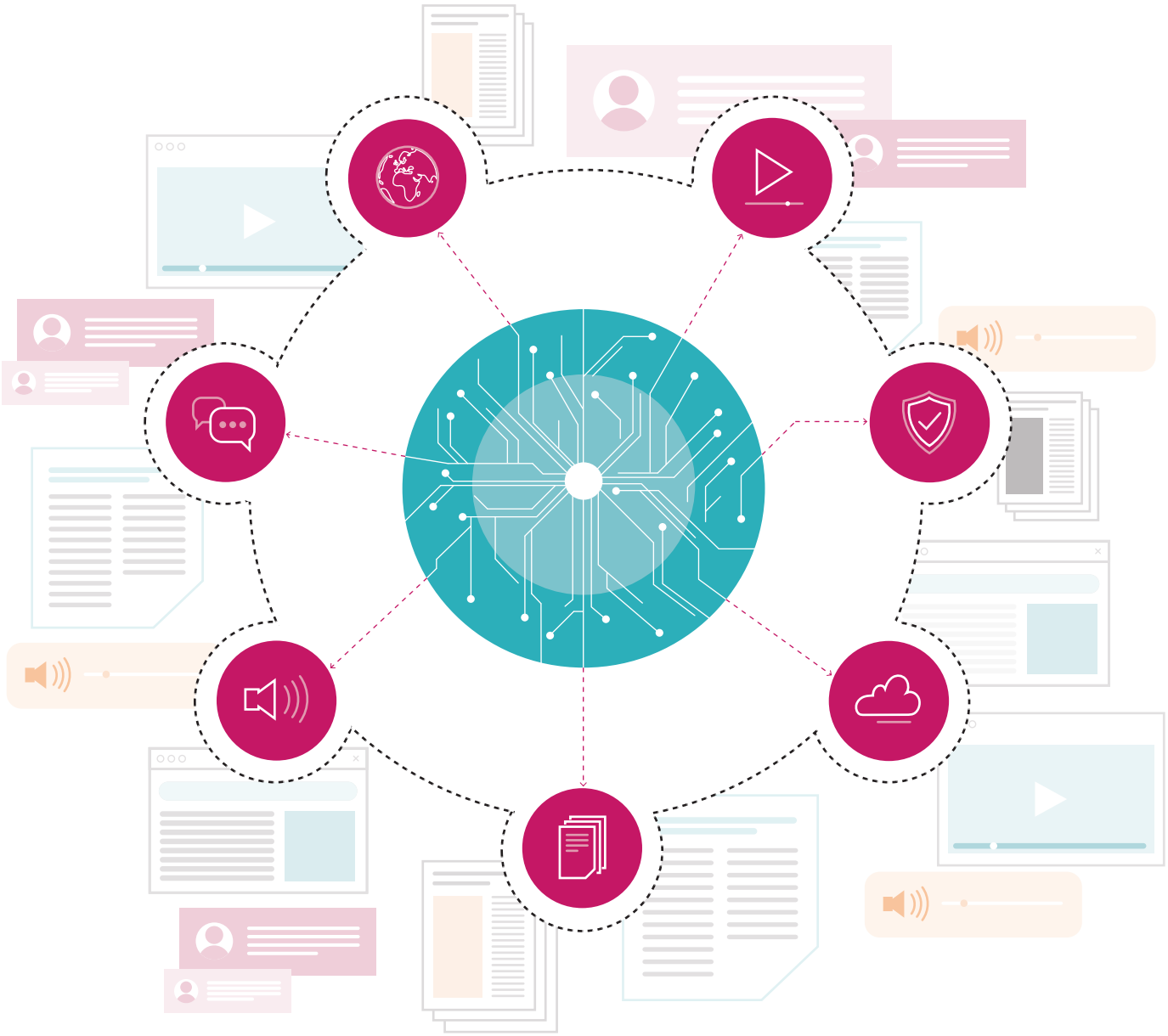
How to de-risk GenAI across your enterprise

A white paper by Oliver Cronk



Scott Logic

Contents

Executive summary

Like the internet and the mobile phone before it, generative AI (GenAI) such as ChatGPT promises to be transformative. The tech industry regularly predicts that new technologies will be game-changing, but this is one of those times when it's likely to be true.

