category of the EU AI Act). A score of 3 was grounded in the EU AI Act’s (European Parliament and Council of the European Union 2024) high-risk definitions, and the ethical subcategory (surveillance, privacy, manipulation, discrimination, lack of human oversight, or vulnerable persons) was recorded alongside the score.

Annotation procedure For each startup, the annotator followed a multi-source workflow that took roughly one hour to complete:

- Read the homepage and priority subpages (/about, /technology, /how-it-works, or local equiva-

lents).

- Interacted with the product where a free trial or live demo was available, to verify whether the marketed AI behaviour was observable rather than merely described.
- Reviewed company and founder activity on LinkedIn, X, and other public social media for product descriptions, founder background, and shifts in framing over time.
- Consulted external sources (Crunchbase, Tracxn, investor press releases) for funding stage, team size, and prior product positioning.
- Scored each dimension only against the codebook anchors and only on the basis of publicly visible material; coded N/A with a reason where a dimension could not be scored.

Each annotation also recorded fixed metadata for later cross-sectional analysis: sector, product type, business model, founding year, location, funding stage, team size, founder background, and whether “AI” appears in the company name.

Beyond scores and metadata, several free-text fields recorded substantive qualitative evidence about each startup:

- **AI-washing hints:** we annotated exaggeration signals lifted verbatim from each website, including superlative capability claims, vague autonomy or intelligence boasts, and unsubstantiated quantitative comparisons.
- **Ethical concerns:** we annotated the type of application when it met the EU AI Act’s high-risk criteria, including surveillance, biometric categorisation, credit assessment, and behavioural manipulation, regardless of how the marketing dimensions were scored.
- **Value proposition and product description:** we annotated the headline value proposition as advertised on the homepage alongside a sober description of what the product actually does.
- Finally, we documented edge cases and scoring difficulties that fed into periodic codebook refinement.

Sample composition The 100 startups divide by business model into 77 Business-to-Business (B2B), 15 mixed Business-to-Business and Business-to-Consumer (B2B/B2C), 7 Business-to-Consumer (B2C), and 1 Business-to-Business-to-Consumer (B2B2C). Sectors span finance, e-commerce, healthcare, marketing technology, legal technology, entertainment, industrial automation, logistics, and human resources. Product types break down into 49 software-as-a-service tools, 30 platforms, 14 hardware-software combinations, 5 consultancies, and 2 API products. Founding years range from 2014 to 2025; 50 of the 100 startups were founded in or after 2022. Berlin (29) and Munich (19) form the largest geographic nodes, followed by Hamburg (8) and Stuttgart (5).

Definitions In this paper, we use the following conventions. $\rho(D_X, D_Y)$ denotes the Spearman rank correlation between dimensions D_X and D_Y , and $\rho(D_X, \text{composite})$ the correlation between a dimension and the composite score, computed over the full sample of 100 startups unless

a smaller n is reported. Dimension scales run in different directions (a low D_5 but a high D_6 both indicate greater severity), so for correlations with the composite we align each dimension so that higher values indicate greater AI-washing severity; a positive ρ therefore always means that greater severity on the dimension accompanies a higher composite score. Mann-Whitney U denotes a two-group rank comparison, which we use to compare annotation scores between founder-background groups and Likert ratings between interview-participant groups. All p -values are two-sided, and all group comparisons are exploratory. We write $s \in \{4, 5\}$ to state that a score s takes one of the values listed in the braces. The *high-composite* group denotes the 39 startups with composite scores $\in \{4, 5\}$; all other distributional statistics use the full sample unless stated otherwise.

Interviews

The 53 written interviews followed a structured question set that recorded how people encounter and react to AI marketing claims. We sought participants working in software development, sales and marketing, or management to document the internal pressures behind those claims. The instrument collected demographic information, self-rated AI familiarity (1–5), four 5-point Likert items, two binary items on encounters with AI-washing, and four open-text prompts: an open definition of AI, examples of startup AI claims, an account of any personal AI-washing experience, and a closing comment on AI marketing. The two binary items asked whether the participant had encountered German startups advertising AI capabilities and whether they had personally encountered what they would consider AI-washing as a consumer or at work; the latter item defines the encountered-versus-not grouping used in the analysis of the written interviews. The first Likert item measured reaction to the label “AI-powered” (1=very sceptical, 5=very trustful). The remaining three asked whether current AI is capable of reliably and ethically deploying surveillance, of making complex autonomous decisions, and of taking over the role of therapists or psychologists.

Recruitment and language. We recruited participants through social media such as LinkedIn and Bluesky. Recruitment calls went out in both English and German, and we ran two language versions of the instrument in parallel. German free-text responses were translated into English before analysis. We obtained 53 complete responses.

Sample composition. The 53 written-interview participants are AI-literate professionals and graduate students. 68% self-rated their AI familiarity at 4 or 5 out of 5, 77% held a Bachelor’s or Master’s degree, and 55% fell in the 25–34 age bracket. 47% work or study in technical computing (computer science, NLP, machine learning, data analytics), 17% in business, finance, or marketing, and 36% in journalism, media, regulation, health care, and other domain-specialist roles. This is not a probability sample of German consumers, and we do not treat it as one. The findings describe how AI-literate professionals, the audience best equipped to check an AI claim against the underlying technology, react to AI marketing.

