Demystifying AI: A Two-Part Guide to Strategy and Execution

In a world awash with frantic AI developments and new approaches being unleashed on the unsuspecting business, organizations are simply seeking answers to two questions:

1. *What to do with AI?*
2. *How not to do it wrong?*

As early results have started to trickle in, it is now well established that AI can transform businesses in a deeply meaningful manner that delivers significant business value. Every industry is showcasing diverse use cases where AI helps identify new revenue streams or optimize processes like customer success.

But, the time to achieve these benefits has taken longer than initially expected. Most organizations are still in the experimentation stages and have realized that to achieve the expected benefits in production, they must follow a structured approach that leverages organizational data in a safe, cost-effective, and reliable manner.

While new models and frameworks are released every day, AI's success depends upon a well-thought out strategy and its execution. In fact, strategy and implementation are two sides of the same coin.

This is the first report in a two-part brief to lay out the blueprint for ensuring a successful AI journey. Part 1 of the report examines approaches that your IT architect must engage in with the business that will assist in defining your AI strategy. Part 2 of the series is geared for the IT architects to develop solid execution strategies to meet the strategic AI business imperatives.

AI Strategy Requirement

Let's be honest about our AI strategy.

The success of AI is dependent on how well you understand and manage the underlying data. Hence, AI strategy should be an extension of the data strategy. Organizations that struggle with their data assets should first focus on ensuring their data is high quality, secure and with proper data governance.

However, data and AI strategy diverge very quickly as AI introduces a new set of challenges that were unknown in the data space. As generative AI is based on probabilistic models, unlike the traditional approach of using deterministic languages like SQL, questions don't always produce the same output for a given input. In other words, the very definition of accuracy is different in AI compared to conventional analytics.

In addition, we have typically dealt with structured and semi-structured data. Now, we must extract valuable information from unstructured data, which makes up almost 80% of corporate data, anyway. However, this doesn't mean that we have now a separate stack for unstructured data, perpetuating challenges with disjointed data silos that require an overhead of keeping them in sync. But our goal should be to converge the existing deterministic approaches with probabilistic AI models and get the best of both worlds.

With this goal, we define a framework in figure 1 to develop an AI strategy that should enable rapid prototyping and execution.
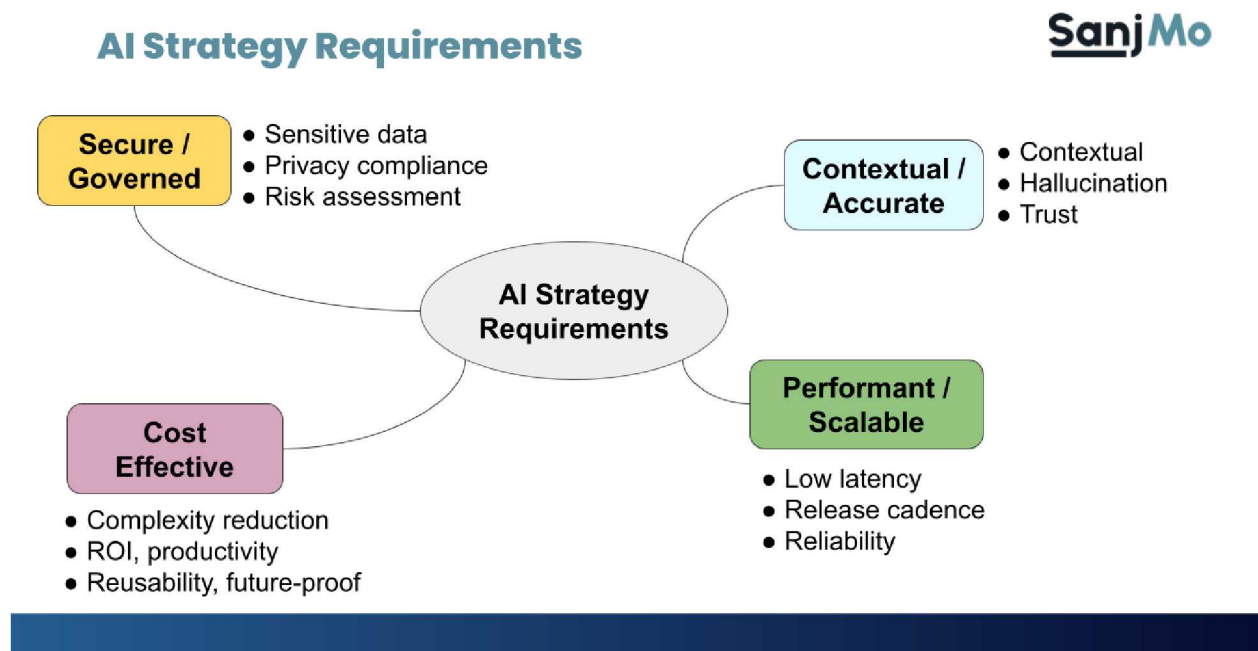

