

Google Cloud's Approach to Trust in Artificial Intelligence

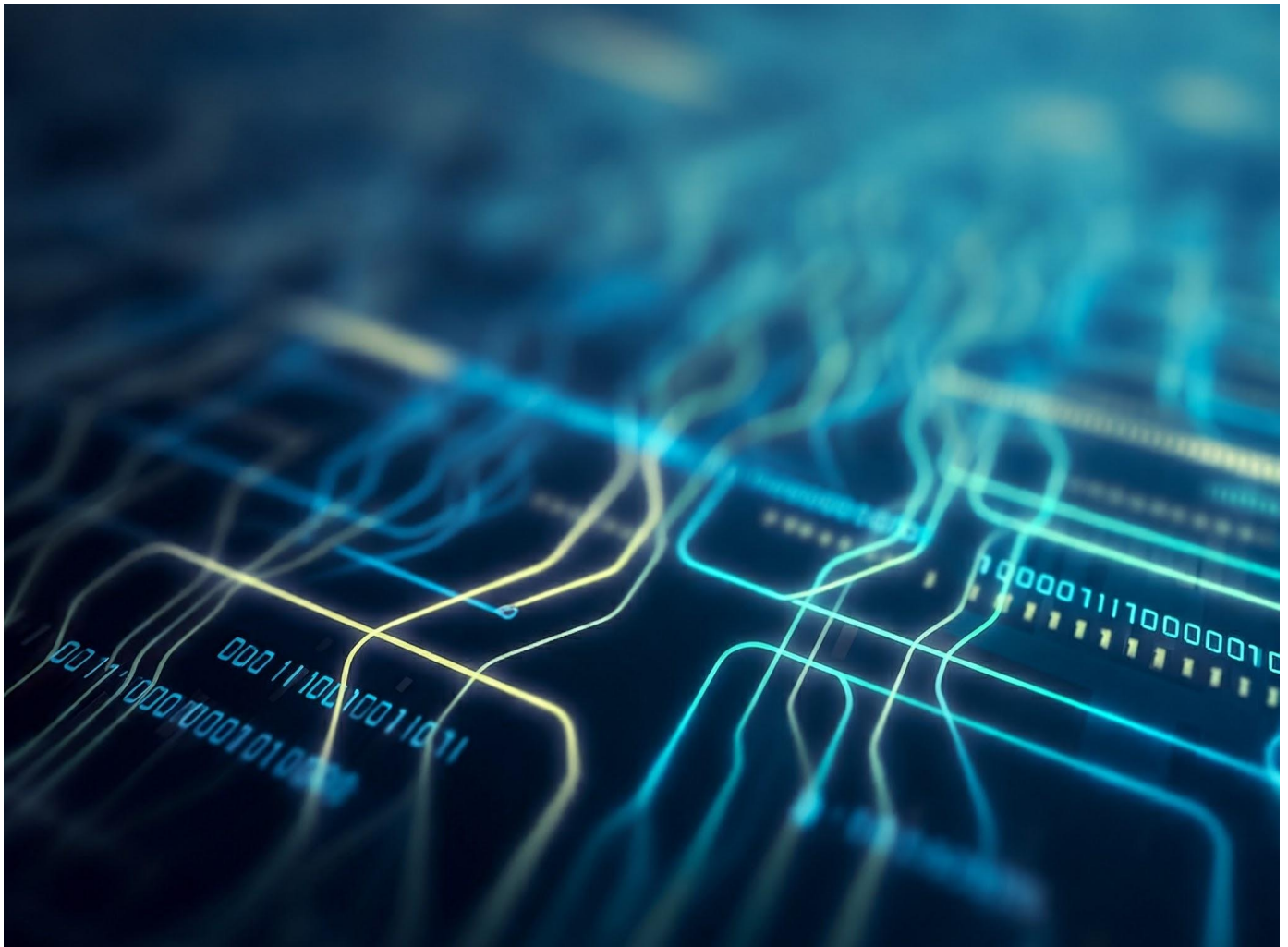


Table of Contents

01	Introduction	3
02	Responsible Innovation	4
	Risk Assessments	5
	Data Governance	6
	Privacy	7
	Security	8
	Compliance	11
	Open Cloud & Portability	12
	Environmental Impact	13
03	Shared Responsibility	15
04	Best Practices	16
	Governance	16
	Acceptable Use	16
	Security	16
	Privacy and Data Governance	17
	Staying on Top of Gen AI Developments	17
05	Conclusion	18

Disclaimer: The content contained herein is correct as of October 2024 and represents the status quo as of the time it was written. Google Cloud's security policies and systems may change going forward, as we continually improve protection for our customers.

