

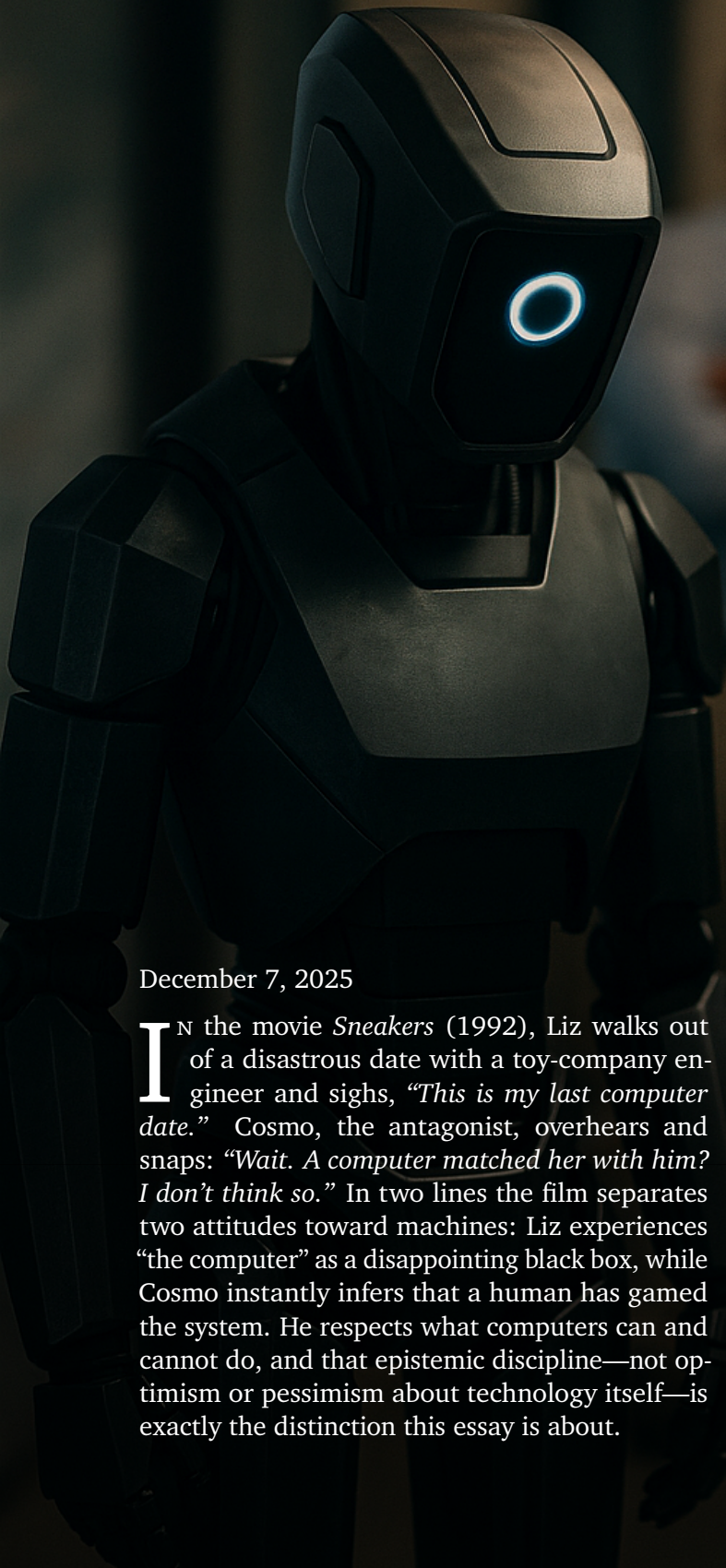
Talking Down The Machine

Status, Attribution, and Coordination in Responses to Generative AI

Michel Blomgren

Software Engineer & Solution Architect

sa6mwa@gmail.com — pkt.systems



December 7, 2025

IN the movie *Sneakers* (1992), Liz walks out of a disastrous date with a toy-company engineer and sighs, “*This is my last computer date.*” Cosmo, the antagonist, overhears and snaps: “*Wait. A computer matched her with him? I don’t think so.*” In two lines the film separates two attitudes toward machines: Liz experiences “the computer” as a disappointing black box, while Cosmo instantly infers that a human has gamed the system. He respects what computers can and cannot do, and that epistemic discipline—not optimism or pessimism about technology itself—is exactly the distinction this essay is about.

