International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Strategic decision-making framework for evaluating and selecting GenAI use cases

Suhas Rumalla [1] and Jafar Salim Mujawar [2,*]

[1] *Senior Director of Product Management, Walmart, United States.*
[2] *Senior Product Manager, Walmart, India.*

International Journal of Science and Research Archive, 2025, 16(01), 170-177

Publication history: Received on 25 May 2025; revised on 01 July 2025; accepted on 04 July 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.16.1.2030

Abstract

Generative-AI (GenAI) rose from boutique research to mass hype in less than three years, but most firms still fail to monetize proofs-of-concept into production worth. Recent surveys show that 69 % of enterprise GenAI initiatives stall before being operationalized while 46 % of PoC initiatives are abandoned outright and Gartner forecasts a 30% abandonment rate by the end of 2025. We contend that a primary cause of this attrition is continued application of classical product-management heuristics—tuned to deterministic feature work—when GenAI problems are probabilistic, socio-technical, and risk-weighted. Drawing on observations from past twelve enterprise GenAI launches (2023-2025), and 25 semi-structured interviews with product managers at multiple firms, we (i) clarify where classical discovery, sizing, and prioritization practices go wrong; and (ii) distill a three-part decision playbook comprising a Three-Gate Decision Funnel, Six-Point Opportunity Scorecard, and Value-Feasibility Prioritization Matrix. Early adopters report cycle-time compression by up to 40%, with an average of 33% and doubled conversion rates from PoC to production after the playbook's use. The paper positions these artefacts as a practitioner-oriented handbook and research-based contribution to the nascent field of GenAI product strategy.

Keywords: Generative AI; Product Management; Value Prioritization; AI Product Strategy; GenAI Adoption Framework; Opportunity Selection; Technology Governance

1. Introduction

Enterprise enthusiasm for Generative-AI (GenAI) now runs well ahead of reliable execution: board directives, internal hackathons and vendor deals produce an expanding backlog of ideas, yet product and business teams still lack a repeatable way to decide which opportunities merit real investment. In practice we observe managers importing familiar heuristics—story-point sizing, fixed ROI hurdles, feature-parity road-mapping—into a domain whose outputs are probabilistic and risk-laden; the result is a pattern of promising pilots that drift, stall or are quietly sunset. To understand this mismatch, our study first documents how teams currently select GenAI initiatives and traces the outcomes of those projects across twelve launches and 25 practitioner interviews conducted between January 2025 and June 2025. We then distil the lessons—where methods failed, where they succeeded and why—into a field-tested handbook designed to align engineering, risk and commercial stakeholders on a single, evidence-based view of value and feasibility.

2. Literature review

2.1. Diffusion of Generative-AI Technologies

Weekly GenAI usage among knowledge workers jumped from 37 % in 2023 to 72 % in 2024, according to IBM's global "AI Readiness" survey [5]. Retail mirrors that curve: 72 % of chains plan to "fundamentally reinvent" operations with

* Corresponding author: Jafar Salim Mujawar.

GenAI by 2026. Financial services lead on deployment scope—two-thirds of digital executives already run models in production, albeit mainly inside-the-firewall assistants rather than customer-facing features. McKinsey's 2025 State-of-AI study finds that 71 % of organizations now apply GenAI in at least one core function, up from 65% the previous year [4]. This diffusion is driven by cloud APIs, foundation-model commoditization, and a fast-growing vendor ecosystem that lowers technical barriers for mid-market adopters.