Introduction

AI responsibility has come to be associated with not only mitigating risks, but also helping improve people's lives and addressing social and scientific challenges. Some elements – such as the need to be good stewards of the data used, and incorporate important societal values including privacy, fairness, and transparency – are widely agreed, although striking the right balance between them can be complex. We recognize that advanced technologies can raise important challenges that must be addressed clearly, thoughtfully, and affirmatively.

In this paper, we focus on providing insight into Google Cloud's approach to building enterprise-grade generative AI (hereinafter "gen AI") responsibly, with an overview of how we approach AI data governance, privacy, security, and compliance when developing [gen AI](#), which customers can interact with through the [Vertex AI platform](#). For context as used throughout this paper, gen AI refers to the use of AI to create new content such as text, images, music, audio, and video, or some variation thereof as enabled by [multimodal gen AI](#).

Gen AI works by using an ML model to learn the patterns and relationships in the provided dataset(s). It then uses the learned patterns to generate new content. Gen AI is powered by foundational models (large AI models) that can multi-task and perform out-of-the-box tasks, including summarization, Q&A, classification, and much more.

The Vertex AI platform is a machine learning platform that enables you to [train and deploy](#) machine learning models and AI applications, and customize large language models (LLMs, a form of foundation models) for use in your AI-powered applications. Vertex AI combines data engineering, data science, and ML engineering workflows, enabling your teams to collaborate using a common toolset and scale your applications using the benefits of Google Cloud. In addition, you can also adapt foundation models for targeted use cases with [minimal training](#) and very little example data.

Google Cloud takes a principled approach to AI development, which is grounded in [AI Principles](#) that describe our commitment to developing technology responsibly and in a manner that is built for safety, enables accountability, and upholds high standards of scientific excellence. In upholding these Principles, Google Cloud has identified specific areas we will not pursue, which include the design or deployment of technologies that are likely to cause overall harm, or whose principal purpose or implementation is to cause or directly facilitate injury to people. Likewise, we won't pursue technologies whose purpose contravenes widely accepted principles of international law and human rights, or which use enables information gathering for surveillance that violates internationally accepted norms.

We've also infused these values into the [Google Cloud Platform Acceptable Use Policy](#) and the [Generative AI Prohibitive Use Policy](#) so that they are transparent and clearly communicated. In addition, when it comes to AI, we recognize the need for both good individual practices and shared industry standards. We've continued evolving our practices, conducting industry-leading [research](#) on AI impacts and risk management, and [assessing proposals](#) for new AI research and applications to ensure they align with our principles. We continuously iterate and [reassess](#) how to build accountability and safety into our work, and publish on our progress to encourage collaboration and advancements in this field.

As AI technologies advance, they provide the opportunity to enhance how we identify, address, and reduce security risks. We've taken a three-pronged approach to [secure, scale, and evolve](#) the security ecosystem by:

- Supporting customers in their AI implementation with controls, [best practices](#), and capabilities
- [Continuing to launch](#) cutting-edge, AI-powered products and services to help organizations achieve better security outcomes at scale
- Continuously [evolving](#) to stay ahead of threats

We are committed to helping our customers leverage the full potential of gen AI by infusing and enabling privacy, security, and compliance capabilities for a safe and secure implementation, as further described below.

Responsible Innovation



As with any transformational and new technology, we understand that gen AI comes with complexities and risks, and these need to be managed as part of a comprehensive risk management framework and governance structure. We recognize that AI presents critical questions and we are [working to build AI responsibly](#) to benefit both our customers and the wider societies in which we operate. The challenge is to do so in a way that is proportionately tailored to mitigate risks and promote reliable, robust, and trustworthy AI applications, while still enabling innovation and the promise of AI for societal benefit.

Responsible AI is woven into the fabric of our work. As part of our [principled](#) approach to building AI technologies, we commit to developing and applying strong safety and security practices, and incorporate our privacy principles in the development and use of AI. [Rigorous evaluations](#) are a critical component of building successful AI and in Google Cloud specialized teams are engaged in analyses and risk assessments for the AI products we build and early-stage customer co-development opportunities involving a customized approach.

