Capgemini

CLOUD ADOPTION:
ENTERPRISE AZURE OPENAI FOR FINANCIAL SERVICES

A practical guide on how your organization can implement OpenAI across your enterprise

Table of contents

Introduction

Artificial intelligence (AI) is everywhere, and its capabilities are continually expanding. Yet for those in the financial services industry, the idea of incorporating AI within your organization raises a red flag - particularly generative AI.

Most chief technology officers (CTOs) are familiar with generative AI and know off-the-shelf products exist, but those options won't suffice due to security concerns and the need to engage with proprietary enterprise data on a regular basis. Instead, what's needed is an innovative, customizable and fortifiable solution that enhances customer operations for teammates, improves customer experiences and remains cost efficient over time.

Generative AI is a type of artificial intelligence that focuses on creating novel and complex image, text, music, or video outputs.

Consider how generative AI could help improve developer productivity:

- Adding data bias checks in generative AI can help level the playing field, correcting and eliminating algorithmic biases stemming from a person's race, gender or disability.
- Training generative AI on senior employees before rolling it out to the entire staff, while improving the model based on continuous feedback, can improve the end user experience across the board.
- Internal engineers will see improved developer productivity.
- Customers will have better experiences that are tailored for their specific needs.
- Streamline human resources management processes with generative AI by using a smaller team to reach the same number of employees, reducing costs.

AI Tools at your fingertips

A variety of AI Tools are now available to help organizations with their digital transformations.

Microsoft Copilot, for example, is designed to provide AI assistance across all Microsoft applications and experiences.

Microsoft Azure OpenAI offers a comprehensive suite of AI tools and services designed to accelerate organizations in their digital transformation and innovative technology solution adoption. Financial services companies can benefit from leveraging Azure OpenAI to streamline processes, enhance decision-making and ensure regulatory compliance.

Once use cases and AI tools are engaged, realizing the true benefits of generative AI will depend on the outcomes your organization sets as a goal. It's up to your CEO and other C-suite executives to build out concrete goals and justify incorporating generative AI across the organization.

To implement generative AI, CTOs need a clear plan of attack to include it in their business models, exploring what works for their business while assessing issues such as data security and regulatory compliance. CTOs know this but need more information on the considerations and requirements to do so safely and successfully. Forging ahead without a plan could cause major missteps in the process.

CTOs know how to get the technology into the hands of their technical teams, but must seek guidance customizing the right solution beyond what's easily obtained off-the-shelf. And yet, the cost of devoting the time, energy and team members required internally to get up to speed on generative AI solutions, and/or to design a custom application for the business, is often out of reach. That's where Capgemini can help.

Join us on the generative AI adoption journey:

1	Awareness	• Researching Azure OpenAI capabilities • Identifying potential use cases and applications • Educating internal stakeholders
2	Experimentation	• Building and deploying simple AI models • Testing AI application in controlled environments • Learning from initial successes and failures
3	Operationalization	• Implementing AI solutions in production environments • Establishing AI regulatory and governance best practices • Training staff in AI-related skills and tools
4	Optimization	• Continuously monitoring and improving AI performance • Identifying new opportunities for AI application • Collaborating with other departments to expand AI adoption
5	Innovation	• Developing cutting-edge AI applications • Collaborating with external partners and researchers • Becoming a thought leader in AI and driving industry-wide change

Beginning with **awareness**, the Azure OpenAI adoption journey progresses through five phases of maturity. Next is **experimentation**, using software development kits (SDKs) and built-in tools from Microsoft. Move onto the **operationalization** level with the adoption of security requirements and regulatory compliance and eventually to the ultimate end game: **optimization** and **innovation**.

Level 1: Awareness

Getting to know Azure OpenAI Services and Azure OpenAI in financial services

Central to obtaining buy-in from internal stakeholders about the benefits of generative AI, is to educate them about it. Here's some information CTOs can share with their internal stakeholders about Microsoft's Azure AI Services and its generative AI platform, Azure OpenAI.

