

Robustheit statt Revolution

Designprinzipien für erfolgreiche KI-Implementierung im Controlling

Generative Künstliche Intelligenz (GenAI) im Controlling verspricht Effizienzgewinne und Automatisierung – doch in der Praxis scheitern viele Projekte. Anhand von vier realen GenAI-Initiativen eines der größten deutschen Unternehmen zeigt dieser Beitrag, warum KI-Projekte im Controlling scheitern und wie sich erfolgreiche Projekte davon systematisch unterscheiden.

Erik Strauß und Solin Rafiq

1. Von der Euphorie zur Evidenz: GenAI-Implementierung im Controlling

Seit der öffentlichen Verfügbarkeit generativer Sprachmodelle wie ChatGPT Ende 2022 erlebt die Diskussion um Künstliche Intelligenz – und allen voran GenAI – im Controlling eine neue Dynamik. In Fachmedien (Englisch et al., 2026) und Praxisbeiträgen (Challapally et al., 2025; Demyttenaere et al., 2023) dominieren Begriffe wie „Automatisierung“, „selbstlernende Systeme“ oder „Predictive Finance“. Viele Unternehmen verbinden damit die Erwartung, Routinen zu automatisieren, Analysen zu beschleunigen und Entscheidungen zu verbessern. Zugleich mehren sich erste Erfahrungen, die ein anderes Bild zeichnen (Challapally et al., 2025): Zahlreiche GenAI-Initiativen scheitern, weil sie nicht auf tragfähigen technischen, organisatorischen und kulturellen Grundlagen beruhen.

Dieser Beitrag zeigt, warum gerade im Controlling GenAI keine Ad-hoc-Revolution auslöst, da dort Informationsqualität, Nachvollziehbarkeit und Governance zentrale Anforderungen sind. Vielmehr zeigt sich, dass Robustheit – verstanden als die Fähigkeit, Technologie stabil, kontrollierbar und lernfähig zu implementieren – zur entscheidenden Erfolgsvoraussetzung wird. Es wird aufgezeigt, welche Designprinzipien für eine robuste KI-Implementierung im Controlling maßgeblich sind. Aufbauend auf einer kontrastierenden Mehrfallstudie in einem der größten deutschen Industrieunternehmen werden drei gescheiterte und ein erfolgreiches Projekt vergleichend analysiert. Die kontrastierende Betrachtung erlaubt es, wiederkehrende Muster des Scheiterns und Erfolgsfaktoren zu identifizieren.

Das Ziel des Beitrags besteht darin, aus wissenschaftlicher Perspektive zu zeigen, wie sich tech-

nologische Innovation und organisatorische Einbettung gegenseitig bedingen. Gleichzeitig werden auf dieser Grundlage praxisorientierte Gestaltungsempfehlungen abgeleitet, die Führungskräfte bei der Planung und Umsetzung von GenAI-Projekten im Controlling unterstützen sollen.

Die Struktur des Artikels ist wie folgt aufgebaut: Abschnitt 2 erläutert den theoretischen Bezugsrahmen und entwickelt das Konzept der Robustheit. Abschnitt 3 beschreibt das Forschungsdesign und die empirische Fallbasis. Abschnitt 4 stellt die Ergebnisse der Fallanalyse dar, bevor Abschnitt 5 daraus zentrale Designprinzipien ableitet. Abschnitt 6 schließt mit einem Fazit.

2. Disruptive vs. robuste Einbettung von genAI

Die Diskussion um den Einsatz von GenAI wie ChatGPT oder Claude im Controlling (verstanden als Einsatz von LLMs und den dahinterliegenden Modellen) ist bislang stark von technologischen und disruptiven Narrativen geprägt. Begriffe wie „Predictive Analytics“, „Cognitive Finance“ oder „Autonomous Forecasting“ verheißen einen tiefgreifenden Wandel der Steuerungsfunktion.

Gleichwohl zeigen empirische Befunde aus der jüngeren Controlling-Forschung, dass technologische Innovationen nur dann nachhaltige Wirkung entfalten, wenn sie in robuste organisatorische, prozessuale und kulturelle Strukturen eingebettet werden (Franke/Hiebl, 2023; Papiorek/Hiebl, 2024). Franke und Hiebl (2023) verdeutlichen in ihrer Studie zu Big Data und Entscheidungsqualität, dass die Qualität von Entscheidungen nicht allein von der Menge oder Vielfalt der Daten abhängt, sondern von der Analytikkompetenz und Kontextintegration der Controlling-Abteilung. Big-Data-Systeme



Prof. Dr. **Erik Strauß** ist Inhaber des Lehrstuhls für Management Control an der ESCP Business School in Berlin und Geschäftsführer von Strauss MindTech. E-Mail: estrauss@escp.eu



Solin Rafiq, M.Sc., ist wissenschaftliche Mitarbeiterin am Dr. Werner Jackstädt-Stiftungslehrstuhl für Controlling und Unternehmenssteuerung an der Universität Witten/Herdecke.

Zentrale Aussagen

- GenAI ist im Controlling keine Ad-hoc-Revolution: Entscheidend ist Robustheit über Technik, Daten, Prozesse, Rollen und Kultur.
- Scheitern entsteht v. a. durch instabile Outputs, schlechte Daten und fehlende Prozess-/Rollenintegration – nicht primär durch das Modell.
- Erfolgreiche GenAI-Implementierung im Controlling basiert auf klarem Scope, kuratiertem Korpus, RAG, Soft-Go-Live und Governance.

generieren nur dann einen Mehrwert, wenn explizites Wissen aus Daten mit dem „tacit knowledge“ der handelnden Controlling-Akteure interagiert. In ähnlicher Weise zeigen Papiorek und Hiebl (2024), dass Informationssystemqualität in der Controlling-Funktion nur dann zu einer höheren Effektivität von Controlling-Systemen führt, wenn Prozesse standardisiert und zugleich an das Automatisierungsniveau der Organisation angepasst sind. Diese Ergebnisse unterstreichen, dass die Digitalisierung des Controllings vor allem dann produktiv wirkt, wenn sie von klaren Rollen, standardisierten Prozessen und digitaler Kompetenz begleitet wird (Fähndrich/Pedell, 2025). Dementsprechend erlangt ein systemisches Verständnis von Managementsteuerung noch größere Bedeutung, da smarte Technologien wie GenAI ihr Potenzial erst entfalten, wenn Unternehmen analytische Fähigkeiten entwickeln, Datenarchitekturen zu professionalisieren und eine lernorientierte Governance zu etablieren (Lombardi et al., 2021).