In fact, here at Scott Logic, we believe the world is about to enter a new technology epoch.[1] For instance, research by McKinsey reveals that AI has huge potential for enterprises, predicting that the technology could deliver value equal to an additional $200 billion to $340 billion annually in the banking industry alone.[2]

All aboard

It explains why the potential of the technology is already being considered within enterprises and the public sector. For example, according to a recent survey of senior execs by The Harris Poll, 39% said they had already set up policies and strategies relating to GenAI while 59% were either developing or planning to develop them. That leaves only 2% with no plans at all.[3]

In the midst of all this excitement and speculation, organisations should avoid getting carried away by the hype. Harnessing GenAI's potential will introduce a range of organisational challenges to overcome; judicious decisions will be required about where to invest time and money.

Let business value lead

Among enterprises currently investigating the potential of GenAI, the most successful pioneers are those focusing on demonstrating business potential rather than what's technologically possible. Organisations should engage multidisciplinary teams from across the enterprise to prioritise high-value use cases and run lean experiments to identify where the technology has the most potential to deliver business value.

Tread carefully

As the use of GenAI begins to proliferate across the business and public sector spheres, so too does the potential harm it can cause—from ethical and privacy issues to explainability, repeatability and cybersecurity challenges. The pioneering organisations that first successfully harness this technology will likely gain a competitive advantage; but they'll also be the first to have to navigate the risks and challenges presented by deploying GenAI.

That's why Scott Logic has created this white paper, to help you navigate the potential risks and challenges ahead, and to present a robust solution architecture within which GenAI can be deployed and significantly de-risked.

[1] https://blog.scottlogic.com/2023/03/31/the-new-ai-platform.html
[2] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
[3] https://www.cio.com/article/482235/7-key-questions-cios-need-to-answer-before-committing-to-generative-ai.html

Going beyond the hype to the genuine value

It's important to tune out the speculation and focus on how your organisation can gain its own first-hand insights into where the technology can deliver value.

From AI-generated imagery and coding to text creation and next-generation chatbots, the technology has astonished the world with its human-like content and capabilities. Such powerful functionality promises to introduce big opportunities for business—and high business value, estimated to be in the hundreds of billions if the anticipated use cases are fully implemented.[4]

Speculation is accordingly rife about the scale of the transformation that AI might bring about in the economy, society and in our personal and professional lives—and the forms the transformation will take.

For example, our first-hand experience at Scott Logic of working with GenAI, and the conversations we're having with clients, lead us to speculate that use cases might include:

- Customer experiences that feature greatly enhanced explanations of product features and personalisation of recommendations and advice
- The automation of financial guidance that's currently only available to the few, democratising access by the wider population to financial literacy
- The transformation of employee onboarding through personalised employee chatbots that draw on information from an organisation's systems

In a report of June 2023, McKinsey estimates that 75% of the value that GenAI use cases could deliver falls within four areas: customer operations, marketing and sales, software engineering, and R&D. They suggest that there's the potential to automate activities that take up 60%–70% of employees' time, freeing up time for them to deliver value in other ways.[5]

However, at Scott Logic we always aim to go beyond the hype of claims made about new and emerging technologies. As I said earlier, it's important to let business value guide your investment decisions rather than technology hype. AI is evolving with unprecedented rapidity, and so speculation about its potential business applications—while helpful—should be taken with a pinch of salt. You should trust the outcomes of your organisation's own lean experiments with the technology over the claims made in the technology press.