Azure AI Services make up a wide swath of available intelligence solutions that Microsoft makes available within Azure. The latest addition to these services — Azure OpenAI — adds language generation capabilities to the suite. The following language services represent a niche within AI and are offered through Azure Artificial Intelligence services:

- Bot Services
- Content Safety
- Document Intelligence
- Immersive Reader
- Language
- Speech
- Translator

Utilizing generative AI in a financial services business provides many general benefits, including simplified decision-making. AI models can analyze large volumes of data quickly and accurately, enabling human users the ability to make better and faster decisions in areas such as credit risk assessment, investment strategies and fraud detection.

Harnessing Azure OpenAI's cloud-based infrastructure as the basis for a generative AI solution allows companies to easily scale their AI solutions as their needs grow, ensuring they can continue to innovate.

Once an organization achieves this awareness, the first step in progressing through the levels of maturity has begun.

A primer: How Azure OpenAI protects an organization's data

When incorporating generative AI into your organization, safeguarding company data is one of the first steps — and greatest concerns. Microsoft won't use your company's data to feed its larger AI data pools.

Here are three ways how Azure OpenAI data protection is designed to maintain security:

1. **Data and prompts are private -**
   Azure OpenAI as the basis for generative AI in financial services is private. Most importantly, the data and prompts supplied to Azure OpenAI are never available to other customers in any capacity. Microsoft does not use data to fine tune the models or prompts supplied by the users as training for the OpenAI model itself. As such, the customized models, prompts and outputs are more secure for each organization. Fine-tuned models generated by each organization are only available for the use of the organization that trains these models.

2. **Only company-specific data is used -**
   When an organization is fine tuning AI models to fit its needs, the available data is exclusively for the organization that uploads it — no outside parties have access to it — and may be deleted by your company at any time.
   This data is stored in the same region at the Azure OpenAI resource and may be double-encrypted at rest. This approach for double encryption utilizes a Microsoft-managed key and a customer-managed key for encryption.

3. **Content filtering and abuse monitoring controls provided -**
   To reduce harmful use of the service and ensure that it is used in a manner consistent with Microsoft's code of conduct, Microsoft provides content filtering and abuse monitoring controls as part of the Azure OpenAI service.
   For the content filtering capability, no prompt or generated result data is stored. In addition, because of this ephemeral functionality, the data is not used to train, retrain or improve the models.
   The abuse monitoring capability is asynchronous and the prompts and generated results are stored by default for up to 30 days. This monitoring can be configured not to store data upon exemption approval from Microsoft. Even if this monitoring is not disabled, the data is stored in a logically separated location from other users.
   Finally, human review is only triggered when the content is flagged.

Once internal stakeholders are on board, use cases and applications have been identified and potential capabilities have been considered, it's time for the experimentation phase.

Level 2: Experimentation

Learning how to build, deploy and test the AI application in controlled environments

Financial services companies interested in adopting Azure OpenAI need to understand the deployment patterns of the system — how to use it, and the tools and capabilities provided to make it a smooth process. The default models are great for getting started in a sandbox environment and helping your internal engineers understand how the service operates. Once the proof of concept is complete and the team understands the service, the next step is to deploy a fine-tuned model. The available application programming interfaces (APIs) allow your organization to:

- Upload training data
- Create a fine tune request
- Check the status of the fine tune request
- Deploy a custom model

These capabilities are enabled automation through a series of scripts in the CI/CD pipeline.

Once data is available, it is easier to move from the sandbox to more sophisticated outputs from the service.

Getting started

To get started in a sandbox environment and progress to the second level of maturity, your organization must engage several team members to ensure a seamless and successful technology development, including the compliance officers, enterprise architects, engineering and the business team. The following chart notes how each team members will be engaged as your organization gets started.