Dieser auf den ersten Blick logisch erscheinende Wandel geht jedoch mit einem fundamentalen Bruch einher: Während klassische Controlling-Systeme auf deduktiver Logik aufbauen, d. h., Daten werden entlang vorgegebener Kennzahlen und Regeln interpretiert, folgt GenAI einer induktiven Logik, d. h. dem Lernen aus Daten. Beispielsweise entwickeln Modelle basierend auf der Machine-Learning-Technologie Muster, ohne dass das Zielsystem vollständig vorgegeben ist. Damit entsteht eine neue Form des Wissens im Controlling, die Unsicherheiten erhöht und die Rolle des Controllings neu definiert: vom Interpretierenden fester Regeln zum Kuratierenden lernender Systeme.

Vor diesem Hintergrund schlägt der vorliegende Beitrag vor, GenAI-Implementierungen im Controlling nicht als disruptives, sondern als robustes Designproblem zu verstehen. Ein robustes Designproblem der GenAI-Implementierung liegt dann vor, wenn GenAI im Controlling als zu gestaltendes soziotechnisches System, dessen Nutzen erst durch verlässliche Einbettung entsteht, verstanden wird. Dadurch stehen klare Zielbilder und Use-Case-Logik, saubere Daten- und Modellgrenzen (z. B. was das Modell darf und was nicht), definierte Rollen und Verantwortlichkeiten (Owner, Reviewer, Freigaben), standardisierte End-to-End-Prozesse (inkl. Kontrollen, Dokumentation und Eskalation) sowie Kompetenzaufbau und eine Governance, die Ler-

nen ermöglicht (Monitoring, Feedback-Loops, Anpassung von Prompts/Policies/Workflows), im Vordergrund. „Robust“ bezieht sich darauf, das System so zu entwerfen, dass es auch bei Halluzinationen, Modellwechseln, Datenlücken, oder veränderten regulatorischen Anforderungen stabil performt und kontrolliert weiterentwickelt werden kann.

Ein disruptiver Ansatz unterscheidet sich davon durch seinen Schwerpunkt auf Substitution und radikalen Funktionsbruch („Autonomous Forecasting ersetzt Planung“, „Menschen in der Controlling-Funktion werden obsolet“) und die damit einhergehende Vernachlässigung von Integrationsarbeit (Prozessdesign, Verantwortlichkeit, Qualitätssicherung, Akzeptanz, Anschlussfähigkeit an Steuerungslogiken). Ein robustes Design zielt dagegen auf schrittweise, kontrollierte Leistungssteigerung durch Gestaltung von Schnittstellen zwischen induktiven GenAI-Outputs und deduktiven Controlling-Regimen (d. h. Logiken, Regeln und Routinen) – mit der Controlling-Funktion als Kurator und Qualitätsinstanz statt als bloßem Nutzer oder als ersetzter Rolle.

Zur Ableitung geeigneter Dimensionen von Designprinzipien für robuste GenAI-Implementierungen im Controlling baut der vorliegende Beitrag auf etablierter Information Systems-Forschung (IS-Forschung) auf. Der IS-Erfolgsforschung folgend (Patel et al., 2026; Wang et al., 2024), ergeben sich durch die analytische Trennung von Systemqualität und Informationsqualität technische Robustheit (Reliabilität/Performanz/Integrationsfähigkeit der Systemarchitektur) und Datenrobustheit (Qualität der bereitgestellten Information) als die beiden ersten wesentlichen Dimensionen der Designprinzipien. Diese Logik wird durch die Datenqualitätsforschung weiter präzisiert, welche zeigt, dass Vertrauen und Verwendbarkeit von Daten gerade an Kriterien wie Konsistenz, Verständlichkeit/Interpretierbarkeit und Kontextpassung hängen, d. h. an einer strukturierten, konsistenten und nachvollziehbaren Datenbasis (Wang/Strong, 1996). Die so entwickelte Datenbasis muss gleichzeitig jedoch auch in die bestehenden Prozesse derart eingebunden sein, dass die Ergebnisse verlässlich und wiederholbar für Entscheidungen genutzt werden können, auch wenn Daten, Modelle oder Rahmenbedingungen schwanken (Dumas et al., 2023). Es muss also das Kriterium der Prozessrobustheit erfüllt sein. Daneben hat die IS-Forschung herausgearbeitet, dass auch die Überprüfung auf Gültigkeit der Daten und Prozesse ein weiteres elementares Designprinzip darstellt (z. B. Grønsund/Aanestad, 2020). Aufgaben, Entscheidungsrechte und Verantwortlichkeiten müssen auch unter Unsicherheit (z. B. schwankende Modellgüte, Black-Box-Outputs) klar, stabil und prüfbar bleiben. Die daraus resultierende Rollenrobustheit liegt vor, wenn (a) eindeutig ist, wer GenAI-Ergebnisse erstellt/konfiguriert, wer sie prüft/validiert, wer sie freigibt und