Oral interviews. We recruited ten additional participants through purposive sampling whose roles span sales, marketing, programming, and management in AI startups, drawing on the authors’ professional networks. Interviews ran 30–60 minutes online by video call. Each session was audio-recorded with the participant’s informed consent and transcribed verbatim. Three interviews conducted in German were translated into English (manually, by a native speaker) and spot-checked against the original transcripts. The interview guide opened with background and AI-familiarity questions and then moved into perceptions of AI startups, first-hand encounters with AI-washing, organisational drivers of AI marketing claims (investor pressure, internal visibility incentives, regulatory gaps), and views on remediation.

Analysis. For the written interviews, we report per-item distributions, means, and the proportion of responses in the lower two and upper two Likert categories. After transcribing each oral interview, the first and second authors independently open-coded the transcripts (Strauss and Corbin 1998). Over three weeks, these authors met to discuss shared patterns across the interview codes and written responses. Throughout, we cite interview participants as P1–P63.

Limitations. The annotation sample covers only German startups, so the findings do not generalise to other jurisdictions. Annotation drew on marketing copy, public demos, and free trials where available, rather than on sustained hands-on product use, so scores largely reflect what firms chose to show. The interview sample is small, so it cannot show what a full-population sample would: lower AI literacy predicts greater receptivity to AI claims (Tully, Longoni, and Appel 2025), so how general consumers receive the same claims remains open.

Finding 1: Startups Market Capabilities Current AI Cannot Deliver

Startups in the high-composite group marketed AI capabilities that current AI systems cannot reliably deliver. Often startups applied AI to product tasks a simpler, deterministic approach would handle more reliably. 59% of the startups described their AI only in buzzwords or superficial terms, without detail about what the AI did [$D_1 \in \{1, 2\}$]. 35% had no actual AI, or relied on an LLM (Large Language Model) wrapper without proprietary contribution [$D_2 \in \{1, 2\}$]. 40% gave AI more marketing weight than its role in the product warranted [$D_5 \in \{1, 2\}$]. Startups whose AI was based on an LLM wrapper were significantly more likely to overstate AI’s role in their product ($\rho(D_2, D_5) = +0.51$, $p < 0.001$, $n = 99$, the strongest pairwise correlation in our codebook). In table 2 we show verbatim public-facing claims from six startups in this group.

Buzzword AI labels without a description of the AI. Startups marketed their AI under labels such as “AI credit assessment”, “real-time AI for the contact centre”, “agentic banking”, and “intelligent autonomy”. None said which part of the product was the AI, or how it differed from rule-based automation or an off-the-shelf model. 31% of the

A legal-research startup. “We develop AI without hallucinations: our AI invents nothing.” [transl.]

A digital-transformation advisory startup. “Our AI provides solutions & best practices across your entire company, enabling smarter decisions without workshops, consultants or IT projects.”

A bio-robotic surveillance startup. “We release cockroaches equipped with a custom-built backpack for control, sensing, and secure communication, enabling precise navigation and real-time data collection.”

A call-centre voice-AI startup. “Digital call center workers are trusted by enterprises to deliver the same or better customer satisfaction as human agents. At scale.”

A specialty-chemicals formulation startup. “We develop AI Twins for complex chemical formulations and materials based on our Deep Knowledge Integration approach.”

An automotive voice-AI startup. “We develop Digital Employees that could think, talk, and act alongside their human counterparts.”

Table 2: Examples of marketing copy showing the exaggerated claims (composite ≥ 4). German quotes are translated and marked [transl.].

startups carried AI descriptions the annotator coded as unclear or technically unspecific. Startups with business-only founders were significantly more likely to overstate AI capabilities than those with technical founders (Mann-Whitney $U = 287$, $p = 0.003$). They also marketed AI more prominently than its actual role in the product warranted ($U = 76$, $p = 0.002$).

P12, a written-interview participant, described their experience of encountering this type of AI-washing:

“Potential business partner claimed their tool would reduce workload by providing **AI powered solution** to assess documents submitted in risk management context. Was basically only keyword search, still required close human supervision.” (P12)

P42 saw the same pattern at consumer-goods scale, where AI labels attach to products whose functions predate the label:

“I often notice that products are marketed as AI-powered, even in areas where it makes absolutely no sense (e.g., Axe shower gel, Bosch washing machine). I think this suggests that the products are ‘intelligent’, even though the functions existed before the AI hype.” [transl.] (P42)

Once the buyer reads the AI label as persuasion, the claim loses its standing as a product fact (Friestad and Wright 1994).

AI applied to tasks that do not call for AI. We observed startups marketing AI for tasks AI is the wrong tool for, and in some cases for tasks a deterministic system would handle more reliably. Such use cases are *technological solutionism*: the assumption that the right answer to a problem must involve technology, even where simpler engineering or human practice would do (Morozov 2013; Selbst et al. 2019). One

startup advertises an outcome (no hallucinations, full source traceability) achievable by deterministic retrieval over the same source corpus but not by current LLMs; another re-frames consultancy work as a self-serve AI platform whose AI label substitutes for the work (Table 2). P31, an AI researcher among the written-interview participants, described from the user side:

“It is very frustrating when generative AI is shoe-horned into applications that it is not ready for, and thus the failure rate in using the product increases exponentially. I am not against AI in principle (I research it), but it is nascent in many of the contexts where it is currently deployed.” (P31)

Current LLMs cannot guarantee zero hallucinations, and an AI label on a consultancy would likely provide homogenised advice.

Wrappers and human labour sold as proprietary AI (D_2). 60% of startups marketed AI prominently without owning a proprietary AI system [$D_2 \in \{1, 2, 3\}$]. 72% of the high-composite group lacked proprietary AI: they had no AI at all, a third-party wrapper, or a combination of existing AI tools (28 of 39).

