Whitepaper

Accelerate ROI With Generative AI Lifecycle Best Practices

Set your business up for success and leverage the benefits of generative AI with this guide.

Deconstructing generative AI

While Artificial Intelligence (AI) has been a mainstream consideration of business leaders for some time, the more recent popularity of generative AI has left many with a lot of questions. Leveraging this powerful technology to drive outcomes is a natural next step for many organisations, but they need support in getting started.

Generative AI is a subset of AI technology that can generate content like text, images or other means of human interaction in response to prompts provided. Generative AI uses neural networks to learn from training data that enables the neural network to identify patterns in the source data, which are then used to produce a probabilistic generative response. Large Language Models (LLMs) are the foundation on which generative AI was built; these are algorithms that use deep learning techniques on large amounts of text to make predictions that mimic human intelligence. If you've ever been shown a suggested ending to the sentence you were typing in an email, you've engaged with a version of generative AI.

But how can this technology support business outcomes? From developers to marketers, generative AI can shift manual, mundane tasks from people to machines. Across any business, the volume of work that can be streamlined in this way is immense, saving time and improving productivity along the way. Additionally, existing solutions like chatbots can be enhanced with this technology by simplifying the typical condition response model to a generative model that is easily extensible and produces more humanlike responses.

Since generative AI is evolving rapidly, it's more critical than ever for organisations to be early adopters and fast followers. Those that do will not only improve efficiency but also create a competitive advantage as the optimisation of the technology progresses. This whitepaper can serve as your practical guide to generative AI, supporting your journey to implementation and accelerated ROI.

By the numbers

of leaders want to adopt generative AI **to improve employee productivity.**

say **quality and control are top concerns** around implementing generative AI.

of leaders say their company has already established or is currently developing **internal generative AI policies.**[1]

Outlining the potential benefits

For organisations looking to establish dominance in the market by adopting impactful technology early, generative AI can provide a wide array of benefits. Here are some examples:

Increased productivity

With this technology, employees can offload repetitive manual tasks to AI. They can prioritise more strategic, skill-based work without having to sacrifice the smaller but essential tasks.

Lower company costs

The increased productivity saves employees time and can even help close labour and skill gaps on teams. With increased efficiency, employees can get more done in less time and therefore produce more work with the same amount of effort.

Improved Employee Experience (EX)

Those repetitive manual tasks can be mundane and removing them from the plates of employees can improve their day to day. Additionally, the saved time can be used for upskilling and completing more rewarding work.

Improved Customer Experience (CX)

The involvement of AI means customers will be served quicker, with a higher quality of experience. For example, a chatbot with generative AI will be able to provide superior customer service experience over current solutions that have more limitations.

Mitigation of risks and errors

With the automation of processes, the chance of human error is removed from the equation. AI can also be fine-tuned to improve over time, further reducing errors and the risks associated with them.

Competitive differentiation

Putting all the potential benefits together can easily carve out a competitive edge for an organisation. Better EX results in retained talent, which can mean more innovation for a business. Improved CX can build loyalty and set brands apart from alternatives. The increased productivity and lower costs can also mean reinvestment into new areas of business that competition hasn't tapped into yet.

Characteristics of practical generative AI

Organisations looking to use generative AI should try to capture certain characteristics to increase their chance of success. For practical use across an enterprise, everything from scalability to privacy will be crucial to the long-term ROI and management of a generative AI solution.

Scalability

For enterprise organisations, their generative AI solution will potentially have to process many prompts per second. There are computational demands for deploying and maintaining generative AI, which should be part of the planning process. There are specialised models designed for specific tasks that can reduce that load and boost the scalability of the solution.

Security & compliance

Any time an organisation wants to pursue a new data or AI solution, security and compliance need to be top of mind. While there are standards in place right now, there is a possibility for policies to change in the future as this technology gains more mainstream use. Additionally, concepts like responsible and ethical AI should be part of the conversation.

Portability

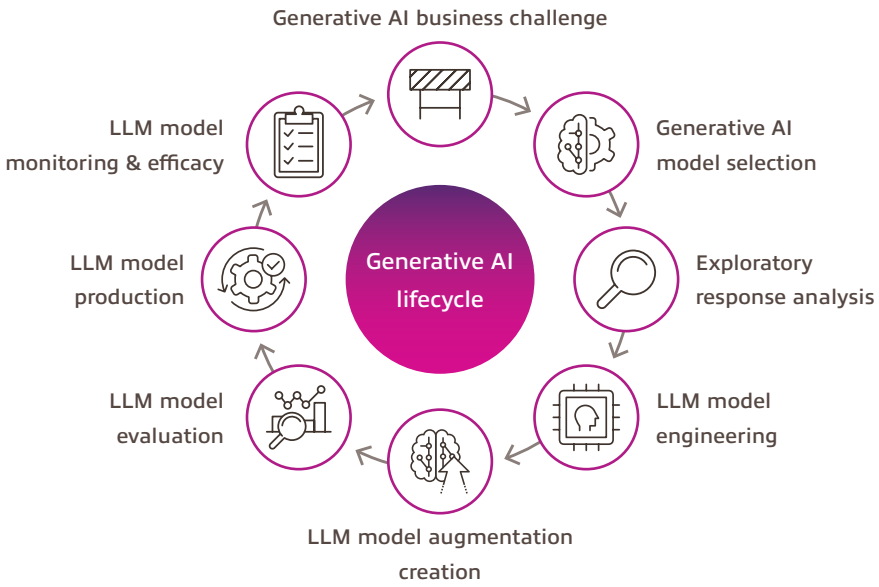
Generative AI solutions are portable, and the techniques that are used to enhance your organisation's specific solution can be used across any LLM domain. If training your own model, it can still be copied and redeployed in another environment of your choice. Additionally, investments in the data ecosystem such as data governance can be leveraged across further generative AI solution deployments. Therefore, the effort put forth to augment even one generative AI solution is easily extensible.