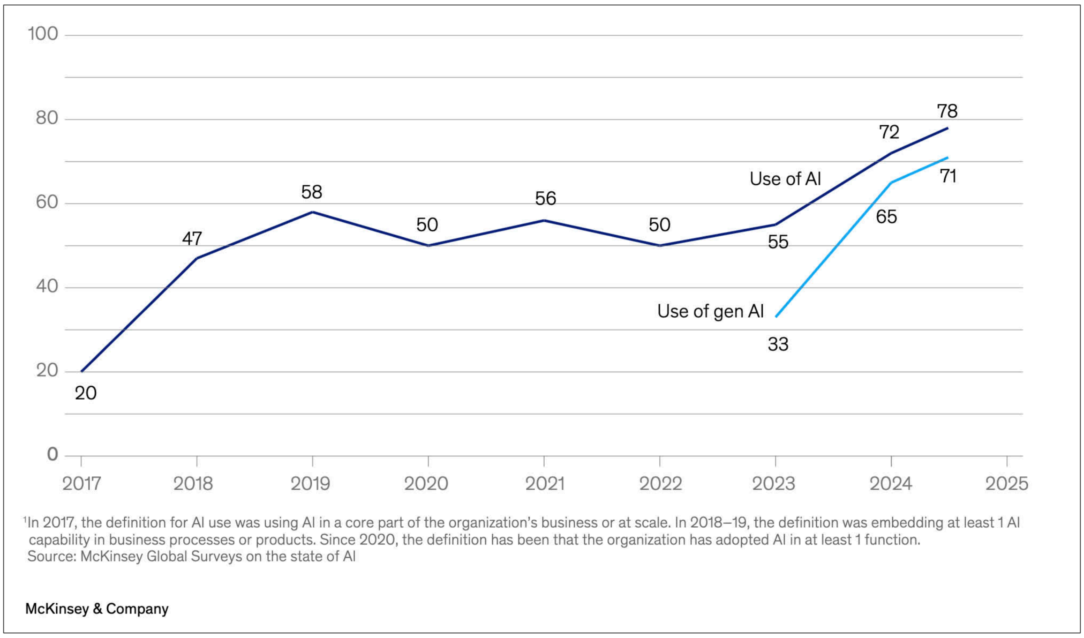


Figure 1 Use of AI

2.2. Opportunity-Selection in Emerging Tech

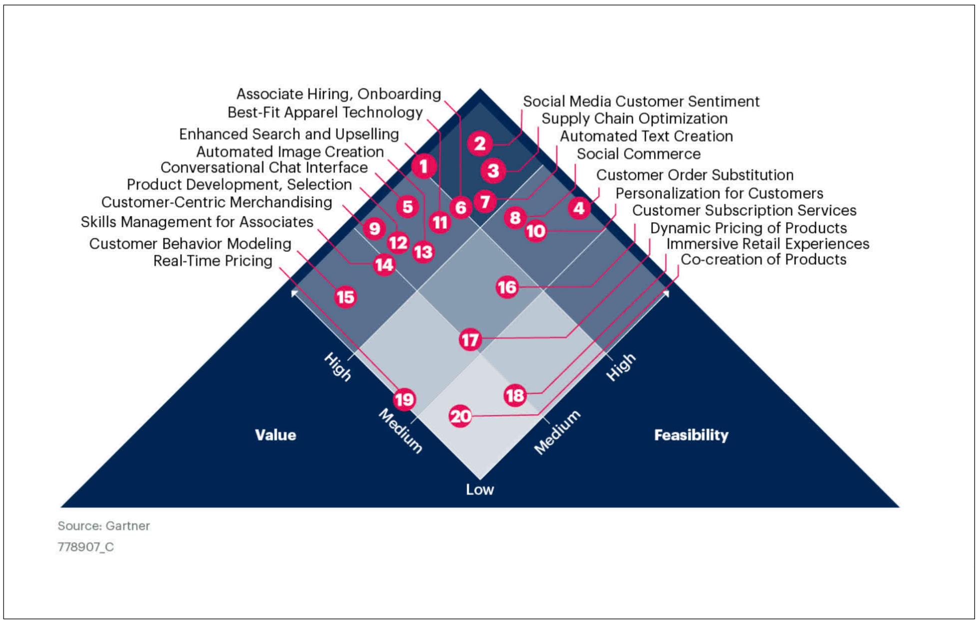


Figure 2 GenAI Use-Case Prism

Early innovation literature champions stage-gate governance and impact-feasibility matrices to balance risk with return; Cooper's 2024 update shows firms embedding AI checkpoints into each stage-gate milestone [7]. GenAI-specific guidance now layers risk axes—truthfulness, intellectual-property leakage, toxicity—on top of traditional demand metrics; HBR's two-by-two "demand vs risk" grid exemplifies this shift [15]. Industry reports extend these models: the "AI Use-Case Prism" plots hundreds of scenarios and has been adapted into sector-specific decision matrices for finance and retail [16]. Practitioners also emphasize qualitative alignment: frameworks advise anchoring ideas in explicit pain points and governance capacity before advancing to build-fund requests.

2.3. Gap Analysis & Research Need

Despite the profusion of tools, two primary shortcomings persist. First, prevailing matrices still assume deterministic outputs and therefore under-weight the cost of guardrails and human-in-loop validation that GenAI mandates. Second, empirical evidence of what actually happens after opportunity selection remains sparse; IDC (a market research firm) finds 88 % of AI pilots never reach production, citing underfunded governance and ambiguous ownership as root causes [9]. Industry analyses corroborate: over half of CEOs list GenAI as an efficiency priority, yet 46 % admit that identifying viable use cases—coupled with data-quality concerns—is their primary barrier to impact [4]. No current framework fuses probabilistic risk, data maturity, and cross-functional alignment into a single, practitioner-friendly gate. Addressing that void is the central contribution of the handbook advanced in this paper.

3. Research methodology

3.1. Research Questions

- **RQ1 Selection Practice –** Which explicit and tacit criteria do product and business leaders apply today when deciding whether a GenAI use case advances beyond ideation?
- **RQ2 Outcome Trace –** What observable results—progression, delay, abandonment—follow those selection decisions across the project life-cycle?
- **RQ3 Prescriptive Design –** How can the observed failure points inform a repeatable decision framework that improves throughput to production and satisfies governance constraints?

