



**Amazon's response to the
European Commission's call
for contributions on
Competition in Generative AI**

March 11, 2024



Amazon welcomes the European Commission’s call for contributions and its open engagement in the initial stages of policy thinking on generative AI. While it is still early days in the development of generative AI technology, we support the European Commission’s ambition, as part of its Digital Decade programme, that 75% of European companies will adopt AI technologies by 2030. We have already seen an extraordinary wave of innovation that in turn has led to intense and growing competition. The degree of innovation in foundation models (FMs) and AI applications and the myriad means of development and deployment, as well as uses, that these technologies can have in numerous areas of life and business means that there will be many opportunities, and many successful players, across all areas of the economy. This includes extraordinary innovation, not just in relation to models themselves, but in the underlying technology, including compute capacity, and in the vast array of applications which they drive. Amazon is committed to further expanding access to this technology in ways that enable competition to continue to flourish, to the benefit of customers of all types, and in all industries.

This submission describes (i) Amazon’s activities in generative AI; (ii) the importance of the responsible development of AI to ensure excellence and engender trust; (iii) how competition and innovation is working to benefit customers in this sector and citizens more generally; and (iv) the potential impact of regulation in light of the sector’s characteristics and rapid evolution.

1. Amazon’s activities in AI: Amazon’s key aim is to continue to democratise access to generative AI

Amazon has been involved in the development and deployment of AI and machine learning (ML) for more than 25 years in both customer-facing services and internal operations—from using ML to train our voice assistant, Alexa, on speech recognition and natural language understanding, utilising population-based differences in speech data to minimise bias and discrimination in responses,¹ to the AI-powered robots that optimise order fulfilment in our warehouses. More recently, our cloud-computing business, Amazon Web Services (AWS), has launched numerous services which offer customers, including developers of all sizes, additional cost-effective and innovative options to develop their own models and to build safe and secure generative AI solutions, using the best models available for their use case from AWS and third parties.

Amazon recognises the transformative potential of generative AI technologies and is committed to continue democratising access to the resources required by developers and deployers of all types to unlock this potential. While there are costs inherent to developing and accessing AI and FMs, Amazon’s key aim is to ensure that customers across all industries can use, develop and commercialise generative AI. In particular:

- For app developers looking to incorporate FMs into their applications, AWS has recently launched a service called Amazon Bedrock, which like other offerings providing access to FMs, allows customers to experiment with a selection of high-performing FMs from leading AI companies, such as AI21 Labs, Cohere, Stability AI, Meta, Anthropic, Mistral, and AWS itself. Amazon Bedrock provides customers access to these models via application programming interfaces (APIs), alongside enterprise-grade security, privacy, and other features that customers can use to build their own generative AI solutions. Amazon Bedrock democratises

¹ See: <https://www.aboutamazon.com/news/devices/how-amazon-works-to-make-machine-learning-tools-fairer-for-everyone>.



AI by making it easy for organisations of all sizes and across industries to access, experiment with different FMs, and integrate the model best suited to their needs into their applications. Amazon Bedrock became generally available on 28 September 2023.² In addition, customers who have built applications using off-the shelf third-party FMs on AWS, including on Amazon Bedrock, can choose to run them on AWS or any other IT environment where the underlying FMs are also accessible. Customers can take and use the application in another environment (e.g., on-premise or another cloud) by accessing the FM either through the alternative IT provider (if available there) or directly from the FM developer. This is true whether customers continue to use AWS for other workloads or migrate off of AWS entirely.

