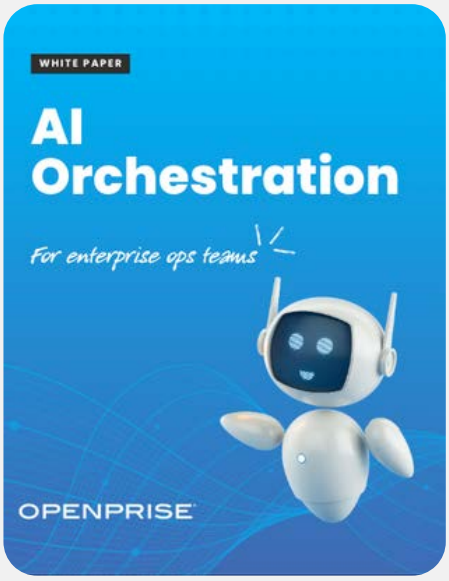
WHITE PAPER

AI Orchestration

For enterprise ops teams

OPENPRISE®

AI Orchestration



AI orchestration makes AI ready for reliable Ops deployment

If you love technology, the last 2+ years may be the most exciting period in our lifetime.

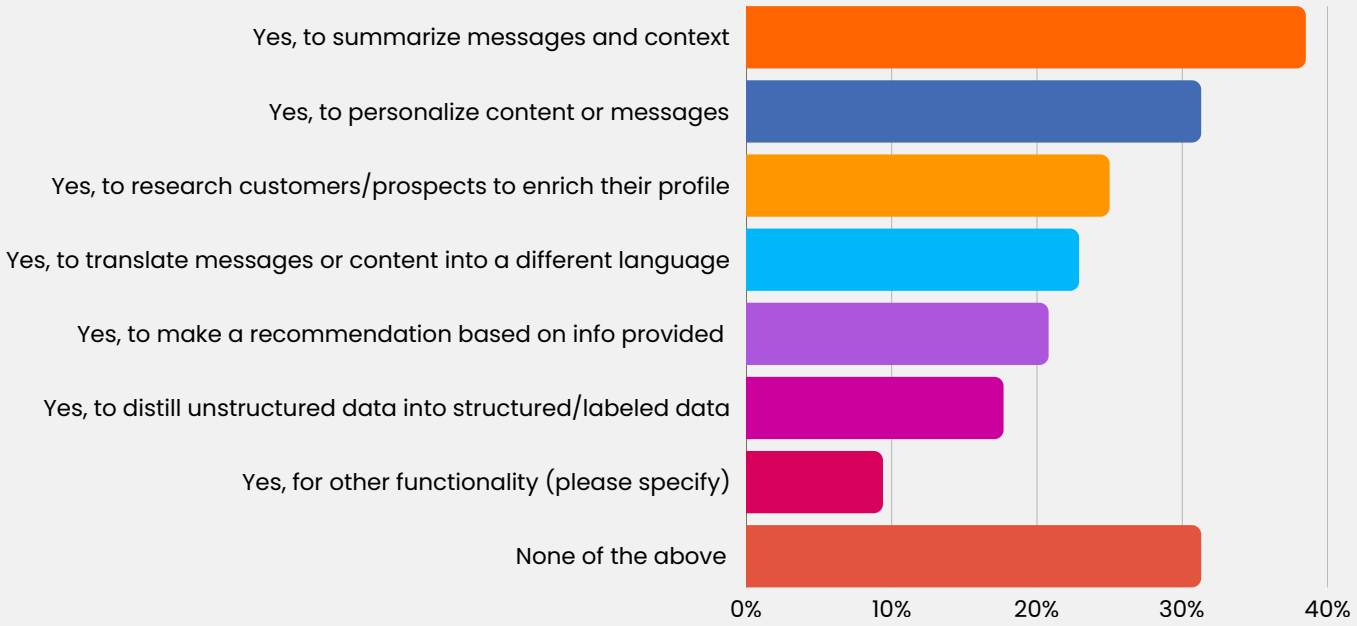
Generative AI (GenAI) came on the scene with a big bang and has evolved rapidly since. As fast as the technology has advanced, the hype around AI has advanced even faster, with some optimistic pundits predicting artificial general intelligence (AGI) and artificial superintelligence within the next two years, followed by robot assistants in every household.

State of AI adoption in enterprise operations

Company boards and management are framing AI adoption as both an existential threat and a market disruption opportunity. The mandate to adopt AI in the enterprise is only increasing. So where does the Ops world stand in 2025 when it comes to AI adoption? Scott Brinker (chiefmartec) and Frans Riemersma (MartechTribe) released The State of Martech 2025 with results that paint the picture below:

Are you using LLMs or agentic AI in any of your marketing workflows or automations?

Source: 2025 AI & Martech Stack Survey, chiefmartec & MartechTribe



The survey results closely match what we have observed in the market. There is much activity around experimentation, proof of concept, and pilots.

These projects share the following characteristics:

- These are limited automations that requires human in the loop
- They are low-risk and non-mission-critical use cases
- Most do not use data from core enterprise systems like CRM
- Solution performance is rarely quantified, often judged purely on the “wow factor”
- Project ROI is still mostly theoretical and not substantiated with hard data
- Solution cost is poorly understood, not measured and modeled

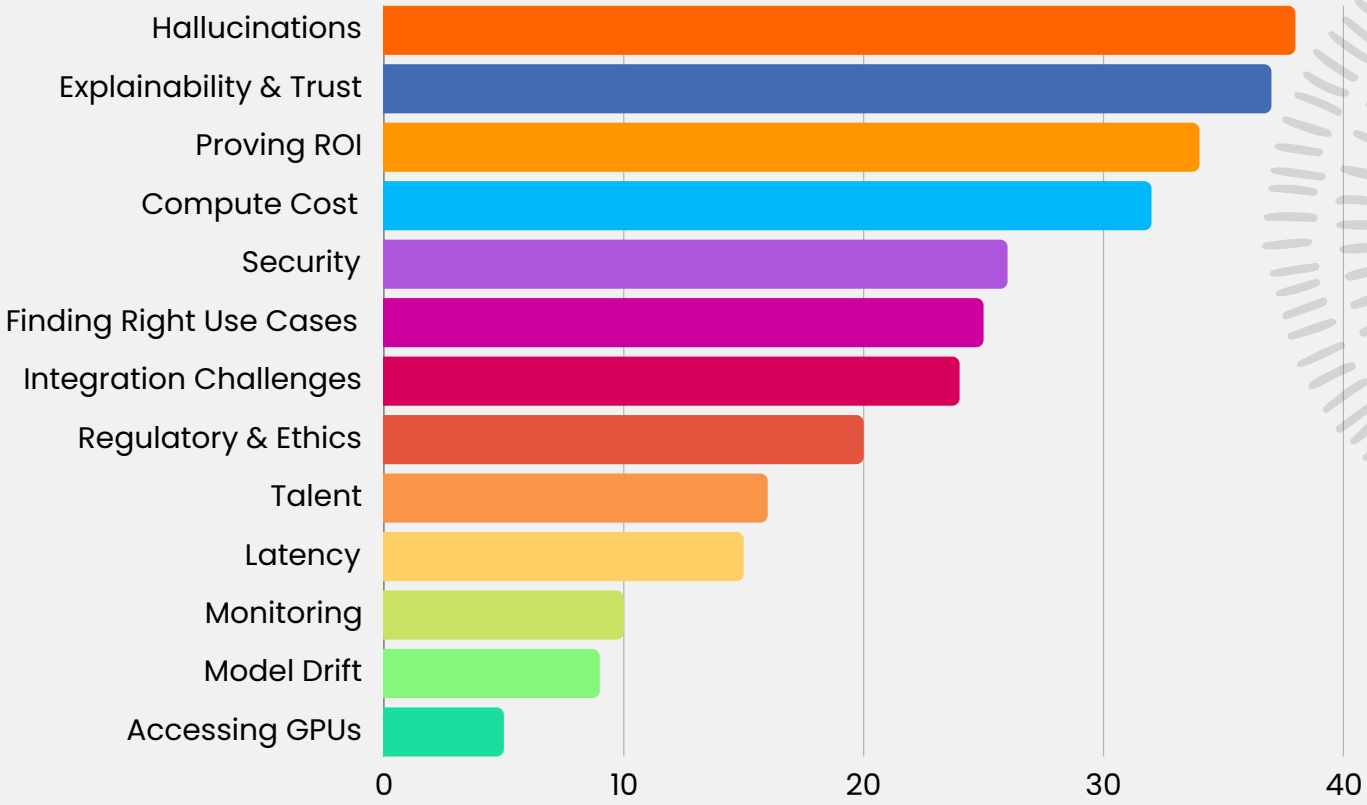
This is further backed by a recent survey by ICONIQ, a leading growth equity firm, on the challenges their portfolio companies face when it comes to adopting AI

While AI is new, adoption challenges for new technology are as old as the invention of electricity and steam engines. All the classic technology adoption challenges still apply to AI, but AI also introduces some new adoption challenges we will discuss.

Top AI Model Deployment Challenges

Percentage of respondants ranking each challenge in their top 3

Source: ICONIQ 2025 State of AI Report | N=273 companies



Barriers to AI adoption in Ops

These are some of the many barriers that have impeded progress toward using AI in mission-critical enterprise applications:

Security roadblocks (InfoSec embargoes)

One of the key takeaways from the usage survey data above is the **lack of use cases that involve sensitive data from core systems like CRM**. This is often because IT, specifically Information Security (InfoSec), has embargoed the sending of sensitive data to AI vendors. If IT does not provide an approved alternative, either an enterprise license or an internally developed solution, GTM use cases will be largely limited to content generation and web data gathering.