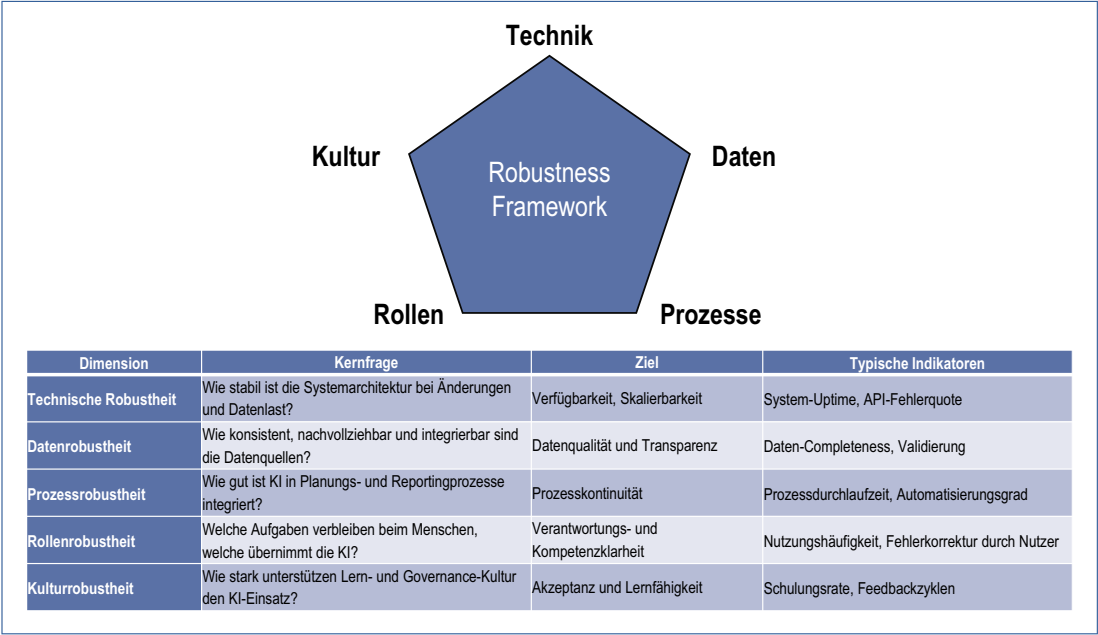


Abb. 1: Fünf Dimensionen robuster KI-Implementierung im Controlling

wer für Folgen haftet bzw. rechenschaftspflichtig ist; (b) die Organisation Rollen für Überwachung, Korrektur und Lernen institutionalisiert (z. B. „Auditing/ground truth“, Prompt-/Policy-Owner, Modell-Monitoring), statt GenAI bloß „in die Linie“ zu geben; und (c) Rollen so definiert sind, dass Verantwortungsdiffusion (keine Zuständigkeit) vermieden wird. Der Umgang mit der Verantwortung und den Rechtfertigungspflichten für Design, Nutzung und Outcomes von IS-Systemen hängt von der unternehmensinternen Akzeptanz, Lernorientierung, und dem Umgang mit Risiken ab (Li et al., 2025), d. h. mit der Kulturobustheit des Unternehmens. Es muss sichergestellt sein, dass auch bei sich verändernden ökonomischen Rahmenbedingungen und Technologien das Unternehmen in der Lage ist, eine Anpassung der Verantwortung vorzunehmen.

Zusammenfassend lassen sich somit aus der bisherigen IS-Forschung fünf Dimensionen organisationaler Robustheit ableiten, die als analytisches Raster für die hier vorliegende empirische Untersuchung dienen (vgl. Abb. 1):

- 1. **Technische Robustheit** (Patel et al., 2026): Stabilität der Systemarchitektur, Schnittstellen und Performanz auch bei steigender Komplexität.
- 2. **Datenrobustheit** (Wang et al., 2024): Strukturierte, konsistente und nachvollziehbare Datenbasis als Grundlage für Vertrauen in KI-Ergebnisse.
- 3. **Prozessrobustheit** (Dumas et al., 2023): Integration von KI-Anwendungen in bestehende Planungs-, Reporting- und Entscheidungsprozesse.
- 4. **Rollenrobustheit** (Grønsund/Aanestad, 2020): Neuverteilung von Aufgaben zwischen Mensch und Maschine bei klarer Verantwortlichkeit („Human-in-the-Loop“).
- 5. **Kulturobustheit** (Li et al., 2025): Lern- und Risikokultur, die experimentelles Arbeiten erlaubt und zugleich Governance-Standards wahrt.

Diese fünf Dimensionen bilden den konzeptionellen Bezugsrahmen des Beitrags und dienen der Systematisierung der empirischen Mehrfallstudie. GenAI wird dabei als soziotechnisches Phänomen (Trist/Emery, 2015) verstanden, dessen Erfolg weniger von disruptiven Technologiesprüngen als von kohärenten, belastbaren Designentscheidungen abhängt.

3. Methodik

Zur Entwicklung der in diesem Beitrag vorgestellten Designprinzipien robuster KI-Implementierung im Controlling wurde eine kontrastierende Mehrfallstudie durchgeführt (Yin, 2014). Ziel war es, aus realen Implementierungsprojekten systematisch zu identifizieren, welche Gestaltungsentscheidungen den Erfolg oder das Scheitern von GenAI-Initiativen im Controlling bestimmen. Dieses qualitative Vorgehen erlaubt es, sowohl technologische als auch organisationale und kulturelle Einflussfaktoren in ihrer Wechselwirkung zu erfassen.

Auswahl der Fälle

Untersuchungsgegenstand ist ein international tätiges Unternehmen der Softwareindustrie mit über 300.000 Mitarbeitenden und mehr als 30 Milliarden Euro Umsatz in 2025, das seit 2022 verschiedene Projekte zur Nutzung generativer KI im Controlling initiiert hat. Das Unternehmen gilt als „Digital Fast Follower“, d. h. es verfügt über eine moderne zentrale ERP- und BI-Infrastruktur, jedoch über ein dezentral organisiertes Controlling mit hoher funktionaler Eigenverantwortung der Geschäftseinheiten.

Im Mittelpunkt der Analyse stehen vier konkrete GenAI-Projekte im Controlling, die jeweils unterschiedliche technologische Ansätze und Implementierungsstrategien repräsentieren (vgl. Abb. 2).