And while GenAI's potential for enterprise is significant, there are many issues that could negatively impact your business—from biased and out-of-date data sets to copyright infringements and 'hallucinations' that see GenAI generating false information, fake sources and bogus insights.

The next two chapters summarise the principal risks presented by GenAI, both from a business perspective and a technology perspective. Let's start with the main business risks.

[4, 5] McKinsey and Company, *The economic potential of generative AI* (pdf), June 2023

How the new epoch introduces threat

To understand the risks of GenAI, you first need to start with the characteristics of trustworthy AI systems.

The NIST's AI Risk Management Framework states that these should include being: "Valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed."[6] This is the idealised view of a GenAI system and is used to refer to the development of the technology itself. However, these characteristics should be sought out by highly regulated enterprises aiming to negate significant risks on two fronts—the technological threats to existing IT infrastructures and the organisational threats to the enterprise itself. The latter include:

Reputational damage

Reputational damage is a major concern because of GenAI's tendency to hallucinate. This is the 'ability' of AI to generate credible-looking output that is actually flawed or outright false. Such falsehoods represent significant legal and regulatory issues, particularly if an application is customer-facing and generating responses in real time. For instance, imagine your GenAI platform were to give inappropriate financial recommendations to customers.

Legal and IP challenges

GenAI is trained on masses of information from across the world wide web. Such training could include data sets that feature commercial intellectual property from an organisation or individual, which the GenAI then uses to create your content without you realising. In some instances, hallmarks or watermarks from such training data could appear in the generated output, leaving organisations open to litigation for copyright infringement.[7]

Without sufficient context, there are multiple ways that generated content could put your organisation at risk of inadvertently breaking laws or acting unethically.

Ethics and privacy

Enterprises must ask themselves whether it is ethical to displace or massively augment people with GenAI's capabilities. Could users start to trust these systems too much? Also, given GenAI's content is based on massive amounts of internet data, any bias inherent in these original data sets could pose a significant problem for your enterprise. The privacy issue is also magnified as most GenAI modelling is cloud-hosted. This means you must ensure you are comfortable with sending data to third parties and be prepared to use pseudonymisation (masking personally identifiable details) or other techniques to address privacy issues.

Darker shadow IT

The interest and hype surrounding GenAI—plus the common perception that internal technology teams can take too long to deliver—means that some employees and teams may be tempted to use publicly available platforms for work directly. In effect, this is the latest evolution of the bring-your-own-device trend—but one whose ramifications are far more serious. Enterprises will be rightly concerned to hear of research showing that as of April 2023, sensitive data made up 11% of what employees had pasted into ChatGPT.[8] Major enterprises are taking this threat seriously, with Bank of America, Citigroup, Deutsche Bank and Goldman Sachs among the many financial institutions that have now implemented company-wide bans on ChatGPT.[9]

[6] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
[7] https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion
[8] https://www.cyberhaven.com/lp-eb-data-report-chatgpt/ | [9] https://www.sciencealert.com/many-companies-are-banning-chatgpt-this-is-why

Mapping the technological risks of GenAI

There are several challenges during GenAI implementation that threaten to create serious issues for your entire technology infrastructure.

These don't solely relate to machine learning and data science but go far deeper, encompassing deployment and systems integration, the two vital cornerstones of any successful IT strategy.

Explainability and repeatability challenges

Generative AI outputs are probabilistic responses to complex inputs. This creates consistency, reliability and repeatability challenges. This means how we design, manage and test GenAI requires different thinking from more traditional technologies.

The maths and algorithms used by GenAI are not only inaccessible to the average person, they are of such complexity that the creators of AI models often require further AI to interpret and understand them.[10] Add to this the complexity of enterprise architecture, and explainability becomes even more problematic.

Scaling and data integration issues

Scaling systems that are powered by machine learning presents fresh challenges. For example, creating an algorithm in a lab environment that comes up with an answer in several hours (or even minutes) is acceptable for a one-off exercise. However, adopting such an approach for real-time customer interactions at scale is impractical.

