ESCP
BUSINESS SCHOOL

IT ALL STARTS HERE

BERLIN | LONDON | MADRID | PARIS | TURIN | WARSAW

© Adriyanca-AdobeStock

LIGHTS - Sustainability

Responsible RAG systems in practice: Lessons from commercial and pharma use case

ESCP Impact Paper No.2025-43-EN

Alara TASCIOGLU
ESCP Business School

ESCP RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ESCP Impact Paper No.2025-43-EN

Responsible RAG systems in practice: Lessons from commercial and pharma use case

Alara Tascioglu*

ESCP Business School

Abstract

As generative AI systems evolve from tools that deliver information to agents that speak and act on behalf of organizations, their design challenges become not only technical, but also ethical, organizational, and strategic. This paper explores two hypothetical cases developed in 2025 to examine how retrieval-augmented generation (RAG) systems with agentic capabilities can be responsibly deployed in practice. The first case, co-developed with Klart AI, focuses on dual-purpose chatbots for both internal and customer-facing use in the airline industry. The second, in collaboration with Pfizer, considers a digital companion designed to support clinical trial participants in a high-trust healthcare environment. Together, the cases highlight the key risks of hallucination, tone misalignment, and weak fallback mechanisms of such solutions. Based on the analysis of these cases, we suggest a refined framework for responsible AI productization. The proposed principles emphasize traceable retrieval, calibrated tone, ethical fallback behavior, confidence moderation, and agentic restraint, offering a foundation for building AI systems that are not only able, but also trusted to act, or pause, with integrity.

Keywords: generative AI, RAG, pharma, chatbot, customer experience

*Postdoctoral Research Fellow, ESCP Business School

Responsible RAG systems in practice: Lessons from commercial and pharma use cases

Once considered a back-office convenience, chatbots are now stepping into more central roles in business operations (Urbani et al., 2024). Internally, they can support complex data tasks and lighten repetitive workloads. Externally, they can interact directly with customers and often become the first voice of the brand. As their roles expand, so do expectations around tone, trust, and contextual understanding (Finn et al., 2025).

Retrieval-augmented generation (RAG) systems have emerged in response to these demands. By grounding their answers in curated, up-to-date documents, they combine the fluency of large language models with the reliability of retrieval (Lin, 2024). Many now also include agentic capabilities such as making simple decisions, completing multi-step tasks, and initiating actions like bookings or purchases on behalf of the user (Finn et al., 2025). This makes them not only more responsive and potentially more trustworthy, but also more complex to design well.

In 2025, we developed two hypothetical case studies to explore how RAG-based chatbots with agentic capabilities can be designed to communicate, guide, and assist without overstepping. The first, with Klart AI, looks at how companies use these systems as internal tools and customer-facing products. The second, with Pfizer, addresses the more sensitive challenge of supporting clinical trial participants through digital companions. Together, these cases highlight the often invisible design work required to make advanced chatbots feel trustworthy, adaptable, and aligned with human expectations. These two cases were chosen not only for their technical relevance, but for their contrasting stakes: one operates in a commercial setting where user experience shapes brand perception, the other in a medical context where tone and trust can directly influence well-being and participation. Their juxtaposition reveals how the same RAG architecture must be reinterpreted based on use case sensitivity, user vulnerability, and ethical obligations.

Case I: designing RAG systems as products with Klart AI

Klart AI is a generative AI company based at Station F in Paris that develops advanced chatbot systems as commercial products for corporate clients. Built using retrieval-augmented generation (RAG), these bots are designed to integrate directly into core workflows, such as assisting HR teams, supporting customer service operations, or navigating dense policy documents. Klart AI packages its bots not as simple helpers, but as scalable services that can be tailored to both internal and external business needs.

In the hypothetical case study we co-developed, a major airline client requests two bots: one internal assistant to help call center representatives communicate policy details (such as luggage allowances, ticket changes, and pet-in-cabin rules), and one external agent to respond to customers directly through the airline's help portal. Both bots operate on the same RAG framework, drawing from internal policy documents for the employee-facing assistant and FAQ databases for the customer-facing one.