- Amazon also offers an increasing number of apps that incorporate generative AI technology and make it easier for customers to use generative AI in their everyday work and IT solutions. For example, AWS offers Amazon CodeWhisperer,³ a coding companion trained on billions of lines of code which generates code suggestions in real time, helping developers bypass time-consuming coding tasks and accelerate building with unfamiliar APIs, and Amazon Q,⁴ a generative AI-powered assistant, which customers can use to generate guidance tailored to their business through a simple conversational interface. Amazon Q supports over 40 data source connectors,⁵ with 75% of these being connectors to third-party services, such as Microsoft SharePoint, Google Drive, Salesforce, Confluence, and more, interoperating with customers' existing IT systems to answer the customer's unique business questions using relevant data from any source in any location.⁶
- Customers can choose individual tools offered by AWS to assist with specific steps of their AI/ML development process and increase flexibility. For example, Amazon SageMaker⁷ provides customers with tools to build, train, and deploy their own FMs,⁸ including providing features to streamline the management of training data (e.g., features to help with data cleaning, preparation, labelling, removing duplicates) which can also be used to train a model in another IT environment.⁹ At the same time, customers have the flexibility to use and train models initially trained elsewhere on AWS, or take models trained on AWS to other IT environments.
- AWS makes its generative AI services affordable and accessible so that organisations can experiment and innovate without large upfront investments. For example, the AWS Generative AI Studio provides a low-code environment for creating and iterating on generative AI models.

²See AWS's announcement at: <https://www.aboutamazon.com/news/aws/aws-amazon-bedrock-general-availability-generative-ai-innovations>.

³ Details on CodeWhisperer are available on AWS's website at: <https://aws.amazon.com/codewhisperer/>.

⁴ Details on Amazon Q are available on AWS's website at: <https://aws.amazon.com/q/aws/>.

⁵ A *data source connector* is a mechanism for integrating and synchronising data from multiple repositories into one container index. Amazon Q offers multiple data source connectors that customers can connect to their data sources and help them create a generative AI solution with minimal configuration. See: <https://docs.aws.amazon.com/amazonq/latest/business-use-dg/supported-connectors.html>.

⁶ See: <https://docs.aws.amazon.com/amazonq/latest/business-use-dg/connectors-list.html>.

⁷Details on SageMaker JumpStart are available on AWS's website at:

<https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models.html>.

⁸ Details on SageMaker are available on AWS's website at: <https://aws.amazon.com/sagemaker/>.

⁹ Customers can prepare data using such tools, take that data to another IT provider, and use it to train an FM in that environment.



Additionally, AWS offers comprehensive documentation, tutorials, and training resources to help developers build skills in this emerging field.

- In addition to the numerous different FMs available, AWS offers its own FMs, including the Amazon Titan family of models to increase choice for customers using FMs and building AI solutions. The Titan family includes general-purpose large language models (LLMs) for tasks such as creating summaries, text generation, classification, open-ended Q&A, information extraction, translating text into numerical representations, and image generation.¹⁰
- To support locally-relevant AI innovation, AWS has been working with public sector counterparts across the EU to support them with the technical skills and tools necessary to build LLMs in local languages and facilitate the transformation of public services with generative AI. This initiative is part of AWS's overall work to promote the trustworthy adoption of responsible AI, based on local context and culture while amplifying the transformative economic opportunities of AI and the EU's competitiveness. AWS is supporting public sector initiatives by helping local research institutions develop AI systems tailored to their language and needs, with support from computing centres, and fine-tuning these LLMs for specific purposes like healthcare, education, or manufacturing. AWS provides the computing tools and support to power AI innovation from end-to-end of the generative AI lifecycle.
- For developers building or deploying FMs, AWS is increasing the options for a key input in computer processing: AI chips. Today NVIDIA is the leading provider of computer chips for datacentre AI workloads, but AWS (like other companies, including Intel, AMD, Qualcomm, Cerebras, Groq, Meta, Google, and Microsoft) is investing significant resources to develop and deliver additional chip options for customers. Specifically, AWS offers access to AI chips called AWS Trainium (for AI training workloads) and AWS Inferentia (for AI inference workloads). Trainium is purpose built for deep-learning algorithms, and AWS's compute services powered by Trainium are up to 25% more energy efficient, and 50% more cost effective, than comparable AWS compute services, while Inferentia delivers up to 40% better price performance than comparable AWS compute services. AWS is also investing in and provides AWS Neuron,¹¹ a software development kit, to help customers switch back and forth between third-party AI chips and AWS's AI chips.
- AWS has recently launched Amazon EC2 Capacity Blocks¹² to enable customers with high-performance ML workloads to reserve GPU instances for short periods of time, ensuring that developers have predictable and timely access to cost-efficient compute capacity when they need it. This new, innovative solution, alongside similar offerings from other providers,¹³ are increasing access to AI/ML instances and making it easier for customers of all sizes to develop, train, and use FMs.

