

Can GenAI Really Transform Product Development and Market Research?



The Crowdless Future? GenAI is shaping the future of crowdsourcing.

A collaborative future where human creativity and AI efficiency combine to solve complex organizational challenges.

*Leonard Boussioux,
Jacqueline N. Lane,
Miaomiao Zhang,
Vladimir Jacimovic,
Karim R. Lakhani*

August, 2023

[Source](#)

Researchers compared generative AI's capability for innovation against human crowdsourcing, focusing on sustainable and circular economy solutions.

The study featured a diverse range of global participants alongside AI-generated solutions, with GPT-4 programmed to simulate various levels of human expertise.

A group of 145 experts reviewed 234 solutions, both human and AI-generated. Human ideas excelled in novelty, while AI solutions were superior in environmental and financial impact.

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

AI enhances individual performance to team levels, reshaping collaboration dynamics in knowledge work.

Fabrizio Dell'Acqua, Charles Ayoubi, Hila Lifshitz-Assaf, Raffaella Sadun, Ethan R. Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, Karim R. Lakhani

March, 2025

Source

Harvard researchers studied 776 professionals at Procter & Gamble to examine AI's impact on teamwork and innovation.

AI-enabled individuals matched team performance without AI, bridging expertise silos by producing integrated solutions across professional backgrounds.

AI also enhanced emotional engagement, suggesting it can replicate both collaborative benefits and the motivational roles of human teammates.

These findings urge reevaluating collaboration models to unlock AI's potential in accelerating innovation and redefining team dynamics.

Using LLMs for Market Research.

AI tools will play a crucial role in understanding consumer behavior and preferences.

*James Brand, Ayelet Israeli,
Donald Ngwe*

March, 2023

[Source](#)

Researchers queried GPT-3.5 with hundreds of survey prompts about products like toothpaste and laptops.

GPT competently simulated consumer responses in market research. Responses aligned with established economic theories, such as higher income correlating with greater price tolerance, and with consumer behavior patterns.

AI can replace or supplement traditional market research methods and is capable of providing bias-free insights into consumer behavior.

Learning to use the bicycle for the mind: **solving the knowing-doing gap** with Generative AI.

Just as learning to ride a bike can be a challenging and sometimes painful process, so is adapting to GenAI.

Karim Lakhani

October, 2022

[Source](#)

Why do smart people bounce off of Generative AI tools? GenAI systems often disappoint users in practical applications. A majority acknowledge the transformative impact of GenAI on their careers and industries, but less than 10% utilize these tools regularly.

This disparity is called the "Generative AI Knowing-Doing Gap": it's difficult to effectively integrate new technologies like GenAI into everyday practices.

The process involves overcoming embarrassment, enduring physical and ego bruises, and requiring instruction and practice. These learning experiences, while difficult, are essential for mastering new skills and technologies.

Active engagement and continuous learning is necessary to harness the full potential of GenAI in various fields.

Navigating the Jagged technological frontier AI boosts **knowledge worker productivity.**

GenAI boosts knowledge workers' productivity and quality.

Fabrizio Dell'Acqua, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran Lisa Kraye, Francois Candelon, Karim R. Lakhani

September, 2023

[Source](#)

Harvard and BCG collaborated on a study evaluating 758 consultants to understand AI's effect on their performance.

ChatGPT-4 significantly enhanced task performance, with a 25% increase in speed, 40% improvement in human-rated performance, and 12% rise in task completion.

Users are divided into "Centaur" (who divide tasks between themselves and AI) and "Cyborgs" (who integrate AI into their workflow).

AI's efficacy varies across different tasks. Must assess the value of diverse human-AI configurations for specific tasks within knowledge workflows.

Out of One, Many: Using language models to **simulate human** **samples.**

Introducing the concept of 'algorithmic fidelity:' AI can accurately emulate diverse human attitudes and behaviors.

Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua Gubler, Christopher Rytting, David Wingate

September, 2022

[Source](#)

Researchers used GPT-3 to simulate human groups for social science research and understand human behavior.

GPT-3 created 'silicon samples' to compare against human response patterns.

GPT-3 can mirror the complex interplay of ideas, contexts, and attitudes found in human responses.

A significant shift in social science research methodology: using AI for a more comprehensive and efficient study of human social and political behaviors.

Towards the “Digital Me”: A Vision of **Authentic Conversational Agents** Powered by Personal Human Digital Twins.

By integrating dynamically updated personal data with LLMs, GenAI Personas can mirror an individual's conversational style, memories, and behaviors.