In addition to the model's performance, you must also consider how you will integrate such models with enterprise data stores and systems of record without negatively impacting performance. For example, giving the model the ability to look up a customer's recent and historical activity to add context is a crucial part of customer experience (CX)—but will actually damage CX if the data integration process used does not deliver fast enough response times.

Security best practice blind spots

GenAI will likely lead to whole new classes of security issues with any software that leverages the technology. We are already seeing examples of jailbreaking chatbots and there is little doubt that other GenAI-powered applications will be compromised.

Applying sound security practices—and not cutting corners—will be key here. Also, it is vital to ensure that your security team is up to speed and understands this area of technology. If they do not, there is a danger they may demand to shut down projects out of fear and ignorance—and not due to an objective understanding of which risks are real and which are imagined.

Free trial complications

It has been estimated that GenAI models take millions of dollars to train, not to mention the costs of securing sought-after and expensive data science talent. Therefore, it is highly unlikely that platforms which are currently free or low-cost will remain that way. If free platforms do persist, serious questions must be asked about the commercial model that GenAI providers are using to generate earnings, and whether they are making use of the data fed into their platform. Checking provider terms and conditions is important to ensure they are compatible with your organisation's own policies and regulatory regimes.

Such risks are also echoed by the potential harm of adopting a 'bring-your-own-model' (BYOM) approach. Unless governed via strict management processes to ensure bad practice does not permeate throughout your enterprise, BYOM risks creating issues that security cannot track, never mind practically manage.

Use case problems

In future, it is likely that GenAI-based services will become some of the most expensive cloud services offered by hyperscale providers. First, such services will offer significant value to enterprises, so providers can expect demand to be high. Second, GenAI R&D will continue to be expensive, as well as the cost of operating it. Third, any AI service will come with significant risks, which the provider will need to constantly mitigate.

However, at the moment, the excitement and price—often free—means GenAI is being used for 'everything and anything'. As GenAI begins to increase in cost, it is likely that enterprises will focus their investment on using GenAI to solve more challenging, important and high-value problems. These should be ones that cannot easily be solved using traditional technologies; particularly as—in the context of our collective journey to net zero—those traditional technologies tend to have a lower carbon footprint than GenAI.

[10] https://openai.com/research/language-models-can-explain-neurons-in-language-models

Why robust architecture is the answer

To manage risk effectively, enterprises must implement an architecture able to meet all challenges while offering a route to successful GenAI deployment. We recommend the architectural approach below.

The right architecture introduces a series of checks and balances that can help manage the risks of rolling out GenAI across your enterprise. Importantly, Scott Logic's architecture focuses on reducing risk when AI is integrated with existing enterprise systems—think data stores plus transactional and analytical systems. Naturally, systems that are customer-facing present greater challenges as they expose the organisation to deeper public scrutiny and increase the chance of reputational damage if something goes wrong.

Big picture

Taking a macro view reveals that these issues are not simply about GenAI itself but something far deeper and more profound; the data and the overall framework in which your chosen AI model/s sit. Get the architecture and strategy wrong and it not only opens your enterprise up to the additional risks we have already covered—but introduces genuine threats to how your organisation itself is run.

With no coherent strategy, the danger is every department comes up with its own answer and AI solution. In turn, this adds more complexity to your organisation, which is the exact opposite of the efficiency gains you were planning to unlock by introducing GenAI. Worse still, this complexity will become extremely difficult to 'unpick' as AI's reach inevitably finds its way into the many different systems at the heart of your enterprise.

Our GenAI architecture

The right architecture acts against this structural threat, allowing you to plan your rollout as well as introduce guard rails to ensure the GenAI remains contained. This can be achieved by focusing on each stage of the AI process from demand management, inputting and outputting to 'black boxing' and beyond.