[Responsible product development](#) practices span across multiple dimensions – some are technical involving evaluations of data sets and models for bias, some pertain to product experiences, and some are around policy, informing what we will and won't offer from a product perspective. We have developed a four-phase process (consisting of Research, Design, Govern and Share) to review projects against the AI Principles and work with subject matter experts on privacy, security, and compliance, to name a few. The initial Research and Design phases foster innovation, while the Govern and Share phases focus on risk assessment, testing, monitoring, and transparency.

Our research draws on in-house expertise, including computer scientists, social scientists, and user experience researchers. We also regularly publish on the [progress we're making](#) to enable transparency into our work, support safer and more accountable products, earn and keep our customers' trust, and foster a culture of responsible innovation.

Our approach to responsibility by design is guided by our AI Principles and builds upon Google's previous experience with keeping users safe on our platforms. As we build generative AI services, our technical approaches to enforce policies at scale include techniques like fine-tuning and [reinforcement learning from human feedback](#) (RLHF). Other layered protections are invoked both when a person inputs a prompt and again when the model provides the output. Policy improvements are informed by ongoing user [feedback](#) and monitoring. Responsibility by design also involves building security into our products from the very beginning. We've codified this approach in our [Secure AI Framework](#) (SAIF). Applying SAIF, we build on our existing security knowledge and adjust mitigations to these new threats, as further discussed below.

Risk Assessments

For every organization, the decision to leverage the power of gen AI hinges on a myriad of questions, one of the most salient being: How can I help my organization harness the power of AI while minimizing risk?

During reviews, potential risks posed by the AI product being developed are identified and assessed. We assess the potential risk and impact of the gen AI models we're building at both the model level, and the point of their integration into a product or service. We take a socio-technical approach, considering how AI will interact with the world and existing social systems, and assess the potential impacts and risks that may be posed both at the initial release and as time goes on. Reviewers who conduct the risk assessments understand that potential risks and impacts might be different at the model level and at the application level, and consider mitigations accordingly. We draw from various sources, including a wide range of academic literature, external and internal expertise and our in-house ethics and safety research.

Google Cloud has introduced many models through private releases. This release mechanism allows our product teams to gather valuable feedback to support better products before we make them generally available. Once we incorporate their feedback and prepare for customer use in production, we update our product documentation to account for any changes. In our product documentation, we generally provide known limitations of the model and we may issue service-specific terms to further advise customers on proper use of our products. Cloud also continues to invest in tools to support our customers including: Vertex's [Explainable AI](#), [Model Fairness](#), [Model Evaluation](#), [Model Monitoring](#), and [Model Registry](#) to support data and model governance.

Where appropriate, we develop mitigation strategies to address potential risks that may be identified prior to releasing the product for general availability (GA). Mitigations can take various forms. For instance, for gen AI products, mitigations may draw on technical approaches to evaluate and improve models during development, establish policy-driven safety guardrails, or may be enabled by tooling customers can leverage in their own projects for further safety efforts; they may also be a combination of any of the above as best suited for the particular product being released, and are often informed by customer feedback. Policy restrictions are typically guided by the relevant Acceptable Use Policy, Terms of Service, and privacy restrictions, as further discussed in the AI Data Governance and Privacy section below.

In addition, we offer customers technical controls such as [safety filters](#) which can block model responses that violate policy guidelines, for instance, around child safety. We recognize that many customers will have their own

policies, and aim to support this deployment-specific work. For instance, customers can create safety filters leveraging [safety attributes](#), which include “harmful categories” and topics that may be considered sensitive, such as “drugs” or “derogatory.”

For most of these categories, customers can set their own thresholds for blocked responses and control content based on their own business needs. For instance, safety settings can be configured based on both probability and severity scores such that safety attributes can be aligned to your enterprise’s policies. We also recently released the [Responsible Generative AI Toolkit](#) to provide guidance and tools to create safer AI applications with these new open models. The toolkit includes guidance on setting safety policies and methodologies for building robust safety classifiers.

These controls can be further enhanced by tooling customers can use to achieve greater understanding and control of their AI models. For instance, leveraging models similar to those used in our safety filters, customers can use our [text moderation](#) service to scan the entire corpus of training set data for terms that fall within predefined “harmful categories” and topics that may be considered sensitive, enabling ongoing compliance through the identification of content that may violate relevant policies. Our [explainable AI](#) offerings provide a set of tools and frameworks to help customers understand and interpret predictions made by machine learning models, natively integrated with a number of Google Cloud products and services.