Risk & Compliance Officers

1. Define audit processes, practices, and documents to elevate compliance risk and identity violations
2. Create and review policies and internal controls
3. Develop legal compliance requirements for the organization for Generative AI
4. Manage and oversee the implementation to ensure compliance requirements are met

Enterprise, Solution, InfoSec Architects

1. Create whitepaper to identify strategic decisions, best practices, and rationale for adoption
2. Define design patterns and organizational standards for use of Generative AI
3. Obtain approval(s) for utilizing the technology
4. Evaluate and harden security according to organizational policy

Software, Infrastructure, InfoSec Engineers

1. Develop a sandbox prototype and learn the implementation practices using the various pre-built models
2. Develop infrastructure automation and integration APIs
3. Incorporate security and regulatory compliance requirements from architecture

Business Team

1. Identify use cases to benefit customers by leveraging the feaatures of Generative AI
2. Develop business cases with justification that demonstrates the value that will be derived
3. Compile a library of data to support model fine-tuning and evaluation

Once your Azure account is active, the initial step is to set up an OpenAI environment using Azure AI Services. Azure OpenAI is equipped with a set of pre-built AI models that make it easier for developers to get started with generative AI models such as GPT.

Next, your team must choose appropriate AI models to try out and identify which best suit your company's needs. Azure offers pre-built models for tasks such as anomaly detection, fraud prevention and risk assessment, among others.

Critically important is to define security and policies around compliance requirements for your use of AI. While the teams that will eventually utilize the services of OpenAI are researching and understanding the usage patterns, the enterprise architecture group must define AI standards and policies around compliance requirements to enable full adoption of this technology.

Once your team learns the basics and defines policies and standards, the next step is to train and test AI models to verify which fit the needs of your organization. Use your company's historical data to train and fine-tune the AI models. After training, test the models to ensure their accuracy and reliability.

To help code for content generation, summarization, semantic search and natural language, Azure OpenAI Service provides REST API access to OpenAI's language models, including GPT-3, GPT-4, Codex, and embedding models series.

Microsoft's Azure OpenAI Studio user interface accesses the service outside of code-based approaches (APIs or SDKs), providing both playground and management capabilities. On the playground side, both chat and completion options are available, allowing the user to interact with the models for testing and evaluation. On the management side, the user may upload data, provision base and custom models, and create deployments for interaction in playground or via the APIs. The Azure OpenAI Service also may be provisioned using the standard automation tools that Azure provides.

In addition to provisioning the service, several SDKs can interact with the Azure OpenAI APIs, providing a straightforward method to incorporate OpenAI into custom software solutions.

SDKs have been developed for programming languages such as Python, .NET and Java, making it easy to get started by building a wrapper micro-service for the OpenAI service. This wrapper can enforce rules around security, regulations and compliance, and provide a means for FinOps control.

- By the end of the experimentation phase, your team members should have done the following:
- Defined the appropriate policies
- Understand the usage patterns (and may have already started a proof of concept to adopt generative AI)
- Trained and tested models that are fit for use in your organization.

Once the above tasks are complete, your organization can deploy these AI models in production and start reaping the benefits of AI.

While adopting the basics, defining the standards and best practices, and implementing the appropriate policies through experimentation, financial services organizations can progress to the next level of maturity: operationalization.

Azure OpenAI in action

As part of Microsoft's overall offering, an organization gets to adopt software used to consume the platform.

Consider this scenario: An organization implements bicep and SDKs into their customer software so that their customer service reps can chat with an insured. A customer pings the rep to ask about an auto policy issue, providing their coverage type and policy number, as they are wondering if they can earn compensation for paying an auto claim. An AI model helps quickly evaluate the policy and determines the auto claim is covered minus the deductible, saving both the customer and the rep time in the process.

Without generative AI, the customer service rep would have to pull up and read the policy. The rep also may need to confer with internal resources for assistance if they require help, increasing the amount of time needed to help resolve the customer's policy- and claim-related questions.

Level 3: Operationalization

Adding in Security and privacy when enabling Azure OpenAI

While Microsoft may make it easy to provision the infrastructure within Azure to utilize generative AI, financial organizations must consider the security implications of doing so. Some of the key considerations when using products such as OpenAI include:

- Data protection
- Access control
- Network security
- Monitoring and logging
- Vulnerability management
- Compliance
- Secure model development
- Model transparency

As with any other resource, Azure OpenAI is monitored for compliance. Microsoft has built in policies and initiatives that cover most of the audit requirement to meet the frameworks of organizations such as the National Institute of Standards and Technology (NIST), the PCI Data Security Standard (PCI DSS), and the Center for Internet Security (CIS). Organizations should implement policies and initiatives at the management group or subscription level to ensure that any new resources meet the requirements to satisfy these frameworks.