TL;DR: This essay asks why people loudly mock “dumb” AI at the same time as they stream AI-generated music, ship AI-assisted code, and quietly depend on algorithmic systems in everyday work. It argues that much of the trash talk is not about model quality at all, but about human status, identity, and control: AI is held to unrealistically high standards, its rare but salient mistakes are used to justify public skepticism, successes are re-attributed to the human operator, and office banter becomes a way of reassuring one another that we are still in charge. Under the surface, however, teams that treat AI as a fallible but powerful partner—designing for calibrated trust, verification, and clear accountability—already operate at a different level of competence. The essay maps the psychological mechanisms behind the paradox and sketches how adopting this “centaur” style of work can turn noisy resentment into a durable strategic advantage, for individuals and organizations willing to lean into the new coordination problem instead of denying it.

CONTEXT

In many workplaces the same people who denounce an AI system as “dumb” when it produces a bizarre or context-misaligned output will, minutes later, accept an AI-assisted draft as “good enough”—or even superior to what a typical colleague would have produced. This looks like hypocrisy. It is better understood as a *dual-economy problem*: humans evaluate AI in (i) a **performance economy** (does this output help me ship?) and (ii) a **status economy** (what does this imply about my competence, autonomy, and rank?). The two economies reward different behaviors, and the resulting equilibrium can be: *private adoption plus public derogation*.

The coordination literature on human–AI work suggests why this equilibrium is stable. As tools reduce the marginal cost of producing drafts, options, and code, the bottleneck shifts from “*can we produce?*” to “*can we coordinate meaningfully?*” (Blomgren, 2025; Hutchins, 2001). In that regime, talk about AI is not merely opinion; it is a coordination signal about identity, norms, and accountability.

The remainder of this short essay offers an integrated explanation: status threat drives derogation, cognitive biases amplify visible failures, attribution processes reallocate credit and blame, and social norms govern what can be admitted publicly.

MECHANISM I: STATUS AND IDENTITY DEFENSE

Status threats create predictable derogation. Professional identity is sustained through comparison and distinctiveness (Festinger, 1954; Ashforth and Mael, 1989). Generative AI compresses visible differences in output quality, which is precisely the substrate that many knowledge-work status hierarchies are built on. When an external agent suddenly produces credible artifacts, the simplest identity-preserving move is not to deny the artifact; it is to deny the agent: “the tool is worthless” or “it is just stochastic parroting.”

This is a classic identity-threat response. When an identity-relevant domain is threatened, individuals and groups engage in defensive sensemaking, including derogation of the threatening source (Petriglieri, 2011). At the societal level, system-

justification research predicts the same pattern: people, including those disadvantaged by the status quo, often bolster existing hierarchies via rationalizations that make change feel illegitimate or unsafe (Jost and Banaji, 1994; Jost, Banaji, and Nosek, 2004). In organizations, that maps cleanly onto: “real engineers write code” versus “prompting is not engineering.”

Ambivalent stereotypes explain the emotional texture. The stereotype content model predicts that targets perceived as *high competence*, *low warmth* elicit envy and resentment rather than admiration (Fiske et al., 2002). Generative AI is commonly framed as extremely capable but socially cold and morally ungrounded. In this frame, disparagement is not an error; it is the socially licensed affect that accompanies perceived competitive threat.

MECHANISM II: ERROR SALIENCE, INFLATED EXPECTATIONS, AND ALGORITHM AVERSION

If identity threat supplies the motive force, cognitive biases supply the fuel.

Bad is stronger than good. Humans overweight negative, salient events (Baumeister et al., 2001). A single hallucinated API call can dominate memory more than ten correct refactor suggestions. This is not unique to AI; it is a general feature of attention and learning under uncertainty (Tversky and Kahneman, 1974). But the effect is magnified because AI errors are often *confidently phrased* and therefore socially embarrassing when repeated.

Machine heuristic raises the standard, then punishes deviation. People carry a “machine heuristic”—a shortcut expecting mechanical systems to be objective, consistent, and correct (Yang and Sundar, 2024). That heuristic creates an asymmetric threshold: minor human errors are *normal*, minor AI errors are *disqualifying*. The empirical algorithm-aversion literature captures this: after observing an algorithm err, people avoid it even when it remains objectively better than human judgment (Dietvorst, Simmons, and Massey, 2015).

