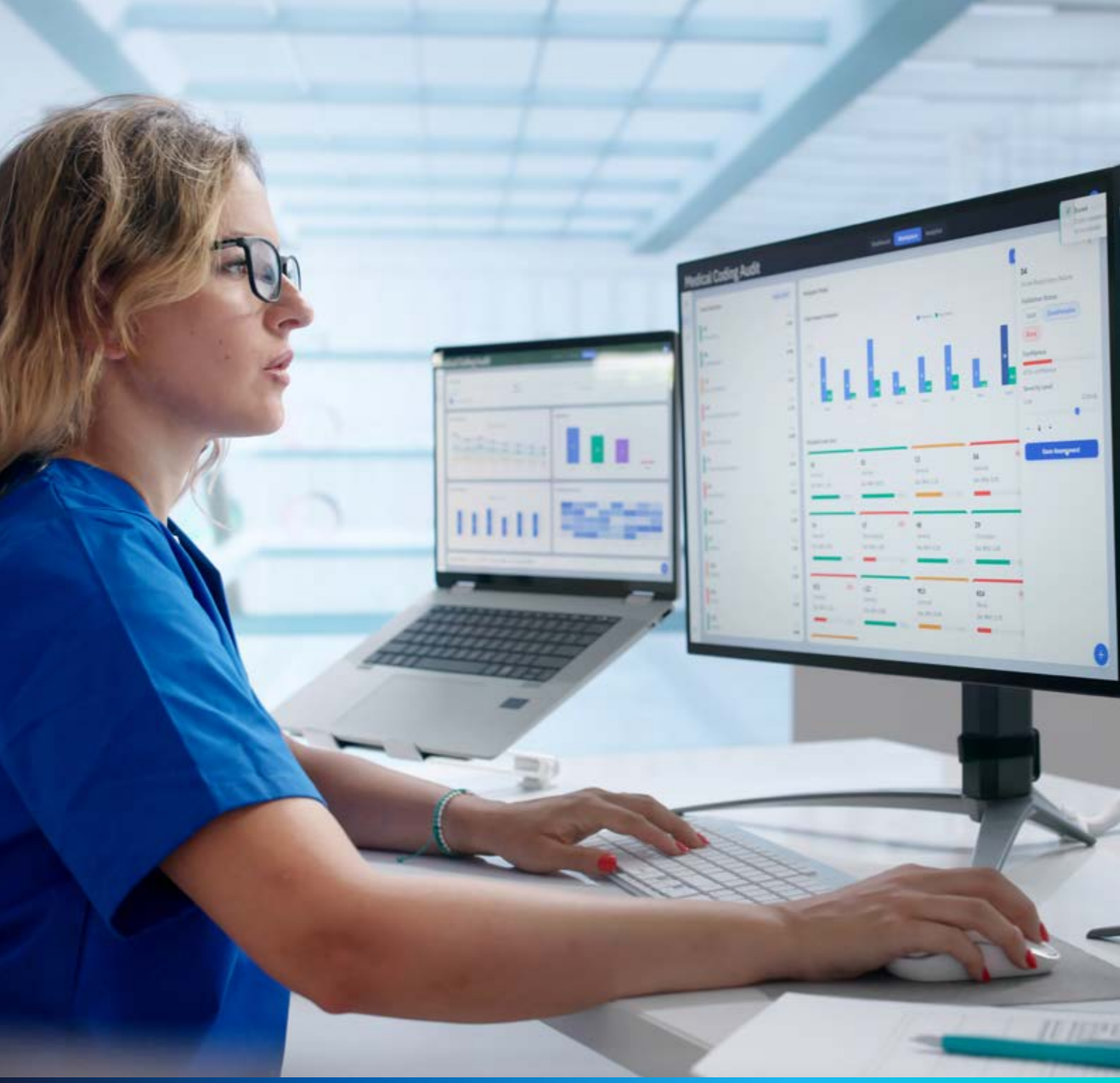


Capgemini invent

Clinical Data Management: From Evolution to Revolution

Building the Autonomous Trial Operating System — A Capgemini Point of View

Executive Summary

Clinical Data Management (CDM) has become the biggest determinant of how fast, compliant, and competitive a drug development organization can be. For two decades CDM sat downstream of the trial, judged only on data cleanliness at database lock. That model is now obsolete, data volume has multiplied across EDC, labs, ePRO/eCOA, wearables, and real-world sources, while regulators have simultaneously tightened data-governance expectations (ICH E6(R3), effective 2025) and opened a credible pathway for AI and real-world evidence in submissions (FDA's January 2025 draft AI guidance; December 2025–January 2026 RWE guidance updates).