The **data protection** policy is already defined for the vast majority of organizations. Organizations should encrypt sensitive data at rest and encrypt all data while in transit. Any further protections defined by the policy should remain in effect for use of Azure OpenAI.

As with any service that is provisioned, organizations must implement strong **access control** policies to limit who can access your Azure OpenAI resources. Use Azure Active Directory (AAD) for identity and access management and apply the principle of least privilege by granting the minimum permissions necessary for users and applications to perform their tasks.

Network security is a large topic but ultimately the tenets are well established for other resources. Organizations should configure **network security** groups and firewalls to control inbound and outbound traffic to its Azure OpenAI resources. Use Azure Virtual Networks to segregate your resources into isolated environments and consider deploying Azure Application Gateway or Azure Front Door for additional security, such as web application firewall and distributed denial of service, or DDoS, protection.

Next, define the **monitoring and logging** standard. Enable monitoring and logging of your Azure OpenAI resources to detect and respond to security incidents quickly. Use Azure Monitor to collect and analyze logs and configure alerts for suspicious activities. Additionally, consider integrating with Azure Sentinel, a cloud-native Security Information and Event Management (SIEM) solution, to gain advanced threat detection capabilities.

To ensure the security of the Azure OpenAI environment, organizations must institute **vulnerability management** practices. Regularly assess your Azure OpenAI environment for vulnerabilities and apply patches as needed. Use Azure Security Center to monitor and manage the security posture of your resources, and follow recommendations provided to mitigate potential risks.

Compliance is another key topic for financial services companies. To ensure that your Azure OpenAI implementations adhere to relevant industry standards and regulations, such as the Payment Card Industry Data Security Standard (PCI-DSS), the General Data Protection Regulation (GDPR), the Bank Secrecy Act/Anti-Money Laundering Act (BSA/AML) and others, financial services companies must implement **compliance** measures. Leverage Azure's built-in compliance offerings, like Azure Policy and Azure Blueprints, to create and enforce compliance policies across your resources.

Secure **AI model development** is a crucial control to apply to the software development lifecycle. Protect your AI models from unauthorized access, tampering and theft by implementing security best practices throughout the development lifecycle. Use Azure Machine Learning's secure development features, such as role-based access control, private workspace networks, and data labeling privacy controls.

Finally, **model transparency**. Putting the above controls into place will advance the model transparency, first through logging and monitoring and then through compliance and regulation. Because the service offerings in generative AI are new, retaining a human in the loop to review the model outputs will help to hone the capabilities, provide data that reinforces the proper outputs, and refine the model outputs to remove unexpected and undesired outputs.

FinOps and regulatory compliance considerations

As pre-requisite enterprise considerations to operationalization, financial organizations should define cost controls and identify regulatory requirements.

Cost controls

It's critical for financial services organizations to have a well-defined FinOps strategy. Because Azure OpenAI is a serverless solution that can scale vertically and horizontally to meet the demands of your organization, controls must be put in place to mitigate costs. At a minimum, this involves creating Azure Budgets, defining cost alerts and adding FinOps dashboard integration to evaluate costs.

Once the budgets are in place, financial organizations can configure alerts to notify the infrastructure teams when the budget allocation has reached a specified threshold. If the costs are regularly meeting or exceeding the expected budget, further analysis can be broken down by the meter, which is incurring that cost. If a specific type of cost accrual is taking up a large portion of the budget, organizations can adjust the services that utilize those models to mitigate costs.

Additionally, a custom service wrapper should be available that can evaluate the requests that are made to Azure OpenAI to determine if the request can be resolved without incurring the cost of sending it to the service.

Finally, the cost data can be exported for finer grain analysis of the usage metrics.