3.2. Framework-Derivation Approach

We employed a three-phase sequence

- **Exploratory Capture.** Drawing on participant-observation inside twelve enterprise launches (from different companies and markets) generated hours of artifacts — such as backlog reviews, gate reviews, and war-room sessions. The action-research model mirrors prior AI-deployment case work in manufacturing and services, enhancing internal validity.
- **Thematic Synthesis.** Twenty-five semi-structured interviews with product managers and risk leads were coded via grounded-theory techniques; axial coding concentrated on selection heuristics, governance blockers, and remediation tactics. This step internalized lessons from stage-gate refinements that now embed AI checkpoints in new-product processes.
- **Iterative Validation.** The draft playbook was continuously applied and refined during real-time opportunity-selection efforts inside the participating organizations. Each feedback cycle sharpened scoring weights and governance gates until stakeholders agreed the framework was both conceptually robust and operationally practical.

3.3. Data Sources and Limitations

Primary data comprised (a) direct observation logs, and (b) interview transcripts. Secondary data included McKinsey macro-surveys on AI pilot conversion, IBM's AI-Readiness pulse, and HBR analyses of risk-adjusted opportunity matrices, etc. Triangulation across these heterogenous sources improved construct validity. Nonetheless, three constraints remain: (1) sector skew toward finance and retail may limit generalizability; (2) the January–June 2025 window omits longer-run adoption dynamics; and (3) voluntary participation introduces self-selection bias, a common limitation in organizational AI research.

4. Integrated opportunity selection framework

Before we turn to the three-step opportunity funnel, it helps to anchor our thinking in the Product Evolution & Maturity pyramid. Products mature vertically and progress horizontally, from well-instrumented source applications (L1) through governed data (L2) and predictive analytics (L3) before successfully reaching GenAI autonomy (L4). Each layer builds on the reliability of the one below it, so ambitions that skip a rung invariably stall. Keeping this hierarchy in mind clarifies why some ideas are prematurely bold and where foundational investment must precede execution.