The technology response has moved in waves - analytics, then AI/ML, then Generative AI copilots, and now Agentic AI, autonomous, goal-directed agents that plan, execute, and learn across CDM workflows, with human oversight designed in rather than bolted on. IQVIA (IQVIA.ai, March 2026) and Veeva (Vault AI Agents, rolling out since October 2025) have already placed platform-scale bets. The question for sponsors is no longer whether agentic AI arrives, but whether their data foundation and governance model can absorb it safely.

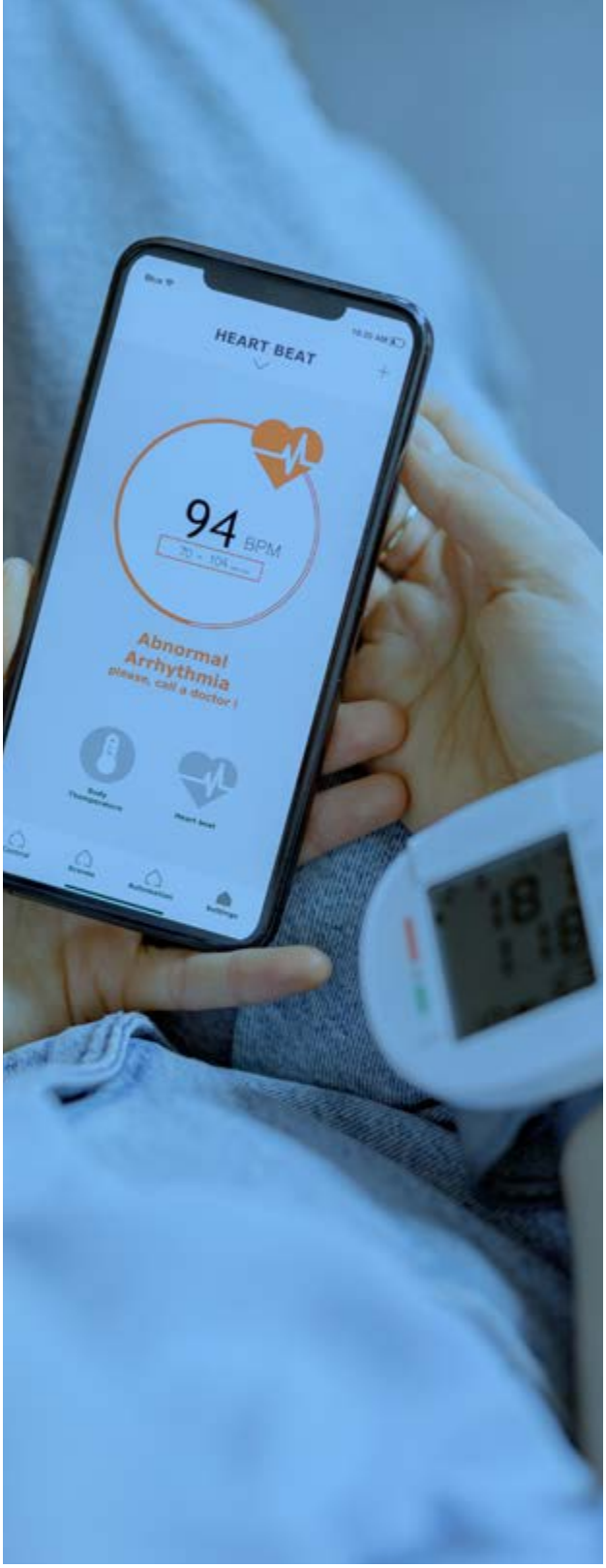
Our Point of View - the fragmented CDM model is structurally incompatible with today's regulatory and data complexity, this is an inflection point, not an optimization exercise. A domain-driven, layered data architecture is the non-negotiable foundation for any agentic capability; agents deployed on fragmented data simply automate the fragmentation faster. Winners will pair automation with disciplined governance such as explainability, human-in-the-loop thresholds, accountability, and scale it in stages - assess data-foundation maturity, pilot a high-volume/low-risk workflow, then scale under a named governance model.

