Nortal

How to create high-value GenAI apps

A Nortal guide

Table of contents

This report may be cited as "How to Create High-Value GenAI Apps: A Nortal Guide (2025)."

Nortal is a global strategic innovation and technology company and a valued partner for governments, healthcare institutions, leading businesses, and Fortune 500 companies.

Introduction

Generative AI (GenAI) has massive potential, but fulfilling that potential takes more than advanced technology. It takes a strategy based on real-world business challenges and guided by practical experience. With both of these in our tool belt, we've been helping organizations design GenAI applications that don't just function — they transform.

This white paper draws from our expertise. We lay out the principles that matter most: building flexible, future-ready architectures, managing data with precision, and identifying the use cases where GenAI delivers measurable value. We also highlight common pitfalls and lessons from successful organizations. These aren't just theories; they're insights we've learned from experience.

Understanding the GenAI opportunity

GenAI is redefining automation, finding data patterns that take extensive human effort to identify. It's most powerful when tasked with specific, high-value problems instead of broad generalizations. The challenge is knowing where to focus.

Our approach starts here: helping clients find the potential that isn't immediately obvious. Not every challenge is a fit for GenAI, so successful use means determining where its impact can be most meaningful. Early wins matter, since each one separates temporary uses from long-term advantages.

Diving into GenAI is like sailing unfamiliar waters. The first journeys need caution and observation. But with practice, the route gets easier. Our approach is to build the ship while helping our clients chart their own course and find hidden opportunities.

The markers of successful GenAI applications

When applied with intention, GenAI reshapes how organizations operate. The most effective implementations have these things in common:

- **Democratized access to data:** Information flows across teams with proper controls in place.
- **Accelerated processes:** Work moves faster and smarter.
- **Reduced costs:** Automation replaces repetitive, manual tasks.
- **Expanded automation:** Efficiency and capability increase.
- **Delivered personalization:** Outputs align precisely with audience and context.

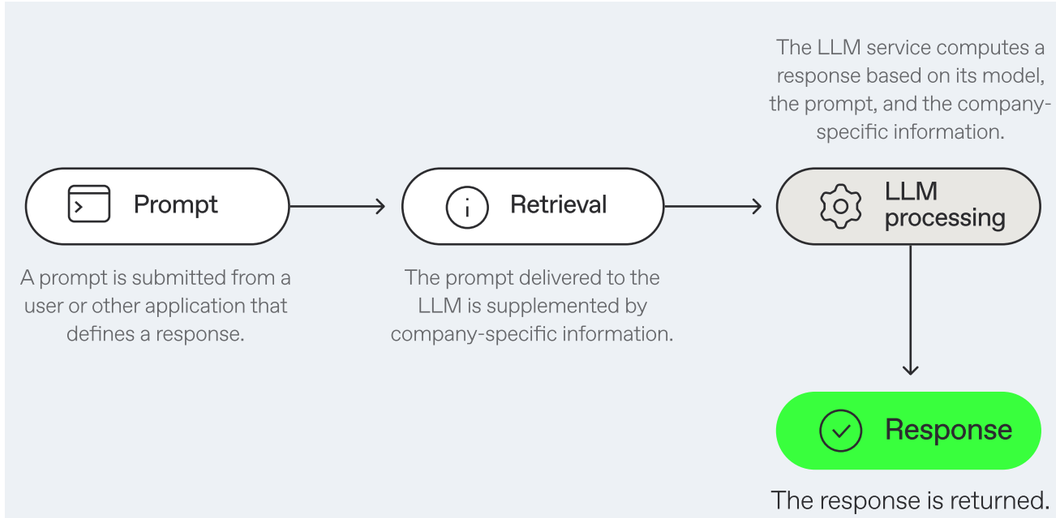
No technology succeeds in isolation — all winning GenAI solutions are founded on alignment with clear business goals. Ask the right questions: What problem are we solving? Do we want to boost productivity, lower costs, or speed up time to market? Without that clarity, even an advanced system can be nothing more than a novelty tool.

We've learned that purpose defines success, and every successful project begins with a clear understanding of *why.* When GenAI yields meaningful outcomes, its potential becomes measurable progress.

The fundamentals of GenAI applications

Across industries, the Retrieval-Augmented Generation (RAG) framework has proven to be one of the most effective approaches for solving real business problems. A basic RAG app is a solid foundation, but impactful enterprise-grade solutions require more layers to become complete systems.