Interviewpartner	Job	Projekt-Involvierung	Rolle im Projekt	Unternehmensfunktion
Michael	Head of Finance IT	Athena, Delphi, & Iris	Projektiator / Testen und Sicherstellen, dass die Gen-AI-Lösung korrekte Antworten liefert	Controlling
Hugo	Web Consultant / Web-Entwickler	Athena, Delphi, & Iris	Projektmitarbeiter / Aufbau der Gen-AI-Lösung / Pilotarbeit	Computer Scientist
Sebastian	Financial Processes Consultant	Athena, Delphi, & Iris	Projektmitarbeiter / Data Scientist	Computer Scientist
Martina	Mitglied eines innovationsorientierten Teams	Iris	Projektmitarbeiterin / Bewertung und Test des Projekts	Controlling
Heinrich	Head of Digital Finance Workplace	Sherlock GPT	Projektiator & Technischer Leiter	Controlling
Frederik	Head of CFR 2	Sherlock GPT	Business Owner / Sicherstellen, dass die Lösung den Finanzabschluss- und Berichtsprozess unterstützt	Controlling

Abb. 2: Interviewpartner

Die drei gescheiterten Projekte wurden von zentralen Digital- oder IT-Einheiten initiiert, die GenAI primär als technologische Effizienzinnovation verstanden. „Sherlock GPT“ hingegen entstand aus dem Controlling selbst, in enger Kooperation mit Fachanwendern, Datenarchitektur-Teams und einer kleinen internen „KI-Werkstatt“.

Datenerhebung und -auswertung

Insgesamt wurden sechs semistrukturierte Interviews mit unterschiedlichen Schlüsselpersonen geführt (n = 6). Die Interviewpartner repräsentierten verschiedene Rollen im Kontext der untersuchten Initiativen (vgl. Abb. 2), darunter Projektleiter, Controlling-Akteure, Data Scientists und IT-Architekten. Jede der sechs Personen wurde in Bezug auf das jeweilige Projekt bzw. die Projekte interviewt, an denen sie unmittelbar beteiligt war. Der Untersuchungszeitraum erstreckte sich vom 10.07.2024 bis 22.10.2024.

Die Datenauswertung erfolgte nach dem Verfahren der themenzentrierten Kodierung (Miles et al., 2020). Dabei wurden alle Projektdokumente und Interviewtranskripte entlang der in Abschnitt 2 definierten fünf Robustheitsdimensionen (technisch, datenbezogen, prozessual, rollenbezogen, kulturell) analysiert. Innerhalb jeder Dimension wurden anschließend Erfolgs- und Misserfolgsmuster identifiziert und zwischen den Fällen verglichen.

4. Ergebnisse der Fallanalyse

Aufbauend auf der kontrastierenden Fallauswahl werden in den folgenden Abschnitten die zentralen Wirkmechanismen herausgearbeitet. Die Bandbreite der Fälle erlaubt es dabei, sowohl auf die Gründe des Scheiterns als auch auf die Designprinzipien für robuste GenAI-Implementierungen im Controlling einzugehen.

Vergleichende Fallübersicht

Zwischen 2022 und 2024 testete das Unternehmen mehrere Chatbot-Projekte (die auf GenAI basieren) zur Unterstützung der Finanz- und Reportingfunktionen. Alle Projekte verfolgten das Ziel, den Informationszugang im Controlling zu erleichtern und wiederkehrende Recherche- oder Analyseaufgaben durch KI zu unterstützen. Alle vier Projekte durchliefen den gleichen Implementierungsprozess, der aus vier zentralen Phasen bestand:

1. **Initiierung und Scoping:** Definition des Anwendungsfalls sowie Festlegung von Nutzungsrahmen und Risiken.
2. **Technische Konfiguration:** Auswahl und Aufbereitung der Dokumente, Aufbau der Retrievalstruktur und Konfiguration des Modells.
3. **Testphase:** Strukturierte Nutzertests zur Bewertung von Qualität und Reproduzierbarkeit.
4. **Entscheidung und Rollout:** Beendigung oder Skalierung auf Basis technischer, organisatorischer und rechtlicher Bewertung.

Drei dieser Initiativen – Project Athena, Project Delphi und Project Iris – wurden eingestellt; nur Project Sherlock GPT konnte erfolgreich implementiert und skaliert werden. Die Analyse der Fälle ergab, dass weniger die formale Abfolge entscheidend war, sondern die Ausgestaltung und organisatorische Einbettung des jeweiligen Prozessschrittes. Die eingestellten Projekte durchliefen die Phasen explorativ und blieben im Prototypenstadium, insbesondere aufgrund begrenzter Datenaufbereitung und fehlender systematischer Retrieval-Optimierung. Project Sherlock GPT hingegen zeichnete sich durch strukturierte Datenaufbereitung, iterative Testzyklen, gezielte Retrieval-Optimierung und einen gestuften „Soft Go-Live“ mit klarer Governance aus, wodurch eine stabile Integration in bestehende Prozesse gelang.

Muster des Scheiterns

Die Analyse der gescheiterten Projekte mit Hilfe des konzeptionellen Bezugsrahmens des Beitrags offenbart eindeutige Muster des Scheiterns. Die analytische Zuordnung zu den fünf Robustheitsdimensionen verdeutlicht, dass Defizite auf mehreren Ebenen gleichzeitig auftraten und sich gegenseitig verstärkten:

1. **Technische Robustheit:** Die Systeme konnten Tabellen und strukturierte Daten nicht zuverlässig verarbeiten; identische Fragen führten zu unterschiedlichen Antworten. Diese Instabilität verweist auf eine mangelnde technische Robustheit im Sinne fehlender Output-Konsistenz und mangelnder Verlässlichkeit unter wiederholten Bedingungen. Gerade im Controlling-Kontext, in dem Reproduzierbarkeit eine zentrale Anforderung darstellt, erwies sich diese technische Unschärfe als kritischer Schwachpunkt.
2. **Datenrobustheit:** Heterogene, teils unstrukturierte Quellen und fehlende Qualitätsmechanismen führten zu unzuverlässigen Ergebnissen und sinkendem Vertrauen der Nutzer. Im Sinne der Datenrobustheit fehlte eine GenAI-fähige, harmonisierte und strukturierte Datenarchitektur. Die empirischen Befunde zeigen, dass GenAI ohne vorgelagerte Standardisierung und systematische Datenaufbereitung in einem finanznahen Umfeld strukturell überfordert ist. Die Defizite in technischer und datenbezogener Robustheit bildeten die primäre Fragilitätsebene der Projekte.
3. **Prozessrobustheit:** Die Anwendungen blieben experimentelle Prototypen ohne Anbindung an bestehende Controlling-Prozesse oder BI-Systeme. Mangels Integration in etablierte Workflows entstanden keine stabilen Nutzungsmuster. Die Projekte blieben außerhalb der formalen Entscheidungs- und Berichtsketten verortet, wodurch sie keinen institutionellen Anker entwickelten. Dies verweist auf eine fehlende Prozessrobustheit. Die GenAI-Anwendungen wurden nicht als Bestandteil der Wertschöpfungslogik des Controllings verankert, sondern als isolierte Experimente behandelt.
4. **Rollenrobustheit:** Verantwortlichkeiten und Feedback-Schleifen waren kaum definiert; Controlling-Mitarbeitende wurden weder geschult noch eingebunden. Es fehlte eine klare Governance-Struktur mit definierter Ownership, Eskalationslogik und Lernzyklen. Die probabilistische Natur generativer KI wurde nicht durch entsprechende Rollen- und Kompetenzzuschreibungen flankiert. Ohne Rollenrobustheit konnten weder Vertrauen noch nachhaltige Weiterentwicklung entstehen.
5. **Kulturelle Robustheit:** Die Projekte wurden neben dem Tagesgeschäft von kleinen Teams mit begrenztem Budget betreut. Fehlende Priorität und Unsicherheit über Risiken behinderten Ak-

zeptanz und Lernfortschritt. Die Initiativen waren organisatorisch nicht als strategische Transformationsprojekte gerahmt, sondern als explorative Nebenaktivitäten positioniert. Diese kulturelle Unterpriorisierung verhinderte die Entwicklung einer lernorientierten Experimentierkultur. Die Erwartungshaltung, insbesondere in Bezug auf hundertprozentige Genauigkeit, stand in einem Spannungsverhältnis zur technologischen Realität probabilistischer Systeme.

Insgesamt zeigt die dimensionsübergreifende Betrachtung, dass insbesondere Defizite in technischer und datenbezogener Robustheit die initiale Fragilitätsebene bildeten. Diese primären Schwächen verhinderten eine stabile Prozessintegration und unterminierten die Entwicklung klarer Rollen und Governance-Strukturen. Kulturelle Defizite wirkten dabei weniger als Auslöser, sondern vielmehr als Verstärker bestehender struktureller Schwächen.

Vergleichbare Initiativen sollten daher insbesondere in der Frühphase technische und datenbezogene Robustheit priorisieren, um Folgeprobleme in anderen Dimensionen zu vermeiden. Die Fehlschläge verdeutlichen, dass der Einsatz von GenAI ohne strukturelle und kulturelle Vorbereitung zu Fragilität statt zu Leistungsfähigkeit führt – eine empirische Bestätigung des im theoretischen Rahmen entwickelten Robustheitskonzepts.

Der Erfolgsfall „Sherlock GPT“

Im Gegensatz zu den gescheiterten Projekten wurde Sherlock GPT nicht als Experiment, sondern als strategisch gesteuertes IT-Produkt konzipiert, mit einem „Soft Go-Live“ und Erfolgskriterien wie Nutzung, Wiederkehrtrate und wahrgenommener Antwortqualität statt rein monetärer Kennzahlen.

Das System ist in das interne Portal des Unternehmens eingebettet und zusätzlich direkt in Microsoft Teams verfügbar („wie ein Kollege“), mit perspektivischen Erweiterungen. Es unterstützt Mitarbeitende mit GenAI-basierten Antworten auf Fragen zu Finanzrichtlinien, Reportingstandards und Abschlussprozessen. Im Vergleich zu den gescheiterten Projekten lassen sich vor allem fünf Erfolgsfaktoren herausarbeiten, die Sherlock GPT von den vorherigen Initiativen unterscheiden:

1. **Klarer, risikoarmer Use Case:** Der Fokus lag auf internem Wissenszugang statt auf externen oder strategischen Inhalten. Die regulatorische Einordnung ist „low risk“; zusätzlich greifen unternehmensinterne Guardrails, die durch interne Hackathons und Team-Tests fortlaufend weiterentwickelt werden.
2. **Technische Robustheit durch RAG-Architektur:** Die Kombination aus Large Language Model und Dokumenten-Retrieval (Retrieval-Augmented Generation, RAG) sorgte für nachvollziehbare, quellengestützte Antworten. Die Retrieval-Qualität wird ständig überwacht und verbessert.

Vergleichbare Initiativen sollten daher insbesondere in der Frühphase technische und datenbezogene Robustheit priorisieren, um Folgeprobleme in anderen Dimensionen zu vermeiden.

Implikationen für die Praxis

- Klein starten: enger, risikoarmer Use Case; dann iterativ erweitern – mit klaren Test-, Qualitäts- und Abbruchkriterien
- Data Discipline: kuratierte, geprüfte Quellen und klare Daten-/Modellgrenzen; Qualität vor Quantität
- RAG mit Quellen einsetzen, Retrieval laufend optimieren; so werden Antworten nachvollziehbar und halluzinationsärmer
- Governance/Human-in-the-loop aufsetzen: Owner, Reviewer, Freigaben, Feedback-Loops, Training – Verantwortung bleibt beim Menschen

3. **Daten- und Prozessrobustheit:** Alle Dokumente stammen aus geprüften, strukturierten internen Quellen. Das Korpus ist kuratiert (z. B. Kontenrahmen, Reporting-Guidelines, Abschlusskalendar/-prozesse) und liefert satzweise referenzierte Antworten; die Anwendung ist in bestehende IT- und Reportingprozesse eingebettet.
4. **Rollenrobustheit mit „Human-in-the-Loop“:** Nutzerfeedback fließt systematisch in Modell- und Prozessverbesserungen ein. Vorgelagert werden mehrwöchige Benutzerakzeptanztest-Phasen und regelmäßige Community-/Monatsrunden etabliert; auch nicht bewertete Anfragen werden ausgewertet, um „stille Zustimmung“ zu erkennen.
5. **Kulturelle Verankerung und Governance:** Das Projekt war in der IT-Abteilung angesiedelt, mit klarer Verantwortung, Budget und methodischer Disziplin. Die Controlling-Teams nahmen die Anwendung als Unterstützung, nicht als Bedrohung, wahr.

