Magic Quadrant for Cloud AI Developer Services

29 April 2024 - ID G00796065 - 43 min read

By Jim Scheibmeir, Arun Batchu, **and 1 more**

Business leaders and boards of directors expect software engineering leaders and their teams to deliver the promised benefits of AI. Cloud AI developer services help deliver these benefits by offering an end-to-end platform to design, develop, deploy and monitor language, vision and tabular AI models.

Market Definition/Description

This document was revised on 14 May 2024. The document you are viewing is the corrected version. For more information, see the Corrections page on gartner.com.

Gartner defines cloud AI developer services (CAIDS) as cloud-hosted or containerized services and products that enable software developers who are not data science experts to use artificial intelligence (AI) models via APIs, software development kits (SDKs) or applications. Core capabilities include automated machine learning (autoML) including automated data preparation, automated feature engineering and automated model building, and model management and operationalization for language, vision and tabular use cases. Optional and important complementary capabilities include AI code models and assistants.

Cloud AI developer services help organizations embed intelligence, such as AI and ML insights, into their applications. While that is *what* cloud AI developer services offer, it is more important to note *how* they accomplish this. These services democratize and increase the availability of AI and ML to software engineers through the automation and features offered. Traditional activities regarding data acquisition, data quality, feature engineering, algorithm selection and model training are augmented by the technology. Cloud AI developer services open up a world of possibilities for software engineers to build AI and ML production capabilities and features for enterprise-built applications.

Must-Have Features

The must-have features for this market include:

- Tabular services: Using autoML services for tabular use cases, developers can create custom models. These services allow developers without significant ML or data science skills to customize the vendor-provided ML services or build purpose-specific ML using enterprise structured data.

- Language services: These services offer developers the ability to cluster documents by topic, analyze sentiment, and summarize and generate text. They can include natural language processing/understanding, speech to text, natural language generation using foundation models, text to speech, translation, sentiment analysis and text analytics.

- Vision services: These services utilize image and video technology as both input and output. Image labeling, segmenting and boundary definition are common capabilities. Other vision services can include image generation using foundation models, video AI and ML-enabled optical character recognition (OCR).

Standard Features

The standard features for this market include:

- Automated data preparation: These services can prepare datasets, such as text, tabular or image data, to be used for training the models. These services can cleanse and augment datasets from raw data provided by the organization and can include data visualization, text embedding and vector database integration.

- Automated feature engineering and model building: These models require developers to provide datasets that can be used to train the models, as well as metadata tags that have the attributes they want the models to be trained to identify and predict. Alternatively, feature engineering can automate data evaluation and metadata tagging. These services can recommend or select the potentially best algorithms that can be used to build and optimize the model.

- Model management/operationalization: These services may offer model factory functionality to automate the building of ML pipelines, including model training, deployment, and monitoring and management of models in production. This service helps to orchestrate across various models, APIs and databases such as retrieval augmented generation (RAG) and agent. The service may also provide for packaging and provisioning of infrastructure for the models created, such as Kubernetes containers. The services should also automate creation of the APIs needed to access the models. This capability allows developers to monitor models in test and production environments to assess model performance and drift. These services may provide automated retraining of the model and provision for replacing an outdated model with a better one. Additional features may include business key performance indicators (KPIs) for model value and the ability to assess the quality of data impacting the model performance. These capabilities may integrate with and enable broader MLOps platforms.

- Responsible AI: These services analyze datasets for potential bias, explainability, interpretability, validity and reliability. They can be used to filter both large language model (LLM) input and output for toxic or violent material. Responsible AI services and capabilities involve creating AI systems that are transparent, explainable, fair and unbiased. Responsible AI also aims to minimize harmful impacts and ensure privacy and security.

- Natural language understanding: A subset of natural language processing, natural language understanding deals with machine comprehension. It takes textual input and extracts metadata from the text.

- Speech to text: This is a subset of computational linguistics that converts analog input to text output. The text output can be the final product, or it can be entered into a natural language understanding model. Many computing devices, such as personal computers and smartphones, have a built-in automated speech recognition (ASR) capability.

- Text to speech: This service converts textual input into analog output/speech. It may offer customizing of the speech with unique vocabulary or applying different voicings such as accents.

- Natural language generation: This service creates humanlike text using LLMs trained on extensive sets of data. Text generation can be used in a wide variety of applications such as drafting emails, writing code, creating written content, translating languages and answering questions.

- Translation: This service takes text input from the source language and converts it to a target language. This is a challenging task because different languages have distinct structures. It is not as simple as translating a word from one language to the corresponding word in another language. Recent advancements in translation services utilize LLMs.

- Image recognition: This service normally identifies what objects or people are contained in an image. Some implementations can also enable multimodal-chat-based Q&A, identify other attributes in the image such as colors or patterns, or offer customizations for object identification.

- Video AI: This service normally combines image recognition and ASR to identify people and other objects in a video and to create a transcript for the audio. Some services also track the movement of people across multiple frames of the video.

- ML-enabled OCR: This service converts electronic images of typed, handwritten or printed text, or text in images or video, into machine-encoded text and adds metadata to the content. The service also uses ML to classify the information in a given field based on its content.

- Image/video generation: This service involves the creation of new, unique images and videos through algorithms and foundation models. This can be done either from inputs such as a series of images, text prompts or other forms of data. The generated images and video, which can be highly realistic or abstract in nature, are often used in a range of fields, including computer vision research, gaming, media and digital art.

- AI code assistance: This service helps developers in software development activities across the life cycle of an application. Developers prompt the assistants with natural language, code snippets and triggers to generate code blocks. The assistants also analyze, explain and refactor

code; generate documentation; help with testing activities; and translate between programming languages.

