Generative AI in Business: Annotated Reading List

Oct 10, 2026 · @Max Killian

Seven free, downloadable studies from Stanford, Harvard, Wharton and MIT, ordered newest first, with relevance to business prompting breaking ties. Each entry credits the institution and authors, summarizes the findings in plain language, says what they mean for a business, and links to the original source.

#	Study	Institution	Latest version
1	AI Index Report 2026	Stanford HAI	April 2026
2	The Cybernetic Teammate	Harvard Business School, Wharton, P&G	Organization Science, June 2026
3	Prompting Science Reports 1–4	Wharton Generative AI Labs	March–December 2025
4	The GenAI Divide: State of AI in Business 2025	MIT NANDA	July 2025
5	Generative AI at Work	Stanford & MIT	Quarterly Journal of Economics, 2025
6	Navigating the Jagged Technological Frontier	Harvard Business School & BCG	Working paper, 2023
7	Experimental Evidence on the Productivity Effects of Generative AI	MIT	Science, July 2023

Summaries are in my own words; figures come from each study's abstract or publisher page. Check the original before quoting a number externally.

1. Stanford AI Index Report 2026

Credit: AI Index Steering Committee, Stanford Institute for Human-Centered Artificial Intelligence (HAI). Published April 2026, ninth edition.

Summary. The AI Index is the leading independent annual data report on AI, covering research, technical performance, economics, policy and public opinion. New this year are estimates of generative AI's economic value and early evidence of its labor-market effects. Stanford reports the value of generative AI to consumers reached about $172 billion a year by early 2026. The report also warns that real-world task performance is getting harder to measure reliably.

Why it matters for business. Use it as the baseline reference for adoption and economic-impact figures when building a case for AI investment. It compiles many underlying studies rather than running one experiment, and its figure notes credit each original source.

Links:

- Full report (PDF)
- Report page, Stanford HAI

2. The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork

Credit: Fabrizio Dell'Acqua (Harvard Business School), Charles Ayoubi (ESSEC), Hila Lifshitz (HBS AI Institute / Warwick), Raffaella Sadun (HBS), Ethan Mollick and Lilach Mollick (Wharton), Karim Lakhani (HBS), with Yi Han, Jeff Goldman, Hari Nair and Stewart Taub of Procter & Gamble. Published in Organization Science, June 2026; earlier as HBS Working Paper 25-043 and NBER Working Paper 33641 (2025).

Summary. A pre-registered field experiment with 776 P&G professionals working on real product-innovation challenges. People were randomly assigned to work alone or in two-person teams, with or without AI. Individuals using AI performed as well as teams without AI. AI also broke down functional silos: without it, R&D staff proposed technical ideas and commercial staff proposed commercial ones; with it, both produced balanced solutions. Teams using AI were the most likely to produce top-tier ideas, and AI users reported more positive emotions while working.

Why it matters for business. This is the strongest recent evidence that AI can act as a working partner, not just a tool. It suggests rethinking team size and composition, and using AI to give specialists perspective from outside their function.

Links:

- NBER Working Paper 33641 (free PDF)

- NBER paper page
- SSRN version (HBS Working Paper 25-043)
- Published article, Organization Science
- Plain-language summary by co-author Ethan Mollick

3. Wharton Prompting Science Reports 1–4

Credit: Generative AI Labs (GAIL), The Wharton School, University of Pennsylvania. Core authors Lennart Meincke, Ethan Mollick, Lilach Mollick and Dan Shapiro; Report 4 adds Savir Basil and Ina Shapiro. Published 2025.

Summary. The most directly relevant research on business prompting. The series rigorously tests popular prompt "tricks" by asking hard questions many times under each prompt. Findings by report:

1. **Prompt engineering is complicated and contingent** (March 2025). No single prompt formula works everywhere. Politeness sometimes helps and sometimes hurts, and how you measure success changes the answer.
2. **The decreasing value of chain of thought** (June 2025). Telling a model to "think step by step" gives small gains on older models but marginal gains on modern reasoning models, while adding roughly 20–80% more time.
3. **Tipping and threatening** (2025). Offering a model money or threatening it generally has no significant effect on performance.
4. **Expert personas don't improve factual accuracy** (December 2025). Prompts like "act as an expert" gave no consistent accuracy boost across six models and sometimes made results worse.

Why it matters for business. Wharton's conclusion is that organizations gain more from how a task is framed, what context and material is fed in, and how outputs are checked than from clever prompt formulas. Train staff on clear task specifications and review processes, not prompt tricks.

Links:

- Report 1 (PDF, arXiv)
- Report 2 (PDF, arXiv)
- Report 3 (PDF, arXiv)
- Report 4 (PDF, arXiv)
- Wharton GAIL summary of Report 2

- Wharton Generative AI Labs home
- Penn Today coverage of Report 4

4. The GenAI Divide: State of AI in Business 2025

Credit: Project NANDA, MIT Media Lab; reviewed by Pradyumna Chari. Preliminary report, July 2025, covering research from January to June 2025.

