AI Reality Check: Turning Strategy into Actionable Results

This is part 2 of the two-part series exploring how to succeed in your AI journey through practical and realistic AI strategy and execution. Part 1 emphasized the crucial link between a well-defined AI strategy and aligning it with business priorities. It is truly the starting point for your AI journey which set out to answer two questions every business is asking:

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

In this report, we examine the practical approach needed for executing our AI strategy requirements. The goal of this document is to provide IT architects with comprehensive and realistic steps to translate AI strategy into action. We explore key considerations for building a robust AI architecture, navigating the evolving AI ecosystem, and establishing effective AI governance practices. And, finally, we take a look at applying our learnings to the *retail banking use case*.

AI Strategy Execution

First, any strategy without execution is just a dream; every strategy needs to be operationalized.

Second, businesses may not be aware of generative AI's potential uses and benefits due to the lack of understanding about AI's capabilities. A close alignment is needed with IT architects to execute AI strategy. In this section, we propose that IT architects should stretch before they run by establishing joint 'hackathons' that can educate their organization on how best to put their AI strategy into action and use figure 1 to lay out essential steps needed to successfully deliver AI initiatives.

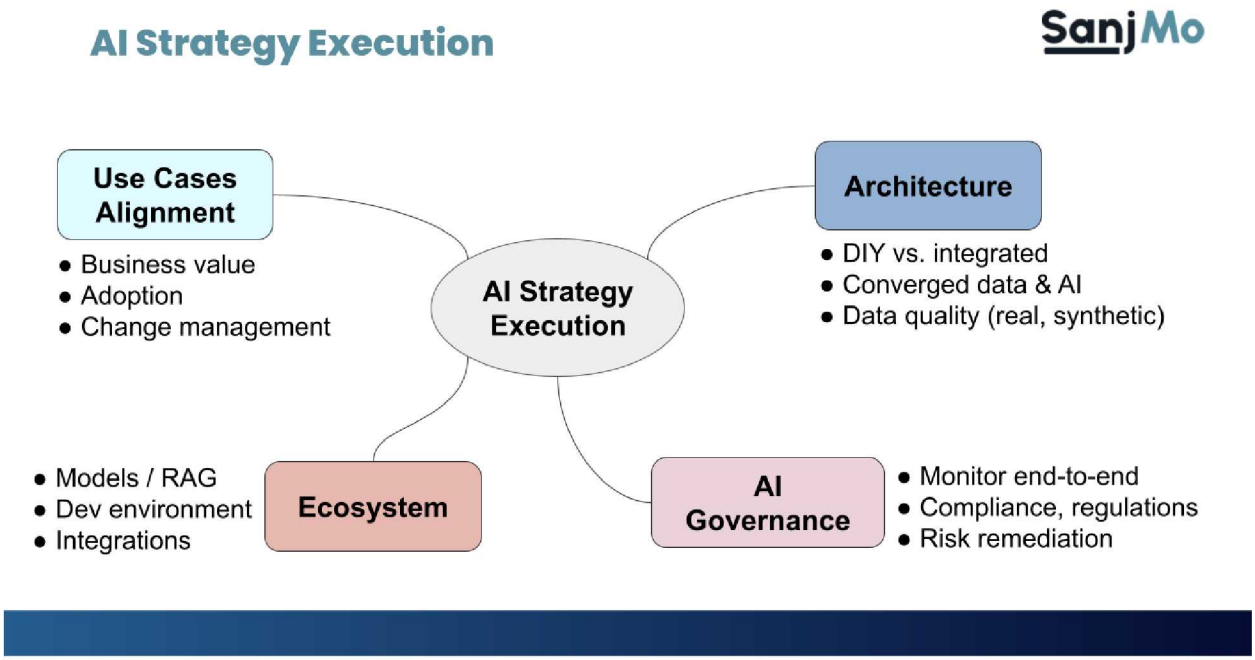

Figure 1: Components of successful execution of AI initiatives

IT architects play a key role in ensuring success of their business AI initiatives. Let's explore each area in more detail.

- **Use Cases Alignment**

  IT architects and business teams should work together to effectively shortlist promising AI projects. This collaborative approach leverages the business acumen of stakeholders to identify areas where AI can deliver value, while also ensuring the chosen projects are technically feasible based on the IT architect's expertise and cost-effective to stakeholders.

  - **ROI and productivity gains**. Use AI to optimize functional areas such as customer success. Take the example of Klarna, a Swedish Fintech company that saved $40M through the use of an AI customer service chat agent and is now taking this success further by launching an AI shopping assistant.
  - **Competitive advantage**. AI use cases can be transformative. For example, in health care, they can accelerate drug discovery, or in retail, they can perform a more situationally appropriate product recommendations that not only incorporates historical profile preferences, but also handles complex matching for color coordination, quality rating, price ranges, brand preferences and local availability without pestering the users for these inputs.
  - **Risk mitigation**. AI provides a deeper semantic understanding of data to identify trends and patterns. Financial Services and retail organizations use AI to detect and reduce fraud and anomalies faster.

As mentioned earlier, the AI governance foundation should be used to track AI use cases.

- **Architecture**

  IT architects should design their AI architecture to build applications that meet the enterprise standards of trustworthiness, reliability, scale and security. AI presents unique challenges compared to traditional data management and software due to complexity risks of using inconsistent or incorrect data when training and serving models. AI does not like data complexity.

  Hence, in designing your IT architecture, the goal should be to bring IT to your data. Not the other way around. Minimize assembling prompt variables from multiple data and AI sources and stop proliferating data silos. Imagine the nightmare of trying to unwind an AI hallucination that was triggered from data values that originates from multiple, probably inconsistent, data sources.