Without dedicated resources to assist with AI adoption progress has been slow and uneven

Lack of support: the figure-it-out-yourself approach

Management mandate to use AI often doesn't come with any assistance. This not only leaves progress up to employee initiative to learn and experiment, it also makes the erroneous assumption that AI is easy to figure out. Anyone who has spent time working with AI will tell you that getting AI to perform to your exact specifications is not a trivial task and requires more technical understanding than what AI marketers would lead you to believe. **Without dedicated resources to assist with AI adoption, it's no wonder that, in most companies, progress has been slow and uneven**.

Cutting through the hype

Few technologies have been hyped as much as AI. AI as it stands today is already amazingly powerful and future possibilities are even more exciting.

However, **AI is still not a miracle technology that is good at or can replace everything**. Part of the current learning and experimenting phase is to separate hype from reality. You may have to develop ten pilots to figure out that only two are good use cases for AI, but that is part of the learning process when it comes to new technology. Gaining a good technical understanding of AI's capabilities can help with focusing on use cases that are likely the best fit and primed for success.

Integration challenges with existing tech stacks

Few, if any, companies will be able to replace existing technologies with AI wholesale in the next five to ten years, if ever.

AI adoption means injecting AI capabilities into existing processes and integrating with existing tech stacks. This leads to integration challenges and change management that can be complex and time consuming, especially in larger enterprises

The challenge of "good enough"

Before the introduction of GenAI, if you put good data into a system, you'd get predictable good results. If you put bad data in, you'd get bad results. GenAI introduces a unique challenge: **even if you put good quality data into AI, it will output 20% garbage, seamlessly mixed in with 80% good stuff, due to hallucination and other challenges.**

This limits AI's application to processes that can tolerate this level of error. Getting a good handle on what that 20% error is, how it can be improved on, and whether it is a viable solution for a particular business process is not trivial.

Gaining a good technical understanding of AI's capabilities can help prime you for success.

Trust is the missing ingredient

For a team to deploy AI into mission-critical processes, it needs to develop trust in the technology, and AI has not earned that trust yet. **This white paper will describe a framework and the technologies necessary for the Ops team to provide the necessary support infrastructure around AI so they can develop the necessary trust in AI.** We call this framework and associated technologies **"AI orchestration."**

20%

percent of garbage outputed by AI despite good quality data

Understanding AI's limitations

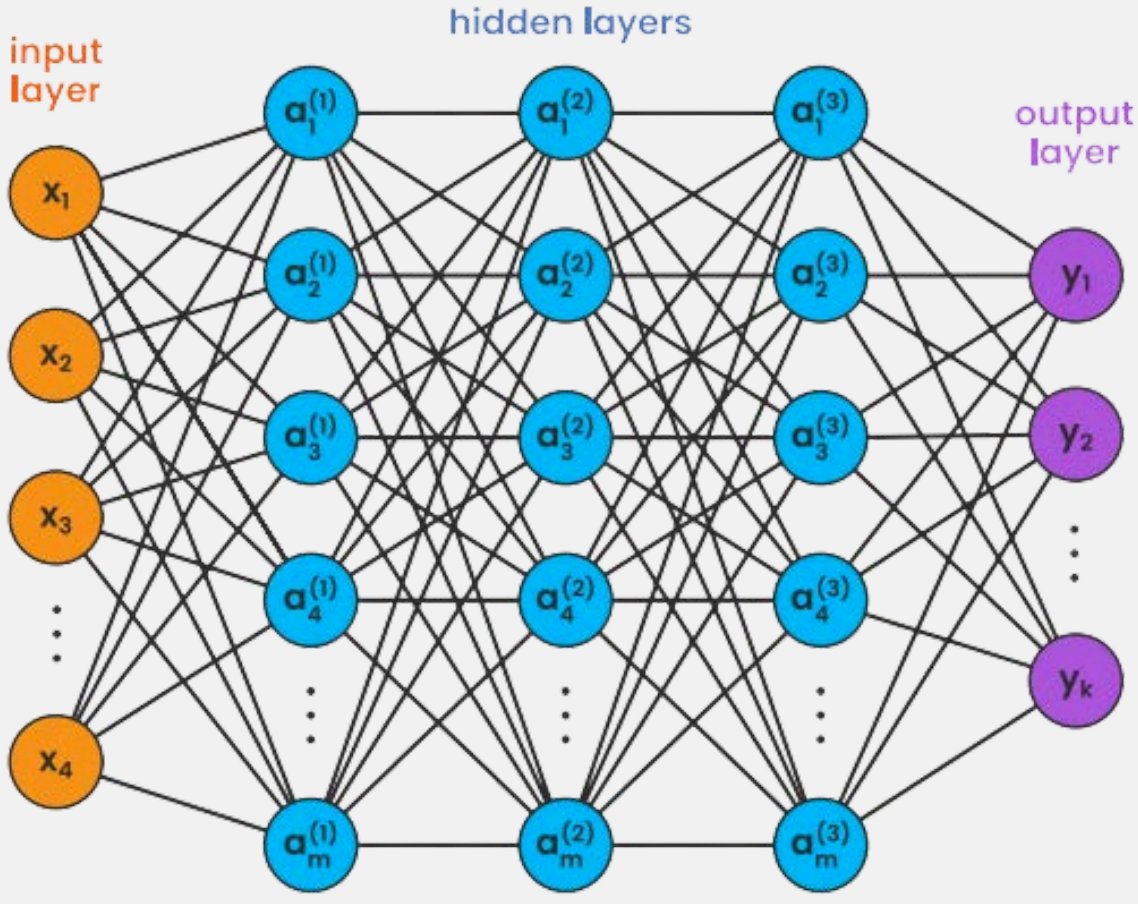
In order to understand what AI needs to be successful in mission-critical enterprise environments, we must understand the technology's limitations first. This is by no means a data science academic paper, so we will cover only the absolute essentials with the least amount of technical jargon and the right level of abstraction.

Language models 101: how neural networks work

Today's GenAI is based on neural network technology invented in the 1940's. According to Google's AI Gemini:

Neural networks are a type of machine learning model inspired by the structure and function of the human brain. They are made up of interconnected nodes called neurons that process information and learn from data.

The diagram below illustrates the basic concepts of how a neural network operates



Our description here is for a language model—the LM in LLM.

Tokenization

Your input text is broken down into tokens. Each token can be a word, part of a word, or a punctuation, so there are always more tokens than there are words in your prompt. For basic understanding, it's sufficient to equate a token with a word. Each token is then encoded into a set of numbers, or a vector for the mathematically inclined. This is the orange input layer on the left in the diagram.

Transformation

This numerical vector goes through a sequence of transformations through the nodes of the neural network. This is the blue section in the diagram. The extent of transformations is dictated by the number of nodes and the parameters associated with the nodes. The number of parameters is often used to indicate the size of the neural network, thus the size of the model. The state-of-the-art "frontier" large language model (LLM) now has more than one trillion parameters. A small language model has around ten billion parameters.

Output

The final transformation, the purple output layer in the diagram, outputs a list of words with associated probabilities that the model believes is the best next word. A word is chosen based on its probability to be the next word.

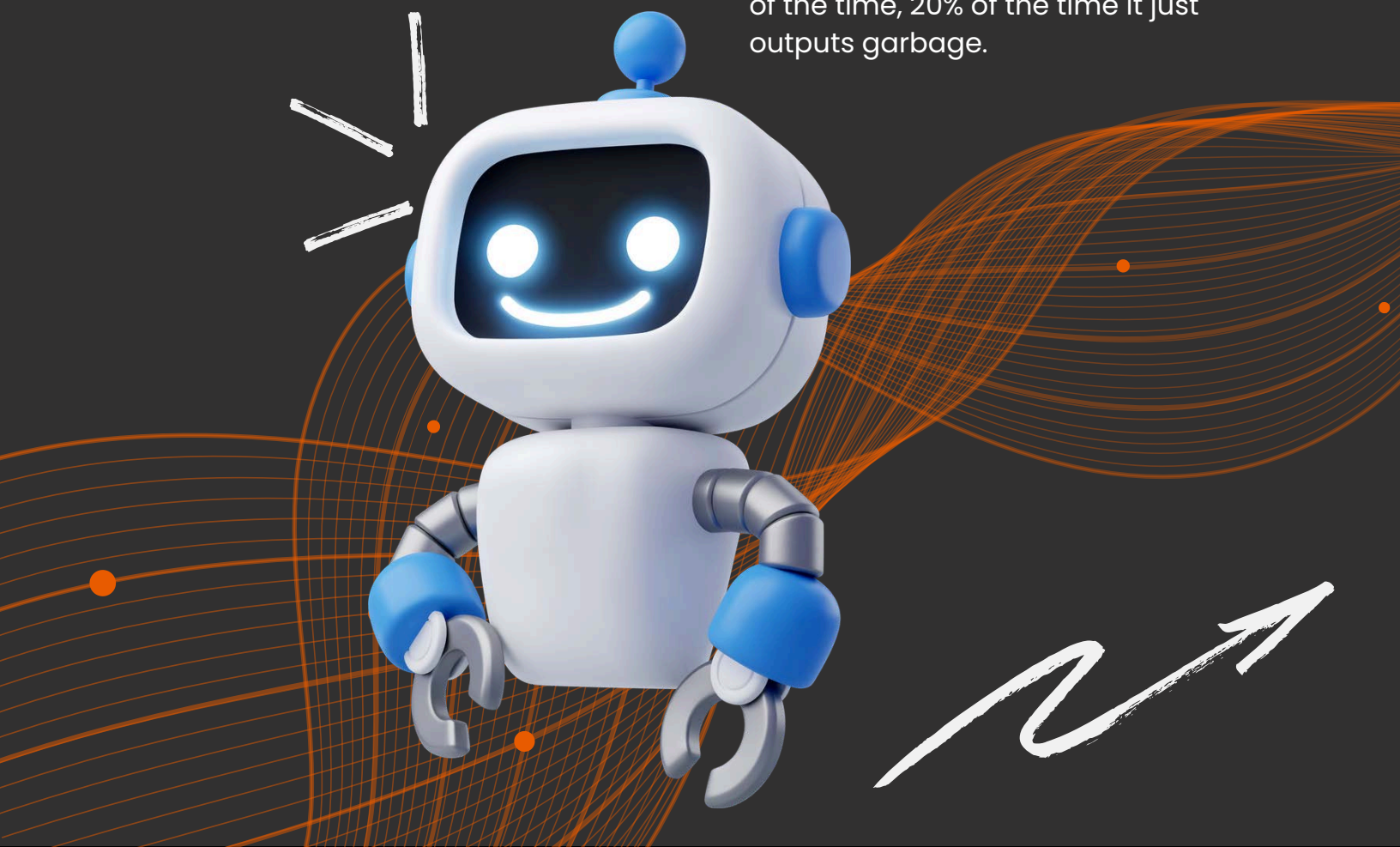
This basic process is run over and over using different strategies to generate each of the words that forms the text output to your prompt. Although it's a gross simplification, it's useful to think of LM as autocomplete on steroids.