Sherlock GPT zeigt, dass GenAI im Controlling dann Wirkung entfaltet, wenn sie fokussiert, kontrolliert und lernorientiert implementiert wird. Die erwartete Wirkung ist Zeitgewinn für komplexe Aufgaben und Qualitätssteigerung durch das Zusammenführen verteilter Quellen in einem Schritt.

Sherlock GPT wurde von der IT-Abteilung des Fallunternehmens als zu gestaltendes sozio-technisches System verstanden, dessen Nutzen durch verlässliche Einbettung entstehen würde. Statt auf revolutionäre Substitution („Autonomous Forecasting“) zielte der gewählte Ansatz von vornherein auf eine klar abgegrenzte, robuste Use-Case-Logik. Sherlock GPT sollte den Zugang zu internem Wissen (Kontenrahmen/Chart of Accounts, Reporting-Guidelines, Closing-Informationen) über Natural-Language-Fragen ermöglichen, ohne dabei jedoch Entscheidungen zu automatisieren. Diese bewusste Begrenzung des Projektfokus reduzierte das Risiko und die Komplexität. Damit wurde die zentrale Robustheitsanforderung, auch bei Unsicherheit (z. B. Halluzinationen) stabil zu performen, in das Zentrum des Designs gestellt. In Bezug auf die fünf Dimensionen zeigt der Fall ausgeprägte technische Robustheit durch eine starke Systemarchitektur (sichere Hosting-Umgebung innerhalb der Unternehmensinfrastruktur, Integration in das interne Por-

tal, hohe Zugänglichkeit über Regionen hinweg) sowie durch ein LLM in Kombination mit einem Retrieval-Design, das die Performance und Zuverlässigkeit gegenüber früheren, instabilen Chatbot-Initiativen erhöht. Die Datenrobustheit wurde explizit über „saubere Daten- und Modellgrenzen“ operationalisiert. Antworten wurden mittels strukturierter Dokumenten-Retrieval-Prozesse (Chunking, semantische Suche) in verlässlichen internen Dokumenten verankert, wodurch Konsistenz, Nachvollziehbarkeit und Kontextpassung der Informationen gestärkt und Halluzinationen reduziert wurden. Prozessrobustheit wurde insofern sichergestellt, dass Sherlock GPT nicht isoliert als Tool betrieben wurde, sondern direkt in eine bestehende Organisationsroutine eingebunden wurde. Sherlock GPT erlaubte den alltäglichen Zugriff auf Controlling-relevante Richtlinien und Closing-Wissen über das unternehmensinterne System und konnte damit GenAI-Outputs wiederholbar in Arbeitsabläufe überführen (Informationssuche/Interpretation), anstatt als „One-off“-Experiment zu verbleiben. Bei der Entwicklung von Sherlock GPT wurde zudem auf eine Rollenrobustheit geachtet, indem eine klare Aufgabenverteilung von Anfang an sichergestellt wurde. Die IT fungierte als „Besitzer“ des Systems (Aufbau, Betrieb, Weiterentwicklung), während Nutzende primär als Anwendende und Feedback-Gebende auftraten. Zugleich blieb die Verantwortlichkeit für Entscheidungen implizit beim Menschen, da die Anwendung Informationsunterstützung statt Entscheidungsautomatisierung darstellte (Human-in-the-Loop durch Design der Entscheidungsgrenzen). Schließlich wird Kulturrobustheit über zwei Mechanismen sichtbar: eine risikobewusste Lernorientierung („careful scope“, „low risk“, Investitionen in Training/Qualität) und eine Governance des Lernens durch Feedback-Loops, die kontrollierte Verbesserung im Live-Betrieb ermöglichen.

Insgesamt illustriert Sherlock GPT, wie robuste GenAI-Implementierungen im Controlling-Kontext vor allem über Integrationsarbeit (Scope, Daten-/Modellgrenzen, Architektur, Prozessanschluss, feedbackbasiertes Lernen) gelingen und warum frühere Initiativen an genau der mangelnden Berücksichtigung (z. B. durch instabile Konstruktion, geringe Genauigkeit, Halluzinationen) der Robustheitsdimensionen scheiterten.

5. Diskussion: Designprinzipien robuster GenAI-Implementierungen im Controlling

Die kontrastierende Fallanalyse zeigt deutlich, dass sich erfolgreiche von gescheiterten KI-Initiativen weniger durch technologische Exzellenz als durch robustes Design unterscheiden. Robuste GenAI-Implementierungen zeichnen sich dadurch aus, dass sie technologische Möglichkeiten mit klaren orga-

nisatorischen Grenzen, stabilen Prozessen und einer lernorientierten Kultur verbinden. Während die nicht erfolgreichen Projekte auf einen schnellen Einsatz von GenAI ohne ausreichende Einbettung in Datenarchitektur, Governance und Nutzerstrukturen setzten, verfolgte das erfolgreiche Projekt einen umgekehrten Ansatz: schrittweise Entwicklung, klarer Fokus und kontrolliertes Lernen. Daraus lassen sich fünf übergreifende Designprinzipien für den Einsatz von GenAI im Controlling ableiten, die im Folgenden diskutiert werden.

Strategic Embedding – GenAI als Teil der Steuerungsarchitektur

GenAI-Anwendungen müssen in die bestehende Logik der Unternehmenssteuerung integriert werden. Erfolgreiche Implementierungen zeichnen sich dadurch aus, dass sie nicht „neben“ dem Controlling laufen, sondern in die Kernprozesse der Planung, Analyse und Berichterstattung eingebettet sind. Dies ermöglicht eine Rückkopplung zwischen GenAI-gestützten Erkenntnissen und Managemententscheidungen. Dementsprechend sollten GenAI-Projekte stets in der Verantwortung von Funktionen liegen, die selbst Teil der Steuerungsarchitektur sind (z. B. Controlling, Accounting oder Data Governance), und nicht isoliert in Innovationslaboren betrieben werden.