Optional Features

The optional features for this market includes:

- AI code assistants: These assistants help developers by leveraging foundation models trained on millions of lines of code. Developers prompt the assistants with natural language, code snippets and triggers to generate code blocks. They also analyze, explain and refactor code; generate documentation; and translate between programming languages. Assistants support multiple languages and integrate into programming environments such as code editors, command-line and chat.

Magic Quadrant

Figure 1: Magic Quadrant for Cloud AI Developer Services

Vendor Strengths and Cautions

Alibaba Cloud

Alibaba Cloud is a Challenger in this Magic Quadrant. Alibaba Cloud's CAIDS offering spans language, vision, tabular and AI coding assistant services. The company's Visual Intelligence Platform offers both image recognition and video generation capabilities. Alibaba Cloud is a subsidiary of Alibaba, a public company with 48 global offices that is headquartered in Hangzhou, China. Its operations are primarily in China and other Southeast Asia/Pacific countries. It has a limited presence outside of the Asia/Pacific region, but it has expanded its South America presence in the last year.

Strengths

- **Product:** Alibaba Cloud's CAIDS platform is noteworthy for its multimodal ability to generate video from text. This feature distinguishes it from many competitors and offers significant innovation and potential for users.
- **Sales execution and pricing:** Alibaba Cloud offers freemium tools, tiered plans and intuitive, consumption-based pricing. These pricing strategies make its services both accessible and affordable for a wide range of customers.
- **Overall viability:** With the backing of its parent company, Alibaba Group, Alibaba Cloud is well-positioned in the tech industry. It has a large workforce across major regions and makes substantial investments in R&D.

Cautions

- **Customer experience:** Alibaba Cloud could provide its customers with more comprehensive and user-friendly developer documentation. Its community outreach programs could also be improved, as Alibaba Cloud is not as effective at reaching customers as Leaders in this market.
- **Marketing execution:** Alibaba Cloud's marketing execution could be more cohesive and impactful. Its messaging does not fully communicate the value of its services or engage potential customers as effectively as possible. Alibaba Cloud needs to increase marketing around the security of its platform and the ease of developing AI agents.
- **Market understanding:** Alibaba Cloud needs to develop a more in-depth vertical industry understanding of its target audience's needs and preferences so it can provide more tailored product offerings and marketing strategies. Additionally, Alibaba Cloud needs more responsible AI capabilities, such as offering an automated bias detection service, which is still a roadmap item.

Amazon Web Services

Amazon Web Services (AWS) is a Leader in this Magic Quadrant. Its CAIDS offering, which includes Amazon SageMaker, Amazon Bedrock and Amazon CodeWhisperer, helps developers to use AI across the use cases of image, language, tabular and AI code assistance. AWS provides prebuilt and vertical-specific AI and ML solutions, such as AWS HealthScribe, within its SageMaker JumpStart library. AWS is headquartered in Seattle, Washington, and its operations are global.

Strengths

- **Geographic strategy:** AWS offers an extensive global cloud infrastructure. As of the end of 2023, AWS had 105 availability zones in 33 geographic regions. Its CAIDS are available in Europe, the Middle East and Africa (EMEA), the Americas and Asia/Pacific. Revenue for the AWS segment of Amazon increased 13% year over year to $90.8 billion for 2023, driven in part by its long-standing success at attracting developers across the world.

- **Vertical/industry strategy:** AWS showcases an excellent understanding of the developers' world by offering industry-specific tooling across many sectors, including finance, insurance, healthcare and manufacturing. This targeted approach allows AWS to meet the unique needs of clients in these industries effectively.

- **Overall viability:** AWS is backed by its parent company, Amazon, which boasted more than $578 billion in revenue and double-digit growth for its cloud business in 2023. This financial stability and growth trajectory bolster AWS' strong position in the market.

Cautions

- **Marketing execution:** To align with a partner-neutral approach, AWS highlights a choice of multiple models on Amazon Bedrock without highlighting its own Amazon Titan model, making Titan's positioning unclear within the overall large language model narrative.

- **Market understanding:** While AWS acknowledges the need for applying GenAI to the software development life cycle, it shows gaps in its market understanding. AWS must prioritize the emergence and development of AI agents so that organizations can leverage GenAI-powered orchestration across systems and processes.

- **Innovation:** AWS should better emphasize how its capable services already in market (such as Agents for Bedrock) and its free technologies (such as PartyRock, an Amazon Bedrock Playground) drive its overall vision for AI. This will make it more understandable to customers.

Google

Google is a Leader in this Magic Quadrant. Its CAIDS offering, Vertex AI, provides language, vision, video, tabular data and AI code assistant services. Google offers multiple LLMs, including multimodal models and model families for enterprise workloads. Its vertical-specific offerings comprise models, AI applications and APIs designed for specific industry use cases. These include MedLM, a family of LLMs for medical use cases; Medical Imaging Suite, an application to accelerate imaging diagnostics with AI; and Healthcare APIs to ingest data in standardized healthcare formats and unstructured text. The SecLM model helps to power the Google Cloud Security AI Workbench product. Google's headquarters is in Mountain View, California, and its operations are global.

Some of Google's recent advancements, such as the release of Gemini 1.5 models and its open-source Gemma models, occurred after our research cutoff date and were **not** *included in the scoring for this Magic Quadrant.*

Strengths

- **Product:** Google's Vertex AI stands out for its extensive language support offered through a portfolio of prebuilt APIs and capabilities for summarization, classification and generation of natural language via Google's language models. For example, Google's Translation API remains popular. Vertex AI also excels in its vision capability, with features like creating custom entity labels with video intelligence and gaining near-real-time insights with streaming video annotation and object-based event triggers.

- **Market responsiveness:** Google focuses on key areas of this market, such as Model Garden for model discoverability, retrieval augmented generation (RAG) and search, preconfigured Colab Enterprise notebooks, services such as indemnity protection, data privacy, and responsible AI with functionality like safety filters and citation metadata. Google's sharp focus in these areas shows that it is committed to, and capable of, meeting the evolving needs of developers.