The Inflection Point

Bringing a therapy to market still takes roughly a decade or more and can cost $1–2 billion per approval including failures (Tufts CSDD), putting a premium on adaptive, data-driven operations. Most CDM systems - EDC, CTMS, labs, ePRO/eCOA, EHRs have evolved independently and are stitched together through manual reconciliation, as trials grow more complex and decentralized, this fragmentation compounds, burying safety signals and protocol deviations until they surface too late to act on cheaply. The same gap drives two of the industry's most visible failures - most studies miss enrolment timelines, and phase-dependent attrition frequently reaches 20–30%, not because of one mistake, but because small failures (missed signals, slow responses) compound into a systems problem, not a patient problem.

Regulatory change now reinforces this. **ICH E6**(R3) (adopted January 2025, EU-effective July 2025) raises expectations for proportionate, risk-based oversight and sponsor accountability for data governance even when execution is outsourced. The **FDA's January 2025 draft AI guidance** (final expected mid-2026, aligned to a January 2026 joint FDA–EMA statement) gives sponsors a first credibility framework for AI in regulatory submissions, while the **FDA's December 2025 RWE guidance** update allows de-identified real-world data in submissions while raising the bar on provenance and traceability. Together, incremental fixes - another dashboard, another manual check - no longer close this gap; CDM must be re-architected as the foundation of trial execution, not a downstream function of it.

From Manual to Agentic – The CDM Evolution



CDM has moved through six recognizable stages, and most sponsors sit across at least three at once: **manual** (paper CRFs, slow reconciliation) ⟶ **analytics** (dashboards describing what already happened) ⟶ **adaptive designs** (data-driven interim analyses, still constrained by data lag) ⟶**AI/ML** (forecasting enrolment and risk, still human-coordinated) ⟶ **Gen AI** (drafting protocols and documents, conversational data access, still human-led) ⟶ **Agentic AI** (autonomous agents executing multi-step workflows - query triage, reconciliation, next-best-action - with defined human checkpoints). This is a re-architecture of the trial operating system, with autonomy, explainability, and governance at its core, not autonomy alone.

The evidence is now concrete, Pfizer's ML enabled "Smart Data Query" compressed data cleaning from ~30 days to ~22 hours in its COVID-19 trials; Novartis's Data42 initiative unifies multi-trial datasets for predictive site selection and protocol optimization; Medidata, with partners like PHASTAR, has compressed review cycles and standardized TLF oversight. Most recently, IQVIA launched IQVIA.ai (March 2026), a unified agentic platform spanning clinical, commercial, and real-world domains, while Veeva's Vault AI Agents (from October 2025) put Clinical Data agents on its roadmap for December 2026 - evidence that the vendors underpinning most sponsors' existing estates are building agentic capability into the platforms already in use. (Results are as publicly disclosed; outcomes vary by trial complexity and data maturity)

Technology	Complexity	Cost Impact	Adaptability
Manual CRFs	Low	Low per task, high total cost via rework	Low
Analytics Dashboards	Medium	Medium	High
Adaptive Designs	High	Medium-High (reduced amendments, optimized trial execution)	Very High
AI/ML	High	Medium–high savings (forecasting, RBQM)	Medium
Gen AI	Medium	Medium–high (authoring automation)	Medium–High
Agentic AI	Medium	High–very high potential, governance-dependent	Medium–High

The Domain-Driven CDM Platform

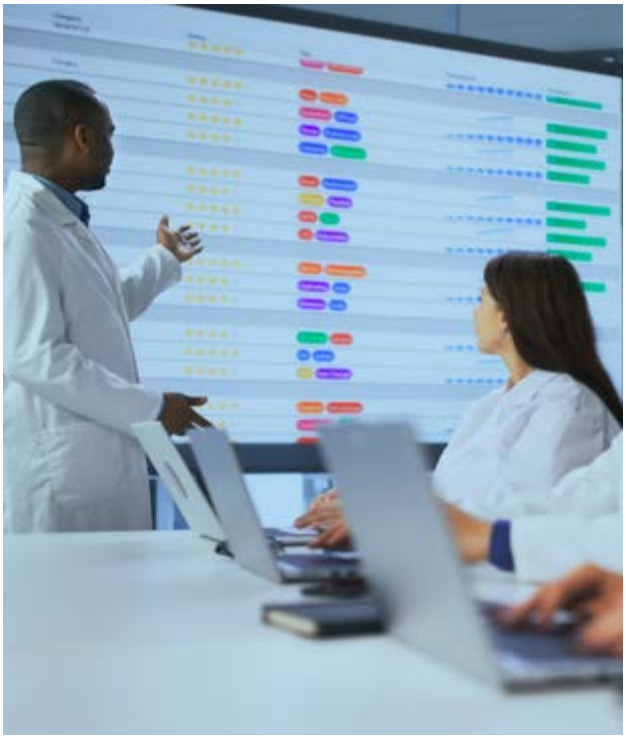
Modern platforms separate concerns into layers so they can scale without losing oversight, ingestion (EDC, CTMS, labs, ePRO/eCOA, wearables - structured and unstructured), standardization (CDISC and controlled terminologies, ensuring cross-site/vendor consistency), analytics (ML forecasting and anomaly detection, plus Gen AI natural-language querying), visualization (role-based, real-time dashboards), and governance, which for an agentic platform must go beyond audit trails to define which decisions agents may execute autonomously, how agent reasoning is exposed for review (explainability), and how model drift is monitored (AI model risk management) - aligned to GCP, ICH E6(R3), 21 CFR Part 11, GDPR, and HIPAA. At the core sits a single source of truth with persona-based access for sponsors, CROs, and sites, configurable across study designs rather than custom-built each time - reducing manual reconciliation, deviations, and cross-region blind spots while strengthening sponsor-CRO-site collaboration.