A basic RAG app



Large language models (LLMs) are built on massive public datasets, making them powerful but general; they don't inherently understand your business. RAG bridges that gap by supplementing the LLM with organization-specific content, so responses reflect the right context.

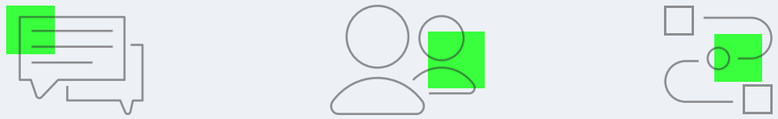
The key to making the RAG approach work is deciding what company-specific content to add to the prompt. By adding proprietary data into a vector database, relevant information can be dynamically retrieved and built into the prompt. The LLM then delivers insights based on the company's unique operations and goals.

A mature RAG app

UX layer	The User Experience (UX) layer adapts the application to the needs of the users in a business context the output from the RAG and any other layers needed are combined to best serve the user.
Orchestration and integration layer	The orchestration layer takes instructions from the UX layer and invokes the RAG apps and other services, integrates their responses, and provides the result to the UX layer.
RAG layer	The RAG layer consists of the various RAG applications that are needed by the application.
Service integration layer	The service integration layer manages the conversation between all of these applications and services.
Data layer	The data layer provides a gateway to this data to support RAG and the rest of the application.
Security and operations layer	When an application makes a prompt for a user, it is vital that they have permission to access any supplemental data or documents provided to RAG or accessed by other services.

High-value business applications extend the RAG model with layers that address production readiness, governance, and domain-specific requirements. Without those additional components, GenAI apps can't reliably support mission-critical use cases.

Chat vs. copilot vs. agent



GenAI shows its versatility in its many interfaces.

- **Chat:** The most familiar interface. Users enter prompts and receive responses in text, image, or audio format.
- **Copilot:** Observes user actions and provides guidance or suggestions, such as acting as an assistant in a car or watching a coder and making remarks. The user stays in control.
- **Agent:** Operates autonomously toward defined goals, often teaming up with other agents for trickier tasks.

The most advanced GenAI systems can combine these approaches, creating layered experiences that adapt to user needs.

Nortal's guardrails for successful GenAI app development

Drawing on experience across industries, we've refined guiding principles for building effective GenAI applications. These guardrails keep projects focused, realistic, and aligned with business goals.

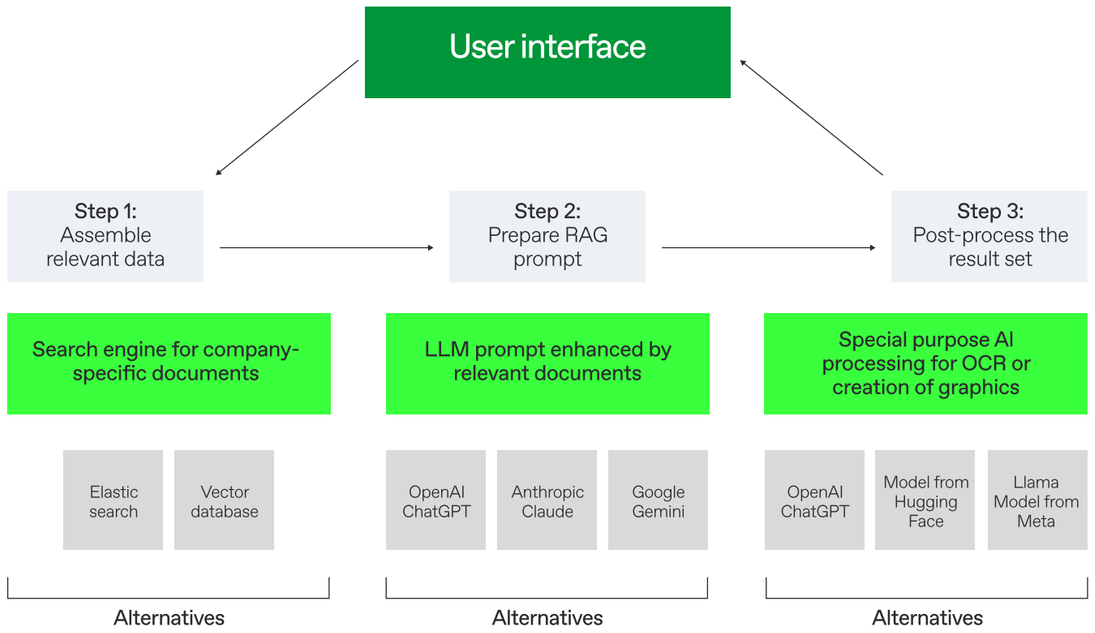
Guardrails for successful GenAI app development

Utilize tech-agnostic architecture

The architecture should allow for choices between alternatives at as many stages as possible.

