Optima Consulting

Research publication

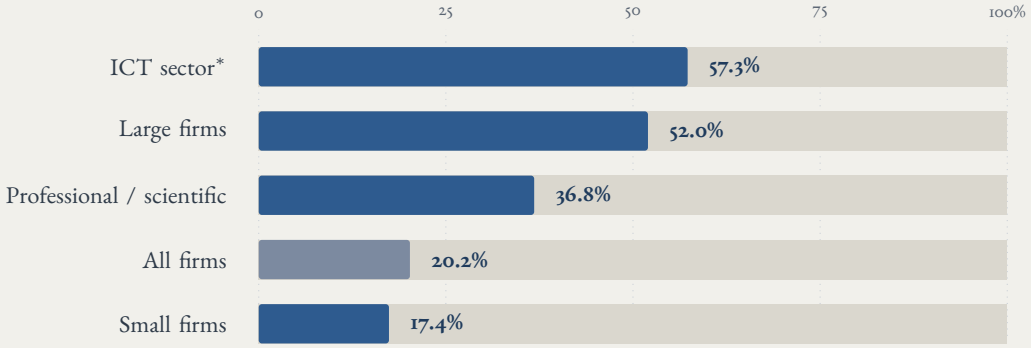
AI in Financial Modelling: What the 2025–2026 Evidence Shows

Adoption, reliability and the continuing role of professional judgement.



Firm AI use in 2025: adoption is real, but uneven

Share of firms reporting AI use, OECD countries with available data, 2025

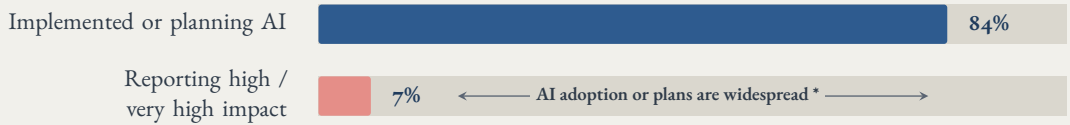


*ICT = Information and Communication Technology — the OECD sector covering software, IT services, telecom and data-processing firms

Source: OECD firm-level AI-use statistics, 2025 snapshot.

Deployment is not value

Finance organisations and AI — bars: Gartner CFO survey, June 2025 (n = 183)



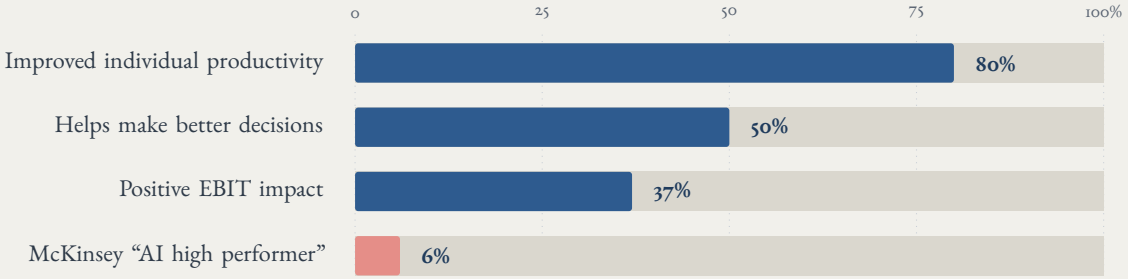
* However, high reported impact remains rare.

Gartner finance AI research, March 2026 (n = 204): 66% of adopters cite efficiency / productivity as the leading benefit
63% report implementation ran slower than expected

Source: Gartner, CFO survey, June 2025 (n = 183); finance AI research, March 2026 (n = 204).

Productivity gains are more widespread than reported financial impact

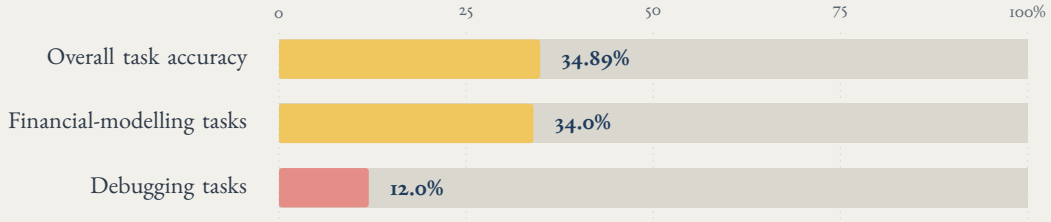
Share of respondents, McKinsey State of AI global survey 2026 (n = 1,719, May–June 2026)



Source: McKinsey, The State of AI: Global Survey 2026 (archived).

End-to-end spreadsheet workflows remain unreliable

Highest task accuracy reported in the original paper (Claude Opus 4.6) — SpreadsheetBench 2, 321 expert-annotated business workflows

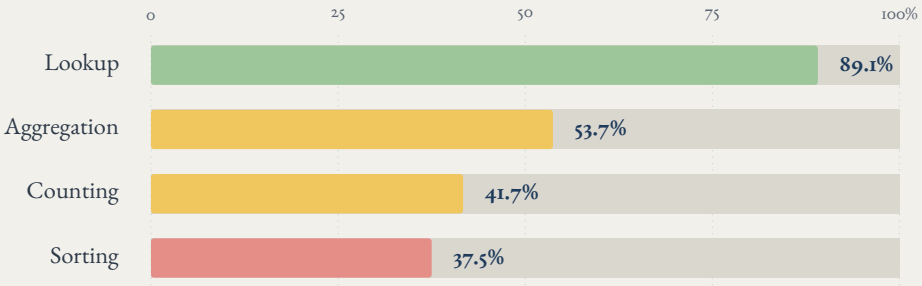


Successful cell-level edits do not equal a correct end-to-end workbook — the dominant failure modes are insufficient spreadsheet inspection and wrong target-cell selection.