- **Overall viability:** Google is backed by Alphabet, its parent company. Alphabet's 2023 financial statement showed impressive growth, including a double-digit growth rate for Google Cloud. Google's investment and expertise in its research organization, Google DeepMind, enables the company to offer differentiated innovation at every layer in the stack.

Cautions

- **Customer experience:** While Google offers support for customers via Google Groups and Slack channels, multiple support tiers and 24/7 support, its "service and support" score on Gartner's Peer Insights platform is below other vendors in this Magic Quadrant. Also, Google's product portfolio is vast and comprehensive, as Google owns and develops IP at every layer of the stack, which can make it difficult to navigate. Organizations that are adept at assembling these capable "puzzle pieces" are well-positioned to create value from the stack, while organizations that are less experienced with Google Cloud products may find it challenging.

- **Vertical/industry strategy:** Google's approach to vertical-specific solutions is to provide a portfolio of models, data platforms, AI applications and APIs designed for specific industry use cases. It has some strong offerings, such as SecLM and MedLM, along with AI applications and industry-specific APIs.

- **Marketing execution:** Google Cloud has consistently delivered on its CAIDS strategy and portfolio for the enterprise buyer. Google's other consumer-oriented innovations, such as Bard (now Gemini), are maturing.

H2O.ai

H2O.ai is a Visionary in this Magic Quadrant. Its CAIDS offering, H2O AI Cloud, provides end-to-end AI life cycle development, management and deployment of AI models and applications. The company supports flexible deployment options across its language, vision and tabular services. H2O.ai is a major open-source contributor. Its open-source h2oGPT platform enables developers to host local LLMs and vector databases. h2oGPT is available on GitHub and includes access to LLMs, embedding capabilities, a vector database, and both a command-line interface (CLI) and UI

for developers to use in building AI-powered applications. H2O.ai is headquartered in Mountain View, California, and its operations and support are consistent across regions.

Strengths

- **Geographic strategy:** H2O.ai provides broad geographic coverage via its open-source tools and by offering customers the flexibility to decide where models and applications are deployed, such as on-premises, hosted by H2O.ai, or deployed to other cloud providers. H2O.ai also supports a moderate number of languages in its speech-to-text and text-to-speech services.

- **Operations:** H2O.ai allows customers to choose when to adopt new versions of APIs and services, instead of mandating updates. It also offers 24/7 support for enterprise customers so they can further enhance their operational efficiency.

- **Vertical/industry strategy:** H2O.ai has go-to-market teams aligned to support verticals and use cases by verticals. The company has launched initiatives to serve specific industries. It introduced in-person GenAI Masterclass, with certification training days tailored for financial services/banking and public sector verticals, and plans to roll out more. H2O.ai launched a public GenAI App Store and a Government GenAI App Store curated with public-sector-specific apps. H2O.ai's approach to serving these industries underscores its commitment to meeting the unique needs of customers.

Cautions

- **Overall viability:** H2O.ai generates less revenue than the large cloud providers in this market. It has fewer resources to invest in R&D, and it has limited staff across its supported regions. These constraints may impact its ability to compete effectively. However, its focus on innovation through open source enables it to compete and deliver AI technologies.

- **Customer experience:** H2O.ai's documentation is not comprehensive for all of its service offerings. Also, customers could do with access to a wider range of communication methods for customer support and engagement than H2O.ai currently offers.

- **Innovation:** H2O.ai does not offer plug-ins to Visual Studio Code or other common integrated development environments (IDEs), but it does provide an AI code assistance experience through its Jupyter notebook lab (JupyterLab) and sketch2app technology. The H2O GenAI App Store has a relatively limited number of available apps. H2O.ai needs to deliver more innovation in these areas to stay competitive in the market.

Huawei Cloud

Huawei Cloud is a Niche Player in this Magic Quadrant. Its CAIDS offering, ModelArts, provides language, vision, tabular and AI code assistant services. Huawei Cloud's language, vision, and tabular services and models can be hosted in its cloud, at the edge or deployed on-premises. It offers the Pangu family of industry-specific foundation models, as well as enterprise knowledge management and intelligent decision making. Huawei Cloud is one of the few vendors in this evaluation that offers language services for Arabic. Based in Guangdong, China, Huawei Cloud's

operations are primarily concentrated in China, with some presence in Southeast Asia and recent investments in EMEA.

Strengths

- **Market responsiveness:** Huawei Cloud differentiates its platform by offering strong vision services capabilities for emotion detection and the ability to generate answers to questions about images using its Pangu multimodal models. Huawei Cloud has also released some AI code assistant features, such as code models, plug-ins and an IDE; additional features are in beta.

- **Operations:** Huawei Cloud offers 24/7 customer support, tiered support plans and strong SLAs. This operational efficiency ensures that customers receive timely and effective support.

- **Vertical/industry strategy:** Huawei Cloud's Pangu models are targeted at the finance, government and mining sectors. Each model has a single architecture composed of five foundational models — natural language processing (NLP), computer vision, multimodal, prediction and scientific computing — with additional layers of industry-specific models. Huawei Cloud is also a thought leader across numerous industries, as demonstrated by its healthcare innovation summit, autonomous driving summit, and initiatives in energy and government.

Cautions

- **Business model:** Huawei Cloud's platform lacks some privacy features that enterprises expect from AI solutions — these are, however, being addressed in beta and as part of a strategic product roadmap. At the cutoff date of this research, the Huawei Cloud AI code assistant code citations, and protection from code suggestions that originate from copyrighted code features, were still under development with beta capabilities.