A LinkedIn-growth tool advertises “Leverage the power of AI for LinkedIn Growth” in the main text, whereas a footnote on the same page says the product is “powered by OpenAI’s ChatGPT 4 technology” ($D_2 = 1$, $D_5 = 1$). Startups have been building wrappers around closed-source LLMs rather than engineering systems with fine-tuned models of their own. P2, an AI researcher tracking the German startup scene, reaches the same conclusion across many such websites:

“In the end I tend to come to the conclusion: okay, that’s just a nice interface with an API key to ChatGPT, a nice function. It gets copied and done. It’s just a quick cash grab for a year or so.” [transl.] (P2)

P62 recognised the same wrapper pattern from the consumer side:

“Often, for specific cases, a prompt is simply forwarded to ChatGPT. You think it is a custom, trained model, but in the end it is just a ChatGPT context that offers no unique selling point.” [transl.] (P62)

P4 described a case where a no-code app-builder had marketed an AI assistant capable of producing finished client apps from a single prompt at a fraction of normal development cost. When the company collapsed publicly in 2025, reporting confirmed that outsourced developers had done much of the underlying coding (Orosz 2025). When a company sells human work as AI, the workers behind the system become invisible to customers, and their labour falls outside the wage and protection norms that would apply if the same work were sold as software development (Gray and Suri 2019; Casilli 2025). In both cases, what is sold as AI is something else: a third-party model the startup does not own, or human labour standing in for the advertised AI.

The AI label is attached to products whose function does not match the claim. At a former employer of P6, a

An AI-agent startup. “We develop The Autonomous Agent that gives you superpowers.”

A manufacturing-optimisation startup. “Our AI saves companies up to 96 % of production costs.”

An autonomous-retail startup. “10× lower theft than the industry average, with the highest prosecution rates.”

A field-sales-AI startup. “Thinks ahead, learns from corrections, and never drops the ball.”

A knowledge-management startup. “Your employees can ask anything and get a perfect, validated answer instantly.”

Table 3: Verbatim marketing quotes from five exemplar startups demonstrating exaggerated claims.

sustainability-reporting workflow was relabelled as “AI reporting” although the underlying work was data aggregation and dashboarding (P6, paraphrased). P5 reported an advertised capability that did not work in use:

“I cannot dial any phone number that contains a 2, because the system does not understand the way I say 2. I have never managed it. But it is advertised prominently: voice control, AI-driven, and so on. Why do we need this kind of AI when it is biased against German accents?” [transl.] (P5)

The automotive voice-AI startup in Table 2 promises “Digital Employees” that can think, which no current voice system can do [$D_1 = 2$, $D_2 = 3$, $D_5 = 2$]. In each case the advertised AI is either absent behind a relabelled non-AI workflow or unable to perform the advertised function.

Finding 2: Startups Exaggerate AI Capabilities And Customers Are Losing Trust

Finding 1 documented startups claiming AI capabilities their products do not have. Startups produce the discrepancy by exaggerating: presenting an AI capability with greater scope, certainty, or sophistication than the underlying technology supports. **Three forms of exaggeration recur across the scoping sample: capability claims with no inspectable evidence behind them, numerical claims that name no methodology, and capability claims that acknowledge no limits.** Disproportionate marketing weight (D_5), the mismatch between AI prominence in marketing copy and the actual role of AI in the product, was the strongest correlate of the composite AI-washing score ($\rho(D_5, \text{composite}) = +0.57$, $p < 0.001$). The interviews trace these patterns to financing and visibility incentives, and participants who had previously encountered exaggerated claims reported lower trust in subsequent AI marketing. Table 3 shows verbatim marketing copy from five exemplar startups.

Capability claims with no inspectable evidence. 43% of the startups claimed AI capabilities without publicly inspectable evidence behind them [$D_3 \in \{1, 2\}$], and only 2% offered evidence at the highest tier of the codebook.

We found that most websites offering any evidence presented it as client logos, anonymised case studies, or controlled product demos. None of these forms lets an outside reader audit the claimed AI capability, because the seller controls what the artefact shows: a logo proves a customer relationship rather than what the AI did for that customer, an anonymised case study cannot be checked with the source firm, and a controlled demo runs on inputs the seller selected. The startup in Row 1 of Table 3 names its product as an “Autonomous Agent” that confers “superpowers” on the user, without specifying what task the agent performs autonomously, what the superpowers consist of operationally, or how either claim could be verified. P16 reported the same absence of checkable content in the claims they had encountered: “It is often left to the reader’s imagination. Specific examples refer to data volumes or speed.” (P16). By offering self-controlled artefacts in place of independent verification, these startups foreclose the basic check a stakeholder needs to evaluate the product.

Numerical claims with no methodology. Among startups that make any quantitative claim, methodology is almost never disclosed: 79% of quantitative claimants (49 of 62) show no substantiation for their claims ($D_4 = 1$), and none are fully transparent ($D_4 = 3$). The high-composite group follows the same pattern: 95% make either no quantitative claims or only unsubstantiated ones. Among startups that make quantitative claims, weaker substantiation accompanied higher AI-washing severity ($\rho(D_4, \text{composite}) = +0.25$, $p = 0.05$, $n = 62$). The most common form is a multiplier or “up to X%” phrasing with no baseline, no customer reference, and no methodology. Rows 2 and 3 of Table 3 illustrate two examples within the pattern. The manufacturing-optimisation firm advertises a 96% saving with no comparison baseline disclosed; the autonomous-retail firm names a baseline (“the industry average”) but does not define which industry, which measurement, or which time window. P33 traces the practice to capital pressure: “AI companies are desperately seeking use cases so they can justify to their funders the capital that they have invested. Because of this, they try to make their AI approaches seems [sic] far more sophisticated and ‘human’ than they are.” (P33). Without disclosed baselines or methodology, the numbers function as branding: they invite comparison but provide nothing to compare against.