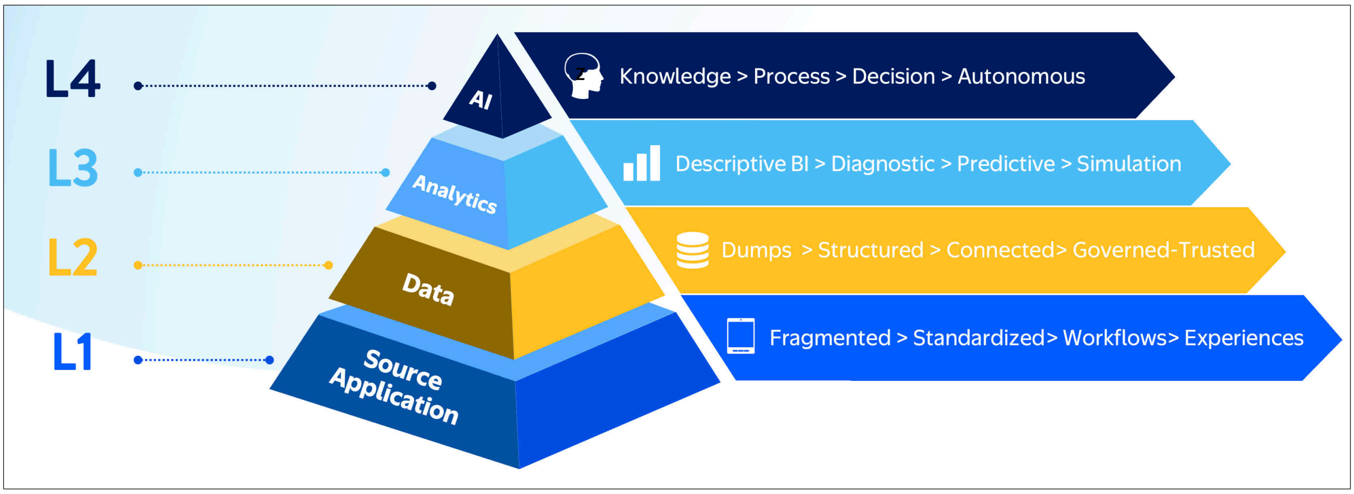
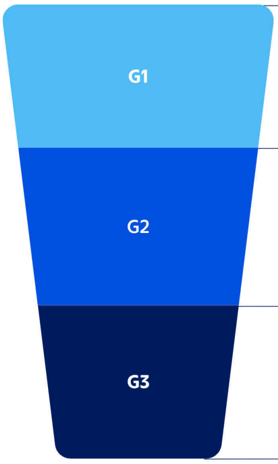


Figure 3 Product Evolution & Maturity pyramid

The framework operates as a three-step funnel-and-score sequence. Each stage answers a distinct question—Should we even consider this idea? → Can we deliver it under current constraints? → Where does it sit in the execution queue?—and passes only qualified candidates to the next gate.

4.1. Step 1: The Three-Gate Decision Funnel – "Select or Eliminate"

Step 1 runs each proposal through a three-gate screen. Structural Fit checks whether the task is inherently probabilistic; deterministic chores revert to classic code or RPA. If it passes, we test Outcome Variability (does judgment truly vary?) and Tolerance & Controls—only ideas whose stochastic answers can be cost-effectively guarded advance to feasibility scoring.



Gate	Focal Question	Exit Route if "No"	Illustrative Notes
Structural Fit	Is the task inherently probabilistic or ambiguous?	Classic code / rule engine / RPA	Deterministic rules (e.g., FX tables, GL posting) gain no lift from GenAI.
Outcome Variability	Does the "right" answer vary or require judgment?	Traditional ML / analytics	Fraud scoring and cash flow forecasting vary; regression may suffice if variance is low.
Tolerance & Controls	Can we afford stochastic answers and enforce guardrails?	Hybrid AI + human guardrails, or defer	High stakes domains (SOX, PCI) need human in loop review; if review cost erodes ROI, shelve the idea.