Figure 1: Essential requirements for embarking on AI initiatives

A well-thought-out AI strategy hinges on addressing several critical enterprise needs. These needs encompass not only ensuring the accuracy and security of your data, but also maximizing its potential to deliver real-world value for your organization. The objective components of AI strategy include:

- **Secure / Governed**

  No organization wants to expose its crown jewels, aka proprietary data, to the outside world for multiple reasons, like regulatory compliance, customer trust and maintaining competitive advantage. Hence, AI strategy should ensure data does not leak.

  IT architects should establish a comprehensive AI governance practice that tracks the entire lifecycle - from ideation, to consumption with continuous monitoring of AI models and usage in production.

- **Contextual / Accurate**

  Accuracy isn't just about suggesting the "correct" item. It's about understanding user preferences and situational context, such as time, location & current events to generate personalized responses.

  Achieving high accuracy depends upon both the quality of underlying data as well as the process used to incorporate existing data stores that already store vast amounts of corporate data. Also, this data must be timely and real-time so that decisions can be made on the most up-to-date and fresh data.

  IT architects should use retrieval-augmented generation (RAG), fine-tuning models, creating domain-specific small language models, and synthetic training data to improve contextual accuracy.

- **Performant / Scalable**

  Gen AI applications should facilitate real-time interactions for fast decision making, maximizing impact, and delivering high quality user experiences. If searches take orders of seconds to return results, then user adoption will suffer. Besides low latency, these applications should be capable of high throughput and handle a large volume of requests simultaneously.

  Gen AI training tasks tend to be computationally expensive and hence IT architects should explore performance optimization options like parallelization, distributed computing and elasticity.

- **Cost**

  Gen AI cost requirements are also more complex as they often require specialized hardware for training and inference. Specialized hardware is in short supply. So, the cost of running AI workloads has been much higher. In addition, acquiring high-quality data for training AI models can be expensive, especially if it requires cleaning, curation, preparation, manual labeling or annotation.

  IT architects should be aware of how the technology landscape is evolving and keep the AI strategy in sync. For example, new silicon chips and model optimization techniques are now in

development that will lower the cost for developing, deploying and maintaining AI applications. Another example of cost management is consolidation / doing more in one platform vs. the cost and complexity/risk of multiple point solutions.

To achieve success in your Gen AI deployment a balanced approach consisting of accuracy, security, performance and cost is needed.

Conclusion

The world of AI is brimming with potential, but navigating its complexities can be daunting. This two-part series aims to equip you with the tools to embark on a successful AI journey. In this report, we examined the importance of creating an AI strategy that is closely aligned to your business imperative. Adopting a balanced approach, can help you unlock the transformative power of AI for your organization.

Remember, the key lies in understanding your data, crafting a clear strategy, and meticulously executing that plan. With the right approach, AI can become a powerful tool for driving innovation, optimizing processes, and achieving significant business value.

This report lays the groundwork for your AI journey and builds anticipation for part 2 which delves deeper into strategic execution for success in AI.