Data Discipline – Qualität vor Quantität

In allen gescheiterten Projekten zeigte sich, dass inkonsistente, unstrukturierte oder unzureichend gepflegte Datenquellen zu Halluzinationen und Fehlinterpretationen führten. Der erfolgreiche Fall hingegen basierte auf kuratierten, geprüften internen Dokumenten, deren Strukturierung und Aktualisierung klar geregelt waren. Vor diesem Hintergrund wird deutlich, dass Organisationen verbindliche Standards für Datenqualität, -pflege und -zugriff benötigen. „Data Discipline“ bedeutet, dass weniger, aber zuverlässige Daten besser sind als große, unkontrollierte Datenmengen. Damit ist es Unternehmen möglich, Datenrobustheit sicherstellen zu können.

Iterative Scope Control – klein starten, gezielt lernen

Ein zentrales Erfolgsprinzip robuster GenAI-Implementierungen ist die bewusste Begrenzung des Anwendungsumfangs. Statt umfassende, schwer steuerbare „GenAI-for-everything“-Ansätze zu verfolgen, empfiehlt sich ein inkrementelles Vorgehen: kleine, klar definierte Use Cases, die in kurzen Lernzyklen getestet und verbessert werden. Im erfolgreichen Fall wurde die GenAI zunächst auf ein eng umrissenes internes Wissensgebiet angewendet. Die Stabilität des Systems ermöglichte es anschließend, den Funktionsumfang kontrolliert auszuweiten. Es empfiehlt sich, GenAI-Implementierungen im Controlling wie Lernsysteme zu behan-

deln, d. h. mit iterativer Erweiterung, klaren Evaluationskriterien und der Bereitschaft, nicht funktionierende Features konsequent zu verwerfen.

Human-in-the-Loop – Controlling-Mitarbeitende als Kuratierende und Übersetzende

Der Mensch bleibt zentrale Instanz im Zusammenspiel mit GenAI. Robustheit entsteht, wenn Menschen die GenAI beaufsichtigen, korrigieren und interpretieren, anstatt Entscheidungen vollständig zu delegieren. In erfolgreichen Implementierungen fungiert die Controlling-Funktion als „Kuratierende“ von GenAI-Antworten, d. h. sie prüft, bewertet und übersetzt die Ergebnisse in betriebswirtschaftliche Bedeutung mit Hilfe ihres unternehmensspezifischen Kontextverständnisses. Deshalb sollten Unternehmen gezielt in AI-Literacy und Controlling-Kompetenzen investieren, um die Fähigkeit zur kritischen Nutzung von GenAI zu fördern. Governance-Regeln („Review before use“) sichern zusätzlich, dass Entscheidungen nachvollziehbar bleiben. Indem sie festlegen, welche Quellen und Versionen zulässig sind, wer Ergebnisse prüft und freigibt und wie Prompting, Output sowie Abweichungen dokumentiert werden, entsteht ein prüfbarer Entscheidungs- und Audit-Pfad. Dadurch werden GenAI-Beiträge nicht als „Black Box“ genutzt, sondern als kontrollierter Input, dessen Zustandekommen und Verantwortlichkeit transparent gemacht werden.

Cultural Anchoring – Lern- und Governance-Kultur als Erfolgsbedingung

Technische und prozessuale Robustheit sind wirkungslos, wenn sie nicht in eine entsprechende Kultur eingebettet sind. Das erfolgreiche Projekt war von Beginn an strategisch verankert, mit klarer Governance, Budget und Ownership. Die Projekte mit experimentellem Charakter hingegen litten unter Unsicherheit, fehlender Legitimation und mangelndem Vertrauen. Insofern sollten Unternehmen den GenAI-Einsatz im Controlling als Teil ihrer Kulturentwicklung begreifen. Eine „Robustheitskultur“ fördert Lernen, Verantwortlichkeit und den offenen Umgang mit Fehlern, anstatt Innovation allein an kurzfristige Effizienzgewinne zu knüpfen.

6. Fazit

Die gegenwärtigen Erfolgsversprechen rund um die Nutzung von GenAI im Controlling erzeugen hohe Erwartungen, doch die Realität zeigt, dass nicht die Technologie selbst, sondern ihre robuste Einbettung über den Erfolg entscheidet. Die untersuchten Fallbeispiele belegen, dass viele GenAI-Initiativen an mangelnder Datenqualität, unklaren Prozessen und fehlender Governance scheitern – nicht an den Grenzen der Modelle. Nur dort, wo Technik, Organisation und Kultur in Einklang gebracht werden, entsteht nachhaltiger Nutzen. Dementsprechend

Erfolgreiche Implementierungen zeichnen sich dadurch aus, dass sie nicht „neben“ dem Controlling laufen, sondern in die Kernprozesse der Planung, Analyse und Berichterstattung eingebettet sind.

kann aktuell konstatiert werden, dass GenAI das Controlling (vorerst) nicht revolutionieren, sondern weiterentwickeln wird, sofern sie als Bestandteil eines disziplinierten Lern- und Governance-Systems gestaltet wird. CFOs und Controlling-Abteilungen sollten GenAI als Baustein einer robusten Steuerungsarchitektur begreifen und mit klaren Prozessen, Verantwortlichkeiten und Feedbackschleifen in ihre bestehenden Prozesse integrieren.