- **Innovation:** Huawei Cloud supports 22 languages and four dialects (Shanghainese, Sichuanese, Cantonese and Egyptian Arabic) for its speech-to-text service; multiple vendors offer the service for over 100 languages. Huawei Cloud currently lacks sufficient multimodal capabilities for video, although its text- and image-to-video generation service with stitching and editing is set to be released in June 2024.

- **Marketing execution:** Huawei Cloud must communicate the value of its offerings more effectively to improve its ability to engage potential customers. In particular, it needs to increase investment in promotion of its services, available APIs, multimodal enhancements, security and AI agents to match the capabilities and breadth of those technologies.

IBM

IBM is a Leader in this Magic Quadrant. Its CAIDS offering, watsonx, combines language, vision, tabular data, and AI code assistance services for building and deploying AI solutions in hybrid, multicloud environments. IBM continues to invest in the watsonx platform in areas such as AI governance, automation and support for enterprise-level integrations, as well as in products such as watsonx Code Assistant. Based in Armonk, New York, IBM's operations are global.

Strengths

- **Geographic strategy:** IBM's diverse customer base and robust global presence are commendable. Its technology solutions reach across different regions, effectively meeting customers' unique needs and regulatory requirements.

- **Marketing strategy:** IBM's marketing strategy effectively communicates the value of its offerings, which attract a diverse range of customers. IBM achieves this by focusing on content creation, actively participating in events and communities, contributing to open source and supporting cloud neutrality.

- **Overall viability:** IBM is a large, diversified company with a large team of support and product staff across the world. Revenue growth for fiscal year 2023 came in at 2.2%. IBM has the resources to continue providing reliable services for its global customer base.

Cautions

- **Product:** IBM watsonx Code Assistant products are currently limited to targeted use cases (COBOL to Java, and Ansible) when compared with the capabilities of leading AI code assistants. watsonx Code Assistant was purpose-built for COBOL, Java coding and modernization efforts, and for Ansible configuration. Also, IBM's language translation service only supports a subset of the languages and dialects that other Leaders in this Magic Quadrant offer.

- **Innovation:** IBM's innovation is not consistent across its services and capabilities, particularly in the area of vision services where image and video generation are lagging. These feature gaps limit the applicability of IBM's solutions for clients with use cases such as creative work or synthetic test data generation of images.

- **Market understanding:** To date, IBM has been focused on offering smaller, fit-for-purpose foundation models that address challenges around cost, as well as concerns related to sustainability and accessibility. The decision to not focus on large, general purpose, multimodal models may be a missed market opportunity.

Microsoft

Microsoft is a Leader in this Magic Quadrant. Its CAIDS offerings are Azure AI platform for language, vision and tabular use cases and GitHub Copilot, which has captured much of the early market for AI code assistants. Microsoft's platform offers its partner ecosystem and enterprise customers rich solutions, such as support for spoken languages and dialects in its language tools, as well as video analysis and image model customization in its vision service. Based in Redmond, Washington, Microsoft's operations are global.

Strengths

- **Product:** Microsoft's Azure AI platform is a feature-rich product. It excels in the tabular use case by providing extensive database support, integrations and a data fabric. Microsoft also differentiates itself with GitHub Copilot, which offers content filtering to help ensure that the generated code does not infringe on intellectual property.

- **Geographic strategy:** Microsoft's global reach and presence in China set it apart from most U.S.-based tech companies. It offers some Azure AI services in China, including Azure AI Vision, Azure AI Language and Azure AI Anomaly Detector. Microsoft's strong support for many languages and dialects further enhances its position.

- **Overall viability:** Microsoft's Intelligent Cloud group has demonstrated double-digit growth in 2023. This upward trend underscores the strength and potential of its Azure AI platform.

Cautions

- **Market understanding:** Microsoft does not offer a comprehensive portfolio of vertical and industry solutions when compared with other Leaders in this market. Microsoft could also do more to help customers learn how to integrate AI into applications by democratizing these skills in its services.

- **Marketing execution:** Microsoft's messaging could be more effective at emphasizing its developer education opportunities, product integrations, AI services, APIs and operations. To successfully engage potential customers, Microsoft needs to clearly communicate the value of these offerings.

- **Innovation:** While much of Microsoft's model innovation is based on third-party foundational models, it helped usher in the generative AI wave and has built first-party models such as Phi-2 and Florence. Microsoft needs to invest more resources in innovating its first-party models and services. By focusing its efforts in these areas, Microsoft can enhance its competitive edge.

OpenAI

OpenAI is a Visionary in this Magic Quadrant. Its CAIDS offerings include both models and services such as GPT-3.5 and GPT-4 with Vision, ChatGPT, DALL-E and Whisper. OpenAI has established extensive venture capital and rapid adoption of its technology with startups, enterprises and through partnerships. OpenAI is headquartered in San Francisco, California, and its operations are global.

OpenAI is previewing a red team version of its text-to-video technology, Sora, but this development occurred after our research cutoff date. Thus, it was **not** *included in the scoring for this Magic Quadra*nt.

Strengths

- **Market understanding:** OpenAI's vision is a driving force in this market. Its innovation, particularly in the areas of LLMs, such as GPT-4 with Vision, agents and democratization, has set the stage for the last two years. OpenAI Playground has also set a benchmark for other vendors in this market.

- **Geographic strategy:** OpenAI is supported in most countries and is continuing to expand its availability for customers across a wider range of locations. Its partnership with Microsoft will accelerate OpenAI's efforts to expand its geographic reach via the Azure platform.

- **Marketing strategy:** OpenAI has made many headlines during the past two years, which has created an unshakeable mind share among developers. In particular, its product releases and technical demonstrations for ChatGPT and DALL-E image generation captured the attention of the developer world. OpenAI continues to demonstrate thought leadership among developers through its DevDay conference.