Overall, Amazon's approach to generative AI is to invest and innovate to empower a wide range of stakeholders, including customers of all sizes, and developers of all skill levels, to unlock the benefits of generative AI, while prioritising responsible development and deployment practices. In addition to

¹⁰ Titan FMs only became available in 2023.

¹¹ Details on Neuron are available on AWS's website at: <https://aws.amazon.com/machine-learning/neuron/>.

¹² Details on Amazon EC2 Capacity Blocks are available on AWS's website at: <https://aws.amazon.com/ec2/capacityblocks/>.

¹³ For example, Google's Dynamic Workload Scheduler. See:

<https://cloud.google.com/blog/products/compute/introducing-dynamic-workload-scheduler/>.



the services above, examples at the application level include our AI-powered image generation app, which is part of Amazon Ad Console and helps advertisers create images for their products in an affordable and effective way, and Quicklist, a generative AI tool being piloted to facilitate product detail page creation for third-party sellers on the Amazon Store. Amazon also collaborates with academia, policymakers, and other stakeholders to drive innovation, share knowledge, and contribute to the development of responsible AI solutions, and has committed to provide free AI skills training to 2 million people globally by 2025,¹⁴ furthering trust and understanding in AI in today's workforce and the future generation.

2. Amazon is committed to the responsible development, management, and use of AI

While generative AI has the potential to unlock significant opportunities benefiting consumers and competition, Amazon acknowledges that it also raises important issues and is committed to the responsible development, management, and use of AI across all of its services, which is vital to build and maintain customer trust. This, in turn, will drive adoption and further innovation. Last year, Amazon endorsed the White House Voluntary AI Commitments,¹⁵ which set out ambitious and concrete objectives for managing many of the unique risks of powerful generative AI systems, and for building trust with the public, including through red teaming, transparency, mechanisms such as watermarking to disclose AI generated content, and research prioritisation, and we have taken a number of important steps to operationalise these commitments. Through processes like the UK AI Safety Summit, Amazon is collaborating with other industry leaders and policymakers and is committed to knowledge sharing to help shape key commitments and principles around responsible AI.

Amazon is guided by important responsible AI principles when developing and using AI, including safety, veracity, controllability, fairness, explainability, robustness, privacy and security, and transparency, which are key building blocks in creating a safe and innovation-friendly environment for users, developers and deployers. Amazon engages in rigorous testing, assessments, and improvement of its services. We have also implemented important safeguards into our generative AI offerings, including (i) combatting risks around synthetic content by offering built-in watermarking with Titan Image Generator, (ii) applying filters on user inputs and model outputs for our Titan models to reduce the likelihood and spread of harmful content; and (iii) automated abuse detection in Amazon Bedrock to detect and block potentially harmful content.

Finally, Amazon is committed to offering tools and resources to help its customers use AI/ML responsibly. For instance, services like Amazon SageMaker¹⁶ incorporate tools that help customers identify and mitigate bias, explain predictions, and continuously monitor system performance to identify new bias risks. Amazon also provides customers with resources and tools to help them manage risks and responsibilities associated with their deployment of generative AI and the underlying FMs. For example, Amazon Bedrock Model Evaluation¹⁷ offers customers model evaluation tools, and with Amazon Bedrock Guardrails,¹⁸ customers are able to implement safeguards across their FMs to help deliver more consistent and safe user experiences aligned with their company policies and principles.

¹⁴ See: <https://www.aboutamazon.com/news/aws/aws-free-ai-skills-training-courses>.

¹⁵ See: <https://www.whitehouse.gov/wp-content/uploads/2023/09/Voluntary-AI-Commitments-September-2023.pdf>.