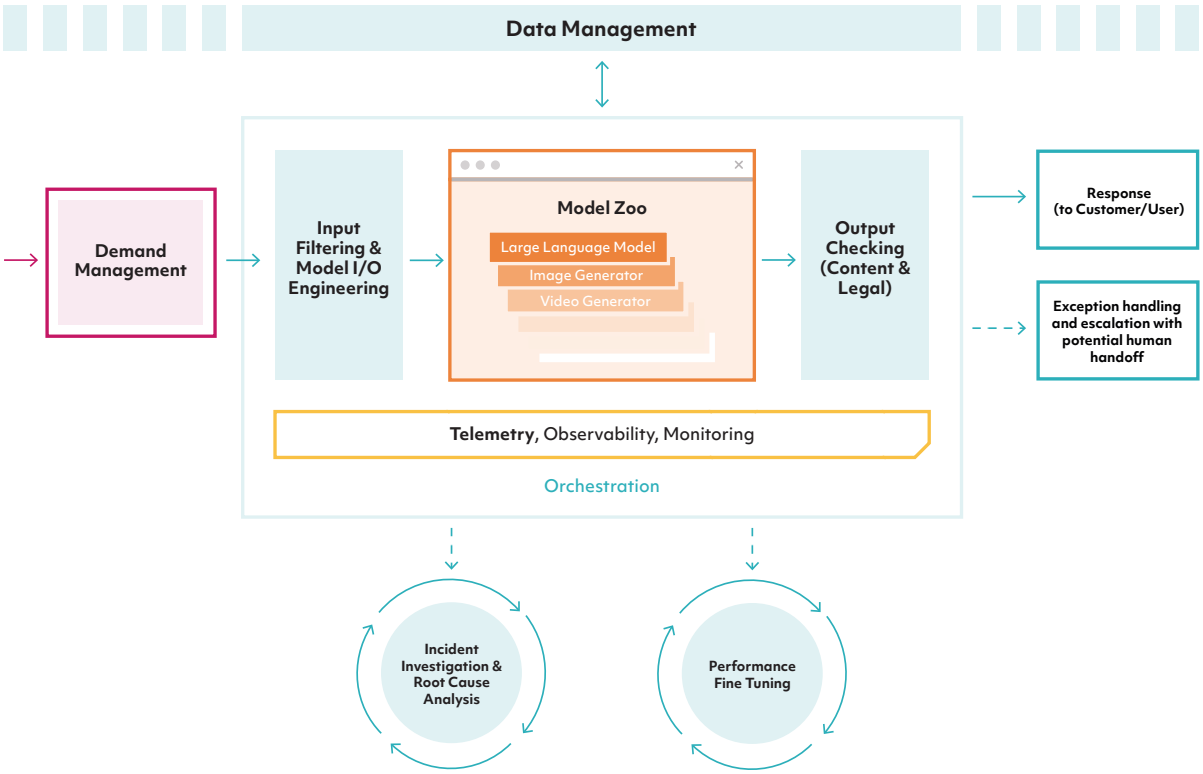
We have identified some key aspects of architecture to consider for customer-facing applications (such as chatbots and real-time content generation). What follows is a high-level summary of the diagram, explaining how the different components contribute to governance and mitigating the risks of deploying GenAI.

Demand management

Demand management mechanisms—e.g., queuing of requests—ensures the platform is isolated from any spikes in customer demand and can manage them. A robust architecture will offer genuine scalability to meet any demand, all while ensuring back-end systems can deliver results at speed to address internal or customer needs in real-time.

Orchestration

This is the wrapper around the different components, managing the GenAI models (see Model Zoo below) and providing the framework to add elements such as telemetry capture and input and output checking. With this approach, you can scale or swap out components within the orchestrator while keeping the upstream and downstream interfaces consistent.



Input filtering and Model Input/Output engineering

While GenAI already filters its results to at least try to avoid offence, the right architecture will enable the introduction of further filtering. This will give you a better oversight of what is being produced by the AI plus, if planned correctly, give you the option to introduce different modelling systems to your Model Zoo, tailored to your specific needs.