Privacy

For some organisations, privacy may be a concern as open-source generative AI solutions will store inputs and use them to improve future results. Considerations need to be made and an appropriate solution identified for information that is not public, is Personal Identifiable Information (PII) or is considered intellectual property. For example, your organisation could use a model designed for non-public information, train its own model, or leverage a containerised version of a generative AI model.

Best practices: A lifecycle for generative AI

No matter the industry or use case, this lifecycle can guide the process. The lifecycle is iterative; there may be times when an organisation must revisit certain steps or start the process over as their generative AI needs evolve. To establish a generative AI Centre of Excellence (CoE), business leaders should start with this expert-developed journey from idea and development, to deployment, management and governance.



Generative AI business challenge

First and foremost, determine the business goal(s). What does your organisation want to accomplish? Part of this process is ensuring the desired outcome aligns with what an AI solution can realistically solve. Additionally, there needs to be a way to measure the success of generative AI's impact on the goal.

Generative AI model selection

Once your organisation has defined its business goals with generative AI, it should start exploring what model will make the most sense for the use case. While training your own model is on the table, a pretrained model will be suitable for the vast majority of applications. Picking a pretrained model also means saving the time and expenses required to do the heavy computational work of training a model. For those situations where an organisation wants extensive control of the model, training its own will be the best way to accomplish this.

There are two types of pretrained models: general purpose and specific use. As the names imply, general-purpose AI models are created to learn and do a variety of tasks, while specific use is focused on one task.

Some examples of popular pretrained models include GPT-4, LaMDA and Titan. While pretrained models are a viable choice, decision-makers should keep in mind the age and types of data that have been used to train the model. For example, many popular pretrained options are only trained on data before 2021, which may limit usability or require augmentation. Determining the right model for your use case might be something that requires expert guidance — and leveraging a strategic partner at this point can streamline the process.

Ethical considerations are important — but should not be a deterrent to leveraging this impactful technology. Working with a solutions provider can give you access to guidance in these areas that streamline deployment and usability.

Ethical considerations of generative AI

Bias: For pretrained models, there isn't a definitive way to know what data has been used to develop it; therefore, the amount or type of biases can be harder to counteract. However, even for enterprise-trained models, great care needs to be taken to ensure unintentional bias isn't present.

Fairness: This term describes the concept that the data used to train a model should be representative of the population that will be using it. Mitigation of fairness concerns can be harder to achieve with complex data and potentially impact model performance.

Exploratory response analysis

Once a model is selected, there should be an assumption that the model isn't perfect. What additional information will you need the AI to have in order to be valuable for your use case? While a base model has been chosen, it doesn't have all the context it needs to be truly practical for your business needs. Determine what collateral it can leverage to support answering the business problem. It's important to ensure the collateral is sufficient and good quality since it will be the information the AI uses to generate responses.

LLM model engineering

With the previous two steps completed, model engineering is what puts them together. At this point, a technique will need to be selected that will support using this AI to solve the business problem. Organisations aren't restricted to only picking one of the techniques, and it may be beneficial to employ multiple.

- **Prompt engineering:** refinement of the model to ensure the expected prompts produce the desired outputs
- **Knowledge embedding:** imparting relevant information into the model that will be needed to solve the business problem
- **Transfer learning:** leveraging knowledge learned from one task to improve the completion of another
- **Fine-tuning:** uses what's already part of an existing model to start training a new portion into the model. This requires a smaller amount of data and computational resources.
- **Model retraining:** providing new data to an existing model to improve performance as it relates to the defined business problem

LLM model augmentation creation

Once a technique is chosen, your organisation can focus on applying it. Training internal teams who will be using the AI solution should also begin at this stage. Ensuring teams understand prompting and what the AI is capable of will make their use more valuable. With the model being set for use, the people who will be using it need to understand how they can do so effectively.

LLM model evaluation

This step is where the generative AI model will be tested. After all the previous decisions have been made and implemented, did they all come together to help address the needs determined in the first step? Part of this evaluation is determining to what extent the model addresses the needs. For example, things like efficiency and the amount of time saved should be considered at this point. If there are gaps that can be closed to improve these aspects, they should be prioritised. Once the model is determined to be fit for use, the next step can be pursued.