Exaggeration by hiding limitations. 76% of the startups named no edge cases, accuracy bounds, data dependencies, or human-in-the-loop requirements anywhere in their marketing material [$D_7 \in \{1, 2\}$], and only 7% reached the top disclosure tier [$D_7 = 4$]. Hiding limits works the same way as the low claim specificity (D_1) and disproportionate marketing weight (D_5) covered earlier: each makes a system sound more capable than it is. Two startups go further and promise that the AI never fails: a field-sales AI “thinks ahead, learns from corrections, and never drops the ball”; a knowledge platform invites employees to ask “anything” and receive a “perfect, validated answer instantly” (Table 3, Rows 4 and 5). Neither line names a failure mode, an accuracy bound, a domain boundary, or a human review step.

Hiding limits is a baseline practice even outside the high-composite group, where 72% of startups disclosed no limitations, and the high-composite group disclosed still less (82%; $\rho(D_7, \text{composite}) = +0.24, p < 0.05$). P21 described this from their former employment, where a product marketed as fully automated depended on a continuous human review team:

“The classification system of cybersecurity alerts I was working with was presented to clients as being fully capable of delivering automatic alerts. In reality, a whole team of humans was working each day to analyze, correct, refine and package results. I discovered the discrepancy as I was working in the company, both with the ‘refining team’ and with the commercial team selling the product as fully automated to clients. To my knowledge, no specific consequences happened to the company.” (P21)

What P21 saw is the inside view of the pattern in Rows 4 and 5: capabilities advertised as perfect did have limitations the marketing did not disclose. P9 noted that academic conferences have begun to require an explicit limitations paragraph as a structural mitigation against the same pattern in published research.

AI-generated marketing material as a correlated signal.

We observed apparent AI-generated images or text on 13% of the sites. We flagged sites whose copy carried the lexical signatures of LLM-assisted writing (Kobak et al. 2024), or whose avatars and product images looked synthetic. We observed that the use of AI-generated marketing material is correlated with AI-washing severity: 77% of the flagged sites (10 of 13) sit in the high-composite group, against 33% of the 87 unflagged sites ($\rho(\text{AI-gen flag, composite}) = +0.25, p = 0.011$). LLM-generated text tends toward superlatives and infallibility framing, the same pattern observed in the exaggerated copy in Table 3, which may explain the correlation.

The mechanism of exaggeration. Participants trace the exaggeration to two pressures: an external one from investors and public funders, and an internal one from management and marketing on the people building the product. P2, who works in German startup financing, connects the AI label to fundraising: *“AI is currently the buzzword for financing. That’s why everyone does it.” (P2)*. Participants who worked inside the firms report the internal pressure. At P3’s firm, the instruction came from management: *“We were really caught up in the hype, where the word was: AI has to go in everywhere.” (P3)*. P4, on the marketing side, saw the website claim run ahead of the product: while a small development team was still in its first three planning meetings on an AI feature, the company website already described the firm as *“AI First”*, presenting a prototype as a deployed product. P6 gives the budget logic behind such labels: top management *“would rather put extra funding to any project which has the name AI because they believe that it might bring in huge innovation” (P6)*, so employees who want funding for a project label it AI. P7 recalls the same renaming with an earlier buzzword: companies added

“Blockchain” to their names to attract investment without changing their products, and AMD now sells its processor line as *“Ryzen AI”*. Across the interviews, the AI claim comes from people who do not build the product: founders speaking to investors, top management, and marketing. The developers closest to the technical work (P1, P4) do not decide how the product is presented.

Demand-side recognition and trust withdrawal. In the written interviews, we examined how participants perceive exaggeration. Participants reported personally encountering AI-washing as consumers or in their work (53%). On the trust-in-*“AI-powered”* item, participants who had encountered AI-washing were markedly more sceptical than those who had not (encountered: median 1, mean 1.71; not encountered: median 3, mean 2.48; Mann-Whitney $U = 190, p = 0.003$). The two groups did not differ in self-rated AI familiarity ($U = 448, p = 0.069$). Participants distrust *“AI-powered”* claims after they have seen exaggeration, not because they know more about AI. P18 saw the pattern at two employers before graduate school:

“Over the course of my career, AI capabilities have always been systematically overclaimed – the benefits often oversold, and the costs understated. At both companies I was at pre-grad school, AI was often presented in ways that led with the potential, rather than the actual truth of the product being provided. AI tools were decontextualized of the training data and attendant limitations, and performance was reported in ways that did not take into account generalization to end-user usecases.” (P18)

Consumers can detect exaggeration, and once they have, they trust later *“AI-powered”* claims less. Friestad and Wright (1994) call this *persuasion knowledge*: a consumer who recognises a persuasion attempt assigns the message a different meaning. *Finding 3* examines startups whose overselling reaches domains where the application itself raises ethical concern: surveillance, autonomous decision-making, and behavioural manipulation.