Model Zoo

The Model Zoo houses and isolates the approved GenAI models (e.g., LLMs, image generators, etc.) that can be called on by the orchestrator. This allows for the management and governance of models used in the application. You will probably need bespoke models tailored to your systems and requirements—and trained on unique customer data—to create an AI model capable of servicing your unique needs. The right architecture will allow you to 'curate', manage and deploy such models effectively and at speed.

Telemetry

When applications leveraging machine learning and GenAI fail, we need an audit trail—a data source that captures the input data, decisions and outputs. This way, lessons can be learned and decisions made to tune or change models on the basis of data and evidence. From a regulatory perspective, it's possible that this will be made a requirement when using GenAI technologies for customer processes, with regulators demanding to see your application telemetry.

Output checking

In order to prevent brand damage, misselling or other mishaps from generating inappropriate content, output checks and filtering will be required. This is likely to be a blend of traditional logic-based filtering and machine learning models that generate a confidence percentage that outputs are aligned with company policies and/or regulatory standards. Responses back to the customer can then be altered or held back and escalated to a human employee to respond to the customer instead.

Ongoing maintenance components

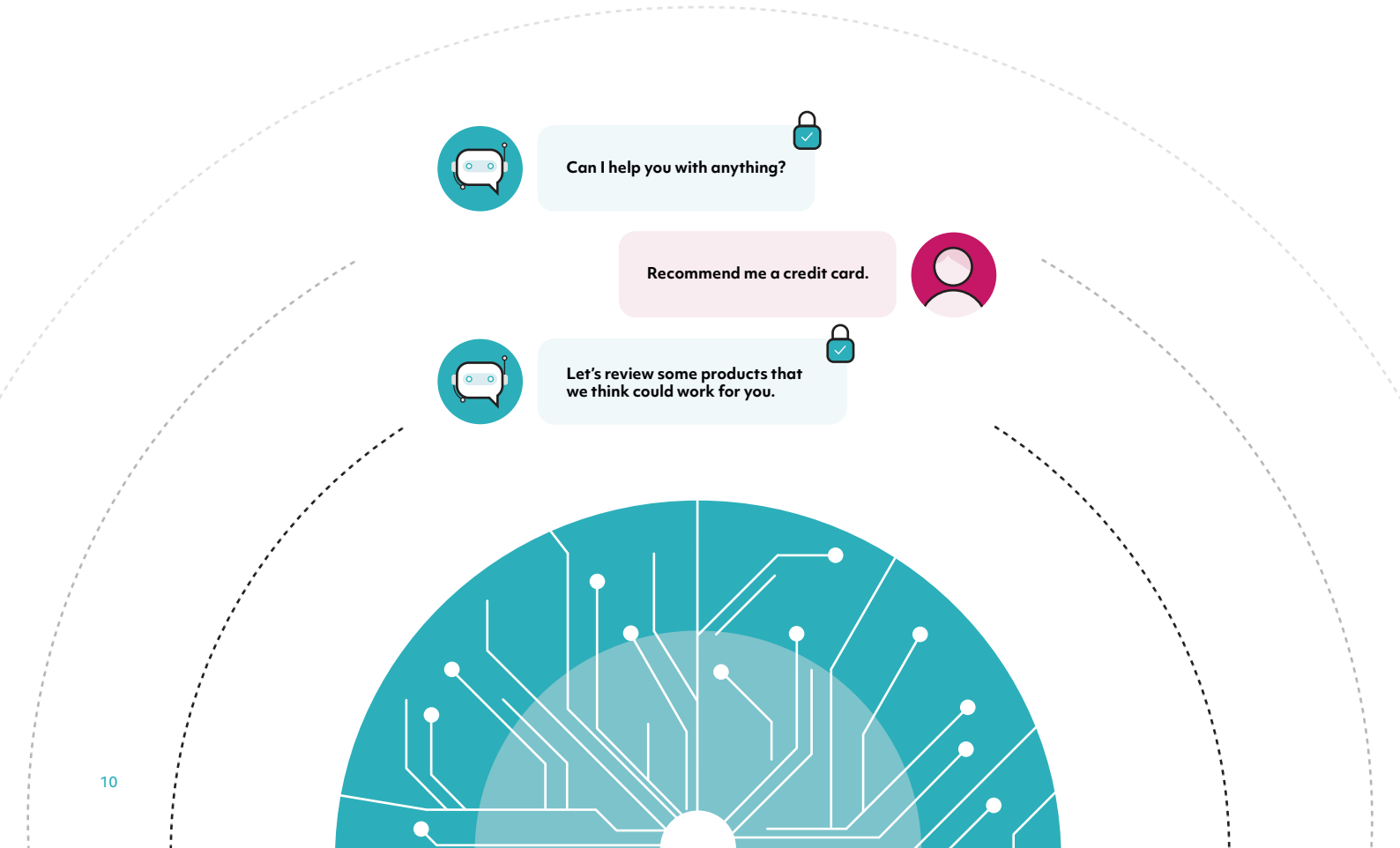
Underpinning the above processes will be components that support platform improvements. These will ingest telemetry and performance data and assist with updates to the Model Zoo, input and output filtering, and other supporting components. The performance of the platform can be evaluated and fine-tuned with any issues and incidents investigated using the captured telemetry data.