Cautions

- **Customer experience:** OpenAI offers few methods for customers to request support, and current and potential customers have cited concerns about its response times. However, OpenAI has been building out its customer success and solution teams, and now staffs enterprise accounts with customer service managers and technical support, such as account engineers and solutions architects. On Gartner's Peer Insights platform, OpenAI received below-average customer experience scores from users, especially with regard to its associated costs.

- **Vertical/industry strategy:** OpenAI does not provide industry-specific solutions. As an alternative, OpenAI offers customers the option to fine-tune its foundational models with industry-specific knowledge or create a new custom-trained model that is proprietary to the customer, but this process requires customer effort and resources. Additionally, OpenAI is not currently listed in the FedRAMP database, which may restrict its use in U.S. government organizations.

- **Marketing execution:** OpenAI offers a novel approach using function calls for its tabular data capabilities, but it should do more to help customers understand how its tool can help them unlock opportunities from structured data. OpenAI also needs to do more to promote its open-source contributions, such as its LLM evaluation technology, OpenAI Evals.

Oracle

Oracle is a Niche Player in this Magic Quadrant. Its CAIDS offering, Oracle Cloud Infrastructure (OCI) AI, offers vision, image, tabular and AI code assistance. In 2023, Oracle released new services for language that use generative AI and also released an AI code assistant for SQL for database administration and querying. Oracle's ML-enabled optical character recognition (OCR) service supports all basic functionalities, including table comprehension and key value extraction across document and file formats. Oracle is headquartered in Austin, Texas, and its operations are global.

Strengths

- **Customer experience:** Oracle offers free training and certifications for the community, along with reference architectures and tech exchange articles. It also provides customers with a Slack channel for communication, strong SLAs and 24/7 support, all of which enhance the customer experience.

- **Geographic strategy:** Oracle offers strong support in every region except China for most of its services. Potential customers should verify if Oracle's recent GenAI offerings are available in their region.

- **Overall viability**: Oracle demonstrated double-digit growth in revenue for 2023 year over year. This positive momentum underscores the strength and potential of its OCI AI services.

Cautions

- **Product**: Oracle should improve its language services by investing more in first-party models and expanding availability across more cloud regions. It also needs to develop more capabilities for its AI code assistant to support more programming languages, software development use cases and integrations into common development environments.

- **Market understanding**: Oracle does not offer text-to-speech capabilities in its language services. Also, Oracle has yet to establish strong video analysis and image generation capabilities in its vision services.

- **Business model**: Oracle's product roadmap and use cases for AI are currently selective, rather than broad and uniform across its industries. Oracle needs to deliver a wider range of targeted solutions to meet the unique needs of customers in different industries.

Tencent Cloud

Tencent Cloud is a Niche Player in this Magic Quadrant. Its CAIDS offering, Tencent AI, provides an AI code assistant, vision, language and tabular data services. Its AI vision capabilities have been accelerated by its knowledge from and access to its WeChat social media superapp. Tencent Cloud has increased its AI capabilities for vertical-specific use cases, with solutions for child care, cybersecurity and manufacturing. Tencent Cloud is headquartered in Shenzhen, China, and its operations are mostly in China. It is growing its customer base in EMEA and North America, but does not offer its platform and services in South America.

Strengths

- **Sales execution/pricing**: Tencent Cloud demonstrates strong sales execution, and the pricing of its AI services are competitive. Tencent Cloud succeeds in making its offerings accessible and attractive to a wide range of customers.

- **Operations**: Tencent Cloud offers its customers the flexibility to decide when to adopt new versions of AI services. It also provides 24/7 support to meet customer needs promptly and professionally, and has established SLAs for its AI services to ensure that these are delivered reliably.

- **Market responsiveness**: Tencent AI delivers text-to-image and image-to-image capabilities, facial detection in video analysis and one-click deployment of LLM-enabled apps. These innovative features demonstrate its ability to respond to market trends and customer needs.

Cautions

- **Market understanding**: Tencent Cloud needs to place more emphasis on AI agents and verticals. While it offers capabilities for some vertical-specific use cases, Tencent Cloud has room to expand its portfolio of industry solutions. Tencent Cloud has already begun focusing on improving its responsible AI and safety to better meet the evolving needs of the market.

- **Innovation:** Tencent Cloud needs to develop additional capabilities in bias detection and responsible AI. Although it has not been implemented yet, Tencent Cloud is preparing to provide customers with the capability of video generation through its foundation model, Hunyuan.
- **Marketing execution:** Tencent Cloud's positioning of its services to the developer community recognizes the importance of developer experience, but there is still room for improvement. Tencent Cloud could better emphasize its platform's intuitive capabilities and its integrations. Tencent Cloud also needs to improve its messaging on how its technology democratizes the integration of AI/ML into application development.

Vendors Added and Dropped

We review and adjust our inclusion criteria for Magic Quadrants as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant may change over time. A vendor's appearance in a Magic Quadrant one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. It may be a reflection of a change in the market and, therefore, changed evaluation criteria, or of a change of focus by that vendor.

Added

- OpenAI

Dropped

- **Baidu:** Baidu was dropped because its geographic reach of customers across regions did not satisfy inclusion criteria.
- **Clarifai:** Clarifai was dropped because its geographic reach of customers across regions did not satisfy inclusion criteria.

Inclusion and Exclusion Criteria

To qualify for inclusion in this Magic Quadrant, each provider needed to:

- Demonstrate a go-to-market strategy that focuses on the persona of a professional software developer.
- Include all the must-have capabilities of a CAIDS platform as per the Market Definition.
- Have generated at least $15 million in revenue from its CAIDS offerings (excluding professional services revenue) in 2023.
- Have at least 15 paying customers (logos) as of 1 December 2023 for its cloud AI developer services in at least three of the following regions:
  - North America
  - South America
  - Europe, Middle East and Africa

- China
- Central Asia
- Asia/Pacific

We excluded any vendor that:

- Offered CAIDS only as part of a professional services contract, where the services are used exclusively by the vendor's consultants.
- Markets and messages exclusively to target users who are not professional software developers.