While the underlying architecture is similar, the way each bot must behave and be perceived is very different. The internal system is expected to deliver fast, factual answers to support operational efficiency. Tone is secondary to speed and consistency, and small slips in phrasing are generally tolerated. The external bot, by contrast, must reflect the airline's brand values. It must respond with empathy, accuracy, and tone sensitivity, especially when

delivering policy decisions. It also includes agentic capabilities, such as offering booking options or enabling users to purchase extras like additional luggage or pet-in-cabin services directly through the chat.

RAG systems bring considerable benefits in these contexts, but also carry three persistent risks: hallucinated content, tone inconsistency, and inadequate fallback strategies. These are not new problems but become business-critical in a productized AI setting. Drawing from this case, we worked with Klart AI to outline a preliminary design framework to guide the deployment of dual-use RAG systems. The initial version included five key elements:

1. **Retrieval grounding:** Responses must be based on curated, current, and role-specific documents, and not general LLM knowledge.
2. **Tone by design:** RAG systems must be trained and instructed to adopt context-appropriate tone (e.g., policy vs. FAQ).
3. **Fallback with clarity:** RAG systems should be equipped with fallback strategies that clearly communicate uncertainty or redirect users to human support, without undermining user trust.
4. **Confidence moderation:** When a RAG system is unsure or the source material is ambiguous, it should signal limitations rather than improvise or overcommit.
5. **Agentic boundary-setting:** For RAG systems with action-triggering capabilities (e.g., bookings), safeguards must be in place to verify input clarity, prevent misuse, and respect system constraints.

While this framework served as a solid starting point, it raised further questions about what happens when tone, trust, and responsibility aren't just nice-to-have features but essential design requirements. We explored this deeper in the second case.

Case II: RAG systems for sensitive interactions with Pfizer

The second hypothetical case centers on the more sensitive and ethically complex challenge of supporting clinical trial participants through a digital companion chatbot at Pfizer. Unlike the Klart AI case, where bots are built as commercial products, this case explores what happens when a generative AI system becomes part of a human medical experience. Given the high dropout rates often seen in clinical research, this RAG system is designed to provide logistical guidance, clarify procedures, and offer psychological reassurance throughout the participant journey.

On the surface, the system performs similar tasks to those in the Klart AI case by answering questions, interpreting emotionally charged inputs, and maintaining a consistent tone. But the stakes here are entirely different because a factually accurate response delivered in the wrong tone could heighten anxiety or even cause patient withdrawal, making tone not just a branding concern but a vector of trust. Therefore, the bot isn't merely representing an organization but mediating a person's engagement with a high-stakes medical process. As a result, the design framework developed with Klart AI still applies, but it comes under pressure in the following areas:

1. **Tone by design:** now means knowing when not to speak at all.
2. **Fallback with clarity:** becomes a tool not only for handling uncertainty, but also for signaling care and ethical boundaries.
3. **Confidence moderation:** shifts from a business risk safeguard to an ethical imperative.

The digital companion is explicitly instructed to avoid any form of medical advice beyond its scope, flag emotionally charged queries for human follow-up, and defer to clinical staff whenever ambiguity arises. Its purpose is not to provide every answer, but to recognize when it should not try, making fallback a form of ethical care.

This case also reframes the notion of agentic behavior. While the Klart AI bot can perform actions like bookings and add-ons, the Pfizer bot's value lies in presence, not productivity, and its outer reach is deliberately constrained. Since its primary function is to hold emotional space and avoid harm, restraint in this case is a design strength and not a limitation. The digital companion exercises agentic behavior not to act externally, but to make real-time judgments (e.g., recognizing extreme anxiety or hypochondria) and escalate these moments to human professionals.

In high-risk domains like healthcare, AI systems must comply with evolving requirements around data privacy, explainability, and human oversight (Freyer et al., 2024). Pfizer's hypothetical decision to build this system internally reflects a growing trend among companies seeking full control over tone calibration, data governance, and ethical guardrails, especially when stakes are personal, consequences irreversible, and regulatory expectations increasing. Internal development enables organizations to align with both institutional standards of care and emerging frameworks for Responsible AI, where accountability is not just a policy layer, but a design constraint (Stix et al., 2025). In this context, system design becomes a form of governance, and a commitment to ethical responsibility.