Finding 3: AI-washing Leads To Ethical Risks

Startups in the high-composite group were more likely to market applications that raised ethical concerns ($\rho(D_6, \text{composite}) = +0.41, p < 0.001$). 19% of the 100 startups marketed applications raising serious ethical concerns [$D_6 = 3$], and a further 19% marketed applications raising moderate concerns [$D_6 = 2$]. All of these applications fall within categories the EU AI Act classifies as high-risk or prohibits outright (European Parliament and Council of the European Union 2024). The three subcategories with the largest counts were surveillance, autonomous decisions in regulated domains, and behavioural tracking and manipulation. Table 4 reproduces marketing copy from five exemplar startups in the $D_6 = 3$ cluster.

Surveillance. Three of the five exemplar startups market surveillance applications. The EU AI Act classifies law-enforcement and public-space monitoring as high-risk under Annex III §6 of Regulation 2024/1689 and addition-

A geospatial-AI startup (<i>Surveillance</i>) <i>“Our AI can detect unusual activities, monitor crowd movements, and identify security threats, aiding law enforcement and security agencies in maintaining public order and safety.”</i>
A parking-surveillance startup (<i>Surveillance</i>) <i>“Persons, animals, and personal data are tracked using our in-house intelligent software.”</i>
A defence bio-robotics startup (<i>Surveillance</i>) <i>“Deploy silent sensors and conduct covert surveillance in dense, unstable, and politically sensitive urban environments.”</i>
A voice-AI call-centre startup (<i>Manipulation</i>) <i>“We use LLMs, human-like voices and background call center music to deliver >90 % CSAT score.”</i>
An AI fertility-clinic startup (<i>Autonomous decisions</i>) <i>“We use our computer vision-based tool to select your highest-quality embryo based on comprehensive clinical and lab data.”</i>

Table 4: Examples of marketing copy from startups with serious ethical concerns ($D_6 = 3$).

ally restricts real-time remote biometric identification under Article 5(1)(h). Municipalities market these systems as “smart city” projects; the term often covers surveillance rather than urban infrastructure and keeps public scepticism away (Sadowski and Pasquale 2015; Kitchin 2014). Facial recognition, the technology behind many of these systems, misidentifies people at rates that differ by factors of ten or more across racial and sex categories (Buolamwini and Gebru 2018; Grother, Ngan, and Hanaoka 2019). Many such systems train on person re-identification datasets whose subjects never consented to collection (Peng, Mathur, and Narayanan 2021). As in Finding 1, the startups advertise capabilities current AI cannot deliver.

Written-interview participants doubted that current AI can be deployed reliably or ethically for surveillance. 70% rated capability in the lower two scores of a five-point scale (mean 2.17). Participants trained in computer science, AI, NLP, or business informatics rated capability lower still, at 1.70 on average against 2.45 for the rest of the sample (Mann-Whitney $U = 197, p = 0.011$).

P1 worked inside an AI-surveillance startup that sold camera-based football analysis to shopping malls; the firm pitched capabilities its cameras could not support:

“We were mainly AI systems behind it. I never understood the point of doing retail analytics with surveillance. We cannot even determine their age or gender via surveillance cameras. These cameras are really bad and they try to get your video from really far away, so you cannot really do that kind of stuff. But we were trying to oversell to the investors that we can do that kind of stuff. We were trying to oversell them like ‘okay, we can figure out who is doing robbery in your mall,’ but that’s not really possible to do that.” (P1)

Asked whether AI surveillance can currently be deployed ethically and reliably, P10 said it can be neither:

“Of course not ethically and reliably. Firstly, it does not work well. AI based surveillance have a lot of false positives as they work on low-quality security cameras. So even on a responsible level do you want to deploy AI when it does not work well at all? And also what kind of knowledge will they gain from it? It’s just a huge data mess.” (P10)

In P1’s account, only people working inside the firm could see that the marketing did not match the engineering. Buyers (police forces, security agencies, parking operators) cannot test such claims before deployment, and the people the cameras record are never asked at all. An exaggerated surveillance product therefore harms people who never bought it.

Autonomous decisions in regulated domains. Two of the five exemplar startups market AI for decisions normally made by a regulated professional: an AI fertility-clinic startup with an embryo-selection tool (the fifth exemplar in Table 4), and a startup that offers autism screening for children and an AI therapy platform. The fertility-clinic startup claims its AI makes “highest-quality” embryo selections. However, the leading commercial system in this class did not outperform trained embryologists in a randomised clinical trial (Illingworth et al. 2024).

Written-interview participants doubted that current AI can make decisions of this kind. 57% of participants rated AI as not capable of such decisions and 32% rated it capable (mean 2.57). Participants trained in computer science, AI, NLP, or business informatics gave a mean rating of 2.15 against 2.82 for the rest of the sample (Mann-Whitney $U = 230, p = 0.058$); as on the surveillance item, the technically trained group was the more sceptical one.

P23 rated AI as fully capable on this item and still opposed the use:

“I think AI is capable of making decisions like medical and legal ones but I don’t think it should.” (P23)

P1, asked in the oral interview about complex decision-making, also opposed it:

“I definitely don’t think that AI should be used to do complex decision making processes. I don’t think that AI should be used for, let’s say, crime and policing stuff, giving out loans, or determining if you have cancer or not. ... Even if you try to properly train your AI, AI will have biases. AI does not provide proper reasoning.” (P1)

The second startup screens children for autism spectrum disorder from voice recordings and pairs neurodivergent families with therapists through an LLM platform. The screening falls under Annex III §2 as a medical-device safety component; the therapy platform under Annex III §5 as access to an essential service. Participants were most sceptical on the therapy item: 81% placed AI in the lower two scores in the written interviews (mean 1.85). Current language models lack the contextual and situational grounding clinical practice requires, and audits report risks of privacy violation, sycophancy, persuasion, overtrust, and emotional dependence (Moore et al. 2025).

