

Should you own your Generative AI model?

3 considerations to guide your decision

Generative AI is the hottest topic in technology — and with good reason. Generative AI has the potential to radically transform the work of enterprises and governments, but their needs are vastly different from the consumer use cases that have been gaining headlines.

Most importantly, before embarking on this AI journey, organizations have a critical decision to make: whether to use a model built and owned by a vendor, or a model that is owned and controlled by the customer organization. Based on decades of cumulative experience, we believe this decision should be informed by three fundamental considerations.



1 What are the laws, regulations, and expectations of your industry and customers?

In many industries, transparency and explainability of model training is already required. Both **GDPR** and **FINRA Rule 3110 (Supervision)** require transparency and explainability in the decision-making process, whether manual or automated.

However, in all industries, being able to document the model's training lineage and model weights is essential to retain customer trust, facilitate audits, and help identify potential biases or ethical concerns. As an example, model governance and lack of explainability were cited as the highest concerns among 84 percent of financial executives responding to a LendIt annual survey on AI.

Which of the following concerns does your organization have about adopting AI for credit risk?



Source: LendIt and Brighterion, "AI perspectives: Credit risk and lending," 2022.
*Responses do not total 100% as more than one answer may apply

2 How accurate does your model have to be?

Achieving the highest levels of accuracy requires ‘fine tuning’—that is, additional training specific to the industry or organization. For instance, the word ‘quarter’ means something very different in banking than it does in sports.

Fine-tuning also reduces the chance of dangerous “hallucinations” (when an AI response sounds confidently correct but is not factual)—less important in recreational AI tools, but potentially catastrophic in those focused on enterprise tasks.

Since shared models are trained on massive public datasets, a single model can solve dozens of consumer tasks with strong general accuracy but is far less accurate for domain-specific or company-specific tasks.



However, almost by definition, the most important data for fine tuning is also the most valuable and sensitive—customer data, product design information, competitive intelligence, financial or patient data, personnel, supplier, or legal material.

Passing this data to a shared model means that the resultant model would not be owned by the originating organization, and potentially exposes an organization to concerning security issues and privacy breaches.



The Power of Domain-Specific Training

Bloomberg developed BloombergGPT—a model tailored for the financial industry that focuses on financial natural language processing tasks, such as sentiment analysis. This delivered benchmarking scores for financial tasks 27-33 points ahead of models trained with general-purpose data alone.¹

1 – Bloomberg Professional Services, March 30, 2023, “Introducing BloombergGPT, Bloomberg’s 50-billion parameter large language model, purpose-built from scratch for finance,” <https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/>

3 Should you use a tool or build an asset?

In the 20th century, organizations were valued primarily on their capital and personnel. Recently, we have understood more clearly the value of intangible assets like brand, reputation, and market position. Looking toward the future, how organizations retain, understand, and analyze their data will be a differentiator critical to market value.

Specific AI tools (point solutions) may increase efficiency of specific tasks such as writing or data retrieval, but will not be able to fundamentally shift the business model underpinning the organizations that use them. The value created will mainly benefit the vendors of the shared model, not its users.

By contrast, organizations that invest in core AI assets—using foundation models and dedicated systems to build capability and value—will experience myriad benefits across the organization, and retain value for themselves, their customers, and their partners. Generative AI offers a strategic opportunity for the enterprise to build core IP with large language models (LLMs) that have been adapted with the company’s data, unlocking long-term competitive business value across a wide range of workflows.

Conclusion

Deploying Generative AI requires careful consideration of model development and ownership. Understanding and owning your AI model is critical to transparency, explainability, customer trust, and meeting laws and regulations.

Looking forward, organizations may be defined by how well they adopt and utilize Generative AI—and building AI as an asset will create substantial and increasing ROI over time. To reap maximum value, these models should be developed as critical business assets—utilizing state of the art technology and foundation models, while also being fine-tuned, refined, deployed, and owned by the organization.

Learn more at [SambaNova.ai](https://www.sambanova.ai)

 [linkedin.com/company/sambanova](https://www.linkedin.com/company/sambanova)

 @SambaNovaAI

 info@sambanova.ai



Building Asset Value

A manufacturing company that builds complex systems relies on a highly skilled but aging workforce. A major concern for leadership is the increasing difficulty of training and upskilling new team members.

Training an LLM with the company’s own private data would help to transfer this knowledge, retain top talent and expertise, and support a variety of business processes such as supply chain optimization, system design, manufacturing optimization and field service. Most importantly, the company will be able to leverage Generative AI to accelerate onboarding and upskilling of new employees.

The resulting LLM is an IP asset that grows in value, intelligence, and relevance over time—increasing organizational capability, delivering greater value across multiple domains, and increasing competitive advantage and differentiation for the organization, its customers, stakeholders, and investors.

About SambaNova Systems

Customers turn to SambaNova to quickly deploy state-of-the-art AI capabilities to meet the demands of the AI-enabled world. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing. We enable customers to unlock the valuable business insights trapped in their data. Our flagship offering, SambaNova Suite, overcomes the limitations of legacy technology to power the large complex foundation models that enable customers to discover new services and revenue streams, and boost operational efficiency.

[Contact Us](#)