Honorable Mentions

Baidu: Baidu is a prominent technology company based in China, known for its internet-related services and AI products. It offers a comprehensive suite of AI solutions, including a cloud-based AI platform, autonomous driving technologies and advanced search algorithms. Baidu was not included because its geographic reach of customers across regions did not satisfy inclusion criteria.

Clarifai: Clarifai provides a powerful AI platform to build enterprise AI faster. It specializes in modern AI technologies, including computer vision, LLMs, RAG, data labeling, inference, NLP, automatic speech recognition, and more. Clarifai was not included because its geographic reach of customers across regions did not satisfy inclusion criteria.

Evaluation Criteria

Ability to Execute

The Ability to Execute criteria used in this Magic Quadrant are as follows (for the sources of information that informed Gartner's evaluations using these criteria, see the Evidence section):

- **Product or Service:** This criterion assesses the competitiveness and success of a vendor's CAIDS offering(s) with regard to the critical capability areas, in light of the vendor's RFP response and video submission. Product weight is high due to the critical end-user value generated by the ability to learn, adopt and leverage tabular, language, and vision AI and ML services into innovative and intelligent enterprise applications and systems. The AI code assistant products are a compelling area offering improved developer experience and productivity.
- **Overall Viability:** This criterion concerns the organization's financial status and business model as it relates to CAIDS. It also takes account of existing and prospective customers' views about the vendor's likely future relevance.
- **Sales Execution/Pricing**: This criterion covers the vendor's capabilities in sales activities. It includes the overall evaluation and contract negotiation/flexibility with a vendor, as well as the

value the customer receives.

- **Market Responsiveness/Record:** This criterion addresses the extent to which a vendor has momentum and success in the worldwide market. Software engineering leaders' expectations have grown quickly with the recent hype in generative AI technology.
- **Marketing Execution:** This criterion reviews how vendors target software development personas, execute and communicate regarding trends and product capabilities, and leverage communities, events, social media and partnerships to drive awareness.
- **Customer Experience:** This criterion concerns customers' experience of working with a vendor after a purchase. Factors include the availability of quality third-party resources (such as integrators and service providers), the quality and availability of end-user training and certification, and the quality of the peer-user community.
- **Operations:** This criterion concerns how well a vendor supports its customers, and how trouble-free its software is.

Table 1: Ability to Execute Evaluation Criteria

Evaluation Criteria ↓	*Weighting* ↓
Product or Service	High
Overall Viability	Medium
Sales Execution/Pricing	Medium
Market Responsiveness/Record	Medium
Marketing Execution	Medium
Customer Experience	Medium
Operations	Medium

Source: Gartner (April 2024)

Completeness of Vision

The Completeness of Vision criteria used in this Magic Quadrant are as follows (for the sources of information that informed Gartner's evaluations using these criteria, see the Evidence section):

- **Market Understanding:** This criterion concerns how closely aligned a CAIDS vendor is with the shifting needs of software engineering leaders and how widely its customers use recent and emerging capabilities. Due to the interest in generative AI technology, the weighting for market understanding increased from medium to high with this iteration of the Magic Quadrant research.
- **Marketing Strategy:** This criterion considers whether a vendor has a clear set of messages that communicate its value and differentiation in the CAIDS market, and whether that vendor is generating awareness of its differentiation.
- **Sales Strategy:** This criterion concerns the extent to which a vendor's sales approach benefits from a range of options and drivers that encourage customers to evaluate its CAIDS offering.
- **Offering (Product) Strategy:** Gartner evaluates a vendor's ability to support key trends that will create business value in the future. Existing and planned products and functions that contribute to these trends are factored into each vendor's score for this criterion, based on its presented roadmap. Vendor roadmaps offered distinction and showed separation, and this criterion is weighted highly due to the fast pace of change and need for prioritized execution in this market.
- **Business Model:** Gartner evaluates a vendor's evolution of their business model, revenue and indications of market share, interaction in open-source technology and community, and partner ecosystems.
- **Vertical/Industry Strategy:** This criterion assesses how well a vendor can meet the needs of various industries through templates or packaged data and analytics content. Vendors have been investing in their vertical strategies as organizations are seeking AI built to solve problems within their specific business domains. Thus, this is a highly weighted criterion.
- **Innovation:** This criterion gauges the extent to which a vendor is investing in generative AI capabilities, extending the value of its autoML technology and delivering unique capabilities. It considers whether a vendor is setting standards for innovation that others are emulating. Due to the speed at which innovation is occurring in this market, this is a highly weighted criterion.
- **Geographic Strategy:** This criterion considers how well-represented a vendor is around the world.

Table 2: Completeness of Vision Evaluation Criteria

Evaluation Criteria ↓	*Weighting* ↓
Market Understanding	High
Marketing Strategy	Medium
Sales Strategy	Medium
Offering (Product) Strategy	High
Business Model	Medium
Vertical/Industry Strategy	High
Innovation	High
Geographic Strategy	Medium

Source: Gartner (April 2024)

Quadrant Descriptions

Leaders

Leaders have robust offerings in all three key service areas: language, vision and tabular. Their CAIDS offerings are accessible via APIs and do not require developers to have data science expertise. Leaders also provide supporting capabilities to enhance their core services, including automated bias detection and mitigation, feature engineering, NLP, image labeling, MLOps, and AI that is explainable and interpretable. Leaders serve multiple regions and support multiple languages.

Challengers

Challengers are typically large businesses with substantial assets. Challengers have the resources to invest in developing their CAIDS services, but may lack the ability to articulate and map their

product vision to market needs. Challengers may operate regionally or globally. They may even dominate in one region.