The paradox is compatible with algorithm appreciation. Humans can simultaneously dis-

trust *delegation* to algorithms and appreciate *advice* from algorithms (Logg, Minson, and Moore, 2019). In practice, office behavior often becomes: “I will use the model as a drafting engine, but I will not grant it authority.” This mixed stance can look like contempt paired with dependence, but it is often a crude attempt at appropriate reliance (Lee and See, 2004).

Control is the release valve. Providing even slight ability to modify an algorithm’s output increases willingness to use it (Dietvorst, Simmons, and Massey, 2018). This aligns with autonomy theory: perceived loss of control evokes reactance, and reactance predicts resistance and derogation (Deci and Ryan, 2000; Rains, 2013; Dillard and Shen, 2005). “Back talk” is frequently a linguistic form of reactance aimed at reasserting agency: *I am still the principal; the machine is the servant.*

MECHANISM III: ATTRIBUTION, BLAME, AND MORALIZATION

The most important asymmetry is not technical; it is *moral*.

Attribution shifts credit to self and blame to other. Self-serving attribution biases are large and robust (Mezulis et al., 2004), and attribution theory explains how people assign responsibility to protect self-worth (Weiner, 1985). In AI-assisted work this yields a stable pattern:

- If the result is good, the human “used the tool well” (credit to operator).
- If the result is bad, the tool “is dumb” (blame to system).

This asymmetric mapping is socially useful because it lowers reputational risk for adopters: one can take credit for wins while externalizing losses.

Mind perception turns engineering artifacts into moral actors. Humans readily apply social cognition to nonhumans (Nass and Moon, 2000; Waytz, Epley, and Cacioppo, 2010). When an AI is perceived to have agency and experience, it becomes eligible for moral appraisal (Gray, Gray, and Wegner, 2007). Moral typecasting further predicts that agents are judged primarily as doers (eligible for blame) rather than as patients (eligible for sympathy) (Gray and Wegner, 2009). The result: users feel licensed to ridicule and punish the system for “transgressions”.

Recent evidence makes the scapegoating mechanism explicit: perceiving more “mind” in AI increases blame directed toward AI and can reduce blame toward human stakeholders (Joo, 2024). In other words, anthropomorphism does not guarantee empathy; it can increase punitive blame.

Moral domains amplify aversion. People are especially resistant to machines making moral decisions (Bigman and Gray, 2018; Haidt, 2001). This matters in office settings because professional identity is moralized: craftsmanship, originality, and responsibility are treated as sacred values. Once the discourse becomes sacred, technical counter-arguments (benchmarks, defect rates) become less persuasive than signals of loyalty and purity.

MECHANISM IV: PUBLIC TALK VS NORMATIVE COORDINATION

A final ingredient explains why even “smart” people participate in disparagement while privately benefiting.

Pluralistic ignorance makes the room louder than the truth. Groups routinely misperceive what the group believes, leading individuals to conform outwardly to a norm that few privately endorse (Prentice and Miller, 1993). When AI use is simultaneously valuable (productivity) and threatening (status), the socially safe public posture is often mild contempt. That posture communicates: “I am not replaceable; I am not cheating; I remain competent.”

In this sense, trash talk is a coordination device: it enforces a boundary around what counts as legitimate skill, regulates who receives credit, and defines what failures are tolerable. It is less a belief about AI than a negotiation over rank in the new production system (Ashforth and Mael, 1989; Jost, Banaji, and Nosek, 2004).

THE “CENTAURS”

The “one with the machine” minority is usually not experiencing a different technology. They are operating a different *socio-technical system*.

First, they treat the unit of cognition as distributed across human and artifacts (Hutchins, 2001). Second, they calibrate trust and responsibility: knowing when to rely, when to verify, and when to refuse (Lee and See, 2004; Parasuraman, Sheridan, and Wickens, 2000; Hancock et al., 2011; Glikson and Woolley, 2020). Third, they build control loops: tests, eval harnesses, constraints, and review rituals that turn a stochastic generator into a predictable production process (Dietvorst, Simmons, and Massey, 2018).

In the coordination-shift framing, the “10×” effect is rarely raw typing speed. It is the ability to (i) specify intent, (ii) allocate work to machine strengths, and (iii) perform fast verification and integration in a high-option environment (Blomgren, 2025). Many organizations have not yet made these practices normal, teachable, or rewarded—so the advantage concentrates.