Additionally, [model evaluation](#) on Vertex AI includes metrics to understand model performance as well as [evaluate potential bias](#) using common data and model bias metrics. These tools can promote fairness by evaluating data and model outputs during training and over time, highlighting areas of concern and providing suggestions for remediation. In addition, we also make documentation on our [API models](#), [open models](#) and [large language foundation models](#) available on our [model card hub](#), articulating our model’s strengths and limitations.

Data Governance

Using generative AI in a business setting can pose various risks around accuracy, privacy and security, regulatory compliance, and intellectual property. As with other forms of digital innovation, creating a programmatic, repeatable structure can help achieve a consistent approach to evaluating AI use cases.

We’ve implemented robust data governance reviews designed to ensure data privacy commitments are considered when developing and deploying gen AI. One of the questions we are frequently asked is whether our foundation models are trained on customer data, and by extension, whether customer data may as a result be exposed to Google Cloud, Google Cloud’s other customers, or the public. To address this question, we outline some key aspects of our [model tuning and deployment](#) and [data governance practices](#) and our approach to privacy in our Vertex AI offerings below. Additional information regarding foundation model adaptation can be found [here](#).

- Google Cloud processes customer data to provide our services. Google Cloud does not use customer data to train our foundation models without the customer’s prior permission or instruction.
- The foundation models on Vertex AI are developed to handle general use cases. Customers can customize foundation models for specific use cases by tuning them using our tuning APIs. This approach combines our research and product development expertise to enable world-class AI without compromising customers’ control over their data.

- Vertex AI offers a Parameter Efficient Tuning service that enables tuning of the foundation model to specific tasks without having to rebuild the entire model. Each tuning job results in the creation of a few additional learned parameters, called “adapter weights” that are outside the foundation model. The adapter weights are specific to the customer, and only available to the customer who tuned those weights. During inference, the foundation model receives the adapter weights, runs through the request and returns the results, without modifying the foundation model or storing the request.
- Input data, including prompts and adapter weights, are considered customer data and are stored securely at every step along the way – encrypted at rest and in transit. Customers can control the encryption of the stored adapter weights by using customer-managed encryption keys (CMEK) and can delete their adapter weights at any time. Customer data used to train adapter models will not be logged or used for improving the foundation model without the customer’s permission.

Privacy

In this section, we provide an initial set of [considerations](#) for how we apply foundational privacy principles, such as accountability, transparency, and data minimization to enable the responsible collection and use of data for model training and safeguards for model outputs. Our [approach](#) includes incorporating privacy design principles, designing architectures with privacy safeguards, and providing appropriate transparency and control over the use of data. When bringing new offerings to the market, we incorporate these principles throughout the product lifecycle and design architectures with comprehensive privacy safeguards such as data encryption.



Accountability: AI governance at Google is built around our [AI Principles](#), which we released in 2018, and our [AI Privacy Practices](#). These help ensure that the way we develop AI is aligned with core human values.

Our fifth AI Principle, “incorporate privacy by design,” helps guide our privacy practices in the development of AI technologies. To carry out this principle in gen AI, we use structured product launch processes to ensure that, wherever appropriate, gen AI products provide an opportunity for notice and consent, encourage architectures with privacy safeguards, include appropriate transparency and control over the use of data, and employ data anonymization techniques and other privacy protections.

Our launch reviews rely on teams of engineers, product specialists, and other specialists in privacy, legal, and safety charged with taking reasonable steps to assess relevant privacy risks. Google has a robust process for AI development, which includes risk assessment frameworks, ethics reviews, and executive accountability processes in place to implement the AI Principles and practices at both the development and deployment stages.



Transparency: Google Cloud builds privacy protections into its architecture and provides meaningful transparency over the use of data, including clear disclosures and commitments regarding access to a customer’s data. In addition, Google Cloud does not use data provided to it by its customers to train its own models without the customer’s permission. Our teams assess products for compliance with data privacy and transparency requirements. These reviews also consider adherence to the [commitments](#) made to our customers regarding data privacy and protection – specifically, the ability to control how customer data is accessed, used and processed – as articulated in the [Google Cloud Platform Terms of Service](#) and [Cloud Data Processing Addendum](#). In addition, we provide customers with the option to see [who can access their data and why](#).



User Controls: Empowering users to manage their personal data helps establish a foundation of trust and provides tools for people to exercise their rights. Google Cloud's enterprise customers have the ability to control their data. On our Vertex AI Platform, customer data prompted into a model and the output generated by Vertex AI from those prompts is "Customer Data," and Google processes Customer Data only according to the customer's instructions.