P17, who reported conscious daily use of generative AI, worried most about crisis scenarios:

“AI cannot replace the specifically human element of interaction. This is why claims for products like AI therapists greatly worry me. Yes, there is anecdotal evidence that some people feel better talking to them in the short term, but AI is in no way equipped to handle and direct a person in a crisis scenario.” (P17)

Privacy issues and tracking. Startups present manipulation, behavioural tracking, and AI-generated outputs built on personal data collected without consent as product features (Table 4). The call-centre product places automated outbound calls in which the synthesised voice and dialogue are AI-generated, with no disclosure to the recipient who answers. Matz et al. (2024) found that LLM messages personalised to a recipient were more persuasive than generic ones across consumer, political, and health domains. A behavioural-analytics product reads visitor behaviour during a web session and reshapes the page in real time to lift conversion, with no disclosure to the visitor; Mathur et al. (2019) catalogued interface designs of this kind as dark patterns across 11,000 shopping sites. Susser, Roessler, and Nissenbaum (2019a) call this online manipulation: hidden influence on a person’s choices, made cheap to deploy at scale by digital infrastructure.

P5, an investor and AI-strategy adviser, encountered undisclosed data collection in person at a German trade fair, where a startup offered passers-by a “psychological assessment” generated from biometric input collected on the spot. P5 asked the vendor how the system was supposed to work:

“There was one person saying: do you want a psychological assessment of yourself? I said, how is that supposed to work? You have to put your finger here. Your eye has to look in here, and then I have to put my hand somewhere else. Then I said: so you want all my biometric data and you give me some kind of three-page printout.” (P5)

The visitor hands over fingerprint, iris, and palm data on the strength of an AI label, and receives a printout for which no scientific basis exists (Stark and Hutson 2021; Barrett et al. 2019). The pitch describes the assessment and says nothing about what happens to the collected data.

P34 worked barista shifts at a multinational coffee chain that deployed two AI systems on the shop floor, a computer-vision inventory counter and a workforce-planning model:

“Shift supervisors had to increasingly use AI technologies for tasks that shouldn’t have necessarily required AI. ... they had to use AI to plan the layout of the store during our peak rush, a model that often miscalculated where people should go and led to strain on certain baristas who had to do more work than others. ... they had to use a computer vision system to do the stock inventory for them as it was supposed to make counting inventory quicker. In reality, it just made it more frustrating and longer due to the amount of false positives.” (P34)

Both systems underperformed against the marketed claim, and the underperformance made the shifts harder: the planning model dictated where each barista worked during peak rush and loaded some with more work than others, and the vision system’s false positives put the counting back on the staff. Sold as efficiency tools, the systems in practice placed workers under algorithmic decisions and monitoring during every shift, and collected tracking data about them.

AI-washing in all three cases conceals the regulatory class to which the underlying application belongs. The EU AI Act places each application in a risk category: high-risk biometric identification, regulated medical device, prohibited manipulation. None of this appears in the marketing that customers and investors read. Findings 1 and 2 showed AI-washing costing buyers money and trust; here, startups sell systems the Act treats as high-risk or prohibited, and the harm falls on the people these systems watch.

Discussion

This paper documented AI-washing in German startups across three positions: the marketing copy on startup websites, the accounts of experts who advise and invest in the market, and the responses of consumers who buy AI products. German AI startups marketed products as autonomous or proprietary AI systems when those products wrapped third-party models, ran rule-based logic, or contained no AI at all (*Finding 1*). Consumers who had encountered exaggerated claims trusted subsequent AI marketing less; startups lose the trust of the consumers who detect the exaggeration (*Finding 2*). Startups also marketed surveillance, autonomous decisions in regulated domains, and behavioural tracking and manipulation as ordinary AI products, so these applications reached customers and investors with no mention of the high-risk or prohibited status the EU AI Act assigns them (*Finding 3*). We provide recommendations for policymakers and auditors that target the marketing copy rather than the systems behind it, because AI-washing is marketing that claims more than the startup has built.

Recommendations. First, any company that commercially advertises AI capabilities should publish, on the product page, a short structured statement of what data the system collects about the people who use it and what the system cannot reliably do; datasheets (Gebu et al. 2021) and model cards (Mitchell et al. 2019) already specify the content categories such a statement can reuse, and this content is the part marketing copy omits. Second, because a disclosure statement is still self-reported, an independent body should certify commercial AI marketing claims by inspecting a system’s technical documentation and performance benchmarks under confidentiality; the EU AI Act’s conformity assessment already delegates a comparable inspection of high-risk systems to designated third parties (European Parliament and Council of the European Union 2024), and this proposal extends that assessment from the system to the marketing claim made about it. Third, both measures require deciding which startups warrant scrutiny, and the codebook can serve as the first-pass screen: a regulator or investor can score a startup’s public marketing copy without access to the

underlying system, and escalate to system-level review only the startups whose marketing the screen flags.