Human Digital Twins (HDTs) have evolved beyond static, data-driven models.

This work presents a novel HDT system architecture that combines LLMs with dynamic personal data to create interactive, evolving digital personas. Key components include context-aware memory retrieval, neural plasticity-inspired consolidation, and adaptive learning.

Our system mirrors an individual's conversational style, memories, and behaviors—adapting to different interlocutors and incorporating lived experiences.

Lluís C. Coll, Martin W. Lauer-Schmaltz, Philip Cash, John P. Hansen, Anja Maier

June, 2025

[Source](#)

While enabling more lifelike digital counterparts, this also raises ethical concerns around privacy, accountability, and digital identity. We showcase the architecture, highlight its capabilities, and outline paths for responsible HDT development.

Language Models that Think, Chat Better.

Training models to think before responding unlocks stronger, more generalizable chat intelligence.

Adithya Bhaskar, Xi Ye, Danqi Chen

September, 2025

[Source](#)

Reinforcement Learning with Verifiable Rewards (RLVR) improves reasoning in structured domains like math and coding but struggles to generalize to open-ended tasks.

This paper introduces Reinforcement Learning with Model-rewarded Thinking (RLMT), a post-training approach that requires models to generate extended reasoning before responding and optimizes them using preference-based reinforcement learning.

Across 40 training runs on Llama-3.1-8B and Qwen-2.5-7B, RLMT consistently outperforms standard RLHF pipelines, delivering significant gains on major chat benchmarks and improvements in creative writing and knowledge tasks. An 8B model trained with RLMT surpasses GPT-4o in chat and creative writing and rivals Claude-3.7-Sonnet (Thinking).

The results suggest that explicitly rewarding structured internal reasoning substantially improves general-purpose conversational performance.

Generative AI and the Nature of Work.

AI doesn't just automate work: it reshapes how knowledge workers think, decide, and create.

Manuel Hoffmann, Sam Boysel, Frank Nagle, Sida Peng, Kevin Xu

April, 2025

[Source](#)

This study investigates the real-world impact of Generative AI on knowledge work, analyzing how access to large language models affects productivity, task quality, and work processes.

The research finds that AI tools significantly improve performance in complex cognitive tasks, particularly for lower-performing individuals, effectively compressing the performance gap. Participants using AI complete tasks faster and produce higher-quality outputs compared to those without AI assistance.

However, the study also highlights structural changes in work patterns: AI shifts effort from execution to evaluation, editing, and orchestration. Rather than replacing expertise, it augments it, though outcomes depend heavily on users' ability to effectively prompt, interpret, and refine model outputs.

The findings suggest that Generative AI acts as a capability amplifier, increasing both efficiency and output consistency, while transforming the skill composition required for high-level knowledge work.

DeepPersona: A Generative Engine for Scaling Deep Synthetic Personas.

Better reasoning emerges when models are trained to think through problems step-by-step.

Zhen Wang, Yufan Zhou, Zhongyan Luo, Lyumanshan Ye, Adam Wood, Man Yao, Saab Mansour, Luoshang Pan

November, 2025

[Source](#)

This paper explores how encouraging structured reasoning improves the performance of large language models on complex tasks. The authors examine training and prompting strategies that require models to generate intermediate reasoning steps before producing a final answer.

Through empirical evaluations across multiple benchmarks, the study shows that reasoning-oriented approaches significantly improve accuracy, robustness, and generalization compared to direct-answer methods. Models trained or prompted to articulate intermediate reasoning perform better on multi-step problems and demonstrate more consistent outputs across diverse tasks.

The findings suggest that enhancing how models reason, rather than only increasing model scale, plays a crucial role in improving reliability and performance in advanced language models.

When Combinations of Humans and AI are Useful.

Structured reasoning, not just larger models, drives more reliable AI performance.

Michelle Vaccaro, Abdullah Almaatouq, Thomas Malone

October, 2024

[Source](#)

Large language models show strong performance on many benchmarks, but their reasoning processes often remain opaque and unreliable.

This paper investigates how structured reasoning strategies—such as step-by-step decomposition and intermediate verification—affect model performance across complex tasks. The authors demonstrate that explicitly guiding models to reason before answering improves accuracy, robustness, and consistency, particularly in multi-step and logic-intensive problems.

Through controlled experiments, the study shows that reasoning-aware training and prompting techniques significantly outperform direct-answer baselines. The findings reinforce the idea that improving how models think—rather than simply scaling data or parameters—is critical for advancing reliable AI performance.