Vertex AI already [provides robust data commitments](#) to ensure customer data and models are not used to train our foundation models without permission or leaked to other customers. However, there are two additional concerns that organizations have: (1) protecting data, and (2) reducing the risk of customizing and training models with their own data that may include sensitive elements such as personal information (PI) or personally identifiable information (PII). Often, this personal data is surrounded by context that the model needs so it can function properly. To effectively segment, anonymize, and help protect data, we have a robust set of tools and services that we are continuously optimizing, including:

- Our [Sensitive Data Protection](#) service, which provides sensitive data detection and transformation options such as masking or tokenization to add additional layers of data protection throughout a generative AI model's lifecycle, from training to tuning to inference.
- [VPC Service Controls](#), which allows for secure deployment within defined data perimeters. With VPC SC, you can define perimeters to isolate resources, control and limit access, and reduce the risk of data exfiltration or leakage.

We are committed to preserving our customers' privacy with our Cloud AI offerings and to supporting their compliance journey. As a global cloud provider, Google Cloud has a long-standing commitment to [GDPR compliance](#) and has compiled comprehensive [DPIA Resource Center documentation](#) to support customers in their data protection impact assessment efforts. We also enable certain AI/ML services to be configured to meet [data residency requirements](#) as noted in our [Service Terms](#).

Security

Privacy is tightly intertwined with security, and both are primary design criteria for all products built on Google Cloud. AI products are no exception, benefitting from Google Cloud's globally distributed and redundant infrastructure, and [inherit the platform's foundational controls](#). AI models running on Google Cloud are protected by layered security controls as we don't rely on any single technology to make our infrastructure secure; rather, our technology stack builds security through progressive layers that deliver defense in depth.

Our AI products are built atop a scalable technical infrastructure designed for maximized availability and reliability, while providing security through the entire information processing lifecycle. Google Cloud's core principles include defense in depth, at scale, and by default. Data and systems are protected through multiple layered defenses using policies and controls that are configured across Identity and access management (IAM), encryption, networking, detection, logging, and monitoring.

The platform is further underpinned by a [secure-by-design foundation](#) supported by operational controls consisting of in-depth security reviews, vulnerability scanning, ongoing threat monitoring, and intrusion detection mechanisms that enable secure service deployment and safeguard customer data. The security controls specific to Vertex AI and generative AI features can be found [here](#).

Building an AI/ML system requires a large corpus of data to appropriately train models and oftentimes the data may be considered sensitive. Securing that data appropriately becomes of paramount importance, and we can protect against data leakage risks with encryption and anonymization techniques. All data stored within Google Cloud is encrypted at rest using the same hardened key management systems that Google uses for our own encrypted data. These key management systems provide strict key access controls and auditing, and encrypt user data at rest using AES-256 encryption standards. No setup, configuration, or management is required. However, you may also use CMEK in the Cloud Key Management Service (Cloud KMS). Customer-managed encryption keys allow you to have more control around key generation, key rotation frequency, and key location, but with added control comes added responsibility in appropriately managing the keys. Because of these added responsibilities, we recommend that you evaluate whether the default encryption is sufficient, or whether you have a compliance requirement that you must use Cloud KMS to manage keys yourself. For more information, see how to meet compliance requirements for encryption at rest.

From a [software supply chain](#) perspective relating to the security of AI, we believe that extending existing software supply chain solutions is an effective way to counter many of the risks associated with AI software supply chains. Rather than creating new solutions, we can approach AI models like traditional software. Across Google's first-party and open source AI development ecosystems, we're adopting the SLSA framework and format to sign model provenance. This metadata document cryptographically binds a model to the service account – an identifying account that represents an application rather than a human user – that was used to train it. It also enables Google to verify all of its models against the expected signing keys, such that an insider cannot overwrite or change the model (including the weights that determine its behavior) without detection. Google invests significantly in securing the open source software world, including support for SLSA and the Sigstore project. In 2023, we [open sourced](#) our work to apply existing solutions such as SLSA and Sigstore to AI.

