

The Ownership Cost of AI Goal-Setting: Empirical Evidence for a Quality-Motivation Dissociation and Differential Harm to Vulnerable Users (Extended Abstract)

Vivienne Bihe Chi¹, Roman Rietsche², Andreas Göldi², Lyle Ungar¹, Sharath Chandra Guntuku¹

¹University of Pennsylvania

²Bern University of Applied Sciences

vchi@seas.upenn.edu, roman.rietsche@bfh.ch, andreas.goeldi@bfh.ch, ungar@cis.upenn.edu, sharathg@seas.upenn.edu

Coaching and wellbeing platforms now generate personal goals on users' behalf, assuming that AI produces better-structured goals than users write themselves and that structure yields better outcomes (Schimpf et al. 2026). Personal goals differ from most AI-assisted artifacts, whose value is realized at the moment of delivery. A goal's value is realized only in the behavior it drives afterward, and it articulates who the person wants to become. Psychological ownership grows through investment of self, intimate knowing, and control over the target (Pierce, Kostova, and Dirks 2001, 2003), and self-determination theory holds that sustained motivation requires behavior to be experienced as self-endorsed rather than externally imposed (Ryan and Deci 2000). Authoring a goal walks through all three ownership routes at once: delegating to an LLM walks through none of them, paralleling the IKEA effect (Norton, Mochon, and Ariely 2012). Goal-setting theory predicts the opposite: clearer, SMART-compliant goals should raise attainability regardless of authorship (Locke and Latham 2002; Doran 1981), though goal effects remain conditional on commitment (Klein et al. 1999). The closest precedent, Schimpf et al. (2026), improved goal quality with LLM chatbots without raising commitment; ours varies the goal's source instead.

In a two-arm design, 470 U.S. Prolific adults completed a trait self-efficacy scale (Chen, Gully, and Eden 2001) and a guided reflection (McAdams 2001), then wrote three SMART goals themselves, with copy-paste disabled, or received three generated from their reflection by an LLM and retyped them verbatim to equate typing effort. Participants rated each goal on commitment, self-efficacy, importance, difficulty, and clarity, and completed an ownership scale and authorship slider. An LLM judge scored all goals against the SMART criteria (Gilardi, Alizadeh, and Kubli 2023). Two weeks later, 332 participants returned (70.6%) for a recall task, action-initiation reports, and re-rated measures.

The manipulation separated cleanly ($d = 2.53$), and LLM goals scored higher on every SMART dimension ($d = 2.26$), yet motivational outcomes moved in the opposite direction on nearly every measure: ownership ($d = 1.38$), commitment ($d = 1.19$), and importance ($d = 1.13$) all favored self-authorship, and the predicted self-efficacy gain reversed

($d = 0.62$). LLM goals felt less difficult, but subjective clarity did not differ despite being objectively clearer. Ownership mediated the authorship effect on every downstream outcome, whereas SMART score accounted only for perceived difficulty. At follow-up, self-authors recalled their goals more accurately and acted on more of them (1.95 vs. 1.43 of three); 72.8% versus 46.6% had acted on two or more goals ($p < .001$), and the ownership and commitment gaps persisted. Baseline ownership, measured before any opportunity to act, mediated later action initiation, a leading indicator of behavior beyond self-report. Trait self-efficacy moderated this path: erosion was approximately 55% larger for users low in trait self-efficacy, though substantial throughout the sample (Bandura 1997; Schunk and DiBenedetto 2020).

Goal quality and ownership operate on different psychological registers, so optimizing one does not advance the other, a boundary condition on goal-setting theory's specificity principle, since specificity supplied by an AI carries no informational value the user did not already possess. Frameworks built around output quality (Green and Chen 2019; Bućinca, Malaya, and Gajos 2021; Dhillon et al. 2024; Alberts, Lyngs, and Lukoff 2026) assume the artifact is the deliverable, and so fail to detect a deployment that scores well at delivery while degrading the conditions its outcomes depend on. We extend the ownership erosion found in AI writing assistance (Draxler et al. 2024; Joshi and Vogel 2025; Kadoma et al. 2024) with a concrete behavioral cost, paralleling harms already documented in AI-managed labor (Dong et al. 2025). This cost concentrates among users lowest in trait self-efficacy, the population AI coaching tools are designed to serve, so a deployment can appear successful on average while concentrating harm among its most vulnerable users (Axelsson et al. 2025; Pi, Turkay, and Bogiatzis-Gibbons 2025). Design should scaffold articulation rather than replace it: the user's own words as the first draft, refinement over generation (Gero et al. 2026), ownership measured directly, and caution against over-assisting users who struggle with articulation (Joseph 2025). Limitations include verbatim retyping, self-reported action, a two-week window, and SMART-only quality (Sheldon and Elliot 1999). Platforms deploying AI in identity-relevant domains should demonstrate ownership and behavioral follow-through beyond quality benchmarks. The full version of our paper can be accessed at: <https://arxiv.org/abs/2605.12344>