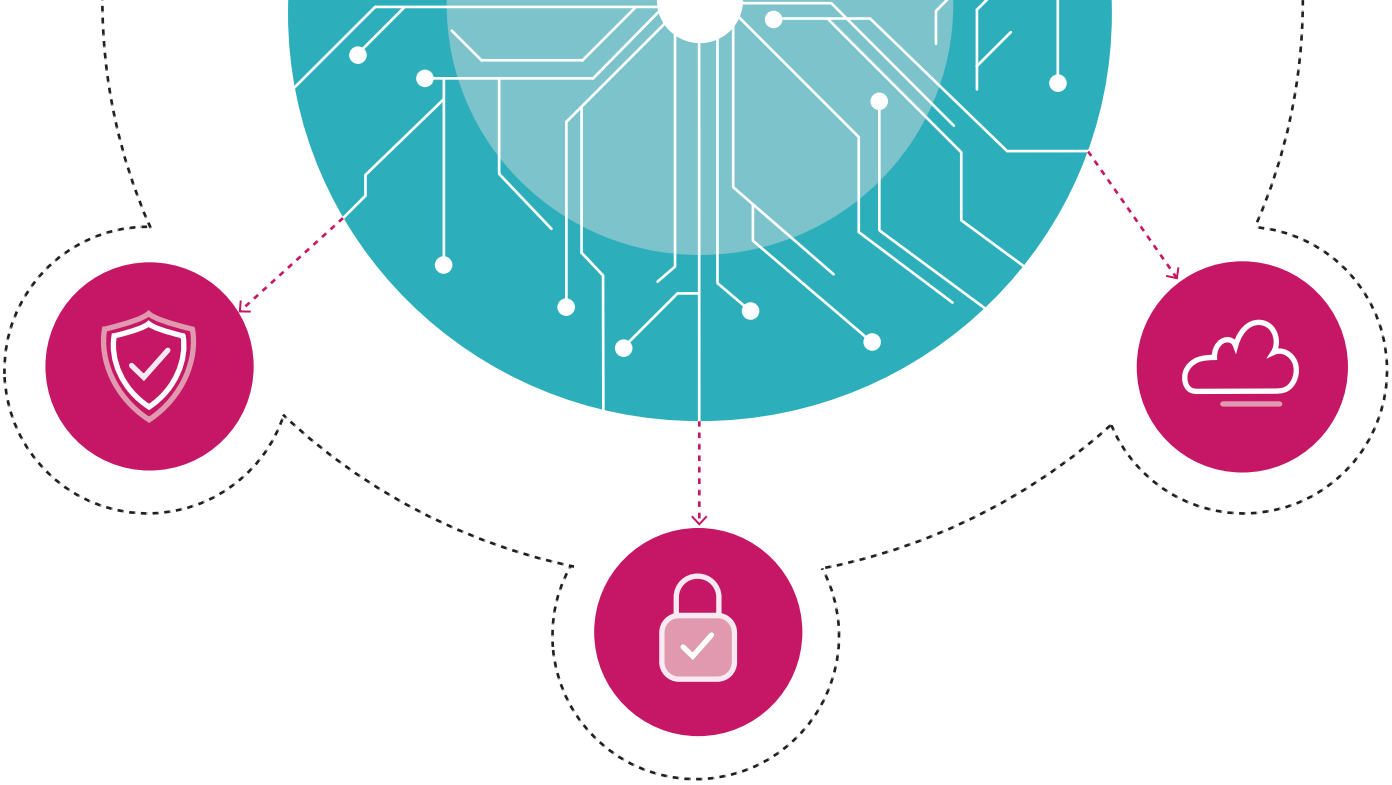
Perspective matters

Another key issue with GenAI and its role within enterprise is how the technology is being sold by providers. The pitch is that AI can be the hub of your operations that everything else feeds in and out of. In other words, simply 'install' your GenAI solution and plug your organisation into it.

This approach is one we strongly advise against adopting. Instead, the right mindset is to see GenAI as a single spoke leading into your enterprise's hub, rather than the hub itself. This way you remain in control as much as possible while using architecture to reduce risk.

Ultimately, this level of control—and how you retain it—must be the most important consideration of your AI strategy. Deploying the right architecture will provide you with the right framework and tools to remain in control of your AI, and not the other way round.





Is GenAI really a risk worth taking?

Every new wave of technology adoption comes with threats—but to ignore GenAI's potential poses a far greater risk to your enterprise in the long term.

Attempting to ignore what is set to become the next major technology platform is not realistic for an enterprise wishing to remain competitive and innovative in the future. The key is to prioritise and find pragmatic trade-offs in the areas of most concern, using solutions including Scott Logic's architecture to chart a safe course through the many challenges that GenAI has created.

In some instances, it won't be possible to make your organisation comfortable with the potential level of risk. If this occurs, consider adopting a 'delay and develop' strategy, where R&D investments are made to explore ways to address the risks or create new control mechanisms prior to customer-facing deployment.