What today's language models can and can't do

Now that we have a basic understanding of the underlying technology, let's first discuss the known limitations and tendencies, then how to create a supporting infrastructure to mitigate them.

AI may be the most data-intensive application humans have ever invented. If you feed AI bad data, you get bad results from AI. This is the classic garbage-in, garbage-out problem that is not new or unique to AI. Providing high-quality data to AI is key to achieving any level of success.

The GIGO idiom implies that if you put quality data into a technology, you get good results out. Assuming a technology is designed and tested correctly, that implication was usually true—until GenAI showed up. With GenAI, even if you put quality data in, you still get some garbage results out, and that outcome cannot be eliminated. While GenAI can perform amazing feats 80% of the time, 20% of the time it just outputs garbage.



What today's language models can and can't do

AI's hallucination problem

It's impossible to talk about language models without talking about their tendency to hallucinate, or to be more technically precise, bullshit. If you haven't read the technical paper ChatGPT is Bullshit, we highly encourage you to give it a read. Not only is it as fun as a technical paper can be, it's extremely informative on what hallucination is, how to think about it correctly, and why bullshit is a more accurate description of this "feature" than hallucination. **If you need to figure out how to deal with the consequences of AI hallucination, you need to understand the nature of the beast.**

While some progress has been made on reducing hallucination, it has been minimal. **Techniques such as retrieval augmented generation (RAG) can also help reduce hallucinations.** Studies such as this one from Open AI on Scaling Laws for Autoregressive Generative Modeling shows there is a theoretical limit, often referred to as the Compute Efficient Frontier.

The TL;DR is that **hallucination can never be zero since GenAI is a probabilistic technology, and driving hallucination down further involves exponentially greater cost and diminishing returns.**

In short, hallucination is not going away anytime soon, if ever, so we must learn to deal with it.

> If you need to figure out how to deal with the **consequences of AI hallucination,** you need to understand the nature of the beast.

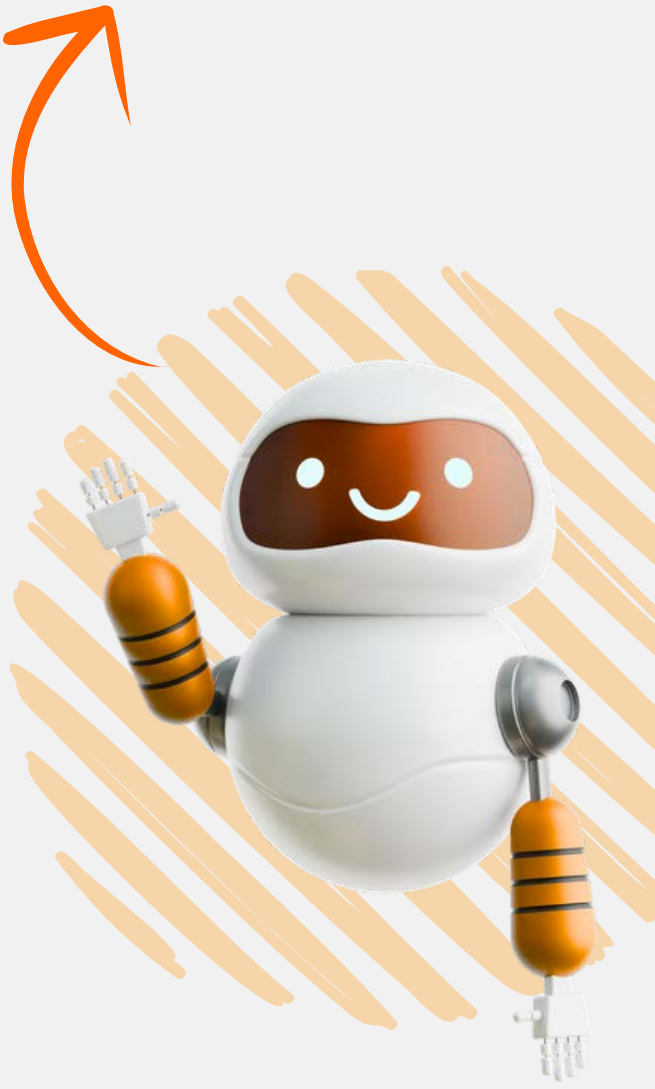
AI does not always follow instructions

While technically this has the same root cause as hallucination, the nature of the problem it creates is different. Hallucination manufactures inaccurate data. However, **even when GenAI generates the correct data, it often doesn't output the data in the format instructed.** This creates data quality issues for downstream systems and workflows.

Even when GenAI generates the correct data, it often doesn't output the data in the format instructed.

AI cannot explain itself; it's a black box

Neural networks are a black box technology, so even the model makers cannot tell you why exactly a specific output is generated. **Techniques such as chain-ofthought try to simulate reasoning**, but be very clear that what is being marketed as reasoning is not what you think reasoning should be, not the way animal brains perform reasoning. Yes, this is an area of debate, but it is the stated position of Yann Le Cun, chief AI Scientist at Meta and one of the three "godfathers of AI."



Instead of showing the symbolic reasoning for how the output is derived, which is not how neural networks work, AI shows how it generates the words that would sound like the most convincing explanation. The consequence of this is that **you cannot rely on the reasoning provided by AI as a means to validate the accuracy of the output.** So not only the final output can be garbage, the reasoning provided is mostly garbage.

Defensive countermeasures: intentional bad data

Another cause of bad AI output data is defensive countermeasures. As more people build AI agents to gather the data they want, that creates costs and lost revenue for the resource owners they are pulling the data from. To mitigate the cost or to protect their information asset, these resource owners will deploy countermeasures that at best will block access, but more likely provide fake data so the agents will happily go away instead of having to deal with persistent attempts.

A great example of this is Cloudflare, a leading cloud infrastructure provider, announcing a new feature in March that will feed fake data to AI agents and crawlers that do not respect a website's no-crawl robot.txt policy.

We are at the start of an AI technology arms race between the parties seeking information using AI and the owners of information seeking to protect their assets and profits. So now, in addition to unintentional bad data from hallucination, you also have to deal with intentional fake data from defensive countermeasures, not to mention bad actors who want to mislead.

We are at the start of an AI technology arms race between the parties seeking information using AI and the owners of information seeking to protect their assets and profits.

Choosing the right AI use cases

No technology is good for everything and GenAI is no exception. The odds of a successful outcome for every technology project are largely determined by the selection of the right technology before the project even starts.