¹⁶ See: <https://aws.amazon.com/sagemaker/>.

¹⁷ See: <https://aws.amazon.com/blogs/aws/evaluate-compare-and-select-the-best-foundation-models-for-your-use-case-in-amazon-bedrock-preview/>.

¹⁸ See: <https://aws.amazon.com/blogs/aws/guardrails-for-amazon-bedrock-helps-implement-safeguards-customized-to-your-use-cases-and-responsible-ai-policies-preview/>.



We also offer AI Service Cards for several of our AI services, which are a type of responsible AI documentation intended to provide customers and interested stakeholders with information on intended uses and limitations, responsible AI design choices, and deployment and performance optimisation best practices.

3. Competition in generative AI is intense and rapidly evolving

It is still early days in the generative AI space and there is a lot of innovation to come

Generative AI is rapidly evolving and there is intense competition at all levels, including in relation to the development of the models themselves, the computing infrastructure necessary for building and training FMs, and the applications that incorporate FMs in diverse ways.

Many companies have been working on FMs for years, and there are already numerous FMs available. These models generate a range of different outputs, including text, images, video, and more, and can be adapted to many downstream tasks. Model developers include well-known companies such as Microsoft, Meta, NVIDIA, and Google, in addition to numerous innovative start-ups such as OpenAI, Anthropic, Hugging Face, Stability AI, Mistral AI, Aleph Alpha, EleutherAI, Cohere, Adept, Character.ai, Midjourney, AI21 Labs, poolside, Model Zoo, Mosaic, Technology Innovation Institute, LightOn, Runway, Jasper, Inflection, and many others. Many of these companies offer both proprietary and more open-source models, and many have launched and gained success within a short period of time. There is significant opportunity in this space and it is impossible to predict at this stage which FMs will gain, and retain, relevance. In fact, Amazon has built its business on the assumption that many—potentially thousands—of different models, big and small, proprietary and more open source, will succeed. No one FM will meet every customer need, and, in the future, bigger may not necessarily be better. Amazon expects customers to choose the models best suited to their specific needs at any given point in time (e.g., with respect to different use cases, cost, size, efficiency, mode of deployment, open source, proprietary, transformer-based large language models, multimodal, etc.). For example, if a model is being used to analyse medical scans to assist in diagnosis, it will be vital that it is optimised for accuracy and performance in that single task, whereas other uses (like a general-purpose chatbot) might require a larger model trained on different data.

There is increasing availability of key inputs to generative AI

Model developers have many options for accessing the resources they need to grow and compete effectively. In particular, the transformational potential of generative AI has attracted an increasing number of options for customers to obtain compute capacity, access large volumes of diverse, high-quality data, and opportunities to collaborate with other industry players.

Developers of generative AI now have more compute options than ever, including on-premises solutions (which represent around 85% of IT solutions used today), online solutions provided by cloud services providers (including AI-specialised providers), solutions deployed in a co-located environment and hybrid solutions combining these options. AWS is one of numerous options available to customers to obtain the compute capacity they need. Other providers include the likes of Microsoft Azure, Google Cloud, IBM, Dell, Oracle, OVHcloud, DigitalOcean, IONOS, Fuga Cloud, Scaleway, Exoscale, gridscale, Alibaba Cloud, HPE, Akamai, Vultr, and JarvisLabs.ai, as well as newer entrants responding to the opportunities created by generative AI, including CoreWeave, Lambda Labs, Denvr Dataworks, G42, Crusoe, Cirrascale, TensorWave, and RunPod. Accordingly, model developers are using a variety of IT providers for their compute capacity needs. For example, Aleph Alpha, a German start-up, is reportedly



training models on HPE servers;¹⁹ Imbue, a startup based in San Francisco, is developing models on a server built by Dell;²⁰ Adept AI and Character AI trained their models on Oracle;²¹ and Mistral AI models were reportedly trained on CoreWeave.²²

Developers can also port models built or trained on one IT provider to another (e.g., from AWS to a cloud service or on-premises provider) as well as make them available across multiple IT providers, including through services like Amazon Bedrock. The models themselves are computer files that are not particularly large; a model with 100 billion parameters, for example, could easily fit on a USB flash drive.