We strive to enable the security of our AI products, and also to support customers in using AI to bolster their security capabilities. Our [use of AI in security offerings](#) can now combine world class threat intelligence with point-in-time incident analysis and threat detections and analytics to help prevent new infections, make security more understandable while helping to improve its effectiveness, and reduce the number of tools organizations need to secure their vast attack surface areas and ultimately, empower systems to secure themselves. Customers using Vertex AI can also benefit from [Sensitive Data Protection](#) which enables the identification of sensitive data such as email addresses, phone numbers, and job titles, to name a few, based on a pattern or a list, and then automatically hides or transforms that data by using methods such as masking or tokenization. This tool can also be used to redact sensitive data, such as social security numbers, from images before ingesting it into a machine learning training environment.

In addition to building our AI products on a secure platform, we've also developed the Secure AI Framework (SAIF) mentioned above, which is a conceptual framework for securing AI systems. It's inspired by the security best practices – like reviewing, testing and controlling the supply chain – that we apply to software development, while incorporating our understanding of [security megatrends](#) and risks specific to AI systems. SAIF offers a practical approach to address the concerns that are top of mind for customers, including security, AI/ML model risk management, privacy, compliance and others. Customers may wish to consider SAIF as they define and refine their approach to adopting AI.

A framework like SAIF which spans across the public and private sectors is essential for safeguarding the technology that supports AI advancements, so that when AI models are implemented, they're secure-by-default. The [six core elements](#) of SAIF are summarized here:



Expand strong security foundations to the AI ecosystem



Extend detection and response to bring AI into an organization's threat model



Automate defenses to keep pace with existing and new threats



Harmonize platform level controls to ensure consistent security across the organization



Adapt controls to adjust mitigations and create faster feedback loops for AI deployment



Contextualize AI system risks in surrounding business processes

These steps aren't simply conceptual. We're [putting them into action](#) to support and advance a framework that works for all. It's also important to consider that while there are novel aspects to securing AI, many of the current approaches to developing, deploying and utilizing AI systems can be adjusted to account for these specificities rather than requiring a completely new approach. In addition to building our AI products on a secure platform, we've also developed the Secure AI Framework (SAIF) mentioned above, which is a conceptual framework for securing AI systems across the four dimensions that make up an AI system: Data, Infrastructure, Application and Model. We explore what changes and what stays the same when it comes to AI cybersecurity in greater depth [here](#) and share [best practices](#) on securely deploying AI.

In addition, we've established a dedicated [Google AI security red team](#) focused on testing for security and privacy risks. Red teaming, also referred to as adversarial testing, is a technique where "ethical hackers" intentionally violate policies for the purpose of discovering and addressing vulnerabilities which could harm users. With the rise of generative AI, it has become a useful tool to help teams systematically improve models and products, and to inform launch decisions.

To expand on these efforts to address content safety risks, we've built a new team to use adversarial testing techniques to identify new and unexpected patterns on generative AI products. This team explores innovative uses of AI to augment and expand existing testing efforts. Applying the adversarial approach we use during pre-launch responsibility evaluations to post-launch evaluations helps us improve model performance based on user feedback and helps us identify emerging risks.

Across the development and deployment lifecycle of our AI technologies, we use robust security and safety controls, which we adapt to risks for specific products and users. This is important as it enables early inclusion of prevention and detection controls, augmented by adversarial testing and red teaming. We also use threat intelligence to stay abreast of novel attacks. Our models are developed, trained, and stored within Google's infrastructure, supported by our global teams of security engineers.

Compliance

Security and privacy in cloud computing are subject to legal, regulatory compliance, and risk management requirements. This is particularly the case in regulated industries such as financial services and healthcare, and for certain critical service or critical infrastructure providers. Organizations running workloads and storing data on Google Cloud rightfully seek assurances as to the platform's controls posture, frequently requiring documentation from an independent third party to validate their existence and efficacy. To provide transparency into its controls, Google Cloud makes compliance documentation, certifications, control attestations, and independent audit reports [readily available](#) to satisfy regional and industry-specific requirements and support customers in their compliance validation efforts of the Google Cloud platform, as well as their assessment of [Vertex AI's compliance and security controls](#).