Choosing the right use cases

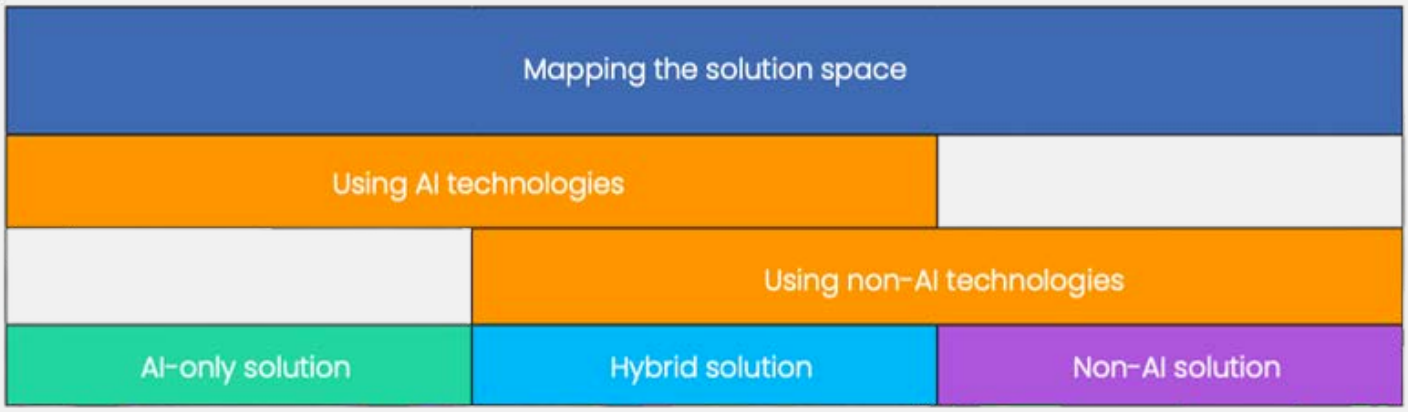
Every use case has its own risk tolerance profile. On one extreme, you have processes like Know Your Customer in the banking industry. These mandatory processes are dictated by law; thus, there is zero tolerance for deviations and errors. **AI agents would not be a good choice for a well-defined process with such low risk tolerance.**

On the other end of the risk tolerance spectrum is the automation of outbound campaigns, using AI agents to crawl the web for data that can be used to personalize cold emails. **Outbound email open and response rates are so low that, even if AI's research data is 20% incorrect, it will likely not materially degrade your response rate.**

The average quality of personalization from human written cold email is so low that the errors in your AI generated emails may not even stand out. If you can get some deals out of the campaign, the ROI may be well worth it, AI error and all.

The average quality of personalization from human written cold email is so low that the errors in your AI generated emails may not even stand out.

AI technology is suitable for some use cases, non-AI technology is suitable for some use cases, and sometimes combining AI and non-AI technologies give you the best outcome:



To determine if AI is the right solution and how big a role AI should play in combination with non-AI deterministic technologies like workflow, use the following decision tree. Hybrid means a solution using a combination of AI and non-AI technologies.

<table>
<tr><td colspan="7">Can the use of AI improve the use case's performance?</td></tr>
<tr><td colspan="6">Yes</td><td>No</td></tr>
<tr><td colspan="6">What is the use case's tolerance for errors?</td><td rowspan="6"></td></tr>
<tr><td colspan="3">High</td><td colspan="3">Low</td></tr>
<tr><td colspan="6">Can errors be detected in real time and at acceptable cost?</td></tr>
<tr><td>Yes</td><td colspan="2">No</td><td colspan="2">Yes</td><td>Yes</td></tr>
<tr><td rowspan="2"></td><td colspan="4">Can errors be detected later and is the feedback time acceptable?</td><td rowspan="2"></td></tr>
<tr><td>Yes</td><td>No</td><td>Yes</td><td>No</td></tr>
<tr><td>Hybrid</td><td>Hybrid</td><td>AI Only</td><td>Hybrid</td><td>Do not use AI</td><td>Hybrid</td><td>Do not use AI</td></tr>
</table>

How to build trust in AI

To use AI to automate mission-critical enterprise use cases, Ops must be able to trust the technology can perform. What does it mean to be able to trust AI?

How to build trust in AI

Comply with InfoSec policy

Compliance with InfoSec policy is never optional. Failure to comply can expose the organization to serious security vulnerabilities, leading to both financial and reputational damages. Make sure:

- Your solution uses only authorized AI services, whether external, internal, or embedded models and agents.
- The commercial solution you use is not using "shadow AI," which means the solution is using a third-party AI service in the back end.

Protect against prompt injection attacks

AI is still so new that few attack vectors have been developed, but it's already starting, such as the recently published EchoLeak against Microsoft 365 Co-pilot and Remote Prompt Injection in GitLab Duo. AI is the first technology where the data and the instruction are not separated, thus making securing the instruction (traditionally code) extra difficult. To protect against prompt injection, you need the ability to:

- Have tight control over the construction of prompts that include dynamic data.
- Extract only specified data from AI responses without exposing the entire response, which may be malicious.

Provide high-quality input and training data

AI is a highly data-driven application. While it may sometimes tolerate lowerquality input data, ensuring high-quality data for both training and runtime (inference) is essential to achieving reliable performance. Classic garbage-in, garbage-out applies.

Ensure accurate outputs

Few, if any, companies will be able to replace existing technologies with AI wholesale in the next five to ten years, if ever. AI adoption means injecting AI capabilities into existing processes and integrating with existing tech stacks. This leads to integration challenges and change management that can be complex and time consuming, especially in larger enterprises

Manage acceptable levels of risk

AI's garbage-out problem is not going away anytime soon, so making AI productive in the enterprise requires managing the risks it introduces. This involves:

- Benchmarking the performance achievable and sustainable by AI
- Defining the performance requirement from the business
- Actively monitoring and managing AI's performance

AI adoption means injecting AI capabilities into existing processes and integrating with existing tech stacks.

Ensure consistent performance

How AI achieves what it can today is still a bit of a mystery. We know many factors can impact AI's performance but there are as many—or more—that we don't understand or can't consistently observe. AI's performance can drift based on these known and unknown factors. To maintain consistency, you need to monitor AI's performance as more data enters the system, the model learns over time, and usage patterns evolve.

Demonstrate ROI

Most organizations are in the experimental and pilot phases of their AI journey. Only a small percentage of pilot projects will receive the green light to move into production. A key decision factor for whether a project is simply a science experiment or an enterprise solution is the ROI it delivers to the organization. To build the business case and sustain funding, you need to measure the performance and impact of the AI solution.

A key decision factor for whether a project is simply a science experiment or an enterprise solution is the ROI it delivers to the organization.

Manage costs

A key component of ROI is cost. A solution can deliver tremendous value but if it also comes with an equally large price tag, the ROI may not pan out. The way AI works today with prompts and tokens, it can be very difficult to model and track cost.

As prompts get longer and more sophisticated, techniques such as chain-of-thought become more widely used, and third-party tools are consumed by agents, the ability to model and manage the cost of the AI solutions can determine the feasibility and degree of adoption within the enterprise.

Making AI work in a hybrid world

It is impossible to predict what AI will look like or be capable of in the next ten years, but one thing any experienced Ops can predict with certainty is that AI will have to live in a messy, hybrid enterprise technology landscape.

Beyond the fact that no single technology can solve every problem, there are many practical reasons and constraints why this will remain the reality as long as humans run enterprises.

This creates the need to orchestrate a portfolio of AI and non-AI solutions. Here are some of the most common reasons why it will always be a hybrid world.

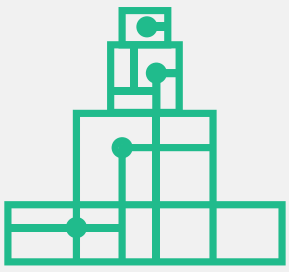
Making AI work in a hybrid world

The rise of specialized AI solutions

While foundational models continue to improve, **specialized solutions have already begun to emerge and will continue to proliferate across content, use cases, verticals, and technologies**; for example, coding, writing marketing copy, and providing customer support. Even within coding, there are specialized solutions for code generation, review, debugging, and testing

Optimizing cost by technology mix

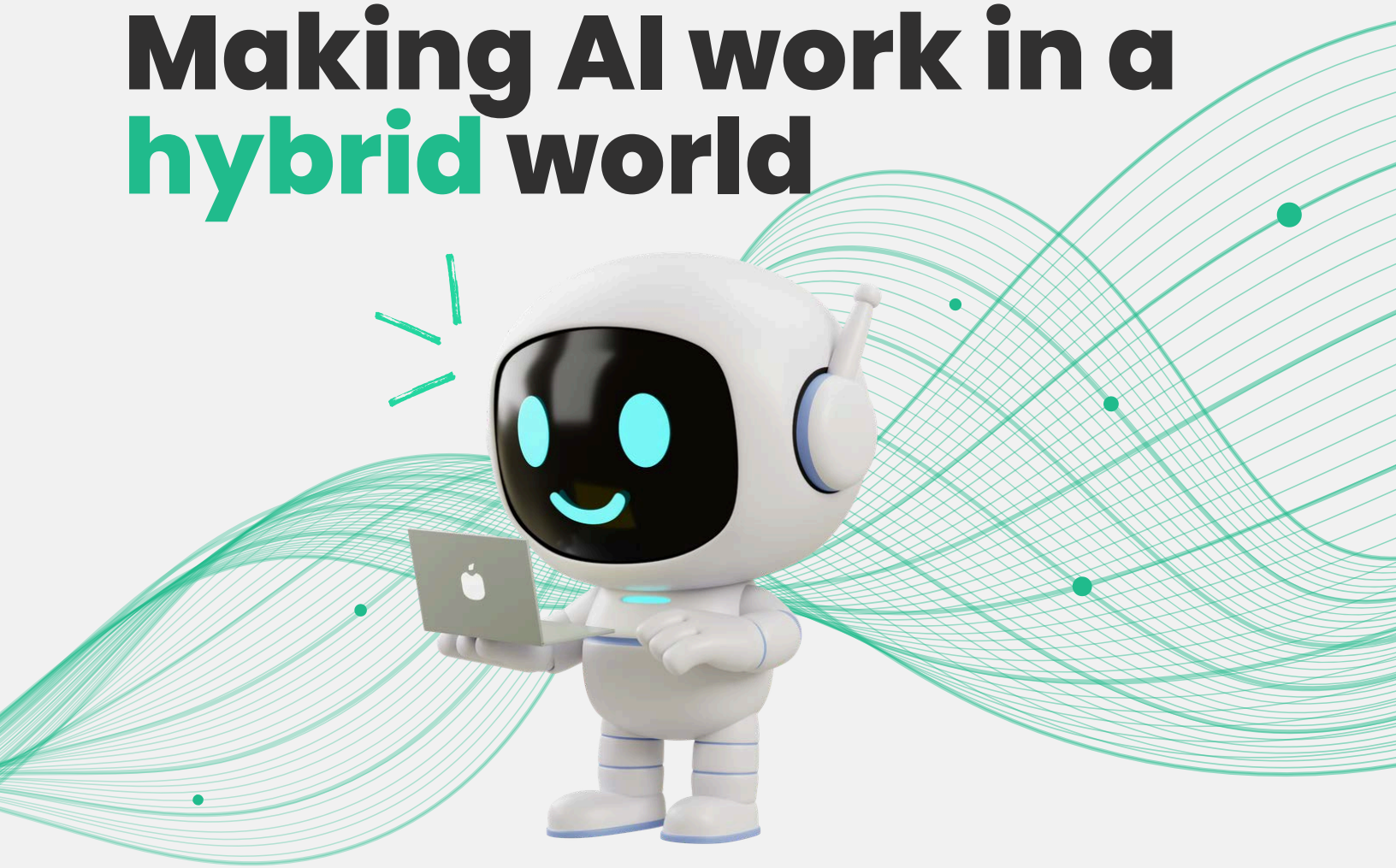
Running AI models is still very expensive, and that's with most AI services still heavily subsidized by either investors or corporate parents. **Costs may rise sharply in the future even with smaller models getting more capable.** To optimize cost, different tasks will be allocated to AI or non-AI technologies, large or small AI models, and services from different vendors.