Conclusion. This paper documents AI-washing in the German startup market through an annotation codebook applied to 100 startups and 63 expert interviews (53 written, ten oral). We demonstrate the incentives that sustain AI-washing: founders exaggerate AI capability to attract investors who reward AI branding, investors fund products whose AI they cannot easily verify, and consumers in unfamiliar domains take the marketing at face value. The pressure that drives exaggeration also pushes some startups into applications that raise ethical concern, in domains such as surveillance, automated medical and legal decisions, and persuasion aimed at vulnerable users, where an overstated capability claim risks direct harm rather than only misdirected funding. AI-washing persists because the parties positioned to check the claims cannot currently do so at low cost: investors lack the technical detail, consumers lack the domain knowledge, and regulators lack the empirical basis. We recommend that AI claims in startup marketing be brought under disclosure requirements and routine audit, and we offer the codebook as a low-cost instrument for regulators, investors, and consumer-protection bodies to apply that audit at population scale rather than case by case.

Acknowledgments

We thank our anonymous reviewers for their feedback. We thank the participants of our study for sharing their perspectives with us. The work of Trustworthy AI lab is funded under the Excellence Strategy of the German Federal Government and States.

Ethics Statements

The interviews constitute human-subjects research, and their protocols were reviewed and approved by the ethics review board of the University of Hamburg Business School before recruitment began.

Positionality. The lead authors have backgrounds in machine learning, AI ethics, business administration, and marketing, and study the German startup market as residents of Germany working at a German university. Before this study, we worked in startups and in marketing roles, and have written and approved the kind of promotional copy this paper audits. We knew from that work that founders and marketers write capability claims under funding and management pressure, so the codebook scores what the public record shows and the interview guide asks about those pressures. The same work taught us to expect exaggeration, and we may have scored ambiguous claims more harshly than annotators without industry experience would. The anchored scales and boundary-case rules limit how far that expectation can move a score, and the interviews describe the same patterns independently of our annotation.

Annotation of public marketing materials. The scoping study coded publicly accessible marketing materials of registered German firms, and we did not collect or store any non-public data about these companies, their founders, or

their employees. We anonymise the 100 startups in the text, receipts tables, and figures, and identify each only by city, sector, and product category. The anonymisation matches the claim we make: our findings describe how often AI-washing patterns recur across the sample, not whether any single named firm has misled a specific investor or customer; that is a legal question the dual-annotation protocol cannot settle. We retain verbatim marketing quotes in the receipts tables because the wording is the analytic object and because a reader can verify the prevalence pattern by sampling the public directories. We did not contact the startups for right-of-reply, because the material is already public and our claims concern population-level prevalence rather than the conduct of any single company, following prior population-level audits of public artefacts (Birhane et al. 2022; Liesenfeld and Dingemanse 2024).

Oral interviews. The ten oral-interview participants gave informed consent before each session; consent covered audio recording, verbatim transcription, anonymised use of quotations in academic publication, and the right to withdraw at any point up to manuscript submission. We stored recordings on encrypted devices, deleted the audio files after transcription and a single accuracy check, and retained only the transcripts under participant codes with no identifying metadata. Participants discussed current and prior employers, and we redacted firm identifiers and other distinguishing details from quoted excerpts where retention would re-identify a specific organisation or person. We translated the three German-language sessions into English and spot-checked the translations against the original transcripts before quoting any passage.

Written interviews. Participants took part voluntarily and without compensation, and we recruited through open posts on LinkedIn and Bluesky with no targeted outreach to individuals. The instrument collected no directly identifying information. We quoted free-text responses under participant codes, after redacting verbatim references to current employers from the excerpts. We ran the instrument in English and German in parallel and translated German free-text responses into English for analysis.

References

- Aditya, R. N. 2001. The psychology of deception in marketing: A conceptual framework for research and practice. *Psychology & Marketing*, 18(7): 735–761.
- Barrett, L. F.; Adolphs, R.; Marsella, S.; Martinez, A. M.; and Pollak, S. D. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1): 1–68.
- Barrios, J. M.; Campbell, J. L.; Johnson, R. G.; and Liu, C. 2025. Artificially Intelligent or Artificially Inflated? Determinants and Informativeness of Corporate AI Disclosures. Working paper, Yale University School of Management / National Bureau of Economic Research.
- Birhane, A.; Ruane, E.; Laurent, T.; Brown, M. S.; Flowers, J.; Ventresque, A.; and Dancy, C. L. 2022. The Forgot-