The rapid advance of AI has captured regulators' attention worldwide, who are increasingly interested in understanding how current regulatory frameworks address AI and what new measures might be necessary in order to ensure AI is developed and deployed in a way that respects laws, norms, and human rights. We believe that AI is too important not to regulate, and too important not to regulate well, and thus advocate for risk-based frameworks that reflect the complexity of the AI ecosystem by building on existing general concepts. We previously published [recommendations for regulating AI](#) which outlined a general approach and some key implementation practicalities for policymakers to consider in developing practical AI regulations. Our top line recommendations included:

01

Taking a sectoral approach that builds on existing regulation

02

Adopting a proportionate, risk-based framework

03

Promoting an interoperable approach to AI standards and governance

04

Ensuring parity in expectations between non-AI and AI systems

05

Recognizing that transparency is a means to an end

Since then, we've further clarified our position, publishing [A Policy Agenda for Responsible Progress in Artificial Intelligence](#), [Applying Model Risk Management Guidance to Artificial Intelligence/Machine Learning-based Risk Models](#), and [Generative AI Risk Management in Financial Institutions](#). In addition, our teams have adopted a risk assessment process to help (1) identify, measure, and analyze ethical risks throughout the life of an AI-powered product, (2) map these risks to appropriate mitigations, and (3) develop clearer standards of acceptable risk.

Google Cloud is a trusted voice in the international and regional standards development community. We actively provide feedback and shape the regulations, standards, and framework. We also closely track, monitor, and actively support industry standards such as the recently published ISO/IEC 42001 AI Management System Standard, NIST AI Risk Management Framework (RMF), as well as global regulatory developments to ensure we continue to develop and deliver tools that serve our customers' needs.

We understand that AI comes with complexities and risks, and to ensure our readiness for the future landscape of AI compliance we proactively benchmark ourselves against emerging AI governance frameworks. To put our commitments into practice, we invited [Coalfire](#), a respected leader in cybersecurity, to examine our current processes, measure alignment and maturity toward the objectives defined in the National Institute of Standards and Technology (NIST) [Artificial Intelligence Risk Management Framework](#) (AI RMF) and International Organization for Standardization (ISO) [ISO/IEC 42001 standard](#). Coalfire's [assessment](#) provided valuable insights, allowing us to enhance our security posture as we continuously work to uphold the highest standards of data protection and privacy. We believe that an independent and external perspective offers critical objectivity, and we are proud to be among the first organizations to perform a third-party AI readiness assessment.

Likewise, we closely monitor regulatory developments, such as the EU AI Act. The AI Act is a legal framework that establishes obligations for AI systems based on their potential risks and levels of impact. It will come into effect in phases and include bans on certain practices, general-purpose AI rules, and obligations for high-risk systems. Google is [actively preparing](#) for AI Act compliance. Internally, our AI Act readiness program is focused on ensuring our products and services align with the Act's requirements while continuing to deliver the innovative solutions our customers expect. This is a company-wide initiative that involves collaboration among a multitude of teams, including:

- **Legal and Policy:** thoroughly analyzing the AI Act's requirements and working to integrate them into our existing policies, practices, and contracts.
- **Risk and Compliance:** assessing and mitigating potential risks associated with AI Act compliance, ensuring robust processes are in place.
- **Product and Engineering:** ensuring our AI systems continue to be designed and built with the AI Act's principles of transparency, accountability, and fairness in mind and constantly improving the user experience, incorporating the AI Act's requirements for testing, monitoring, and documentation.
- **Customer engagement:** Working closely with our customers to understand their needs and concerns regarding the AI Act, and providing guidance and support as needed.