Figure 4 The Three-Gate Decision Funnel

Ideas that clear all three questions advance to feasibility scoring; others are discarded or routed to more suitable technologies.

4.2. Step 2: Feasibility Scorecard – "Can We Execute?" (Max = 65 pts)

Feasibility Scorecard tests, "Can we execute?" by rating readiness across data, algorithms, processes, and know-how. It quantifies delivery risk across four capability pillars plus time-to-proof-of-value (PoV). Each statement is rated on a Likert scale from 0 (strongly disagree) to 5 (strongly agree). The numeric readout converts gut feel into evidence, pinpointing which ideas to advance, re-scope, or shelve.

00-25: High risk / heavy lift – unlikely to fund now
25-40: Medium risk – fund only if strategic
40-55: Manageable – prototype with guardrails
55-65: Ready – fast-track

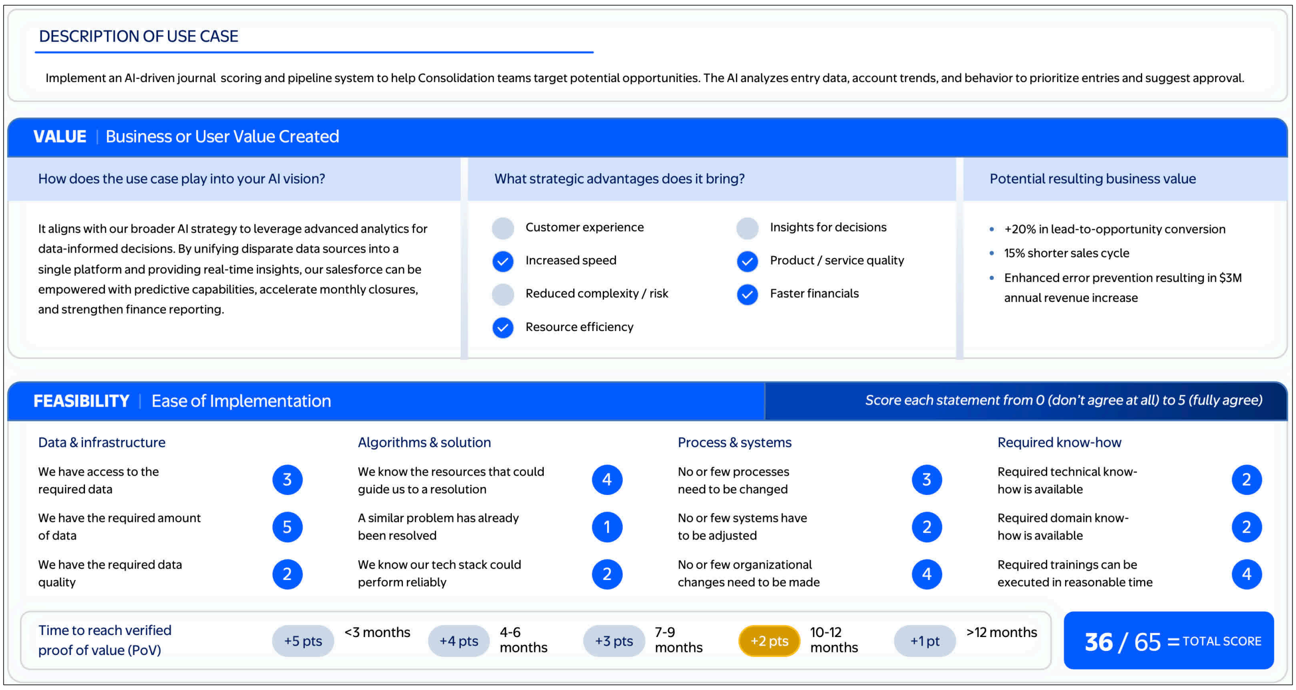


Figure 5 The scorecard quantifies delivery risk to determine if a project should be advanced, re-scoped, or shelved. In the example shown, the total score of 36/65 places the use case into the "Medium risk" category, meaning it should only be funded if it is of high strategic importance