From use case to design principles: productizing AI with responsibility

As generative AI systems move closer to the core of business operations and begin to act on behalf of organizations, they must be designed not only for performance, but for trust, restraint, and long-term accountability. The original framework co-developed with Klart AI provides a solid foundation for thinking about tone, fallback, and action boundaries. But when tested in the more ethically charged context of healthcare, it is clear that these principles need to stretch further toward governance, escalation, and design humility.

What follows is a refined version of the framework, adapted to reflect the demands of RAG-based systems with agentic behavior in both commercial and high-trust environments. It is meant not as a checklist, but as a lens through which to evaluate the responsible productization of generative AI:

1. **Retrieval grounding with traceability:** Design systems to retrieve from curated, context-appropriate sources with transparent links or signals that build user trust and enable accountability.
2. **Tone by design, silence by intent:** Tune tone not just for brand alignment, but for user safety. In sensitive contexts, knowing when to remain silent is a feature, not a flaw.
3. **Fallback as ethical posture:** Build fallback mechanisms that do more than redirect by communicating boundaries, expressing care, and inviting human re-engagement.
4. **Confidence calibration under uncertainty:** Equip systems to flag low-confidence moments, avoid improvisation, and defer decision-making when ambiguity could lead to harm.
5. **Agentic restraint and escalation:** Design agentic behaviors with limits. In high-stakes settings, the bot's ability to act must be paired with clear escalation paths and responsibility handoffs.

These principles were not derived solely from theoretical reflection but from discussions held with Klart AI and Pfizer employees before and during guest speaker sessions. The resulting design framework was further shaped through case study questions where students analyzed risks and recommended improvements in groups. This process surfaced consistent challenges and informed the five design principles proposed here.

As organizations begin to rely on AI systems not just for information delivery but for action, communication, and care, design becomes a matter of responsibility as much as functionality. The cases explored in this paper show that as AI becomes more agentic, design questions evolve into ethical questions, and usability becomes a form of institutional expression. The challenge for organizations is no longer simply how to deploy AI, but how to embed values, boundaries, and trust into systems that increasingly speak and act on their behalf. The principles outlined here are not exhaustive, but a starting point for shaping responsible RAG systems that are not only capable, but trustworthy enough to know when to act, and when to hold back.

For corporations seeking to implement this framework, the process should involve interdisciplinary teams including product managers, UX designers, AI/ML engineers, compliance and ethics leads, and domain experts. Development steps should include: (1) mapping high-value, repeatable use cases; (2) defining role boundaries and tone guidelines; (3) sourcing and curating retrieval documents; (4) prototyping conversation flows with clear fallback strategies; and (5) testing with target users to calibrate confidence thresholds and escalation paths. Governance teams should oversee ongoing evaluation to prevent drift, ensure explainability, and maintain alignment with organizational values.

References

Finn, T., Downie, A. (2025). *Agentic AI vs. generative AI.* Retrieved from https://www.ibm.com/think/topics/agentic-ai-vs-generative-ai? on 07/16/2025

Freyer, N., Groß, D. & Lipprandt, M (2024). *The ethical requirement of explainability for AI-DSS in healthcare: a systematic review of reasons.* BMC Med Ethics 25, 104. https://doi.org/10.1186/s12910-024-01103-2

Lin, B. (2024). From RAGs to vectors: How businesses are customizing AI models. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/from-rags-to-vectors-howbusinessesare-customizingai-models-beea4f11 on 07/16/2025

Stix, C., Pistillo, M., Sastry, G., Hobbhahn, M., Ortega, A., Balesni, M., Hallensleben, A., Goldowsky-Dill, N. & Sharkey, L. (2025). AI Behind Closed Doors: a Primer on The Governance of Internal Deployment. arXiv preprint arXiv:2504.12170.

Urbani, R., Ferreira, C., Lam, J. (2024). *Managerial framework for evaluating AI chatbot integration: Bridging organizational readiness and technological challenges*. Business Horizons, 67, 5. https://doi.org/10.1016/j.bushor.2024.05.004