MARLUC

CASE STUDY · MARLUC SYNOPSIS

McKinsey Lilli internal AI platform

IMPLEMENTED 2023–2026 · PROFESSIONAL SERVICES · UNITED STATES · BAND D · CLAIMED ONLY

VERSION 1.0 · PUBLISHED 17 NOVEMBER 2026 · UPDATED 17 NOVEMBER 2026 · FACTS CHECKED 17 NOVEMBER 2026

McKinsey's in-house generative AI platform, Lilli, searches and summarises the firm's knowledge and drafts slides. By the firm's account over 75% of staff use it monthly and it saves up to 30% of research time. In 2026 a security researcher's agent found an unauthenticated flaw that McKinsey fixed within hours.

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Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date.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 platform with retrieval over the firm's knowledge base, plus agents.

Implemented: 2023–2026. Grade: band D, Task support, Claimed only.

Headline result, as reported: Over 75% of McKinsey's 43,000 employees use Lilli monthly, by the firm's account (June 2025)

1 · The case

The facts, as reported by the 5 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

McKinsey & Company. Global management consultancy. About 40,000 workers, down from more than 45,000 at the end of 2023, per a spokesperson relayed by Entrepreneur; an earlier report gave 43,000 employees. Canvas block: 4. Operations and process. Also 5. People and skills, 6. Data and technology.

Business challenge

The challenge. McKinsey wanted to use its entire body of knowledge and assets to drive new levels of productivity, by its then chief technology and platform officer's account at launch.

How it was addressed. McKinsey launched Lilli in 2023 as a proprietary generative AI platform over more than 40 knowledge sources and 100,000 documents. A prompt returns five to seven

Who implemented it

In-house (In-house): McKinsey built Lilli with a team of over 70 experts and an ecosystem of partners.
OpenAI (Model or cloud provider): Embedding models and GPT-4o used in retrieval and metadata.
Cohere (Model or cloud provider): Re-ranker used to refine search results.

Technology (SEAM pillar E)

SEAM types: T1 Chat assistants, T4 RAG / vector search over your own data, T7 AI agents. Deployment: Cloud APIs. Vendors and products: OpenAI, Cohere, Lilli, OpenAI embedding models, GPT-4o, Cohere re-ranker.

Where it was built in: The firm's knowledge is processed into vector stores with OpenAI embedding models; semantic and keyword search with GPT-4o-generated metadata and a Cohere re-ranker select passages, and answers cite source documents, per the former chief technology and platform officer.

What changed

Before: consultants and analysts searched the archive and built slides by hand. After: many teams generate draft presentations from prompts and search the knowledge base through Lilli, by the firm's account. Roles: McKinsey says it is not necessarily hiring fewer analysts; its headcount fell from more than 45,000 to about 40,000, a change the sources do not attribute to Lilli. [1] [2] [3] [4] [5]

What it cost

Not reported

What it saved

By McKinsey's account: up to 30% of the time consultants spent searching and synthesising information, and 1.5 million hours of work saved in a single year by its AI agents. Not independently verified. [1] [2] [3] [4] [5]

Governance and control

Lilli is the only platform approved for confidential client data. A March 2026 disclosure showed 22 API endpoints without authentication, one open to SQL injection, which security tests had not caught; McKinsey fixed it within hours and says a third-party forensic review found no evidence that client data was accessed. Accountable owner and checks on output used with clients: Not reported.

S2 · Accountable owner · Claimed
Lilli was launched under the firm's chief technology and platform officer; a named owner today is not reported. [1]

A1 · Safe use · Claimed
Lilli is the only platform where staff may enter confidential client data, and staff were trained in prompting and content validation, by the firm's account. [1] [2]

A2 · Human in the loop · Not reported
Checks on Lilli output before it is used with clients: Not reported.

A3 · Monitoring and assurance · Missing (independent source shows it was absent)
Twenty-two API endpoints lacked authentication and a SQL injection flaw went undetected for more than two years, per The Stack; McKinsey fixed it within hours once told. [5]

A4 · Vendor AI · Claimed
Lilli uses OpenAI embedding models, GPT-4o and a Cohere re-ranker, by the former technology chief's account; vendor controls are not reported. [5]

A5 · Agent controls · Not reported
McKinsey says it runs 20,000 agents; limits on what they may do are not reported. [4]

A6 · Standards and regulation · Not reported
No regulator action reported.