Developers have access to significant amounts of data on the open internet and in publicly available datasets. Indeed, many successful FM developers today are new companies that had no existing proprietary data pool to draw from. And the creation of common European data spaces will make even more data available for access and reuse.²³ It is worth highlighting, however, that volume of data alone is not necessarily determinative of the success of a FM—it needs to be the right data for the specific use case. Amazon expects that successful, future open source and proprietary models may depend more on the quality of the data or improved algorithms, not increased volume of data. Often, the most successful solution may not be the largest model; it may instead be a smaller, smarter, more efficient model. Moreover, we are seeing innovation in relation to where FMs can be run and trained. For instance, FMs are being designed and developed specifically for use on mobile devices with efficient model architectures for on-device AI (e.g., Google’s Gemini Nano and Samsung’s Gauss Generative AI Model).²⁴ At the same time, infrastructure providers are supporting these workloads by developing chips for FM deployment on mobile devices. For example, Qualcomm recently announced that its Snapdragon chip helps large models like Stable Diffusion run efficiently on mobile devices.²⁵

Finally, investments and partnerships between industry players, and between the public and private sectors, are bringing together complementary resources and capabilities to enable more rapid and effective innovation than certain companies could achieve on their own. For example, in September 2023, Amazon announced a minority investment and agreement to expand its existing non-exclusive collaboration with Anthropic, an FM developer with a strong focus on responsible AI.²⁶ The collaboration will accelerate Anthropic’s development of its future FMs, facilitate more models being

¹⁹ See: <https://www.hpe.com/us/en/newsroom/press-release/2022/04/hewlett-packard-enterprise-accelerates-ai-journey-from-poc-to-production-with-new-solution-for-ai-development-and-training-at-scale.html> and <https://www.hpe.com/us/en/newsroom/blog-post/2023/11/sovereign-and-trustworthy-ai-is-at-the-core-of-hewlett-packard-enterprises-investment-in-aleph-alpha.html>.

²⁰ See: <https://investors.delltechnologies.com/news-releases/news-release-details/imbue-develop-next-generation-ai-models-150-million-dell-high>.

²¹ See: <https://www.oracle.com/news/announcement/adept-builds-an-ai-teammate-for-everyone-with-oracle-and-nvidia-2022-08-11/> and <https://www.theinformation.com/articles/ai-startups-find-an-unlikely-friend-oracle>.

²² See, e.g.,: <https://venturebeat.com/ai/mistral-shocks-ai-community-as-latest-open-source-model-eclipses-gpt-3-5-performance/>.

²³ See: <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>.

²⁴ See: <https://blog.google/technology/ai/google-gemini-ai/> and <https://developer.samsung.com/sdp/news/en/2023/11/28/samsung-developer-november-newsletter-samsung-electronics-reveals-the-samsung-gauss-generative-ai-model-and-other-latest-news>.

²⁵ See: <https://www.qualcomm.com/news/onq/2023/02/worlds-first-on-device-demonstration-of-stable-diffusion-on-android>.

²⁶ See Amazon’s press release at: <https://press.aboutamazon.com/2023/9/amazon-and-anthropic-announce-strategic-collaboration-to-advance-generative-ai>.



made available to customers, and enable Anthropic and AWS to collaborate in the development of computer chips for AI workloads. More broadly, the availability of public and private investment and regulatory certainty to support businesses' and citizens' ambitions, will be vital in ensuring that Europe can benefit from AI's full potential.