Visionaries

Visionaries aspire to compete in the CAIDS market and have the resources to succeed, but they have yet to deliver a competitive portfolio of services. Visionaries often offer only a subset of the overall services needed in the market and intend to expand their services to compete with Leaders.

Niche Players

Niche Players typically focus on a narrower range of AI services than Leaders and Challengers. In some cases, they are smaller businesses that have limited resources to invest in their services and in expanding beyond their home region. In other cases, they are larger organizations that struggle to create and execute on a compelling product roadmap.

Context

Software engineering leaders should conduct a comprehensive evaluation of vendors in the cloud AI developer services (CAIDS) market across multiple dimensions. These include:

- **Responsible AI:** In line with rising customer expectations, most CAIDS vendors are significantly investing in enhancing responsible AI components.
- **Model portfolio management:** Vendors are progressively transforming AI capabilities into platforms that facilitate governance, reusability, scalability and auditing.
- **Prebuilt models and customizability:** CAIDS vendors are broadening their array of prebuilt models for specific verticals to industrialize AI/ML-augmented software. They are also delivering more API-first models for integration.
- **Model authoring techniques:** Vendors are both utilizing and contributing to open-source and open-data libraries, providing software developers with model authoring capabilities across command line interface (CLI), SDK and low-code approaches.
- **Deployment flexibility:** Vendors vary in the flexibility they offer, with some providing extensive flexibility and others imposing restrictions on where teams can deploy models.
- **Developer experience:** AI code assistants offered by vendors differ in terms of usage metrics, language capabilities across summarization and generation, task capabilities such as testing assistance, and levels of integration across common development environments.

Key differentiators between vendors include their maturity level in responsible AI, generative AI and multimodal capabilities, language and geographic support, vision services for video, and flexibility of model deployment.

When selecting cloud AI developer services, software engineering leaders should prioritize vendors that excel in providing explainable, transparent models with built-in bias detection.

Utilizing CAIDS with these features aids teams in building responsible and ethical AI solutions. The entire life cycle of a model necessitates development time and ongoing investment. Software engineering teams will need to enhance their skills and fluency with these tools, including in areas such as data acquisition, feature engineering, model authoring and model performance improvements. Furthermore, software engineering leaders and their teams must consider the importance of indemnification for potential IP rights violation through the use of generative AI.

Beyond data acquisition and model authoring, software engineering leaders must facilitate their teams in acquiring the skills necessary to support ModelOps. As developers become more conversant with AI and ML models and the functionality that can be added to applications, they will increasingly assume some or all of the responsibilities of ModelOps. This includes:

- Monitoring and managing ML models in the applications where they are deployed.
- Extracting information from, or adding metadata to, unstructured text or data assets.
- Using autoML services to build ML models that suggest next best actions.
- Classifying and automating document processing, or automating decision making for business workflows.
- Evaluating models for ethical concerns and mitigating biases.
- Working within a larger community to scale AI in the enterprise.

Software engineering leaders must also educate themselves and other stakeholders on the need for responsible AI, as well as on the ethical and privacy concerns surrounding AI use. They should spearhead efforts to establish a task force responsible for AI privacy, security and risk.

Market Overview

New advancements in generative AI have sparked a massive surge of investment in AI technologies. By 2027, Gartner projects that spending on AI software will grow to $297.9 billion, with a CAGR of 19.1%. [1]

As organizations spend more on these technologies, boards of directors and C-suite executives expect software engineering leaders and their teams to deliver AI-powered applications that effectively meet the needs and desires of users.

Software engineering leaders are finding it difficult to meet this expectation, as **few software engineers have the skills required to embed AI/ML capabilities into applications.** In the Gartner Software Engineering Survey for 2024, 55% of software engineering leaders said that AI/ML engineer was the most in-demand role in their organization. In that same survey, 65% of software engineering leaders said that applying AI/ML to applications was a high- or medium-level pain point. [2]

Software engineering leaders are in dire need of tools that empower their teams to implement AI/ML capabilities into applications. **Cloud AI developer services (CAIDS) help to bridge the skills**

gap.

CAIDS complement and augment the skill sets of software engineers, enabling them to build applications with predictive, intelligent features that will delight their customers. CAIDS provide capabilities such as:

- Code generation and explanation via models with APIs or through experiences such as IDE plug-ins
- Data preparation to correct data quality concerns, generate missing values when possible, and identify characteristics that prevent responsible AI
- Feature engineering
- Automated ML model building by offering coded and no-code approaches that iterate through algorithms to build and choose the best-performing models
- Model deployment ranges across hosted, on-premises and edge, and monitoring for drift in the model behavior
- LLM fine-tuning and grounding to increase the relevance and accuracy of responses, and to further the use cases where this technology can be implemented
- AI agents that bridge intelligence and actions to assist in customer and employee activities

2024 Advancements in Cloud AI Developer Services

In 2023 and 2024, CAIDS vendors have announced new multimodal capabilities, partnerships and AI code assistants that will continue to revolutionize this market (see Evidence section for more details). [3,4,5,6]

In particular, CAIDS vendors have expanded their repositories of **industry-specific ML models**, while also providing more support for these solutions. More vendors are offering prebuilt models for a wider range of industries, including education, life sciences, professional services, transportation, and banking and finance.

CAIDS vendors are also improving their capabilities across four key use cases:

- **Language:** Vendors are developing LLMs that can deliver an expanded range of high-quality language services. Major cloud vendors are developing proprietary language models, while smaller vendors are using open-source software, data and ML models to try to compete.
- **Vision:** Nearly every CAIDS vendor has improved its vision capabilities. The prolific use of computer vision in China has driven immense improvements among China-based vendors in the past year. OpenAI's red team release of Sora in 2024 occurred outside of the research cutoff date, but underlines the vision and roadmaps of technology providers in this market.