  The "intelligent" data platform spans the entire stack - from using hardware specialized to execute AI workloads on GPUs to AI applications built upon foundation models sourced from trusted data. This single unified store of structured, semistructured and unstructured data may be distributed across a hybrid multi cloud environment, and may need to serve users who are constantly changing their whereabouts. Yes, people move around, therefore your AI workloads should operate from edge to cloud to server. Imagine the quality of interaction when real-time location and situational information can inform your AI functionality.

  Moving up the stack, a disaggregated compute approach that has conventionally used SQL and PySpark is now being augmented by conversational, natural language compute infrastructure that uses foundation models.

  At the top of the stack, where we used to traditionally find reports and dashboards, new AI-driven data products are starting to emerge and these are called AI agents. Imagine in a few years most enterprises and humans will have customized agents, e.g. shopping experience.

- **AI Governance**

  Data governance, sadly, was an afterthought in many organizations. It was only when compliance regulation came into effect, that organizations started to shoehorn governance into rigid architectures and had some painful lessons.

  IT architects should prioritize AI governance and not treat it as an afterthought. The process of AI governance should last the entire lifecycle starting from the identification of candidate use cases documented in a catalog to understanding the accuracy of results delivered by the foundation models. AI governance steps should include:

  - Model evaluation and a mapping of selected models to the prioritized use cases
  - Onboarding of new models in an iterative manner and recording their outputs
  - Assessment and management of risks and risk mitigation of model responses
  - Continuous monitoring of models, their usage and evolution

AI governance is critical to the success of any AI initiative. It can help reduce hallucinations and build trust between businesses and their customers, but it starts with trusting the information that is offered to AI first.

- **Ecosystem**

  The generative AI ecosystem continues to innovate rapidly. IT architects have multiple choices - do-it-yourself (DIY) assembly of multiple best-of-breed tools or choosing homogenous purpose-built and integrated solutions. This is the common tussle between 'build' versus 'buy' and there's no easy answer. Based on the step #1 of the AI strategy execution framework in Figure 1, some use cases may use one approach versus the other.

  A robust ecosystem that spans the entire lifecycle of data - from operational databases to transformation stages to analytical or decision-making stages is needed. The best data platforms are multipurpose ones that allow developers to create robust application functionality without creating excessive architectural complexity that could confuse AI interactions. These platforms should offer plug-and-play access to a plethora of models, because models will change, specialize and evolve rapidly. They should incorporate techniques like RAG and have integrations with external vendors like LangChain, LlamaIndex, HuggingFace, Google Cloud Vertex AI and AWS Bedrock.

  Using open standards like JSON, Parquet, and SQL help future proof technical investments, especially when it is in the fluid space of AI. In addition to lowering the license and skills costs, open standards also gives organizations better control over how their corporate data is used.

If organizations set over-inflated expectations and don't meet them, then we may be in danger of entering yet another AI winter. However, with the proper AI strategy and execution approach, IT architects can help with successful adoption of AI within organizations and make it meaningful and transformative.

Retail Banking Case Study

Let's apply our learnings to see how AI can help retail banking, which expects that 72% of its transactions will soon be digital but their #1 priority is air-tight security and data privacy. This sector has unique challenges as their data is diverse and complex, ranging across transactions, market data, customer data, regulatory, risk, and many other datasets. With 97% of millennials and 89% of consumers using mobile banking apps, scalability, high availability, and performance of the underlying architecture is critical to deliver an excellent real-time usability expectation.

Retail banks recognize they live in constant change and hence, want flexible architecture to meet some of the following use cases:

- Fraud detection and scoring
- E-payments and mobile wallet
- Know Your Customer (KYC) / Customer 360 / Market 360
- Cash Forecasts and Working Capital Funding
- Personal budgeting and expense management

AI-driven architectures can provide these use cases with higher accuracy as they augment existing techniques with a deeper understanding of context. Many retail banks today are developing or fine-tuning models on diverse data that includes relational, JSON, and unstructured data to get a far wider understanding of their customers and products. They are also using retrieval augmented generation (RAG) to customize large language models (LLMs) to their own corporate data and deliver more relevant responses.

The generative AI outcomes and model inferences will be deployed where consumers access data most often-- upon their mobile devices. We expect that in the very near future, all mobile devices will be equipped with GPUs and small language models to make AI ubiquitous. Retail banking will be at the forefront to leverage the power of AI in its servers and on edge devices.

Conclusion

This two-part series has equipped you with a roadmap for navigating the exciting, yet complex, world of AI. We began by emphasizing the critical link between a well-defined AI strategy and its execution. Part 1 explored the foundation of a successful AI journey – understanding your data and crafting a clear, data-driven strategy aligned with business goals.

Part 2 focused on translating strategy into action. We've provided IT architects with actionable steps for execution and wrapped it up by exploring how AI can help the retail banking industry.

By following these steps and maintaining a realistic approach, you can avoid the pitfalls of over-inflated expectations and pave the way for a successful AI adoption journey. Remember, AI is a powerful tool, but its true potential lies in its seamless integration with existing workflows and a commitment to continuous learning and improvement.

SanjMo is an independent analyst firm specializing in delivering insights related to cloud, data, analytics, and AI topics. In the frantic world of rapid development, SanjMo researches new trends and developments to connect the dots and provide pragmatic thought leadership. SanjMo helps elevate awareness to new technologies, brands, and products.

Sanjeev Mohan is the founder and principal analyst at SanjMo. He was a Gartner vice president known for his prolific and detailed research, while directing the research direction for data and analytics. He is a frequent keynote speaker, author of Data Product for Dummies, a frequent blogger on Medium, and the host of It Depends podcast. He can be reached on LinkedIn and at Sanjeev@SanjMo.com.