Geopolitical constraints on AI adoption

AI technologies will likely be subject to some geopolitical constraints. For example, some U.S. organizations may ban the use of DeepSeek and Qwen from China, and China may ban the use of Open AI and Llama from the U.S. **Even if it is not a complete ban, there may be limitations on what data can be used with each service.**

Making AI work in a hybrid world



Security and compliance require compartmentalization

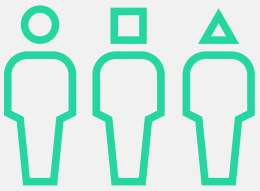
While feeding one AI model with all your enterprise data may sound tempting, the reality is that enterprise data often need to be compartmentalized due to security best practices and compliance mandates.

For example, mixing customer data with employee data would be nearly impossible to justify in most enterprises. AI introduces additional concerns, as it is trickier to control how AI shares data.

With non-AI technologies, you can usually define specific rules, policies, and filters to block sensitive data from exposure, protections that are pretty much bulletproof. With AI, however, once data is embedded in training or provided in real time, it can materialize in outputs in infinite and unpredictable ways.

There is still no bulletproof way to ensure AI will not share sensitive data, either accidentally or as a result of deliberate prompt attack.

Making AI work in a hybrid world

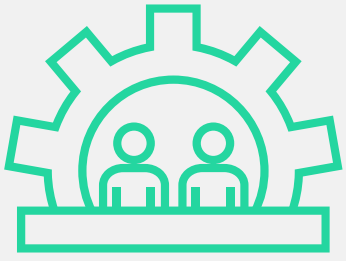


Corporate silos won't disappear

Enterprises will always have silos across business units, departments, regions, etc. Different groups will have their own budgets, priorities, and level of autonomy, which will lead to technology silos, and AI will not change that the slightest bit.

Change management hurdles and legacy integrations

Technology changes in an enterprise can take a long time due to obstacles related to change management and compatibility with legacy systems. These obstacles lead to uneven adoption and technology silos.



The messy mix of AI deployment options

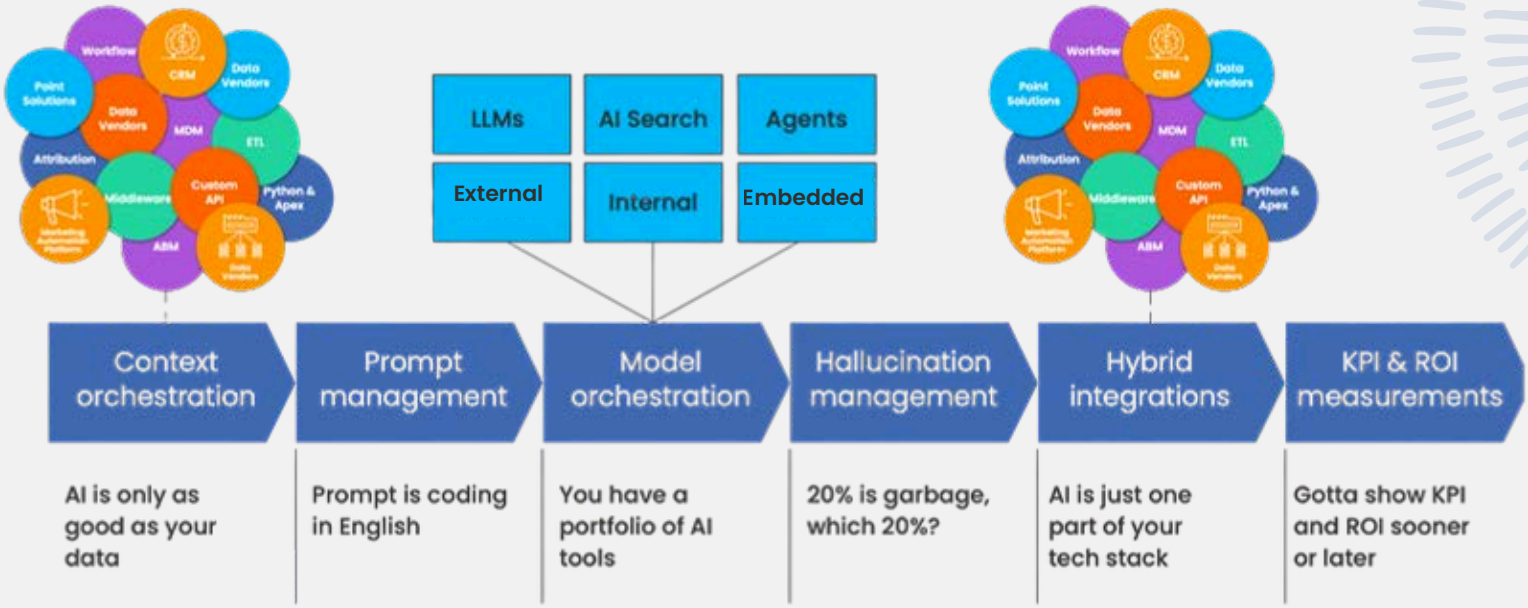
At the end of the day, the enterprise technology landscape is always highly heterogeneous and quite messy, so different versions of AI from external, internal, and embedded options will likely exist over time and will need to be orchestrated.

The AI orchestration framework

The irony about AI is that it is designed to behave like humans, so one of the consequences is that it needs to be supervised and supported like a very smart junior employee if you want to fully leverage its capabilities.

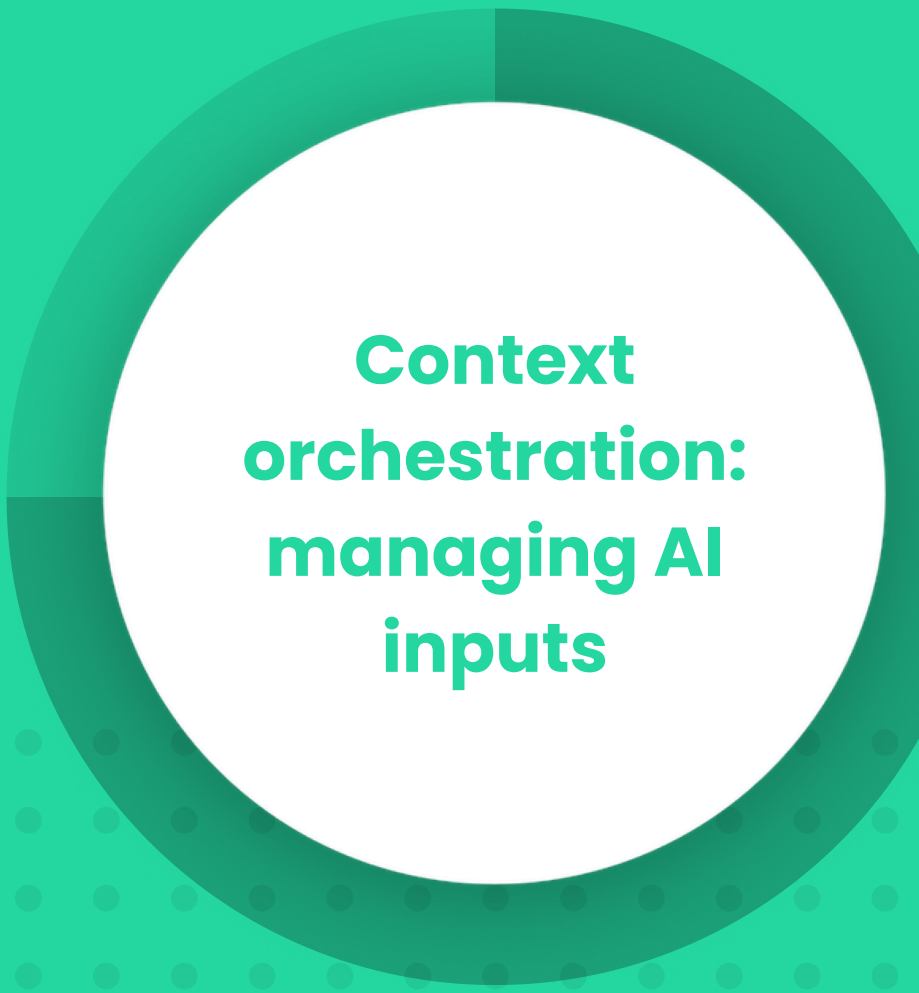
The AI Orchestration framework

Just as you would pair a junior employee with a seasoned employee and review the intern's work product before sharing it, you will need to provide AI the necessary support in order to use it in mission-critical, lights-out enterprise automations. We call this set of support capabilities "AI orchestration."