Damit scheint sich auch eine der zentralen Aufgaben des Controllings weiterzuentwickeln – allerdings nicht in jedem Unternehmen und nicht in jeder Ausgestaltung gleichermaßen. Während ein Teil der Literatur das Controlling als Gestaltende und Kuratierende von GenAI-Systemen beschreibt, zeigen andere Studien (Wixom et al., 2024), dass GenAI-Initiativen an der Schnittstelle zum Controlling zunehmend funktionsübergreifend getragen werden (z. B. durch IT, Data/Analytics, Finance Transformation oder zentrale AI-Teams). Für die Analyse wird damit entscheidend, welche Controlling-Aufgaben (z. B. Daten-/Definitionsicherheit, Modell- und Ergebnisvalidierung, Übersetzung in betriebswirtschaftliche Bedeutung, Prozessintegration und Governance) tatsächlich im Controlling verbleiben und welche in andere Funktionen wandern. Die Erfolge robuster GenAI-Implementierungen treten insbesondere dort ein, wo Controlling nicht als „Automatisierungsziel“, sondern als Qualitäts- und Verantwortungsinstantz ausgestaltet ist: KI liefert Beiträge und Entwürfe, während Controlling – ggf. in enger Arbeitsteilung mit IT/Analytics – die Plausibilisierung, Kontextualisierung und Freigabe in die Steuerungsprozesse hinein sicherstellt. Controlling der Zukunft ist damit weniger Revolution als robuste Evolution, deren Wirkung von der konkreten Rollen- und Aufgabenarchitektur an der Schnittstelle zwischen Controlling, IT und Data/AI abhängt.

Der vorliegende Beitrag baut auf einer Mehrfallstudie mit nur einem erfolgreichen Case innerhalb eines Unternehmens auf. Die damit einhergehenden Limitationen stellen gleichzeitig den Startpunkt für weitere Forschung dar. Zukünftige Studien sollten bspw. untersuchen, welche Erfolgsfaktoren außerhalb der Softwarebranche für die GenAI-Implementierung im Controlling vorliegen, welche Relevanz Datenrobustheit in weniger volatilen Märkten zugeschrieben wird und ob sich die Erfolgsfaktoren über die Zeit hin verändern.

Literatur

- Challapally, A., Pease, C., Raskar, R., Chari, P., The GenAI Divide. State of AI in Business 2025, 2025, https://www.artificialintelligence-news.com/wp-content/uploads/2025/08/ai_report_2025.pdf, Stand: 08.03.2026.
- Demyttenaere, M., Roos, A., Sheth, H., Rodt, M., Harris, M., Khodabandeh, S., Fehling, R., Marti-

nes, D., Grabowski, J., Generative AI in the Finance Function of the Future, 2023, <https://www.bcg.com/publications/2023/generative-ai-in-finance-and-accounting>, Stand: 08.03.2026.

- Dumas, M., Fournier, F., Limonad, L., Marrella, A., Montali, M., Rehse, J. R., Weber, I., AI-augmented business process management systems: a research manifesto, in: *ACM Transactions on Management Information Systems*, 14. Jg. (2023), H. 1, S. 1–19.
- Englisch, D., Hasebrook, J. P., Holländer, D., KI im Controlling von Finanzdienstleistern, in: *Controlling*, 38. Jg. (2026), H. 1, S. 46–51.
- Franke, F., Hiebl, M. R., Big data and decision quality: the role of management accountants' data analytics skills, in: *International Journal of Accounting & Information Management*, 31. Jg. (2023), H. 1, S. 93–127.
- Grønsund, T., Aanestad, M., Augmenting the algorithm: Emerging human-in-the-loop work configurations, in: *The Journal of Strategic Information Systems*, 29. Jg. (2020), H. 2, Art. 101614.
- Li, J., Zhu, F., Hua, P., Overcoming the Organizational Barriers to AI Adoption, 2025, <https://hbr.org/2025/11/overcoming-the-organizational-barriers-to-ai-adoption>, Stand: 08.03.2026.
- Lombardi, R., Cano-Rubio, M., Schimperna, F., Trequattrini, R., The impact of smart technologies on the management and strategic control: a structured literature review, in: *Management Control, Special issue 1* (2021), S. 11–30.
- Miles, M., Huberman, M., Saldeña, J., *Qualitative Data Analysis*, 3. Aufl., New York 2020.
- Papiorek, K. L., Hiebl, M. R., Information systems quality in management accounting and management control effectiveness, in: *Journal of Accounting & Organizational Change*, 20. Jg. (2024), H. 3, S. 433–458.
- Patel, B., Vasa, J., Patel, R., Parmar, M., Maheriya, K., Enhancing Software Robustness: A Comprehensive Guide to Quality Attributes and Optimization Tools, in: Senjyu, T./So-In, C./Joshi, A. (Hrsg.), *Smart Trends in Computing and Communications*. SmartCom 2025, Singapore 2025.
- Trist, E., Emery, F., Sociotechnical systems theory, in: *Organizational Behavior*, 2. Jg. (2015), S. 169–194.
- Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., Aggarwal, S., Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality, in: *Journal of the Knowledge Economy*, 15. Jg. (2024), H. 1, S. 1159–1178.
- Wang, R. Y., Strong, D. M., Beyond accuracy: What data quality means to data consumers, in: *Journal of Management Information Systems*, 12. Jg. (1996), H. 4, S. 5–33.
- Wixom, B. H., Someh, I. A., Beath, C. M., Achieving AI at Scale: Cemex's Learning Journey,

2024, https://cisr.mit.edu/publication/2024_0701_CemexScalingAI_WixomSomehBeath, Stand: 08.03.2026.

- Yin, R., Case Study Research and Design, 5. Aufl., New York 2014.

Literaturtipps auf inlibra.com:

- Generative Künstliche Intelligenz im Industriel-
len Service, Jean Paul Potthoff, Lukas Budde,
Thomas Friedli, 2024/5, S. 24–31
- Digitale Transformation des Controlling/Finan-
ce-Bereichs, Goran Sejdić, 2020/4, S. 61–67
- Künstliche Intelligenz in der Finanztransfor-
mation, Leonhard Brinster, Daniele Bühler, 2024/6,
S. 73–75

Stichwörter

Controllingimplementierung # Daten- und Sys-
temqualität # Designprinzipien # GenAI # Hu-
man-in-the-loop # Robustheit

Keywords

Controlling implementation # Data and system
quality # Design principles # GenAI # Human-
in-the-loop # Robustness

Summary

Drawing on a multi-case study, this paper shows why GenAI projects in management accounting often fail: a lack of robustness in terms of technology, data, processes, roles, and culture. A success story demonstrates that a clear scope, curated data, RAG, governance, and a human-in-the-loop approach make GenAI both effective and manageable.