What worked, what did not, what is unproven

WORKED

– Adoption: over 75% of employees using Lilli monthly, by the firm's account.

– Time on search, synthesis and slide drafting: up to 30% saved, by the firm's account.

DID NOT

– Security: a researcher's AI agent reached Lilli's production database through 22 unauthenticated endpoints and a SQL injection flaw, per The Stack; CodeWall said this exposed tens of millions of chat messages and hundreds of thousands of files.

– CodeWall's founder said McKinsey's own tools and third-party penetration tests had not found the flaw for more than two years.

UNPROVEN

– Time savings and the 1.5 million hours figure, which are the firm's own and unverified.

– Whether Lilli has changed staffing, pricing or margins.

– How the quality of Lilli output used in client work is checked.

Timeline

Aug 2023 McKinsey announces Lilli; about 7,000 employees have access, with firm-wide access planned that autumn [1]

Jun 2025 McKinsey says over 75% of its 43,000 employees use Lilli monthly [2]

Jan 2026 Chief executive describes the firm as 40,000 humans and 20,000 agents [4]

Mar 2026 Security start-up CodeWall discloses an unauthenticated SQL injection flaw in Lilli; McKinsey fixes it within hours [5]

Jun 2026 Firm says teams now generate slide decks from prompts; agents saved 1.5 million hours in a year [3]

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: B. Grade: D, Task support. Proof: Claimed only. Level of change: Task support.

Task support (Level 1, band D), claimed as operating model change (band B). Consultants do the same work with AI help; the sources do not show changed roles or structure.

Task support: consultants do the same work faster with search, synthesis and slide drafting. Graded D, claimed B: the chief executive describes a workforce of humans and agents, but no change in roles, structure or costs is shown by the sources.

The SEAM view and lessons learnt

Adoption is high because the platform sits on the firm's own knowledge and is the approved route for confidential data. Savings figures are the firm's own and cannot be separated from wider changes in headcount and hiring. Before copying it, a leadership team should ask how output used with clients is checked, and whether the platform holding its most sensitive knowledge has had the same security testing as its other systems.

1. Build the assistant on your own knowledge and make it the approved route for sensitive data.
2. Secure the AI platform's prompts, data and APIs like any other critical system.
3. Measure time savings independently before using them to justify changes in staffing.

Evidence

Score 1 of 4. Self-reported: usage and savings come from McKinsey executives in press interviews; the security flaw is independently reported and confirmed by McKinsey.

Evaluation method. Self-reported company figures: adoption, prompt volumes and time saved, given in interviews and posts by McKinsey executives.

Sources disagree. Reports give McKinsey's headcount as 43,000 employees (June 2025) and about 40,000 workers (spokesperson to Bloomberg, relayed in June 2026). CodeWall's figures on exposed data come from its own account; McKinsey says its forensic review found no evidence client data was accessed.

Status. Ongoing: in firm-wide use; security flaw fixed in March 2026.

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/mckinsey-lilli

[1] CIO Dive: McKinsey rolls out generative AI platform Lilli to internal employees
https://www.ciodive.com/news/McKinsey-generative-AI-Lilli-platform-internal-employees/691231/

[2] Entrepreneur: McKinsey Is Using AI to Create PowerPoints and Take Over Junior Employee Tasks
https://www.entrepreneur.com/business-news/ai-creates-powerpoints-at-mckinsey-replacing-junior-workers/492624

[3] Entrepreneur: McKinsey Consultants Are Letting New Technology Take Over an Essential Part of Their Work
https://www.entrepreneur.com/business-news/mckinsey-consultants-are-letting-ai-take-over-their-work

[4] HRD (HCAMag): McKinsey trials AI-led job interviews as 20,000 AI agents reshape its workforce
https://www.hcamag.com/ca/news/general/mckinsey-trials-ai-led-job-interviews-as-20000-ai-agents-reshape-its-workforce/562017

[5] The Stack: McKinsey's AI agent Lilli hacked by another AI agent
https://www.thestack.technology/mckinsey-ai-agent-hacked-lilli/

McKinsey & Company website: https://www.mckinsey.com

Version 1.0 · Published 17 November 2026 · Updated 17 November 2026 · Facts checked against sources 17 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/mckinsey-lilli

VERSION HISTORY

v1.0 17 November 2026 First published