Regulatory requirements

Financial services companies must adhere to strict regulatory requirements; using generative AI technologies like Azure OpenAI is no exception. To ensure compliance, consider the following:

1. **Keep a human in the loop.** Generative AI is currently bleeding edge; as such it has not been thoroughly vetted for accuracy and capability. Having a human participant in the loop reviewing the outputs and intervening as necessary is still a critical element to success with generative AI.
2. **Maintain data privacy.** When handling sensitive customer data, AI models must comply with data privacy regulations. This may be accomplished through implementing appropriate data management and security measures.
3. **Use AI models that are understandable and documented.** Regulators may require companies to explain how the selected AI models make decisions. Therefore, it is best to select AI models that offer transparency, interpretability and maintain clear documentation on how they work.
4. **Remove bias and ensure fairness.** AI models should be free from biases that could lead to unfair treatment of customers. As such, continuously monitor and evaluate selected AI models for potential biases.
5. **Ensure your AI models are monitored and auditable.** Implement robust auditing and monitoring processes to track the performance of your AI models, ensuring they remain compliant with regulatory requirements.

Once the security and regulatory compliance measures have been considered and incorporated into the solution, the software can be deployed and operationalized. The next step is to build on these foundations to optimize the Azure OpenAI solution.

Governance

As with existing solutions, governance of Azure OpenAI involves the implementation of the following:

- Data protection rules
- Access and network security controls
- Regulatory compliance requirements
- Infrastructure provisioning standards including logging and monitoring strategies
- Secure development practices
- FinOps

Governance as it relates to Azure OpenAI should remain consistent with existing processes, lifecycle management and oversight practices. As such, the Cloud Center of Excellence group, which provides organization-wide leadership and enterprise architecture guidance relating to cloud solutions, should review and approve the management and governance elements described in this section.

Level 4: Optimization

Monitoring and improving AI performance and identifying new opportunities

During the optimization stage, financial organizations can work on iterating over their existing solutions to identify new opportunities and innovate. Optimization involves the continuous monitoring, improvement, development and refinement of the AI models. In addition to technical optimization of the solution, identifying further opportunities for growth and learning from the process of deploying it will lead to further growth.

This will also foster a deep collaboration both internally — with stakeholders of the solution or future candidates — as well as externally within the generative AI community and lead to breakthroughs in the technology enabling further optimization.

Continuous improvement

To ensure the highest efficiency and effectiveness of Azure Open AI solutions, it is essential to monitor and optimize its performance continually. Organizations can facilitate continuous monitoring by the built-in tools, such as Azure Monitor and Application Insights, to collect and optimize performance metrics, identify bottlenecks and detect anomalies. These tools should enable a feedback loop that allows users' actions to be inspected and evaluated to determine if revisions to workflows should take place.

In addition to the built-in tools, a procedure should enable users to report issues, suggest improvements and provide general feedback on how the generative AI features are impacting usage of the software solutions.

Opportunities for growth

Ultimately, the organization and business team will have a high level of understanding of the capabilities of the products to which generative AI is introduced, but will likely require several iterations to grow an understanding of the impact that generative AI has on the product. Some of the opportunities for growth will involve information that can be learned through monitoring and observability while others will involve research and collaboration with external partners.

As the team gains a deeper understanding of the technology and how it is impacting the software portfolio, use the following strategies to identify new opportunities:

1. Analyze existing business processes and workflows to identify areas where AI can enhance efficiency, reduce costs and improve decision making.
2. Monitor industry trends and emerging technologies (or enhancements to the existing technology) to discover new AI applications that can provide a competitive advantage.
3. Engage with stakeholders across the organization to gather insights and ideas on how generative AI can address existing challenges and pain points.
4. Conduct studies on feasibility and pilot projects on the secondary stage products in the portfolio to test the effectiveness.

Note: End users may be less predictable on how usage patterns or satisfaction will form with the new features. As such, it will take some time to analyze the trends produced by the existing capabilities for how generative AI changes the user behavior.

Improving the feedback mechanisms to cover generative AI scenarios will enable the discovery of opportunities for growth.

Collaboration

Collaboration will be key to optimizing the solution as it is initially provided for the end user. Organizations will have many opportunities to incorporate this cutting-edge technology into the existing product portfolio and different teams will have unique experiences when integrating it.