- **Tabular data:** CAIDS vendors are expanding their capabilities in the realm of tabular services, such as increasing data sourcing integrations across native and popular data lake and data fabrics, and releasing compelling services such as synthetic data generation. This allows developers without significant ML or data science skills to customize vendor-provided models or build purpose-specific ML using structured data from the enterprise systems they currently maintain.

- **AI code assistants and code models:** CAIDS vendors are transforming the software development landscape by releasing AI-powered tools like code assistants and code models. Code assistants serve as ready-to-use tools that enhance code efficiency. Code models are LLMs that have been fine-tuned with code data, acting as a springboard for the creation of custom, innovative applications. These AI tools automate repetitive tasks, bridge the skills gap among developers and integrate with developer workflow tools.

Due to the rapid growth, diversification and innovation in the CAIDS market, **Gartner plans to retire the CAIDS market and introduce multiple new Magic Quadrants to cover this technology market**. This includes an upcoming Magic Quadrant for AI Code Assistants later in 2024.

The explosive growth of AI code assistants is especially noteworthy. As of 2024, with 28 million professional software engineers worldwide, the global addressable market for AI code assistants stands at an annual recurring revenue (ARR) of $10.08 billion, with only 14% of the professionals using them. With the global software developer population estimated to grow at an annual rate of 10%, the addressable ARR will be close to $12.20 billion by the end of 2025. [7]

Evidence

[1] Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide

[2] **Gartner Software Engineering Survey for 2024.** This survey was conducted to identify the most important roles and skills for software engineering leaders and the change in their demand and importance since last year; and to understand how talent is sourced generally and for acquiring necessary AI/ML skills, and what tools are seen to increase developer productivity and the metrics used to measure them. The survey examines how software engineering leaders anticipate change in their operating budgets and the cost management steps taken. It further aims to identify the quality and testing techniques and programming languages software engineering leaders currently use and/or plan to use; their frequency of usage of UX design, user research and AI in generating components of user experience; and AI's impact on user satisfaction, accessibility and usability. It also intends to understand the software engineering leaders' responsibilities they find most difficult; the career paths available for senior-level individual contributors and how they are set up; how organizations attract and retain top performers in those career paths; and what management training is offered to staff.

The survey was conducted online from October through December 2023 among 300 respondents from the U.S. (n = 241) and U.K. (n = 59). Qualifying organizations operated in multiple industries (excluding the IT software industry and education sector) and reported enterprisewide revenue for fiscal year 2022 of at least $250 million or equivalent, with 63% over $1 billion in revenue. Qualified

participants were highly involved in managing software engineering/application development teams and the activities they perform.

Disclaimer: Results of this study do not represent global findings or the market as a whole but reflect sentiment of the respondents and companies surveyed.

[3] **Microsoft and Mistral AI Announce New Partnership to Accelerate AI Innovation and Introduce Mistral Large First on Azure**, Microsoft.

[4] **Amazon Bedrock Adds Mistral AI Models**, AWS.

[5] **AI-Powered Coding, Free of Charge With Colab**, Google.

[6] **EMO: Emote Portrait Alive — Generating Expressive Portrait Videos with Audio2Video Diffusion Model under Weak Conditions**, Alibaba Cloud.

[7] Gartner uses public sources of information and works with software vendors to establish estimates for the market. Information from Gartner's secondary research and internal community meetings has also been used to arrive at certain conclusions. The data is published as Gartner estimates and opinion, not as facts.

Evaluation Criteria Definitions

Ability to Execute

Product/Service: Core goods and services offered by the vendor for the defined market. This includes current product/service capabilities, quality, feature sets, skills and so on, whether offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

Overall Viability: Viability includes an assessment of the overall organization's financial health, the financial and practical success of the business unit, and the likelihood that the individual business unit will continue investing in the product, will continue offering the product and will advance the state of the art within the organization's portfolio of products.

Sales Execution/Pricing: The vendor's capabilities in all presales activities and the structure that supports them. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel.

Market Responsiveness/Record: Ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This criterion also considers the vendor's history of responsiveness.

Marketing Execution: The clarity, quality, creativity and efficacy of programs designed to deliver the organization's message to influence the market, promote the brand and business, increase awareness of the products, and establish a positive identification with the product/brand and organization in the minds of buyers. This "mind share" can be driven by a combination of publicity, promotional initiatives, thought leadership, word of mouth and sales activities.

Customer Experience: Relationships, products and services/programs that enable clients to be successful with the products evaluated. Specifically, this includes the ways customers receive technical support or account support. This can also include ancillary tools, customer support programs (and the quality thereof), availability of user groups, service-level agreements and so on.

Operations: The ability of the organization to meet its goals and commitments. Factors include the quality of the organizational structure, including skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently on an ongoing basis.

Completeness of Vision

Market Understanding: Ability of the vendor to understand buyers' wants and needs and to translate those into products and services. Vendors that show the highest degree of vision listen to and understand buyers' wants and needs, and can shape or enhance those with their added vision.

Marketing Strategy: A clear, differentiated set of messages consistently communicated throughout the organization and externalized through the website, advertising, customer programs and positioning statements.

Sales Strategy: The strategy for selling products that uses the appropriate network of direct and indirect sales, marketing, service, and communication affiliates that extend the scope and depth of market reach, skills, expertise, technologies, services and the customer base.

Offering (Product) Strategy: The vendor's approach to product development and delivery that emphasizes differentiation, functionality, methodology and feature sets as they map to current and future requirements.

Business Model: The soundness and logic of the vendor's underlying business proposition.

Vertical/Industry Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including vertical markets.

Innovation: Direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

Geographic Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography, either directly or through partners, channels and subsidiaries as appropriate for that geography and market.