Summary. Based on reviews of 300+ AI initiatives, 52 organizational interviews and 153 leader surveys. Despite an estimated $30–40 billion in enterprise spending on generative AI, about 95% of organizations in the study saw no measurable profit-and-loss impact, while about 5% of integrated pilots produced millions in value. The authors blame a "learning gap": most tools don't retain feedback, adapt to context or improve over time. They also document "shadow AI": only about 40% of firms had official AI subscriptions, yet over 90% of employees reported using personal AI tools for work.

Why it matters for business. The winners tied AI to a specific workflow, measured operating outcomes, kept persistent context and gave a business owner accountability. Treat the 95% figure with care: the sample is anonymized and results are self-reported, so it is not a universal failure rate.

Links:

- Full report (PDF, hosted copy)
- MIT Project NANDA
- Independent critique of the methodology

5. Generative AI at Work

Credit: Erik Brynjolfsson (Stanford), Danielle Li (MIT) and Lindsey Raymond (MIT). Published in the Quarterly Journal of Economics, 2025 (vol. 140, pp. 889–942); first released as NBER Working Paper 31161, April 2023.

Summary. A real-world study of an AI assistant rolled out to 5,172 customer-support agents at a Fortune 500 software company. Productivity, measured as issues resolved per hour, rose about 15% on average. Less experienced and lower-skilled agents improved in both speed and quality, while top performers saw small speed gains and small quality declines. The AI appeared to spread the best agents' know-how to newer staff. Customers were more polite and less likely to ask for a manager, and employee retention improved.

Why it matters for business. Strong evidence for deploying AI where work is high-volume and know-how is unevenly spread, such as support desks and onboarding. Expect the biggest gains among newer staff.

Links:

- NBER Working Paper 31161 (free PDF)
- NBER paper page
- Published article, Quarterly Journal of Economics
- arXiv version (free PDF)

6. Navigating the Jagged Technological Frontier

Credit: Fabrizio Dell'Acqua, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon and Karim Lakhani; Harvard Business School, Wharton, MIT Sloan, Warwick and Boston Consulting Group. HBS Working Paper 24-013, September 2023.

Summary. The landmark field experiment with 758 BCG consultants, about 7% of BCG's individual contributors, working on 18 realistic consulting tasks. For tasks within AI's capabilities, consultants using GPT-4 completed 12.2% more tasks, 25.1% faster, with output rated over 40% higher in quality. On a business-analysis task deliberately outside AI's capabilities, consultants using AI were about 19 percentage points less likely to reach the correct answer. The authors call this uneven boundary the "jagged frontier": AI is strong at some tasks and unreliable at others that look similar.

Why it matters for business. Gains are large but only where the task suits AI. Map which tasks sit inside the frontier, keep human checks on the rest, and don't assume a tool that excels at one task will do well on its neighbor.

Links:

- HBS Working Paper 24-013 page (free PDF)
- SSRN version (free PDF)
- Harvard D^3 Institute overview

7. Experimental Evidence on the Productivity Effects of Generative AI

Credit: Shakked Noy and Whitney Zhang, Department of Economics, MIT. Published in Science, July 14, 2023 (vol. 381, pp. 187–192).

Summary. A pre-registered experiment in which 453 college-educated professionals (marketers, grant writers, consultants, data analysts, HR professionals and managers) completed writing tasks from their own occupations, with half given ChatGPT. Average time fell 40% and graded quality rose 18%. Lower-performing workers gained the most, narrowing the gap between workers. Participants who tried ChatGPT were twice as likely to be using it in their real jobs two weeks later.

Why it matters for business. Clear evidence for routine professional writing: memos, cover letters, sensitive emails, short analyses. Note the limits: tasks were short and self-contained, so results don't speak to long projects or brand consistency.

Links:

- Working paper (free PDF), MIT Economics
- Published article, Science
- Stanford SCALE summary

Villanova and sources

Villanova. I found no published Villanova research study on business uses of generative AI. Villanova's Falvey Library offers a practical prompt-engineering guide by librarian Sara Marx covering clear, specific prompts, stating format, audience and tone, and iterating on results. It is a useful primer, not research.

- Falvey Library: Prompt Engineering guide
- Falvey Library: Artificial Intelligence guide

Credits. All findings belong to the authors and institutions named in each entry: Stanford HAI; Harvard Business School and the HBS AI Institute / Digital Data Design Institute; the Wharton School Generative AI Labs; MIT Media Lab Project NANDA; MIT Department of Economics; Stanford Graduate School of Business; the National Bureau of Economic Research; Boston Consulting Group; and Procter & Gamble. Summaries here are paraphrased; consult the linked originals for exact wording, methods and limitations.