Deploying AI with the support of AI orchestration is how you gain the necessary trust that AI can perform and scale to its potential. AI orchestration consists of six key capabilities.

The AI Orchestration framework: part 1



1. Context orchestration: managing AI inputs

An AI prompt consists of two parts: instruction and context

Context is the data portion. As we collectively figure out how to get the most out of AI, experts are now learning that the quality of the context is even more important than the instruction part of the prompt. This is so important that context engineering has now surpassed prompt engineering as the hot AI skillset.

While context engineering is about more than data quality, it starts with data quality. Enterprise Ops use cases involve enterprise data. AI can be less sensitive to some data quality issues such as non-standardized country names, but having the best possible data quality is still key to AI success. Data quality is an extensive subject in its own right so we will not attempt to cover it in this white paper. You can find an extensive amount of quality content on data quality on the Openprise website.

AI consumes two types of data: training data and prompt data. Quality of training data is extremely important, but it's more of a batch process performed by data engineers and scientists, not within the purview of Ops and context of Ops automation, so here we will focus on prompt data (inference time) only.

> **The quality of the context is even more important than the instruction part of the prompt.**

1. Context orchestration: managing AI inputs

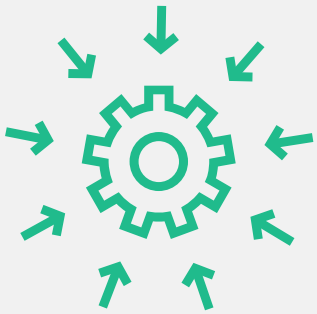
Managing limited context windows

Each AI model has a context window, which is the amount of data you can prompt the model with, including history and additional data payloads. Smaller, less expensive models typically have smaller context windows. Any prompt text beyond the context window will be trimmed. This is why AI may start to "forget" details during long-running chat sessions. Part of context orchestration is ensuring your prompt does not exceed the context window limitation. When using AI via API, there are additional API limitations that need to be managed as well.

Part of context orchestration is ensuring your prompt does not exceed the context window limitation.

Reducing prompt injection risks

Prompt injection is a cybersecurity exploit where malicious inputs are crafted to manipulate language models into producing unintended or harmful outputs. Attackers exploit the model's inability to distinguish between valid and malicious instructions within the prompt. Malicious instruction text can make its way into a data field in the CRM. When text stored in that field is subsequently inserted into a dynamic AI prompt, the prompt becomes compromised. Careful data preparation can greatly reduce the risk of prompt injection.



Meeting compliance requirements

There are many security, compliance, and business requirements that mandate data be compartmentalized, anonymized, or obfuscated even if only used by machines. There is also heightened sensitivity about how context data will be used by both models and model vendors. These compliance requirements will mandate that prompt data be carefully prepared.

Controlling inference costs

AI has a fairly high running cost, known as inference cost. Most services charge by tokens. Longer prompts, deeper reasoning, and more detailed responses all use more tokens and therefore increase cost. By carefully preparing and selecting the data that goes into a prompt, you can keep your prompts efficient and costs low.

Reducing prompt injection risks

Prompt injection is a cybersecurity exploit where malicious inputs are crafted to manipulate language models into producing unintended or harmful outputs. Attackers exploit the model's inability to distinguish between valid and malicious instructions within the prompt. Malicious instruction text can make its way into a data field in the CRM. When text stored in that field is subsequently inserted into a dynamic AI prompt, the prompt becomes compromised. Careful data preparation can greatly reduce the risk of prompt injection.

Avoiding confusion

Early common wisdom suggested giving AI as much data as possible and letting it figure out what's relevant. Recent studies now show that AI can get confused and perform sub-optimally if your context is not well engineered. Common issues include:

- Context with too much data
- Context with irrelevant data
- Context with data at the wrong level of granularity
- Context that presents logical dependencies in anything but "forward" order

The TL;DR is that **hallucination can never be zero since GenAI is a probabilistic technology, and driving hallucination down further involves exponentially greater cost and diminishing returns.**

Speed and throughput limitations

AI is fast compared to manual operations, but slow by automation standards. Typical transactional API calls respond in less than a second and logical steps within a workflow are often order of magnitude faster. AI responses, however, can easily take a few seconds, and if you're using reasoning or chain-of-thought techniques to improve accuracy, AI responses can run into minutes. This slower speed can rule out AI applications for time-sensitive use cases.

Careful context orchestration ensures AI is not wasting time dealing with dirty or excess data so you can maximize speed and throughput.

hallucination is not going away anytime soon, if ever, so we must learn to deal with it.

The AI Orchestration framework: part 2

2. Prompt managment

Using AI in automation means building prompts dynamically by combining:

system prompt + user prompt + first party data

Beyond dynamically stitching prompt text together at runtime, AI orchestration also needs to provide supporting management functions.

Managing system prompts and specialized models

Models can be customized for specific use cases, usually by tailoring a system prompt with general instructions and role descriptions to improve AI's ability to perform specific tasks. While users could copy and paste that portion of the prompt for every single task using a generic model, **it is more scalable to create a catalog of specialized models, each with unique system prompts.** This also follows the best practice of "segregation of duties" to further ensure system integrity and security.

Managing user prompt templates

Writing a good prompt is as much an art as it is a science. **Once a good prompt has been developed, the team should have easy access to it so other team members don't have to reinvent the wheel or use suboptimal prompts.** AI orchestration helps to manage the building, testing, sharing, and versioning of user prompts.

Templates and tools for better prompt development

Prompt engineering is essentially programming using natural language, so it is not surprising that the process of developing a good prompt is similar to the process of writing code. **AI orchestration provides domain- and use case–specific prompt templates so users do not have to start from scratch**, as well as tooling to make prompt engineering and testing more efficient.

The AI Orchestration framework: part 3

3. Model orchestration

As established earlier, the one certainty about AI's future is that it will exist in a hybrid and heterogeneous technology landscape where a collection of AI technologies will have to be orchestrated alongside non-AI technologies to achieve lights-out automation. This means AI orchestration needs to provide the capabilities to manage the following:

Coordinating models and agents

Most of us in Ops already use AI in multiple ways today, including:

- General-purpose and specialized models
- General-purpose and specialized agents
- AI-powered search

As use cases proliferate and become more sophisticated, expect more specialized, platform-specific offerings (e.g., AgentForce from Salesforce) and use case specific solutions (e.g. order-to-cash). A single process will likely utilize multiple AI technologies end to end. This means AI orchestration needs to provide broad integration support for:

- Services (e.g., OpenAI and Gemini)
- Standards (e.g., Model Context Protocol)
- Open-source projects

Supporting multiple deployment types

Just as most enterprise cloud deployments today are some sort of hybrid cloud, AI will also be hybrid, combining commercial AI services, internally developed solutions, and AI embedded in infrastructure or middleware. AI orchestration needs to support all these deployment models.

> **The quality of the context is even more important than the instruction part of the prompt.**

Integrating AI with non-AI technologies

Not every automation task is best handled by AI, so **AI orchestration must seamlessly integrate AI technologies with other automation technologies** such as:

- API / iPaaS / Enterprise Service Bus (ESB)
- Workflows / Business Process Management (BPM)
- Data pipelines / ETL / Reverse ETL

Robotic Process Automation (RPA)

Building composable AI with controlled autonomy

While there is no single agreed definition of an AI agent, the common understanding is **a language model–based technology that does not require explicit step-by-step instructions, can use available tools and resources, and determines what needs to be done to achieve stated objectives.** The exact sequence of tasks executed or tools used may vary, and are up to the discretion of the agent.

However, different business processes have different requirements for repeatability, observability, and risk tolerance. **Automation solutions must have the flexibility to meet different levels of requirements. This means an AI solution with the maximum degree of autonomy can't be the only option.** Commercial AI agent solutions will likely evolve to accommodate configurable levels of autonomy. At the same time, Ops architects will build composable solutions that use a combination of AI technologies to meet process requirements; for example, integrating workflows with direct model access to build an "agentic" solution with less autonomy than commercial agents.

Not every automation task is best handled by AI, so AI orchestration must seamlessly integrate AI technologies with other automation technologies

The AI Orchestration framework part four

4. Hallucination management

Until model developers can eliminate hallucination, the only way AI can be used for mission-cricital enterprise use cases is if there are practical, economical ways to validate and remediate AI responses. This is currently the toughest challenge we see. There are a few options where AI orchestration can help automate validation and remediation tasks.

Real-time validation using non-AI tools

The best option, if available, is using non-AI options to validate AI's output in real time. Here are two simple examples:

Use case 1: AI is used to process out-of-office autoresponse emails and extract the email address and phone number of the person being delegated to.

- **Hallucination problem** - AI can respond with a fake email address crafted from a person's name and an example email address given in the prompt.
- **Detection** - Check if the email address is contained within the email body text.