To the future

However, in spite of the potential risks, the simple fact remains that GenAI is creating genuine opportunities for businesses to leverage as well as making software architecture exciting again.

In recent times, most new applications have been following increasingly standardised cloud platforms. GenAI bucks this trend, giving industries the opportunity to create innovative products, services and ways of working that could never have been considered before. The bottom line? You need to work with your risk, security and regulatory/legal stakeholders as early as possible, collaborating with them so they understand the unique characteristics of GenAI.

Together—and with specialist support from consultants like Scott Logic—you can put yourself in the best possible position to unlock the potential of GenAI without falling foul of the dangers increasingly associated with it.

Let us help you go beyond the hype

At Scott Logic, our expert consultants have decades of experience in helping organisations like yours to seize the opportunities presented by the latest technologies, while carefully managing and mitigating the risks.

Let us support you to go beyond the hype around generative AI to the find the genuine value. From identifying the highest-value use cases, to running GenAI experiments, to designing and deploying GenAI-secure solution architecture, we can help.

Please contact Oliver Cronk on:

+44 333 101 0020

oliver@scottlogic.com

Scott Logic

6th Floor, The Lumen
St James Boulevard
Newcastle Helix
Newcastle upon Tyne
NE4 5BZ

+44 333 101 0020

scottlogic.com