Agentic AI in Action

Five use cases are proving out fastest, each pairing automation with a defined human checkpoint,

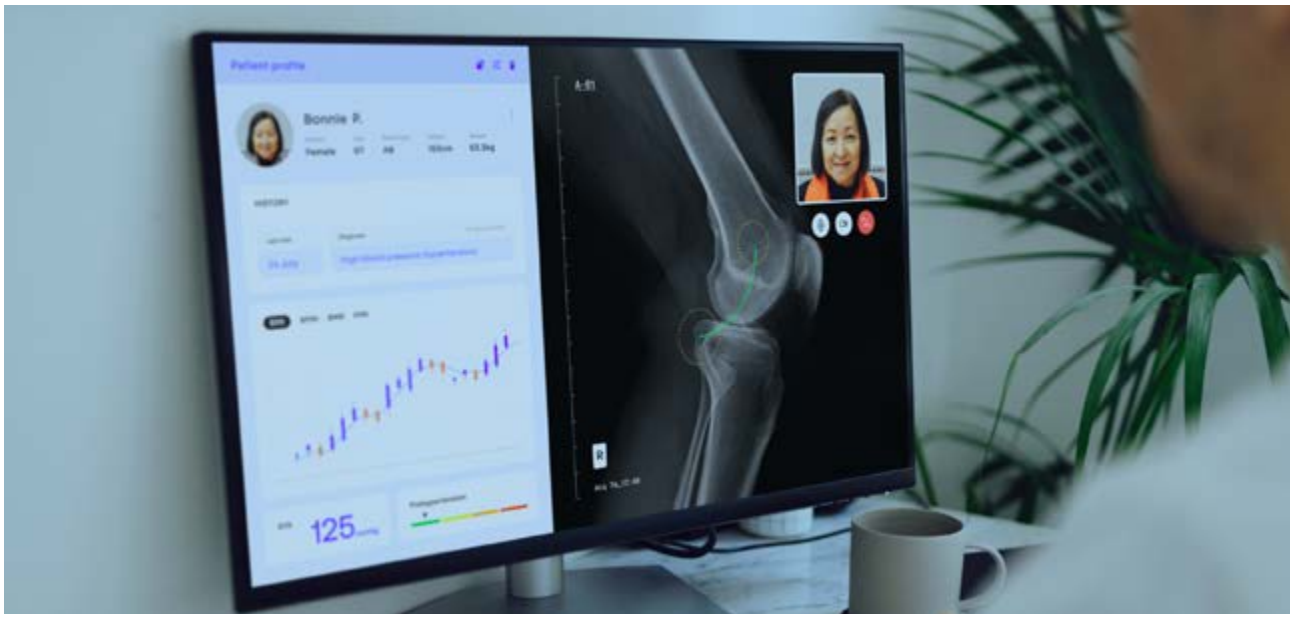
- **Autonomous query management** (agents auto-resolve low-risk queries within set thresholds, routing ambiguous ones to a data manager - early deployments report 60–80% effort reduction, though this varies by therapeutic area and threshold conservatism)
- **End-to-end workflow automation** (discrepancy detection » query generation » site communication » closure as one governed flow)
- **Protocol-to-CRF/SDTM automation** (agents draft eCRFs, edit checks, and specs from protocol text, with reported 30–50% study-build reductions in early pilots, always human-reviewed pre-go-live)
- **Continuous data-quality monitoring** (agents run 24/7 rather than at scheduled checkpoints, cutting mid-study rework), and
- **System interoperability with human oversight** (agents remove duplication across EDC, CTMS, eCOA, safety, and labs, with escalation criteria keeping a human accountable for safety or compliance critical decisions).