Many companies have already successfully entered and expanded in the generative AI space, including numerous innovative start-ups. Overall, entry and expansion is becoming easier, and Amazon expects this to continue in line with rapid innovation. For instance, services like Amazon Bedrock that provide customers with the ability to customise a pre-trained model through fine-tuning have been game-changing. Pre-trained models are designed to perform well at a general range of tasks, but they may not be perfectly suited for each customer's specific use cases. Customers can adapt pre-trained models for their particular tasks via fine-tuning, which is the final step of the ML process. Because the pre-trained models can already perform a wide variety of tasks, they only need a small amount of additional training data and compute to adjust to a customer's specific needs. While fine-tuning data requirements differ based on many factors, effective fine-tuning can be done with as few as 100 samples of the customer's target task. In other words, fine-tuning is substantially faster, requires less computation time, and needs less data than creating a task-specific model, allowing AI companies to enter this space and grow, including through the use of existing open source FMs.

Competition is also thriving at the application level

There is also considerable innovation and competition in AI applications. For example, companies as diverse as adidas, Booking.com, Bridgewater Associates, Clariant, Cox Automotive, GoDaddy, and LexisNexis Legal & Professional have launched generative AI applications on AWS, and Amazon is aware that thousands more will build using inputs from other providers. Amazon itself uses FMs to provide services for customers, vendors, and sellers. For example, a generative AI conversational assistant has been launched for vendors, and is being piloted for third-party sellers, to resolve operational issues with answers and guidance across business topics. In addition, the Alexa service makes use of various kinds of AI models, including to improve Alexa's speech recognition and natural language understanding systems.²⁷ Amazon anticipates that businesses and end consumers will have access to a huge variety of applications and ways to use generative AI.

4. Competition law can address any anticompetitive conduct in the AI space without hindering innovation

Competition and consumer protection laws are adaptable to new technologies and industries, and this remains the case in the AI context. The exponential development of AI and ML across industries and sectors demonstrates how rapidly these technologies are moving and driving innovation (e.g., just two years ago, generative AI was not a part of everyday discourse). Applying the existing competition law framework in a predictable manner provides clarity without hindering innovation. Sector-specific regulation, particularly in a dynamic sector like AI with so much current innovation and so little known today about future benefits, could create unnecessary obstacles, making it difficult for competition and innovation to flourish or, worst case, becoming a barrier to entry in its own right. Indeed, a recent report commissioned by Amazon²⁸ highlighted the importance of clarity and certainty, finding that 21% of businesses consider compliance and legal uncertainties related to digital technology to be a barrier

²⁷ See: <https://www.aboutamazon.com/news/innovation-at-amazon/how-amazon-uses-generative-ai> for more ways in which Amazon is using generative AI to make life easier.

²⁸ Unlocking Europe's AI Potential in the Digital Decade, February 2024, at: https://www.unlockingeuropesaipotential.com/files/ugd/c4ce6f_ecf071799e4c4eba80113648d2b1090b.pdf.



to investment, with those facing uncertainties planning to invest 48% less over a three-year period. This is particularly relevant in the context of competing visions globally on investments into, and regulation of, AI.

While competition regulators have expressed that businesses should be mindful not to engage in anticompetitive conduct in the AI space, the examples of potential anticompetitive behaviour they have raised (such as tying, bundling, exclusive dealing, and anticompetitive M&A transactions) fall squarely within traditional competition doctrine and do not raise different legal issues.

At the same time, there may be instances where the use of AI will create novel questions in other areas that may require adjustments to existing laws and regulations. As recognised in the EU AI Act, however, different use cases carry significantly different risk profiles, so any intervention in this sector should not be one-size-fits-all. The integration of AI into medical devices and autonomous vehicles are two early examples where regulators have updated or issued new regulatory guidance to provide clarity for industry and safeguards for the public. This type of sector-specific regulation that is based on informed determinations about potential risks of using AI for specific purposes is likely to be more effective, and less detrimental to innovation, than attempting to introduce catch-all comprehensive regulation. As most would agree, AI used for bio-tech should be regulated differently than AI used for crossword puzzles.

5. Conclusion

In the rapidly evolving generative AI space, competition is intense and innovation is constant. Amazon is focused on using AI to improve customers' experiences and on expanding access to AI by creating and maintaining an open environment for ML and the underlying compute infrastructure for developers and deployers of all sizes and abilities. Amazon is committed to working with policymakers, researchers, and the AI community to advance the responsible and secure use of AI.