4.3. Step 3: Six-Dimension Opportunity Scorecard – "Prioritize the Executable"

Dimension	0	1	2	3
Business lift	Cosmetic	Nice-to-have	Clear OPEX save	Direct revenue risk $
Nature of solution	Deterministic	Minor nuance	Multi-factor	Open-ended
Human effort	<1 min/task	1-5 min	5-15 min	>15 min
Volume	<10/day	10-100	100-1000	>1000
Error tolerance	Zero	Low	Medium	High (user can retry)
Guardrail cost	Heavy/blocks ROI	High	Medium	Low

Figure 6 Six-Dimension Opportunity Scorecard

Once an idea has been deemed feasible, the third and final step is to prioritize it against other viable opportunities. Each surviving proposal is scored on a scale of 0-3 across six key dimensions, as shown in the accompanying figure: Business lift, Nature of solution, Human effort, Volume, Error tolerance, and Guardrail cost. The summed score, with a maximum of 18, slots the use case into a priority queue. This scoring mechanism is designed to surface high-value, high-volume problems where GenAI can yield a clear impact on revenue or OPEX. Low-scoring concepts are not necessarily discarded; instead, they remain parked or are recycled for future consideration should their projected lift or the economics of applying guardrails improve.

4.4. Operating Rhythm

- Idea intake. Run every proposal through the Decision Funnel in a 30-minute red-team session.
- Feasibility workshop. Assign pillar owners; score collaboratively; capture mitigation actions.
- Scorecard session. Apply the six-dimension rubric; sort backlog by total score; publish roadmap.

This staged approach eliminates low-fit concepts early, exposes hidden delivery blockers before commitment, and concentrates capital on the highest-return, lowest-regret GenAI initiatives. Further, during discovery workshops targeting GenAI opportunities, product managers frequently uncover ideas that don't advance through initial evaluation phases. Rather than dismissing these outright, teams should document all proposals on a quadrant grid comparing "Problem Structure" and "Error Consequence." This reframes rejection as strategic redirection. Deterministic, high-risk tasks unsuitable for GenAI might benefit from traditional code or explainable machine learning, while simpler, low-risk tasks could suit automation or RPA.