Use case 2:
AI is used to segment job title into a selected list of job functions

- **Hallucination problem** - AI can respond with a job function not in the list of acceptable job functions.
- **Detection** - Compare AI's job function response against the approved list.

the only way AI can be used for mission-cricital enterprise use cases is if there are practical, economical ways to validate and remediate AI responses.

Unfortunately, automatable definitive validation is often not feasible, especially for data acquisition and research use cases. A less-than-ideal option is to perform checks that may not provide conclusive validation but can improve confidence that the AI response is likely accurate.

Here are two examples:

Use case 1 - AI is used to research and find a person's email address.

Hallucination problem: - The email address found may be fake.

Detection: Check the deliverability of the email address using a service. If the email is deliverable, while it's no guarantee that the email address is accurate,

- it is less likely that it's fabricated from whole cloth.

Use case 2 - AI is used to research and find a company's address.

Hallucination problem: The address found may be fake.

Detection: Validate the name and address using Google Places. If the address is valid then it's less likely that AI hallucinated. If the business name from Google Places is at least similar, then it's likely the AI result is accurate.

> **Automatable definitive validation is often not feasible, especially for data acquisition and research use cases**

Using AI to validate AI

Theoretically, **it is possible to use AI to validate AI.** Until more concrete solutions and case studies emerge, however, this is still mostly a theoretical approach. There are two possible methods:

Two orthogonal approaches

If a framework or methodology exists to evaluate the quality of certain types of work, AI can be trained to perform assessments using that framework. Since the first AI is used to generate the work and the second AI is trained to validate the work with orthogonal objectives, this can potentially work well.

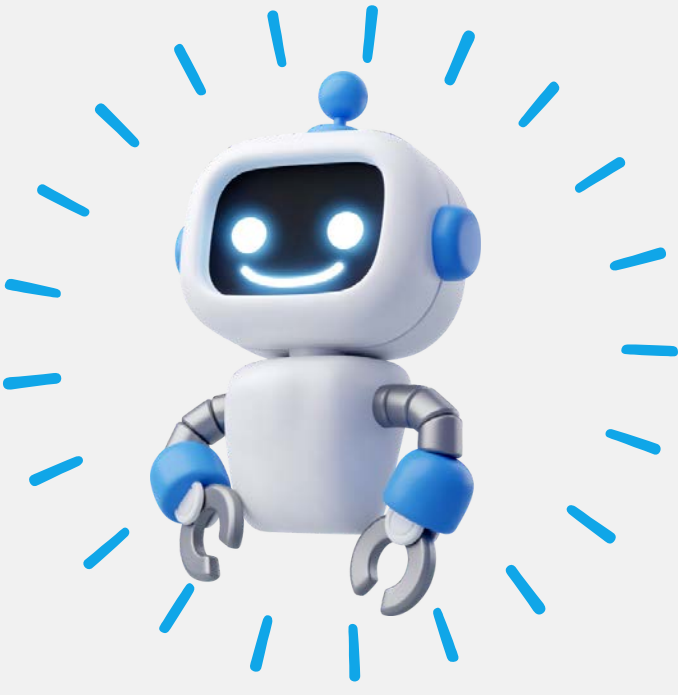
Find consensus among AIs

You can **ask more than one AI service to perform the same task and compare the results**. The level of consensus reached can boost confidence in the result. One significant downside of this method is the increased cost and time. There is also the risk that all the services share the same built-in bias because they are trained on overlapping data sources—in many cases, the entire internet. Models trained on the same data are more likely to exhibit the same bias and produce the same erroneous results. This method should perform better if you can find models trained on different datasets.

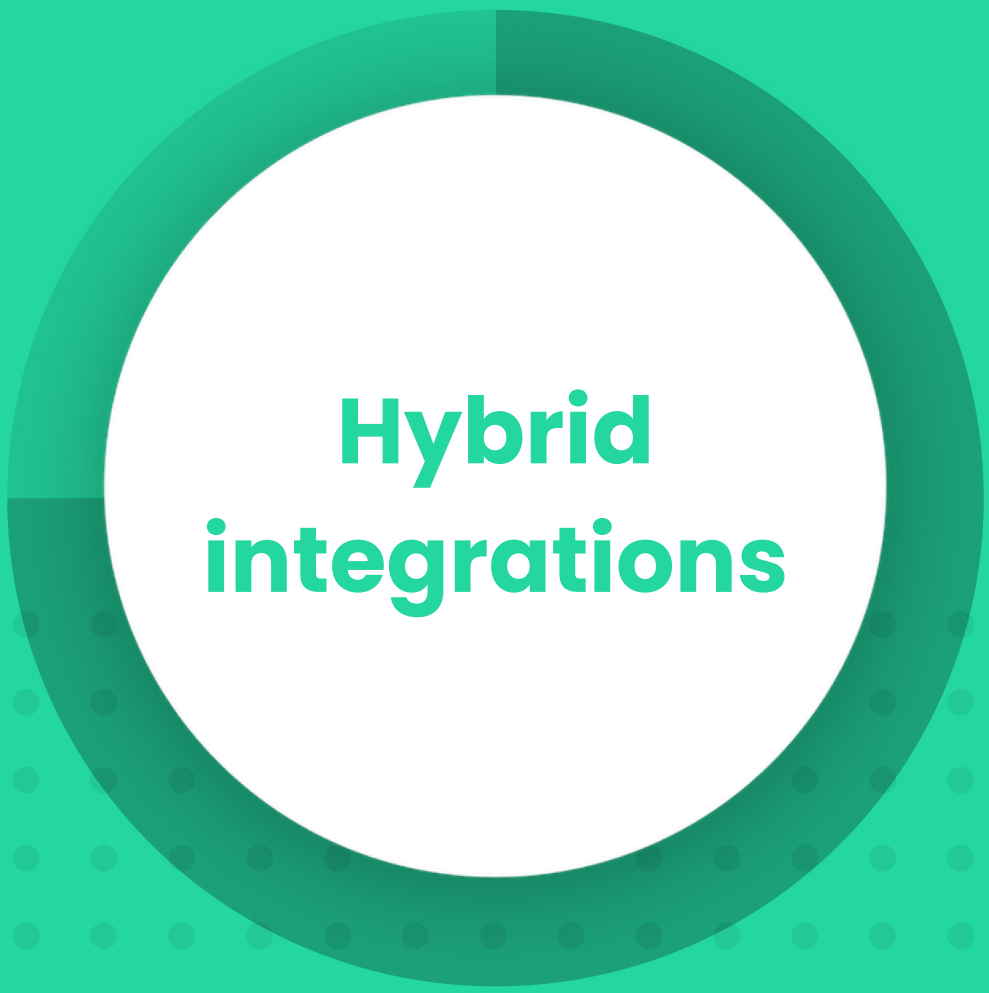
Using feedback loops in real time

If real-time validation is not feasible, **consider whether non-real-time feedback loops may be acceptable**, or at least better than no feedback loop at all. Back to the email address example above: if the deliverability check is either positive or inconclusive, you may decide the confidence is high enough to keep the data and run campaigns against it.

To provide long-term feedback on AI performance, you can **tag how the email address was acquired and run performance reports after a few campaigns**. Bounce and click data can then provide valuable feedback on AI performance and offer guidance for optimizing prompt design or validation criteria.



The AI Orchestration framework part five



5. Hybrid integrations

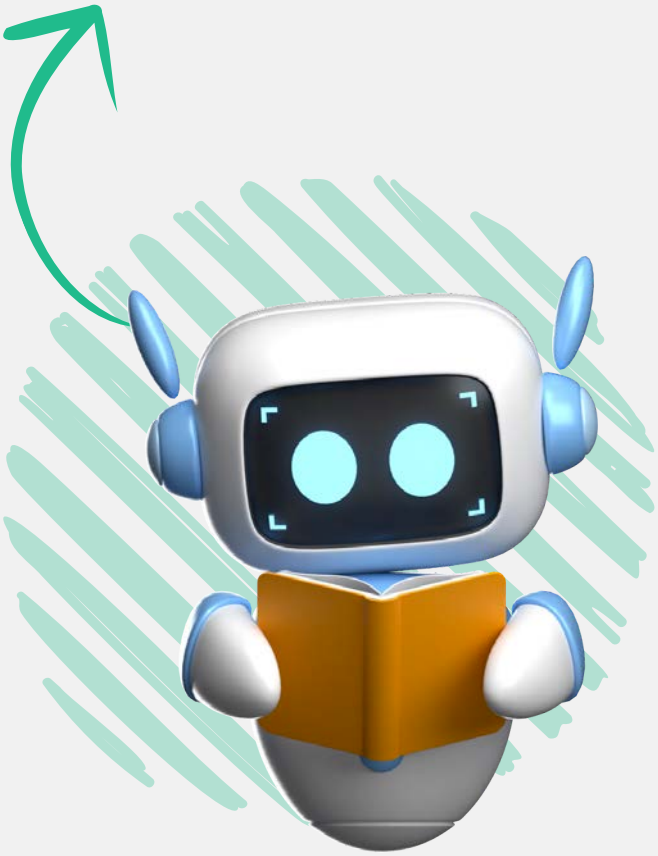
Even if all enterprise technologies could be replaced by AI, which is highly unlikely, the great replacement will not happen overnight. Just look at how much legacy technology your organization is still using: some of it may be well over a decade old. AI will need to work with non-AI technologies, either in a designed hybrid architecture, or simply to accommodate legacy technologies

Today's AI predominantly produces unstructured or semi-structured data, and has a tendency to ignore directions and just do whatever it chooses. Because of this, **AI orchestration must process AI responses and operationalize them to be compatible with downstream systems and workflows**. This can include:

- Extracting structured data from unstructured data
- Extracting data from JSON-formatted responses
- Standardizing values to match a system's picklist values
- Transforming data formats to match a system's formatting requirements
- Integrating with existing systems

While it may be an option to use AI to perform these tasks, it is likely not the best choice, due to AI's lack of consistency and because non-AI technologies can often perform these tasks faster and at a fraction of the cost.