Additionally, different mechanisms will be available, depending on the nature of the product. Some teams will get an enhanced view into the impacts that generative AI technology makes on the software solutions. Because each team will strive to continuously improve, having a mechanism to share the learnings at an enterprise level will be valuable.

Another area for improved collaboration is to establish an AI Center of Excellence, much like the Cloud Center of Excellence, which enables the continuous definition of governance, promotes knowledge sharing, defines best practices as they emerge, and enables cross-functional collaboration around AI initiatives.

Level 5: Innovation

Use cases for Azure OpenAI

Innovation with generative AI must take an iterative approach that incorporates the lessons learned throughout the journey. During this early stage of the technology's deployment, innovation initially can be incorporated into existing software systems and improved over time. Here are some examples of innovative solutions that involve generative AI:

1. **Personalized customer experiences:** Organizations can gain contact center efficiency by incorporating Azure OpenAI into the feedback loop for human customer service representatives. By learning from the top performers, Azure OpenAI can democratize the approaches that these individuals take when working with external customers to improve the experience across the enterprise.

   Consider this scenario: Your organization has a long history of customer communications with varying satisfaction scores. Interactions with high satisfaction scores can be used to fine tune models that will represent the best possible outcomes for the customer. These also can provide the service representatives recommended responses that will increase satisfaction overall.

2. **Enhanced engineering copilot:** Organizations can improve their engineering teams' experience by leveraging copilot solutions — a software development process that's harnessed by AI and was implemented using the best practices defined by the enterprise architecture and principal engineers across the organization. These best practices will help train new recruits on the organization's standards and apply the standards uniformly as new solutions are built.

   Consider this scenario: Your organization has a library of software and DevOps solutions, which can be used to fine-tune a model targeted to new joiners of the organization. This helps to not only educate the new members but provides best practices approaches that have been refined throughout the software development lifecycle as well.

   Drawing on the experience of the developers who built the solutions, the copilot can make code recommendations that are consistent with the standards that have been previously reviewed and adopted. Thus, the wheel does not get reinvented by the next generation; recommendations are provided as a foundation that the team can learn from and incorporate new information over time.

3. **Streamlined human resource management:** Human resource teams can take advantage of indexing support resources to employees across the enterprise while providing an approachable and human-like interface for interactions. Routine tasks from the human resources department to impart policy information can be handled by a generative AI-enabled HR team to ease the workloads of the internal HR staff and provide employees the latest updates in the process.

   Consider this scenario: The human resources group within your organization has representatives who need to accomplish repetitive tasks. Adopting generative AI to support the HR team can reduce the workload on these tasks and improve the overall daily tasks of the team members. There also is a large library of information that the HR team needs to be aware of to service their customers — the employees of your organization. Deploying generative AI to process this data and provide natural language prompt-based responses can improve the experience and better tailor it to meet each individual's needs.

The possibilities for enhancing existing workflows to support typically human-intensive activities can be simplified and streamlined once your technical team has defined its enterprise standards and generative AI has been adopted. In addition, innovation can be applied iteratively to improve these solutions as soon as generative AI is integrated into these workflows and monitored by the experts in these areas.

Conclusion

Azure OpenAI offers numerous benefits to financial services companies, including improved decision-making, enhanced customer experiences, streamlined operations and personalized functionality with fine-tuned models and engineered prompts.

It's easy to get started in a sandbox environment given the tools that Microsoft has available in Azure OpenAI. Just remember to factor in the models' security and regulatory compliance requirements as the solution becomes more robust and moves to production and designate employees to review and intervene in the output to ensure accuracy, fairness and understanding.