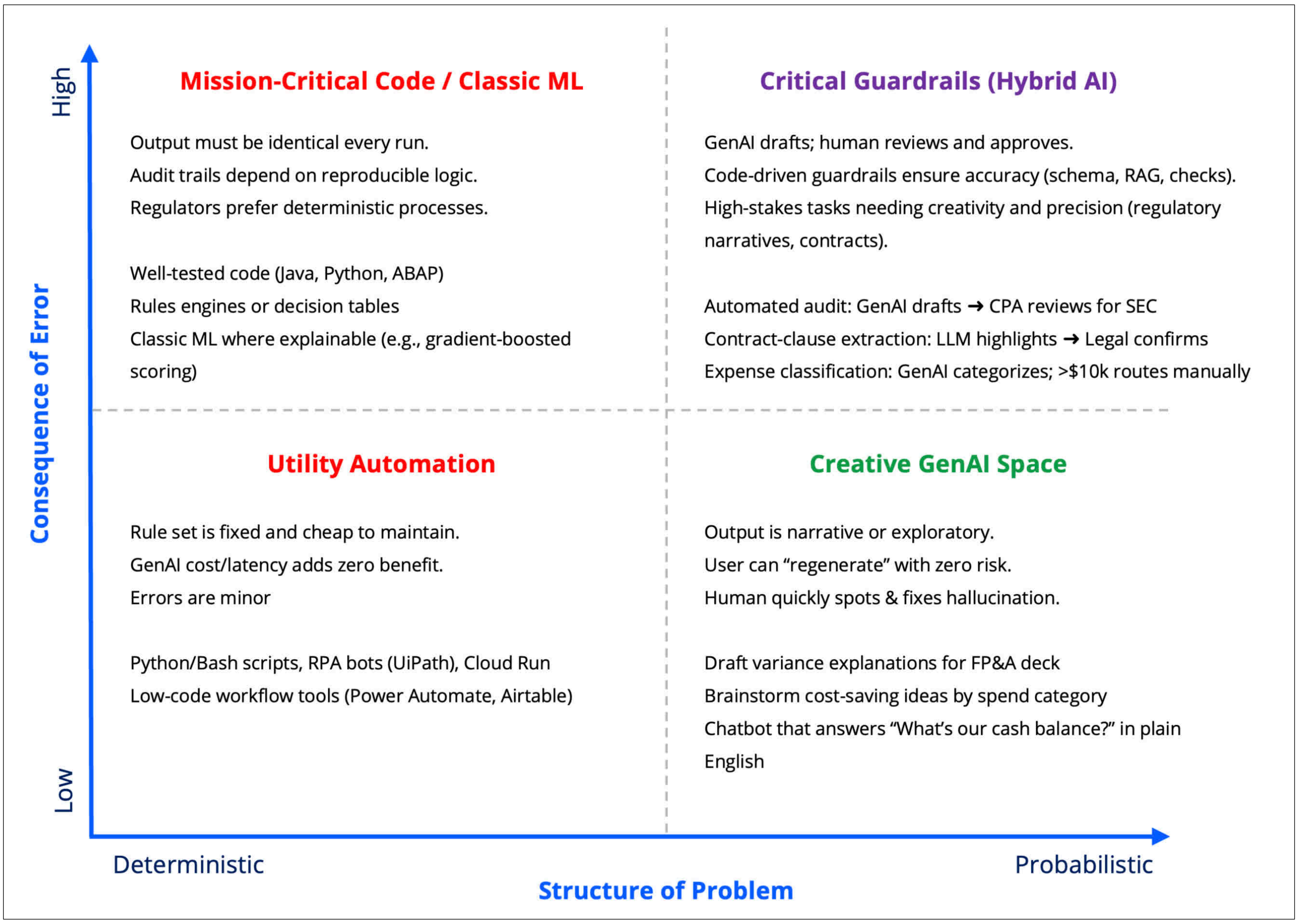


Figure 7 Opportunity Assessment Grid

Mapping ideas ensures comprehensive visibility of potential value, preventing forced adoption of GenAI solutions due to mandates. It encourages parallel exploration: GenAI-ready ideas progress to piloting, while alternative solutions pursue their distinct paths. Ultimately, this structured approach empowers product managers to advocate for decisions driven by value, not hype.

4.5. Minimalism as the Default for GenAI

Product managers must be ruthlessly selective: say “yes” only to the narrowest GenAI opportunity that proves value. Start with a single outcome, feed only the data essential to that outcome, and release to the smallest user cohort that can generate reliable feedback. Every extra feature, dataset, or audience multiplies LLM-specific risks—hallucination, context drift, prompt-injection—so opportunistic, “let’s try everything” bids wait until lean wins and solid guardrails are in place. Incremental successes build organizational trust and sharpen governance muscle, creating a virtuous loop of confidence and capability. Only after several validated wins should the scope widen, ensuring scale follows proven learning rather than speculative ambition.