Anticipate cost pressure

Implement cost and quality metrics so that it is easy to evaluate price and performance. The current cost structure of GenAI systems will change as services become less subsidized.



Compare results

Evaluate the performance of each alternative so you can choose the most appropriate system.

Protect against hallucinations

Implement countermeasures to ensure LLM hallucinations don't cause errors.

Tech-agnostic architecture

Adaptability is the foundation of long-term success. At Nortal, we build architectures independent of any single tech vendor, giving clients the freedom to evolve with the AI landscape.

A modular architecture is adaptable, able to adjust its systems and course as conditions and technologies change. With this flexibility, organizations can:

- **Plan for evolution:** Build systems that adapt as LLMs advance or retire. This ensures seamless transitions when LLM versions change or disappear.
- **Compare and optimize:** Evaluate multiple models from providers like Azure OpenAI, Google Gemini, and Anthropic Claude to find the right balance of performance and cost.
- **Stay ready:** Integrate emerging tools without reengineering the system.
- **Control expenses:** Shift to more efficient or affordable models as markets change.
- **Customize to industry:** Select solutions that align with sector-specific needs.

Our modular approach allows each component to evolve independently for continuous improvement without disruption. And we base it all on the most important factors: data governance, security, performance, cost, and industry-specific requirements.

Comparing results across models

Each AI model has strengths and limitations shaped by its design and training. The right fit depends on your goals. Rigorous benchmarking is key; test different models under real-world conditions and choose based on performance.

In one of our projects for a gas meter company, we focused on character recognition and found that Azure OpenAI ChatGPT-4o, a general-purpose AI model, outperformed a specialized Microsoft OCR system. Controlled testing revealed unexpected advantages and helped keep the technology targeted on business results.

Guarding against hallucinations

Hallucinations happen when LLMs provide unfounded, made-up statements. In a November 2025 update, Vectara's hallucination leaderboard showed the rate of hallucinations for the top 10 commercial LLMs was between 3.3% and 14.7%.

Fortunately, these protective measures mitigate risks:

- Retrieval of high-quality, context-relevant information via RAG.
- Structured prompting methods, like chain-of-thought prompting, which help the model break down complex reasoning into smaller steps.
- Fine-tuning on domain-specific data.
- Confidence scoring and validation checks for accuracy.

Anticipating cost pressure

AI services are currently subsidized, with providers offering competitive pricing to encourage adoption. But as adoption grows, costs will rise. Sustainable GenAI strategies must be prepared for those shifts.

Using specialized open-source models can help manage costs while maintaining performance. For example, we put OLLAMA on virtual machines to generate content, which has delivered strong outcomes at much lower cost.

GenAI is only part of the solution

Chatbots and copilots get most of the attention, but they represent only a slice of GenAI's potential. For true transformation, build AI into core business processes so that it can augment, not just interact.

Agentic AI embodies that idea: intelligent agents work behind the scenes to analyze data, find insights, and automate complex tasks. The chat interface is only the front door to a system capable of transforming the way your business works.

A holistic approach to GenAI success

To get the most of GenAI's potential, look past the technology and to the surrounding ecosystem: user experience, data flow, and integration.

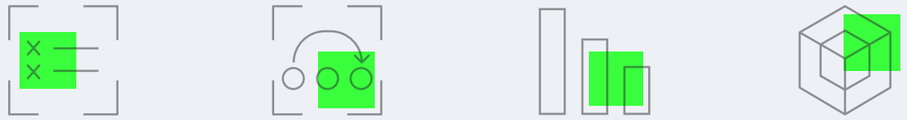
- **User experience:** Design interfaces that reflect how people actually work, whether through chat, dashboards, or embedded analytics.
- **Security and compliance:** Protect sensitive data — such as customer facts or proprietary business info — through encryption, access controls, and adherence to frameworks like GDPR and HIPAA.
- **Integration:** Prevent silos and drive adoption by connecting GenAI to existing tools such as CRM, ERP, or project management systems.