- ten Margins of AI Ethics. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22. ACM.
- Bommasani, R.; Klyman, K.; Kapoor, S.; Longpre, S.; Xiong, B.; Maslej, N.; and Liang, P. 2024. The 2024 Foundation Model Transparency Index. *arXiv preprint arXiv:2407.12929*.
- Buolamwini, J.; and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, 77–91. PMLR.
- Casilli, A. A. 2025. *Waiting for Robots: The Hired Hands of Automation*. Chicago: University of Chicago Press. ISBN 978-0-226-82095-8.
- Cave, S.; and Dihal, K. 2019. Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence*, 1(2): 74–78.
- CFA Institute. 2025. AI Washing: Signs, Symptoms, and Suggested Solutions for Investment Stakeholders. CFA Institute Research & Policy Center Report.
- Crawford, K. 2021. *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Elhajjar, S.; and Itani, O. S. 2025. AI Washing: A Conceptual Exploration. *AMS Review*, 15: 519–538.
- Elsayed, N. 2026. AI Washing and the Erosion of Digital Legitimacy: A Socio-Technical Perspective on Responsible Artificial Intelligence in Business. *arXiv preprint arXiv:2601.06611*.
- European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (EU AI Act). Official Journal of the European Union.
- Floridi, L. 2024. Why the AI Hype Is Another Tech Bubble. *Philosophy & Technology*.
- Friestad, M.; and Wright, P. 1994. The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research*, 21(1): 1–31.
- Garud, R.; Phillips, N.; Snihur, Y.; Thomas, L. D. W.; and Zietsma, C. 2025. Hype in Entrepreneurial Settings. *Journal of Business Venturing*, 106559.
- Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Daumé III, H.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Gray, M. L.; and Suri, S. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Harper Business.
- Grother, P.; Ngan, M.; and Hanaoka, K. 2019. Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. Technical Report NISTIR 8280, National Institute of Standards and Technology.
- Hirschfeld, A.; Gilde, J.; and Walk, V. 2025. Deutscher Startup Monitor 2025. Bundesverband Deutsche Startups e.V.
- Illingworth, P. J.; Venetis, C.; Gardner, D. K.; Nelson, S. M.; Berntsen, J.; Larman, M. G.; et al. 2024. Deep Learning Versus Manual Morphology-Based Embryo Selection in IVF: A Randomized, Double-Blind Noninferiority Trial. *Nature Medicine*, 30(11): 3114–3120.
- Kampmann, D. 2024. Venture Capital, the Fetish of Artificial Intelligence, and the Contradictions of Making Intangible Assets. *Economy and Society*, 53(1): 39–66.
- Kitchin, R. 2014. The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1): 1–14.
- Kobak, D.; González-Márquez, R.; Horvát, E.-A.; and Lause, J. 2024. Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv preprint arXiv:2406.07016*.
- Kozarkiewicz, A. 2025. Artificial Deception: Identifying and Tracing the Phenomenon of “AI Washing”. *Optimum. Economic Studies*, 121(3).
- Liesenfeld, A.; and Dingemanse, M. 2024. Rethinking Open Source Generative AI: Open-Washing and the EU AI Act. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 1774–1787. ACM.
- Mathur, A.; Acar, G.; Friedman, M. J.; Lucherini, E.; Mayer, J.; Chetty, M.; and Narayanan, A. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): Article 81.
- Matz, S. C.; Teeny, J. D.; Vaid, S. S.; Peters, H.; Harari, G. M.; and Cerf, M. 2024. The Potential of Generative AI for Personalized Persuasion at Scale. *Scientific Reports*, 14: 4692.
- Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, 220–229.
- Moore, J.; Grabb, D.; Agnew, W.; Klyman, K.; Chancellor, S.; Ong, D. C.; and Haber, N. 2025. Expressing Stigma and Inappropriate Responses Prevents LLMs from Safely Replacing Mental Health Providers. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 599–627. ACM.
- Morozov, E. 2013. *To save everything, click here: The folly of technological solutionism*. PublicAffairs.
- Nissenbaum, H. 2004. Privacy as contextual integrity. *Washington Law Review*, 79(1): 119–158.
- Nyilasy, G.; and Gangadharbatla, H. 2025. AI-washing: The Asymmetric Effects of Its Two Types on Consumer Moral Judgments. *ArXiv*, abs/2507.04352.
- Orosz, G. 2025. Builder.ai did not “fake AI with 700 engineers”. *The Pragmatic Engineer*. Blog post.
- Peng, K.; Mathur, A.; and Narayanan, A. 2021. Mitigating Dataset Harms Requires Stewardship: Lessons from 1000

Papers. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1.

Rady, J.; Townsend, D.; Hunt, R.; and Simpson, J. 2025. The Expectations Game: The Contingent Value of Hype as a Rhetorical Strategy in Resource Mobilization Processes Among AI Startups. *Journal of Business Venturing*, 40(4): 106499.

Rutherford, M. W.; Buller, P. F.; and Stebbins, J. M. 2009. Ethical Considerations of the Legitimacy Lie. *Entrepreneurship Theory and Practice*, 33(4): 949–964.

Sadasivan, V. S.; Kumar, A.; Balasubramanian, S.; Wang, W.; and Feizi, S. 2023. Can AI-Generated Text be Reliably Detected? arXiv preprint arXiv:2303.11156.

Sadowski, J.; and Pasquale, F. 2015. The spectrum of control: A social theory of the smart city. *First Monday*, 20(7): 5903.

Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; and Vertesi, J. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*, 59–68.

Stark, L.; and Hutson, J. 2021. Physiognomic artificial intelligence. *Fordham Intell. Prop. Media & Ent. LJ*, 32: 922.

Strauss, A.; and Corbin, J. 1998. *Basics of qualitative research techniques*. Thousand oaks, CA: Sage publications.

Susser, D.; Roessler, B.; and Nissenbaum, H. 2019a. Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review*, 4(1): 1–45.

Susser, D.; Roessler, B.; and Nissenbaum, H. 2019b. Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2).

Tully, S. M.; Longoni, C.; and Appel, G. 2025. Lower artificial intelligence literacy predicts greater AI receptivity. *Journal of Marketing*, 89(5): 1–20.

University of Melbourne; and KPMG International. 2025. Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025. KPMG International Report.

Zhang, J.; Blodgett, S. L.; and Markl, N. 2026. “A Natural Language Interface for Everyday Life”: The Social and Political Functions of AI Capabilities Discourses. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’26. ACM.

Zott, C.; and Huy, Q. N. 2007. How Entrepreneurs Use Symbolic Management to Acquire Resources. *Administrative Science Quarterly*, 52(1): 70–105.

Zuboff, S. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1): 75–89.