Source: Zhu et al., SpreadsheetBench 2, arXiv:2606.29955 (2026).

Financial-data accuracy varies substantially by task

Pooled accuracy across ten model configurations on financial-data reasoning tasks (152 companies, 8 funds), FinSheet-Bench 2026

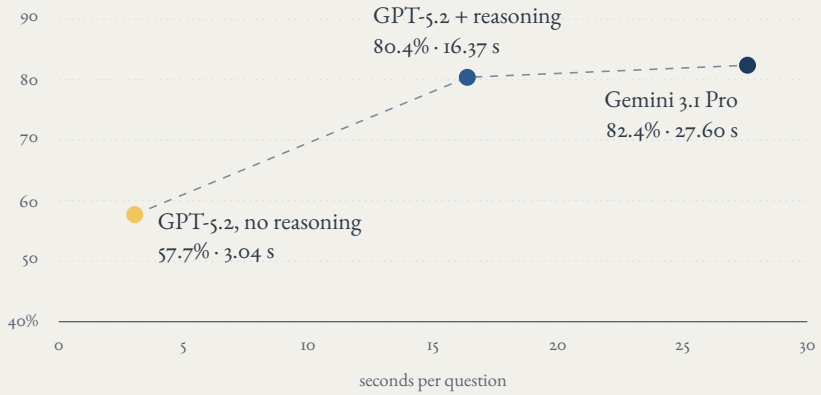


All-model average fell from 86.2% on the highest-scoring file to 48.6% on the largest; difficulty patterns were consistent across all ten models.

Source: Ravnik et al., FinSheet-Bench, arXiv:2603.07316 (2026).

Reasoning improves accuracy, but increases response time

Accuracy versus response time on financial spreadsheets, FinSheet-Bench 2026 — two GPT-5.2 configurations tested; Gemini 3.1 Pro shown for comparison



Source: Ravnik et al., FinSheet-Bench, arXiv:2603.07316 (2026).

The best AI agent passed fewer than one in six finance tasks

MBABench 2.0: a task passes only if the answers are correct and *the workbook meets professional-quality criteria — assumptions, formulas, error checks, documentation*

15.8%

top agent pass rate — 16 of 101 tasks

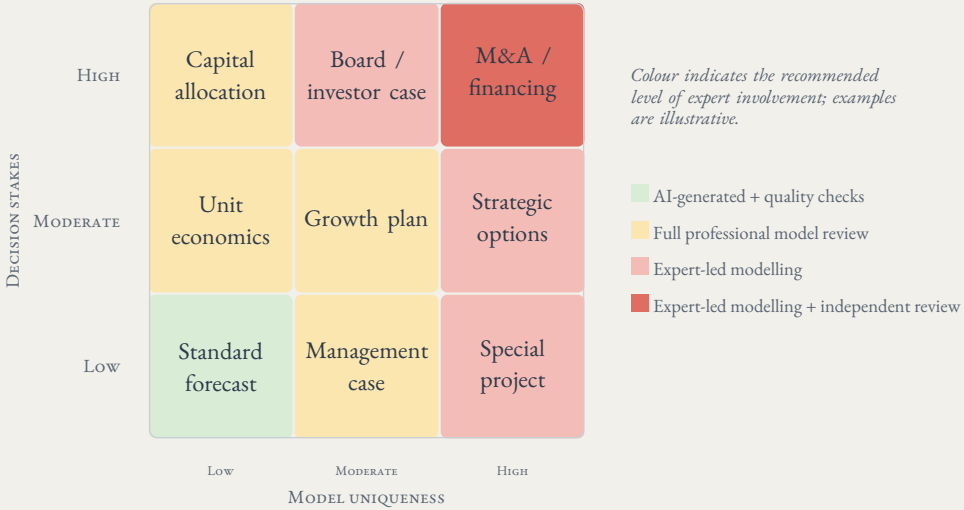


101 expert-written public finance tasks · 20 agents evaluated · 129 quality criteria across 12 categories

Source: MBABench 2.0 leaderboard, October 2026.

Higher stakes & greater model uniqueness call for more expert involvement

An illustrative framework for AI-enabled financial modelling — not an empirical risk score



Optima Consulting

Financial modelling and corporate advisory.

info@optimaconsulting.io

optimaconsulting.io

Auckland CBD
New Zealand
NZBN 9429051066482

Expertise

- Strategy and M&A
- Corporate finance
- Indicative valuation
- Financial modelling
- Process optimisation
- Process automation
- System integration

Industries

- FMCG
- Mining
- Telecom
- Financial
- Agriculture
- Forestry
- Pharma

Explore

- Services
- Approach
- Resources
- About
- Contact

© 2026 Optima Consulting Limited