LLM model production

Once the model has proven to be ready for business use, it's time to release and go to production across the enterprise. At this point, security safeguards will take shape, such as restricting access solely to teammates who should be using it. Other IT and management considerations will plug in, including a framework for Continuous Integration/Continuous Deployment (CI/CD) pipeline release of the generative AI model.

LLM model monitoring & efficacy

After deployment, the organisation should keep a pulse on the efficacy of the model. As time goes on, there should be a way to measure if the solution continues to provide value, and if not, a way to identify what's changed. Additionally, monitoring can occur where prompts and their responses are analyzed to determine where the model may be failing to meet needs. Over time, this process can even be automated to streamline the improvement process of the model.

If gaps have been identified, though, the whole point of the lifecycle is to be able to revisit it. If the model selected earlier turns out to be less effective than an alternative, then an organisation can pick back up at that step. On the other hand, it may turn out that the identified business problem was not defined correctly to be addressed by generative AI and needs to be reevaluated. Or, if an entirely new use case is imagined, then a new lifecycle can start for that business problem. The cycle remains fluid, and generative AI should be recognised as an iterative process for best results.

Generative AI in action: Use cases

Regardless of industry, many will leverage generative AI to augment common business practices in departments like finance or marketing, or for solutions like chatbots. The real-world applications of generative AI are ultimately less about the industry of the business and more about what the business wants to get out of the technology. This will fall into one of five different approaches of generative AI. Let's unpack each:

0-shot

- Suited for simple tasks
- Uses a single clear prompt, without examples
- **Example:** an AI capable of recognising the category of an image without seeing the image beforehand, such as being able to identify an image of a cat that the AI has never seen before

Few-shot

- Suited for simple tasks but will need examples provided
- Uses a prompt that provides examples to guide the LLM
- **Example:** an AI capable of applying the examples to a new context, such as providing example images of dogs and then asking if an image of a fox is a dog

Fine-tuning

- Suited for tasks that require specialised skill or style
- The model will need to have been trained on data that relates to that specialised skill or style
- **Example:** an AI that has been trained on a company's sales emails is capable of generating new sales emails that are a similar style

Embedding

- Suited for tasks that require knowledge of specific, up-to-date or non-public information
- The model combines the relevant information with LLM capabilities

Build your own model

- While this is the most expensive option, it is suited for situations that require a highly customisable solution

Different organisations will have varying needs, and within a single organisation, there may be a variety of needs identified that AI can help address. For example, a business may deploy a few-shot approach for data validation within its Revenue Operations (RevOps) teams — but require fine-tuning for developer tasks. Regardless of the approach, the provided lifecycle can guide organisations to impactful generative AI use.

Use cases across the enterprise

	Sales	Marketing	Finance	Operations	IT	Legal	HR
Find	Customer insights	Trend analysis	Fraud detection	Predictive maintenance	Network security	Contract analysis	Resume screening
Summarise	Sales reports	Social media monitoring	Risk analysis	Quality control	Performance monitoring	Compliance monitoring	Employee feedback
Create	Product recommendations	Ad copy	Financial statements	Supply chain optimisation	Code optimisation	Legal briefs	Onboarding materials
Code	Sales forecasting	Campaign optimisation	Investment strategies	Production planning	Bug detection	Contract generation	Performance metrics

InsightGPT: Our journey to internal generative AI

Looking to empower employees while safeguarding internal data and intellectual property, Insight enabled access to a private, secure instance of ChatGPT. Named InsightGPT, it has the same capabilities as the public version but retains the content and prompts that employees enter into the platform. Across the organisation, it is used to support productivity and save teammates time as they complete their typical job duties.

Accelerate enterprise generative AI with Insight.

Insight helps clients assess their current AI landscape, build AI ideation and data-backed AI solutions, and provide supporting infrastructure — all to deliver exceptional and in-demand business outcomes.

AI Workshop	Accelerate AI	Operationalise AI
Our AI workshop offers a range of benefits for participants, empowering them to explore and harness the creative potential of AI. **The Insight AI workshop is available in two formats:** Generative AI: focused on how to use Generative AI in the enterprise context. AI Journey: focused on all the possible AI solutions that can be implemented, from cognitive Services to Machine Learning (ML).	Our team of experts analyses your existing practices across the entire AI lifecycle, identifies key opportunities for improvement, lays out a step-by-step roadmap to apply best practices to your teams and programmes, and accelerates your AI success.	Empowering your AI journey from visionary concept to seamless execution, we transform ideas into operational excellence

Take control of your organisation's generative AI future.

Capture the value of this fast-moving technology and gain a competitive edge. Our experts can provide your business with a practical, streamlined strategy for implementing your generative AI solution.

Connect with us

Driving innovation with digital transformation

At Insight, we help clients enable innovation with an approach that spans people, processes and technologies. We believe the best path to digital transformation is integrative, responsive and proactively aligned to industry demands. Our client-focused approach delivers best-fit solutions across a scope of services, including the modern workplace, modern applications, modern infrastructures, the intelligent edge, cybersecurity, and data and AI.

Learn more

1] Insight Enterprises. (2023). Beyond Hypotheticals: Understanding the Real Possibilities of Generative AI.

Insight uk.insight.com