Our GenAI design philosophy is holistic. We balance innovation with practicality while ensuring AI serves the business and daily operations seamlessly and securely.

Common pitfalls in GenAI implementation

In our experience, we have observed these mistakes that hinder the success of GenAI adoption:

- **Lack of clear business case:** GenAI should always address a defined problem with measurable ROI. Why are we implementing this technology?
- **Skipping proof of concept:** Pilot first. Before scaling, identify potential issues and validate assumptions with a controlled dataset and clear success metrics.
- **Ignoring performance metrics:** Define KPIs early and measure continuously.
- **Over-customizing:** Avoid building unnecessary models. For most scenarios, LLMs paired with RAG techniques provide powerful, cost-effective solutions.



GenAI requires strong data management

Every GenAI solution depends on high-quality data. A successful implementation depends on solid data management practices, and reliable insights require accurate, secure data sources.

RAG is central to this. It enables accurate, context-aware responses by linking proprietary data to the LLM. This involves building a vector database that acts as a knowledge repository that allows the LLM to speak your business language.

But it also introduces security responsibilities — without proper access controls, RAG-enabled applications could inadvertently expose sensitive information.

Before deploying a GenAI solution, use our best practices to avoid potential security risks and ensure compliance: role-based access, consistent governance across environments, and regular audits.

Selecting the highest-value use cases

Effective GenAI implementations start with defined and measurable tasks. We have identified key applications where this technology consistently proves its value:

- **Automating content creation:** Legal, marketing, and communication teams can shift focus to strategy and revision while GenAI handles drafting and research.
- **Streamlining workflows:** Recruiting teams can use GenAI to optimize tasks: craft job descriptions, sift through resumes, and identify top candidates. Finance and operations benefit from automated analysis of large datasets.
- **Democratizing knowledge:** Internal search and knowledge tools remove bottlenecks and empower employees.
- **Optimizing data management:** Smarter indexing and metadata generation reduce storage and improve retrieval speed, while addressing duplicate and outdated data.

These early wins build confidence and demonstrate real ROI, setting the stage for broader adoption.

Evaluating ROI for GenAI

Traditional ROI models focus on short-term, quantifiable metrics, like cost savings and productivity gains over a set period. GenAI's impact is broader and more dynamic, delivering cumulative, qualitative, and competitive benefits.

Successful measurement considers:

- **Competitive differentiation:** Customer satisfaction and employee engagement.
- **Faster analytics:** Accelerated decision-making and responsiveness.
- **Cumulative gains:** Long-term improvements in efficiency, brand equity, and retention.

Costs should factor in maintenance, updates, and human oversight. ROI for GenAI isn't static — it evolves with the technology.

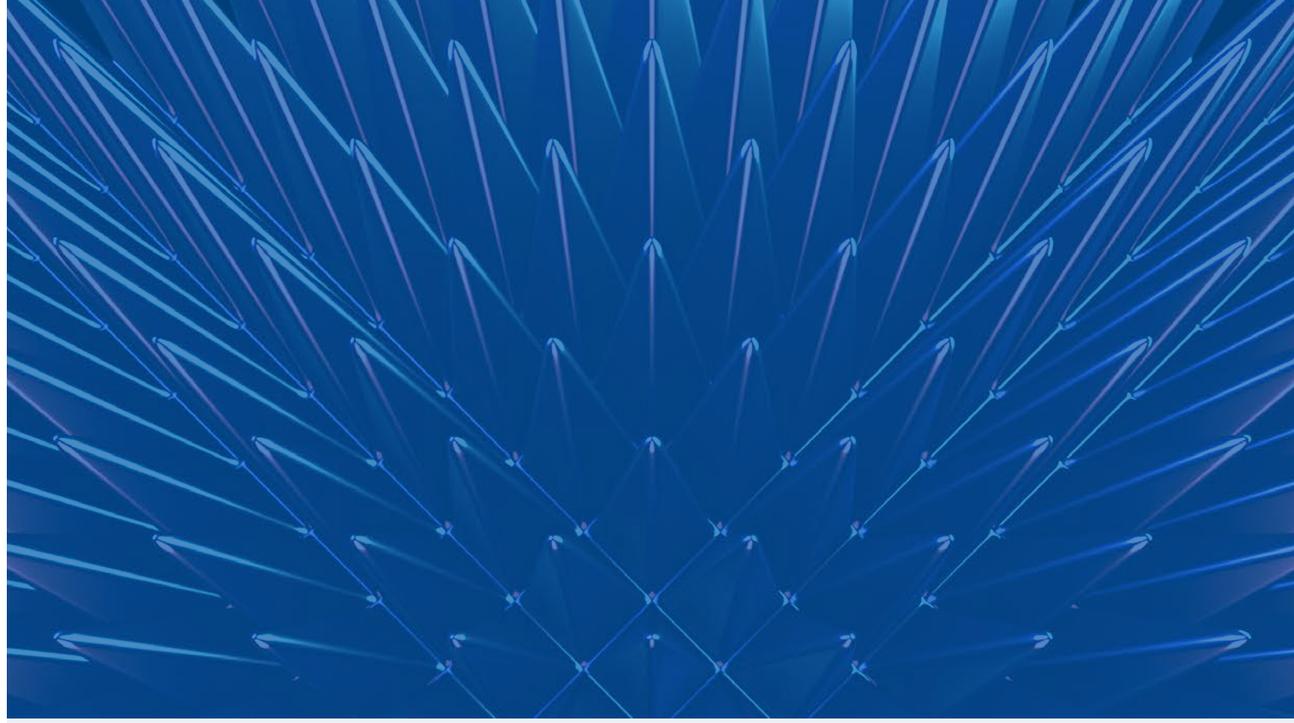
> Example: Call centers
>
> Call centers show GenAI's value clearly:
>
> - Automated transcription and summarization reduce post-call workload.
> - Sentiment analysis enhances follow-up and customer retention while indirectly protecting brand image.
> - Pattern detection plus other techniques identify systemic issues before they escalate.
>
> These direct and indirect gains combine to create measurable, sustained ROI.