AI will need to work with non-AI technologies, either in a designed hybrid architecture, or simply to accommodate legacy technologies.



The AI Orchestration framework part six

6. KPI and ROI measurement

This is probably the least intuitive aspect of AI orchestration. So why do you need to measure AI's performance?

Why you need measurements and benchmarks

The first reason you need performance measures and associated benchmark data is to **determine whether AI is a good fit for your use case or capable of delivering improvement over your existing solution**. Say you have a use case that requires 95% accuracy and your existing solution delivers 97% accuracy. If benchmark data shows that, for this type of use case, AI achieves 80% accuracy at best, you may decide either:

AI is not a viable solution, or
you need to explore how to boost AI's performance before making the final decision.

The second reason is to **know what "good enough" looks like.** With non-AI deterministic technologies, once the proper QA is done, you usually get fairly predictable and consistent performance. So it's usually not difficult to recognize when your development effort has reached a good enough point and you can stop. Writing AI prompts is different. Due to the probabilistic nature of the technology, the still-unpredictable nature of prompt engineering techniques, and the ease with which you can tweak and experiment with the prompts, it's easy to go down a rabbit hole and not know how much more effort should be invested.

In other words, **without concrete performance measurements, it's hard to know when to stop.**

The third reason ties into the recurring theme of this paper: **how do you build trust in AI? The answer is, by providing hard performance data.**

The fourth reason is that if you're using a learning model, **measuring performance is essential to understanding if the model is learning as expected.**

> Due to the probabilistic nature of the technology, the unpredictable nature of prompt engineering techniques, and the ease with which you can tweak and experiment with the prompts, **it's easy to go down a rabbit hole and not know how much more effort should be invested.**

Toolkit for measuring and benchmarking AI

So what does the AI-OSS need to provide to enable measurement and benchmarking? **You need test data, a test rig, prompt templates, and a reference benchmark for each prompt or use case**.

Take job title segmentation as an example. The toolkit would include:

- 100 test records of job titles
- A prompt template to perform job function and level segmentation
- Information on the model used and relevant settings
- The AI output
- Validation results and overall score

With this toolkit, **you now have a baseline performance benchmark.**

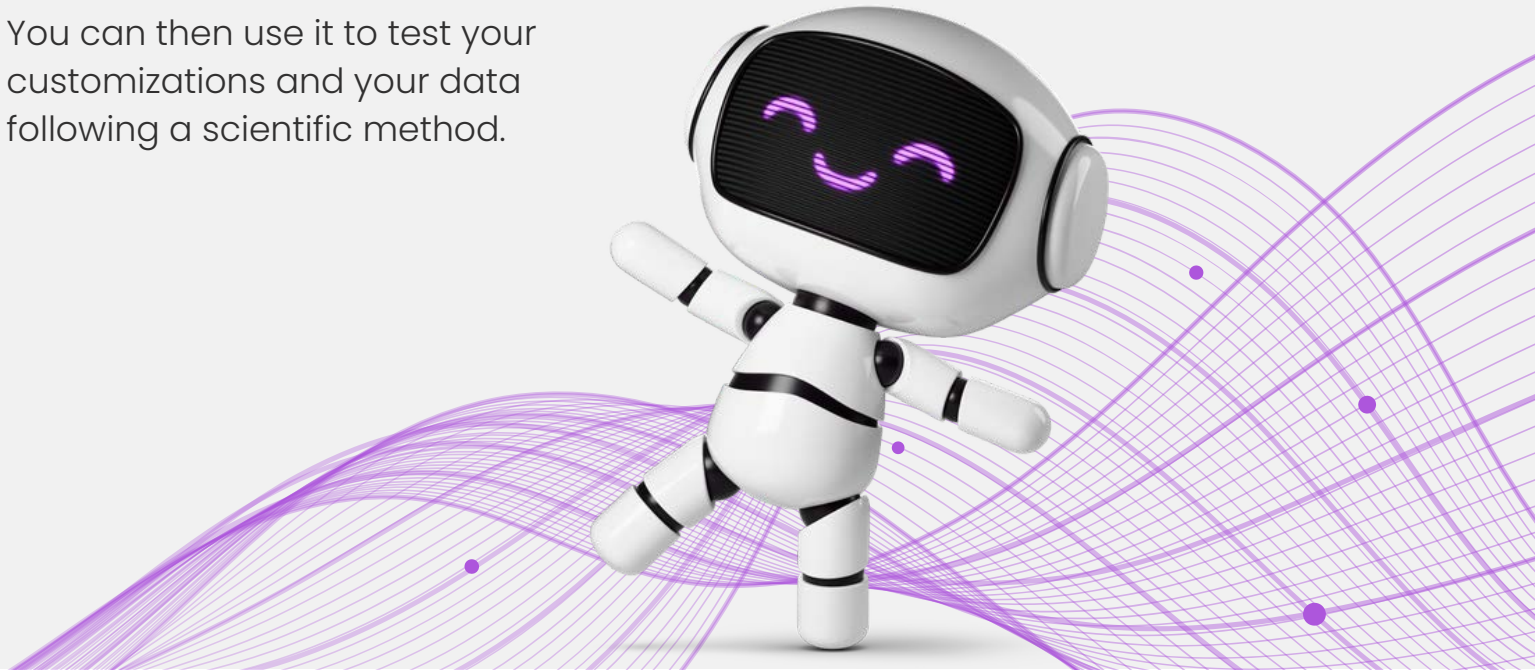
You can then use it to test your customizations and your data following a scientific method.

Cost modeling, estimation, and tracking

Most AI services are priced by token count. While this makes sense for AI vendors, **it makes cost modeling, estimation, and tracking tremendously difficult for users**. If this pricing model persists, AI orchestration should also provide the tooling to support these tasks.

For example:

- Show the token count for each prompt and response
- Show the token count for the entire end-to-end process involving multiple AI tasks
- Show the actual and estimated cost for each process, based on the vendor used and its cost table



Making AI work for ops teams

AI is a powerful technology that can open the door to more advanced Ops automation, but it is not a miracle. The key to getting the most out of AI is to understand how it works, what it excels at, and where its limitations lie. With that understanding, you can wrap AI in the necessary supporting technologies to gain the trust required for mission-critical, lights-out enterprise automation.

AI adoption roadmap

AI orchestration is not a standalone product. It's an infrastructure layer that you can put together using different technologies. The two biggest functional components are data management and automation orchestration, which makes an automation platform with those two core capabilities the most sensible anchor technology. For Ops, the Openprise AI Orchestration is a logical choice. If you're building a more cross-functional infrastructure, you can leverage your existing middleware stack, such as iPaaS and data governance solutions.

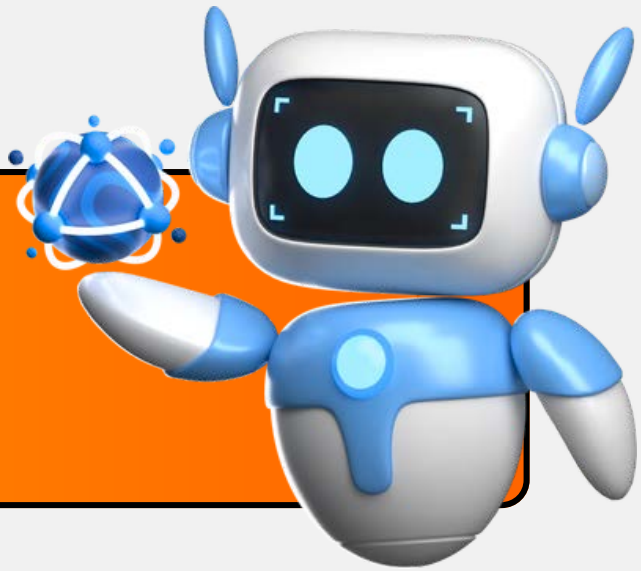
Making AI work for Ops

AI is subject to the same forces that shape enterprise technology adoption. Its adoption cycle will likely take at least a decade and will exist inside a hybrid architecture designed to maximize ROI from technology investments

As AI adoption spreads, new threat vectors and countermeasures will emerge, which means you'll need to remain agile, ready to pivot how you use AI to ensure it remains both safe and productive.

Finally, we are still in the experimental phase of AI adoption. Once organizations gain more experience with the technology and can separate hype from reality, management will be more pragmatic and demand ROI for AI projects. You know, just like every other technology project.

For Ops, the Openprise AI Orchestration is a logical choice.



About Openprise

Openprise **makes your GTM data smarter with AI and automation**. As the only data and AI orchestration platform built for modern go-to-market teams, Openprise automates your processes, unifies your data silos, and consolidates point solutions so Ops leaders can build smarter GTM data — your data, your way, your timeline.

Fortune 500 companies and high-growth enterprises alike rely on Openprise and its partner ecosystem to unlock cleaner data, more efficient operations, and AI-ready pipelines.

See how Openprise makes your GTM data smarter at www.openprisetech.com and follow us on LinkedIn.

BOOK A DEMO

OPENPRISE®