5. Results and Discussion

The framework was trial-run in five in-flight projects in different companies. Four of the five (≈80 %) exhibited rapid operational gains. Average time-to-market fell by roughly one-third and PoC-to-production conversion rates more than doubled. Internal Net-Promoter scores rose by nearly twenty points, rework requests due to 'wrong-tech' selection shrank by more than 70%. Teams credited these gains to early culling of ill-fit ideas, minimalist scoping that curbed over-engineering, and swift redirection of deterministic tasks to rule engines or RPA, delivering higher value with fewer resources. Results underscore the payoff of saying “no” quickly and building only what is essential. A minimalist first release contains risk, simplifies governance, and frees capacity for additional experiments. Equally important, mapping rejected ideas to alternative technology tracks reassures business units that their problems—not the hype—steer solution choice, strengthening trust in product leadership.

6. Conclusion

The proposed GenAI framework—combining red-team intake, feasibility workshops, a six-dimensional scorecard, and a disciplined operating rhythm—was piloted across five live projects, yielding more consistent solution choices and strengthened stakeholder confidence. By enforcing minimalism and iterative validation, it ensures rapid, low-risk delivery of tangible business value.

Future Research Directions

Large-scale validation is the first priority: the framework must be stress-tested across dozens of portfolios to see which gates hold, which weights drift, and how culture alters outcomes. Next, a robust blueprint for hybrid architectures—combining GenAI with rules engines, analytics, and RPA—must be engineered so teams can allocate functions to the technology that maximizes value. Third, research should clarify governance and change-management levers that let data-sensitive organizations adopt LLMs without paralyzing compliance or executive confidence. Finally, the field needs an execution-phase playbook: guidance on model drift, guardrail upkeep, and prompt-management debt that takes product managers from “approved idea” to sustained production impact.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] R. Sallam, "Top Strategic Technology Trends for 2024," Gartner, Oct. 16, 2023.

[2] T. H. Davenport and A. C. Miller, "Want to Scale AI? First, Scale Your Data and Analytics," Harvard Business Review, May 25, 2023.

[3] "IDC Forecasts Worldwide AI Spending Will Double Over Four Years, Reaching $301 Billion in 2026," IDC, Press Release, Oct. 10, 2022.

[4] "The state of AI in 2025: Generative AI's breakout year," McKinsey & Company, Aug. 1, 2025.

[5] "IBM Global AI Adoption Index 2023," IBM, 2023.

[6] T. Fu, S. Garg, and A. Singhvi, "Generative AI Is Here: How Tools Like ChatGPT Could Change Your Business," Harvard Business Review, Mar. 28, 2023.

[7] R. G. Cooper, "The Stage-Gate® process for new products: A game-changer," Journal of Business Strategy, vol. 44, no. 5, pp. 265-271, Sep. 2023.

[8] S. S. Gujjar, N. Damera, and P. Majumdar, "The AI Ladder 2.0: A guide to enterprise AI adoption," IBM, 2022.

[9] D. Vesset, R. C. Blum, and H. Carr, "The Business Opportunity of AI," IDC, Doc # US50939523, Jul. 2023.

[10] A. Kumar, P. K. Sahoo, and B. Kumar, "A Survey on the Hallucination of Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," arXiv preprint arXiv:2401.01313, 2024.

[11] N. F. Liu, K. Lin, D. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the Middle: How Language Models Use Long Contexts," arXiv preprint arXiv:2307.03172, 2023.

[12] K. Greshake, S. Abdelnabi, S. Míšeikis, C. Pade, and M. Pawelec, "Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection," arXiv preprint arXiv:2302.12173, 2023.

[13] M. Iansiti and K. R. Lakhani, Competing in the Age of AI. Boston, MA: Harvard Business Review Press, 2020.

[14] C. Miller, "How to Prioritize Your AI Projects," Harvard Business Review, Dec. 7, 2021.

[15] "A framework for picking the right generative AI project," Harvard Business Review, Mar. 2023.

[16] "Use Case Prisms for Generative AI: A Guide to Emerging Opportunities in Industries," Gartner, 2024.