Boston Consulting Group field experiment

IMPLEMENTED 2023 · PROFESSIONAL SERVICES · GLOBAL · BAND D · PROVED
VERSION 1.0 · PUBLISHED 15 NOVEMBER 2026 · UPDATED 15 NOVEMBER 2026 · FACTS CHECKED 15 NOVEMBER 2026

Consultants were randomly given GPT-4 for set tasks in a preregistered experiment.

ABOUT THIS SYNOPSIS

This is MarLuc's summary and grading of publicly available sources, for information only. It is not investment, legal, financial or professional advice, and should not be relied on for any decision. Readers must validate it against the original sources and, where appropriate, challenge its conclusions.

Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date. Check the original source before relying on any figure.

The impact band, proof marker and SEAM view are MarLuc's opinion. The organisation is not a MarLuc client and has not reviewed or endorsed this synopsis. Names and trademarks belong to their owners.

NOTICE WORDING LAST UPDATED 16 NOVEMBER 2026

AT A GLANCE

AI used: Generative AI: GPT-4.

Implemented: 2023. Grade: band D, Task support, Proved.

Headline result, as reported: Consultants using GPT-4 completed 12.2% more tasks, 25.1% faster, on tasks within AI's capabilities, in a preregistered experiment

1 · The case

The facts, as reported by the 2 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

Boston Consulting Group (Dell'Acqua and colleagues). Global management consulting firm. The experiment involved 758 consultants, about 7% of its individual contributor-level consultants worldwide. Canvas block: 5. People and skills. Also 6. Data and technology, 8. Governance and risk.

Business challenge

The challenge. Not reported in BCG's own terms. The researchers set out to measure how AI affects skilled knowledge workers on realistic, complex tasks, and how it might augment or disrupt their workflows.

How it was addressed. Researchers from Harvard Business School, Wharton, Warwick Business School, MIT Sloan and BCG designed realistic consulting tasks with BCG staff. BCG consultants were offered the chance to spend five hours on the experiment; after a baseline task, they were

Who implemented it

OpenAI (Model or cloud provider): Supplied GPT-4 through its API.
A research experiment, designed by academic researchers with BCG.

Technology (SEAM pillar E)

SEAM types: T1 Chat assistants. Deployment: Cloud APIs. Vendors and products: OpenAI, GPT-4.

Where it was built in: A study platform built on the OpenAI API gave consultants a ChatGPT-like GPT-4 chat for the experiment's tasks.

What changed

Before (baseline task, no AI): consultants worked unaided. After, with GPT-4: on tasks within AI's capabilities they completed 12.2% more tasks, 25.1% faster and at higher quality; on the task outside them, they were less likely to reach a correct answer. [1] [2]

What it cost

Not reported. Participants spent about five hours each on the experiment.

What it saved

By the researchers' analysis: on 18 tasks within AI's capabilities, consultants using AI completed 12.2% more tasks, 25.1% more quickly, with significantly higher quality. On a task outside them, they were 19% less likely to produce correct solutions. Ideas produced with AI were less varied. Achieved cash saving: Not reported; the study was a controlled experiment. [1] [2]

Governance and control

Preregistered, with random assignment, run in collaboration with BCG. Three co-authors work for BCG, and one, Karim Lakhani, advises BCG on AI strategy. Rules for consultants' use of AI on client work: Not reported.

S2 · Accountable owner · Not reported
Not reported.

A1 · Safe use · Not reported
Not reported.

A2 · Human in the loop · Not reported
Not reported.

A3 · Monitoring and assurance · Evidenced
The platform recorded all participants' prompts and the AI's responses, and the preregistered experiment compared AI and non-AI groups on quality and speed. [2]

A4 · Vendor AI · Not reported
Not reported.

A5 · Agent controls · Not reported
No agents described.

A6 · Standards and regulation · Evidenced
The study received institutional review board approval (IRB23-0392) and was preregistered. [2]

What worked, what did not, what is unproven

WORKED

– Higher productivity and quality on tasks within AI's capabilities.

– A short prompt engineering overview improved results further, by the researchers' analysis.

DID NOT

– On a task outside AI's capabilities, consultants using AI were less likely to reach the correct answer.

– Ideas produced with AI were less varied, suggesting more uniform output, by the researchers' analysis.

UNPROVEN

– Whether the gains carry over to real client work over time.

– Effects on cost, staffing or client outcomes, which the experiment did not measure.

Timeline

2023 Experiment run with 758 consultants using GPT-4 [2]

Sept 2023 Findings released as a Harvard working paper [1]

2026 Listed as forthcoming in Organization Science [1]

2 · Grade and SEAM assessment

MarLuc's opinion, made on the evidence above. How cases are graded: marluc.co.uk/case-studies/how-we-rate

Level of change and impact

Claimed: Not applicable. Grade: D, Task support. Proof: Proved. Level of change: Task support.

Task support (Level 1, band D). Individuals used AI on set tasks in an experiment; no process or role was changed.

The SEAM view and lessons learnt

The study's value is the jagged frontier: the same tool helped on some tasks and harmed on others of similar apparent difficulty. Results depended on how people worked with AI, and the authors describe two successful patterns: dividing tasks between person and AI, or working with AI throughout. For a larger organisation, the lesson is to map which tasks sit inside AI's capabilities before scaling, and to train people to spot when they are outside them.

1. Map tasks by whether AI helps or harms before rolling it out widely.
2. Train staff to recognise tasks outside AI's capabilities.
3. Watch for uniform output when AI drafts ideas.

Evidence

Score 2 of 4. Academic, with company co-authors: led by university researchers and preregistered; three co-authors are BCG staff and one advises BCG.

3 · Original sources

Read these in full. To challenge a fact or a grade, use the contact form at marluc.co.uk or the link on https://marluc.co.uk/case-studies/bcg-field-experiment

[1] Dell'Acqua and colleagues, "Navigating the Jagged Technological Frontier"
https://www.hbs.edu/faculty/Pages/item.aspx?num=64700

[2] Dell'Acqua and colleagues, full working paper (PDF, via MIT Sloan)
https://mitsloan.mit.edu/sites/default/files/2023-10/SSRN-id4573321.pdf

Boston Consulting Group (Dell'Acqua and colleagues) website: https://www.bcg.com

Version 1.0 · Published 15 November 2026 · Updated 15 November 2026 · Facts checked against sources 15 November 2026 · Downloaded 26 November 2026. This PDF may be out of date: the latest version, and a way to report an error or a better source, is at https://marluc.co.uk/case-studies/bcg-field-experiment

VERSION HISTORY

v1.0 15 November 2026 First published