IMPLICATIONS & WHAT TO DO

If you lead teams; if you want less hostility and more performance, stop treating this as a persuasion problem and treat it as a *coordination and accountability design* problem:

1. **Redefine competence publicly.** Make “orchestration + verification” an explicit skill standard, not a guilty secret (Blomgren, 2025).
2. **Normalize calibrated reliance.** Train teams on appropriate reliance (when to trust, when to test), rather than demanding unconditional enthusiasm or skepticism (Lee and See, 2004; Parasuraman, Sheridan, and Wickens, 2000).
3. **Give people control.** Adopt workflows where humans can modify, constrain, and iteratively refine outputs; this directly reduces aversion (Dietvorst, Simmons, and Massey, 2018; Deci and Ryan, 2000).
4. **Fix the credit economy.** Establish attribution norms (e.g., logging prompts, crediting review and integration) so that using AI does not imply moral taint, and so that failures are treated as system failures, not personal shame (Petriglieri, 2011; Prentice and Miller, 1993).
5. **Instrument quality.** Replace anecdotal “it hallucinated yesterday” narratives with measured failure modes and targeted mitigations; this counters negativity-driven overgeneralization (Baumeister et al., 2001; Tversky and Kahneman, 1974).

CONCLUSION

The apparent contradiction—ridiculing AI while consuming its outputs—is not primarily a failure of intelligence. It is a predictable human adaptation to a shifting production regime. Status threats motivate derogation (Petriglieri, 2011; Jost, Banaji, and Nosek, 2004); cognitive biases over-index visible errors (Baumeister et al., 2001; Tversky and Kahneman, 1974); machine heuristics raise the expectation bar (Yang and Sundar, 2024); attribution mechanisms reallocate credit and blame to protect the self (Mezulis et al., 2004; Weiner, 1985); and group norms shape what can be safely said aloud (Prentice and Miller, 1993). Other motives clearly exist—simple caution about opaque systems, residual “big design up front” habits, and organizational risk-avoidance—but in environments where AI is already ubiquitous in day-to-day work, contempt paired with dependence is better explained by status, identity, and attribution dynamics than by error rates or safety concerns alone.

“Centaur” practitioners break the loop by designing calibrated, controlled, instrumented workflows that convert stochastic generation into reliable coordination (Dietvorst, Simmons, and Massey, 2018; Lee and See, 2004; Blomgren, 2025). In short: people insult the machine because they are negotiating who they are in the new hierarchy, not because they have accurately evaluated what the machine can do.