The common thread, agentic AI succeeds where data is already standardized and escalation thresholds are explicit - deployed on fragmented data, it simply automates inconsistency faster.



Opportunity Map Across the Trial Lifecycle

Trial Phase	Agentic AI Opportunity
Study start-up	Protocol-to-Spec Agent auto-drafts eCRFs, edit checks, and visit schedules, flags risk, generates UAT scripts
Study set-up	Integration Orchestrator Agent provisions connectors, validates mappings, enforces CDISC conformance
Third-party data tracking	Quality & Controls Agent stress-tests rules, explains failed checks (explainable AI), tracks audit readiness
Study build	Data Reconciliation Agent detects anomalies, triages issues, proposes coding (MedDRA/ WHODrug) for review
Study conduct	RBQM Agent forecasts risk and proposes next-best-action, Medical Review Agent surfaces AE/SAE patterns
Query management	Intelligent Query Agent clusters/auto-resolves low-risk queries, routes complex ones, maintains traceability
Study close-out	Submission Authoring Agent drafts TLF narratives, CSR sections, and data-lineage appendices



Where to start - Query management and Protocol-to-CRF automation offer the fastest, lowest-risk return – high volume, well-bounded tasks with mature standards already in place. Study conduct and Close-out carry higher regulatory visibility and should follow only once the governance model is proven on lower-risk workflows.

Capgemini's Point of View – A Staged Path Forward

Agentic AI is becoming a competitive necessity – however, one earned through data-foundation work and governance discipline, not skipped by buying a platform. Drawing on Capgemini Research Institute's 2026 life sciences research and client transformation programmes, we recommend three stages,

1. Assess data-foundation maturity (standardization, interoperability, governance) before selecting a use case,

2. Pilot one high-volume, lower-risk workflow with explicit human-oversight thresholds, measuring exception rates alongside efficiency, and

3. Scale under governance into higher-risk phases only once explainability, escalation criteria, and audit evidence are proven - under one named governance model, not disconnected pilots.

To address concerns that agentic AI displaces expertise, the evidence to date, and Capgemini's own research position, is that it amplifies scientific and operational judgment rather than replacing it - removing low-value reconciliation work so data managers, monitors, and biostatisticians spend more time on decisions that need human expertise.

Regulatory tightening (ICH E6(R3)) and a newly credible AI/RWE pathway (FDA 2025–2026 guidance) have arrived at the same moment the major CDM vendors are racing to embed agentic capability into existing platforms - standing still is a decision to fall behind.

Enterprises that act now, starting with a clear-eyed data-foundation assessment and a disciplined, governed pilot, will set the pace for quality, speed, and trust in drug development. Capgemini partners with life sciences organizations to make this an engineered, governed capability build and would welcome the opportunity to discuss a tailored data and AI maturity assessment for your organization.

**Disclaimer - This document reflects publicly available information and Capgemini analysis as of September 2026. Company examples are illustrative, not exhaustive, reported efficiency figures are early-deployment ranges, not assured outcomes. This paper does not constitute regulatory, legal, or financial advice.*

Author

Amit Patil
Director
amit.j.patil@capgemini.com

Key Contributor

Dr Sarika Vanarse
Senior Director
sarika.vanarse@capgemini.com

About Capgemini Invent

As the digital innovation, design and transformation brand of the Capgemini Group, at Capgemini Invent we enable CxOs to envision and shape what's next for businesses. Located in over 40 studios and more than 70 offices around the world, we are a 13,500 strong team of strategists, data scientists, product and experience designers, brand experts and technologists who develop new digital services, products, experiences and business models for sustainable growth.

Capgemini Invent is an integral part of Capgemini, an AI-powered global business and technology transformation partner, delivering tangible business value. We imagine the future of organizations and make it real with AI, technology and people. With our strong heritage of nearly 60 years, we are a responsible and diverse group of over 420,000 team members in more than 50 countries. We deliver end-to-end services and solutions with our deep industry expertise and strong partner ecosystem, leveraging our capabilities across strategy, technology, design, engineering and business operations. The Group reported 2025 global revenues of €22.5 billion.

www.capgemini.com

Make it real.