References

- Alberts, L.; Lyngs, U.; and Lukoff, K. 2026. Designing for Sustained Motivation: A Review of Self-Determination Theory in Behaviour Change Technologies. *Interacting with Computers*, 38(3): 447–468. Published online 18 September 2024.
- Axelsson, M.; Cheong, J.; Nyrupe, R.; and Gunes, H. 2025. Who Owns the Robot?: Four Ethical and Socio-Technical Questions About Wellbeing Robots in the Real World Through Community Engagement. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(1): 249–264.
- Bandura, A. 1997. *Self-Efficacy: The Exercise of Control*. New York: W. H. Freeman.
- Buçinca, Z.; Malaya, M. B.; and Gajos, K. Z. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1).
- Chen, G.; Gully, S. M.; and Eden, D. 2001. Validation of a New General Self-Efficacy Scale. *Organizational Research Methods*, 4(1): 62–83.
- Dhillon, P. S.; Molaei, S.; Li, J.; Golub, M.; Zheng, S.; and Robert, L. P. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-Writing with Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
- Dong, M.; Brinkmann, L.; Sherif, O.; Wang, S.; Zhang, X.; Bonnefon, J.-F.; and Rahwan, I. 2025. Experimental Evidence That AI-Managed Workers Tolerate Lower Pay Without Demotivation. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(1): 798–798.
- Doran, G. T. 1981. There's a S.M.A.R.T. Way to Write Management's Goals and Objectives. *Management Review*, 70(11): 35–36.
- Draxler, F.; Werner, A.; Lehmann, F.; Hoppe, M.; Schmidt, A.; Buschek, D.; and Welsch, R. 2024. The AI Ghostwriter Effect: When Users Do Not Perceive Ownership of AI-Generated Text But Self-Declare as Authors. *ACM Transactions on Computer-Human Interaction*, 31(2): 1–40.
- Gero, K. I.; Long, T.; Schnitzler, C.; and Dhillon, P. 2026. From Planning to Revision: How AI Writing Support at Different Stages Alters Ownership.
- Gilardi, F.; Alizadeh, M.; and Kubli, M. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30).
- Green, B.; and Chen, Y. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW).
- Joseph, J. 2025. The algorithmic self: how AI is reshaping human identity, introspection, and agency. *Frontiers in Psychology*, 16: 1645795.
- Joshi, N.; and Vogel, D. 2025. Writing with AI Lowers Psychological Ownership, but Longer Prompts Can Help. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces*, CUI '25, 1–17. New York, NY, USA: ACM.
- Kadoma, K.; Aubin Le Quere, M.; Fu, X. J.; Munsch, C.; Metaxa, D.; and Naaman, M. 2024. The Role of Inclusion, Control, and Ownership in Workplace AI-Mediated Communication. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM.
- Klein, H. J.; Wesson, M. J.; Hollenbeck, J. R.; and Alge, B. J. 1999. Goal Commitment and the Goal-Setting Process: Conceptual Clarification and Empirical Synthesis. *Journal of Applied Psychology*, 84(6): 885–896.
- Locke, E. A.; and Latham, G. P. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist*, 57(9): 705–717.
- McAdams, D. P. 2001. The Psychology of Life Stories. *Review of General Psychology*, 5(2): 100–122.
- Norton, M. I.; Mochon, D.; and Ariely, D. 2012. The IKEA Effect: When Labor Leads to Love. *Journal of Consumer Psychology*, 22(3): 453–460.
- Pi, Y.; Turkay, C.; and Bogiatzis-Gibbons, D. 2025. Interactive AI and Human Behavior: Challenges and Pathways for AI Governance. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(3): 2016–2029.
- Pierce, J. L.; Kostova, T.; and Dirks, K. T. 2001. Toward a Theory of Psychological Ownership in Organizations. *Academy of Management Review*, 26(2): 298–310.
- Pierce, J. L.; Kostova, T.; and Dirks, K. T. 2003. The State of Psychological Ownership: Integrating and Extending a Century of Research. *Review of General Psychology*, 7(1): 84–107.
- Ryan, R. M.; and Deci, E. L. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55(1): 68–78.
- Schimpf, M.; Maier, S.; Wyrowski, A.; Christoforakos, L.; Feuerriegel, S.; and Bohné, T. 2026. Supporting Effective Goal Setting with LLM-Based Chatbots. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26. New York, NY, USA: Association for Computing Machinery. ArXiv:2602.08636.
- Schunk, D. H.; and DiBenedetto, M. K. 2020. Motivation and social cognitive theory. *Contemporary Educational Psychology*, 60: 101832.
- Sheldon, K. M.; and Elliot, A. J. 1999. Goal Striving, Need Satisfaction, and Longitudinal Well-Being: The Self-Concordance Model. *Journal of Personality and Social Psychology*, 76(3): 482–497.