REFERENCES

- Ashforth, Blake E. and Fred Mael (1989). "Social Identity Theory and the Organization". In: *Academy of Management Review* 14.1, pp. 20–39. doi:10.5465/amr.1989.4278999.
- Baumeister, Roy F. et al. (2001). "Bad Is Stronger than Good". In: *Review of General Psychology* 5.4, pp. 323–370. doi:10.1037/1089-2680.5.4.323.
- Bigman, Yochanan E. and Kurt Gray (2018). "People are averse to machines making moral decisions". In: *Cognition* 181, pp. 21–34. doi:10.1016/j.cognition.2018.08.003.
- Blomgren, Michel (2025). *10× The Coordination Shift: Software Engineering in the Centaur Era*. <https://github.com/sa6mwa/centaurarticles/blob/main/pub/10x.pdf>.
- Deci, Edward L. and Richard M. Ryan (2000). "The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior". In: *Psychological Inquiry* 11.4, pp. 227–268. doi:10.1207/S15327965PLI1104_01.
- Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2015). "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err". In: *Journal of Experimental Psychology: General* 144.1, pp. 114–126. doi:10.1037/xge0000033.
- (2018). "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them". In: *Management Science* 64.3, pp. 1155–1170. doi:10.1287/mnsc.2016.2643.
- Dillard, James Price and Lijiang Shen (2005). "On the Nature of Reactance and Its Role in Persuasive Health Communication". In: *Communication Monographs* 72.2, pp. 144–168. doi:10.1080/03637750500111815.
- Festinger, Leon (1954). "A Theory of Social Comparison Processes". In: *Human Relations* 7.2, pp. 117–140. doi:10.1177/001872675400700202.
- Fiske, Susan T. et al. (2002). "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition". In: *Journal of Personality and Social Psychology* 82.6, pp. 878–902. doi:10.1037/0022-3514.82.6.878.
- Glikson, Ella and Anita Williams Woolley (2020). "Human Trust in Artificial Intelligence: Review of Empirical Research". In: *Academy of Management Annals* 14.2, pp. 627–660. doi:10.5465/annals.2018.0057.
- Gray, Heather M., Kurt Gray, and Daniel M. Wegner (2007). "Dimensions of Mind Perception". In: *Science* 315.5812, p. 619. doi:10.1126/science.1134475.
- Gray, Kurt and Daniel M. Wegner (2009). "Moral Typecasting: Divergent Perceptions of Moral Agents and Moral Patients". In: *Journal of Personality and Social Psychology* 96.3, pp. 505–520. doi:10.1037/a0013748.
- Haidt, Jonathan (2001). "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment". In: *Psychological Review* 108.4, pp. 814–834. doi:10.1037/0033-295X.108.4.814.
- Hancock, Peter A. et al. (2011). "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction". In: *Human Factors* 53.5, pp. 517–527. doi:10.1177/0018720811417254.
- Hutchins, Edwin (2001). "Distributed Cognition". In: *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, pp. 2068–2072. doi:10.1016/B0-08-043076-7/01636-3.
- Joo, Minjoo (2024). "It's the AI's fault, not mine: Mind perception increases blame attribution to AI". In: *PLOS ONE* 19.12, e0314559. doi:10.1371/journal.pone.0314559.
- Jost, John T. and Mahzarin R. Banaji (1994). "The Role of Stereotyping in System-Justification and the Production of False Consciousness". In: *British Journal of Social Psychology* 33.1, pp. 1–27. doi:10.1111/j.2044-8309.1994.tb01008.x.
- Jost, John T., Mahzarin R. Banaji, and Brian A. Nosek (2004). "A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo". In: *Political Psychology* 25.6, pp. 881–919. doi:10.1111/j.1467-9221.2004.00402.x.
- Lee, John D. and Katrina A. See (2004). "Trust in Automation: Designing for Appropriate Reliance". In: *Human Factors* 46.1, pp. 50–80. doi:10.1518/hfes.46.1.50_30392.
- Logg, Jennifer M., Julia A. Minson, and Don A. Moore (2019). "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment". In: *Organizational Behavior and Human Decision Processes* 151, pp. 90–103. doi:10.1016/j.obhdp.2018.12.005.
- Mezulis, Amy H. et al. (2004). "Is There a Universal Positivity Bias in Attributions? A Meta-

- Analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias". In: *Psychological Bulletin* 130.5, pp. 711–747. doi:10.1037/0033-2909.130.5.711.
- Nass, Clifford and Youngme Moon (2000). "Machines and Mindlessness: Social Responses to Computers". In: *Journal of Social Issues* 56.1, pp. 81–103. doi:10.1111/0022-4537.00153.
- Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens (2000). "A Model for Types and Levels of Human Interaction with Automation". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 30.3, pp. 286–297. doi:10.1109/3468.844354.
- Petriglieri, Jennifer L. (2011). "Under Threat: Responses to and the Consequences of Threats to Individuals' Identities". In: *Academy of Management Review* 36.4, pp. 641–662. doi:10.5465/amr.2009.0087.
- Prentice, Deborah A. and Dale T. Miller (1993). "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm". In: *Journal of Personality and Social Psychology* 64.2, pp. 243–256. doi:10.1037/0022-3514.64.2.243.
- Rains, Stephen A. (2013). "The Nature of Psychological Reactance Revisited: A Meta-Analytic Review". In: *Human Communication Research* 39.1, pp. 47–73. doi:10.1111/j.1468-2958.2012.01443.x.
- Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp. 1124–1131. doi:10.1126/science.185.4157.1124.
- Waytz, Adam, Nicholas Epley, and John T. Cacioppo (2010). "Social Cognition Unbound: Psychological Insights into Anthropomorphism and Dehumanization". In: *Current Directions in Psychological Science* 19.1, pp. 58–62. doi:10.1177/0963721409359302.
- Weiner, Bernard (1985). "An Attributional Theory of Achievement Motivation and Emotion". In: *Psychological Review* 92.4, pp. 548–573. doi:10.1037/0033-295X.92.4.548.
- Yang, Hyun and S. Shyam Sundar (2024). "Machine heuristic: concept explication and development of a measurement scale". In: *Journal of Computer-Mediated Communication*. doi:10.1093/jcmc/zmae019.