Final thoughts

Getting real value out of GenAI comes from disciplined execution: selecting the right use cases, basing solutions on high-quality, organized data, and building flexible architectures. To achieve significant outcomes, organizations need to treat GenAI as another business capability and invest time and resources in the foundations that make it dependable, secure, and scalable.

There are three pillars that guide the path forward: defining purpose, validating assumptions via PoC, and measuring results continuously. These actions balance innovation with governance to pair LLMs with strong retrieval systems, integrate AI into workflows, and prevent hallucinations.

While the opportunity is real, so is the complexity. Organizations that take this holistic approach will not only adopt GenAI efficiently but also build robust competitive advantage.



Nortal's competitive edge

We approach GenAI partnerships with transparency, professionalism, and a collaborative mindset. Instead of overpromising, we guide. Our goal is to deliver meaningful results that align with each client's vision and constraints, from planning to deployment.

Trust drives our relationships. We operate as an extension of our clients' teams to make sure every step is practical, ethical, and strategically sound. By being up front about both our strengths and limitations, we build long-term partnerships.

Success stories

Nortal's industry-spanning portfolio showcases how we turn strategy into results.

Using GenAI to find the right talent

Over four weeks, we developed *Vitalia,* an app that accelerates hiring by matching thousands of candidates to job descriptions through:

- Azure OpenAI Studio GPT-4o for natural language processing.
- Text-embedding-3-large for vectorization.
- Azure AI Search for database functions.
- LangChain as an AI application integration framework.

Some of the functionalities:

- Extracting structure characteristics from candidate and job descriptions into a vector database.
- Scoring candidates' qualifications, with an SQL-based search creating a shortlist of the best matches.
- Using GenAI to analyze the shortlist and describe the best candidates.

Vitalia cut the first viable candidate delivery time by 50% and reduced the full-cycle search time by up to 30%.

Additional AI applications by Nortal

GrayMeta: We designed an AI-powered media content management app used by HBO and Disney. Leveraging machine learning models and APIs, the solution introduced features like visual recognition, audio classification, and language detection to navigate media libraries with greater efficiency.

Houston healthcare client: We built a copilot that reduced medical practice incorporation time from weeks to hours. This streamlined automation led to a measurable increase in the number of new practices established, delivering immediate value to the client.

US healthcare provider: We replaced an outdated batch file system with a real-time machine learning model that prioritizes cases that require immediate attention. This improvement significantly streamlined case management workflows.

We also apply these principles internally. Our own scalable data lake integrates information across data sources, improving operational visibility and decision-making.

With our Data and AI Studio, organizations can quickly identify where GenAI creates the greatest impact and move from experimentation to transformation.

Nortal

The future delivered. Seamlessly.

nortal.com