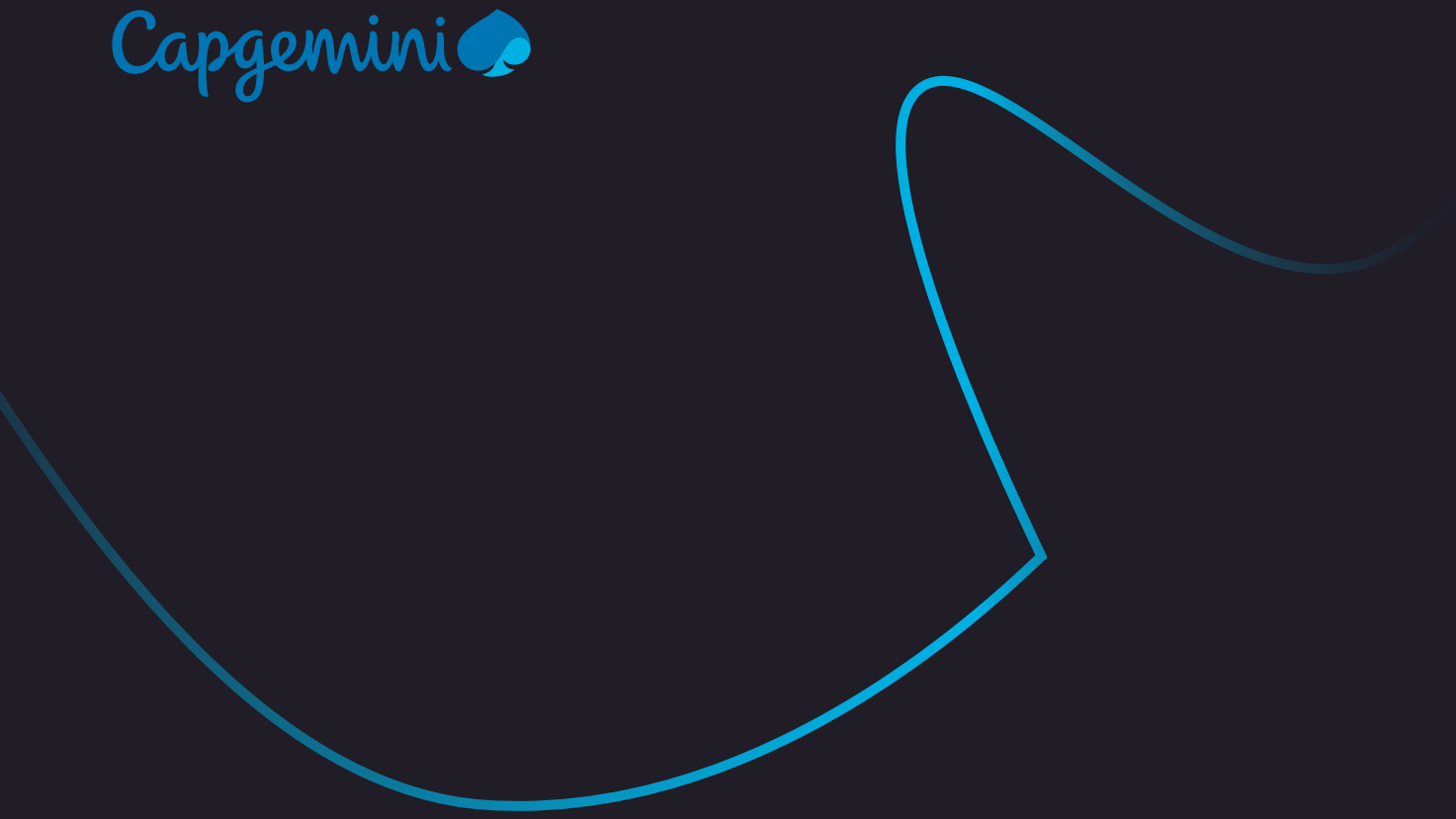
Once these considerations have been incorporated into the architecture and software development teams responsible to bring solutions to the market have leveraged Azure OpenAI across your organization, innovation can begin, and the final stage of maturity will have been reached.

How Capgemini can help

For guidance on customizing Azure OpenAI to meet your organization's needs, reach out to Capgemini. As your organization goes through all levels of implementation, Capgemini will provide insights along the way and can help you integrate AI into your business.

Contact us

Kieran Maltz
Director of Azure Center of Excellence, Financial Services
kieran.maltz@capgemini.com



About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of over 360,000 team members in more than 50 countries. With its strong 55-year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fueled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2022 global revenues of €22 billion.

Get the Future You Want | www.capgemini.com

Disclaimer