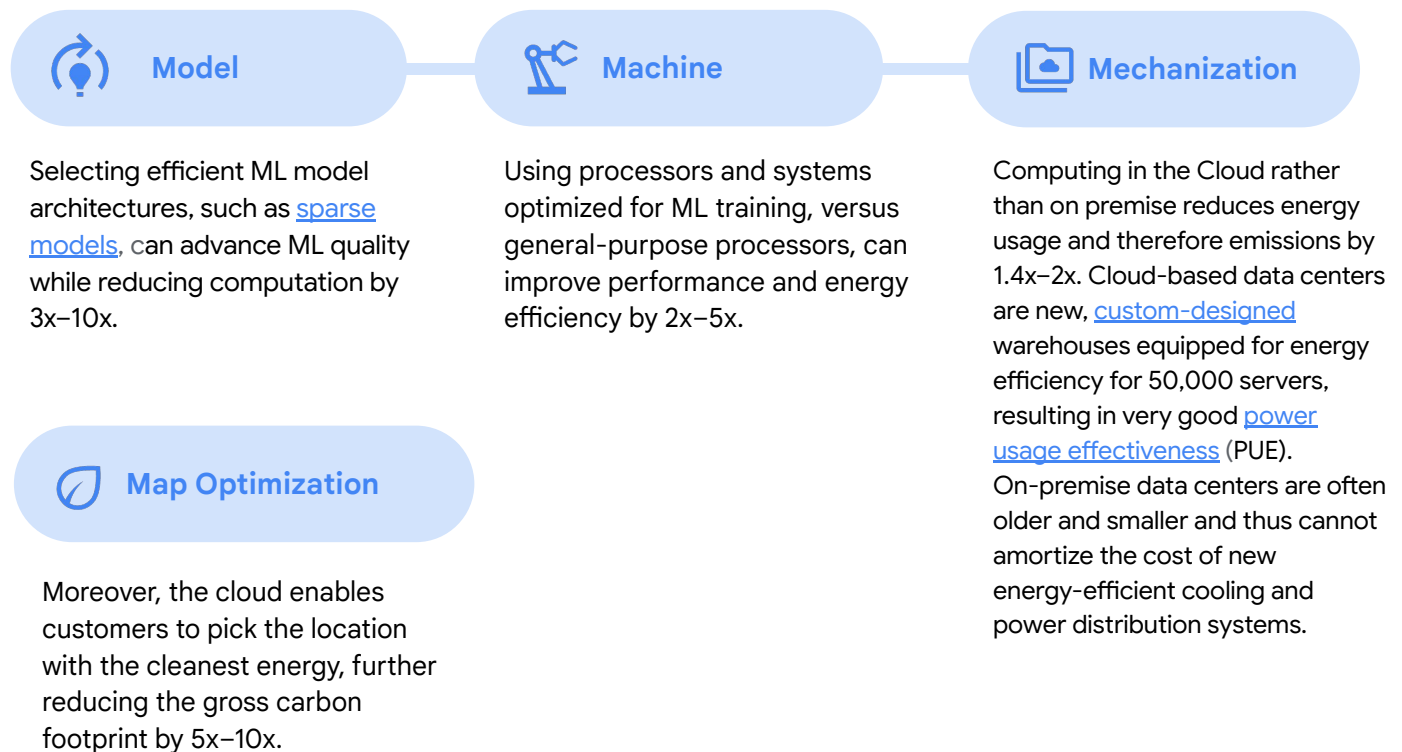
Open Cloud & Portability

Google Cloud offers both first-party and third-party AI models in the Vertex AI Model Garden. This is part of our open philosophy which extends far beyond just the ability to utilize a variety of AI models. It encompasses a core belief in giving customers maximum choice without forced lock-in. This means allowing easy connection to existing on-premises systems, other clouds, SaaS applications, and even supporting businesses' proprietary models. Vertex AI offers a unified platform to manage, monitor, and continuously improve models regardless of their origin, whether that's open-source, partner-developed, or Google's in-house models.

This approach in turn supports portability that enables customers to take custom code, OSS code and containers anywhere. Model Garden simplifies the process of selecting (and switching between) the most options of models from any model provider. Vertex AI integrates with MLOps tools for streamlining the movement of models – be they first-party, third-party, or open-source – from development into production and ensure that changes are well-tracked. Model Garden encourages a modular approach, making it easier to swap out individual components within your ML pipelines, promoting experimentation and faster iteration. Together, this empowers you to make informed decisions based on your specific requirements and risk tolerance.

Environmental Impact

AI models and services can consume vast amounts of energy which raises the responsibility for managing the carbon footprint resulting from the computing power required to train and run foundation models. As part of our ongoing work, we have identified four [best practices](#) that reduce energy and carbon emissions significantly, which we refer to as the “4Ms,” all of which are being used today and are available to anyone using Google Cloud services. These four practices, each of which is briefly noted below, can when implemented together, reduce energy by 100x and emissions by 1000x.



We’ve made significant investments in cleaner cloud computing by making our data centers some of the most efficient in the world and sourcing more carbon-free energy. On average, a Google-owned and -operated data center is more than 1.5 times as energy efficient as a typical enterprise data center and, compared with five years ago, we now deliver approximately three times as much computing power with the same amount of electrical power. To support the next generation of fundamental advances in AI, our latest TPU v4 is proven to be one of the fastest, most efficient, and most sustainable ML infrastructure hubs in the world. The cloud supports many products at a time, so it can more efficiently distribute resources among many users. That means [we can do more with less energy](#).

We're constantly looking for new ways to build products, design out waste and pollution, and keep materials and resources in use for as long as possible. We aim to [maximize the reuse of finite resources](#) across our operations, products, and supply chains and to enable others to do the same. We're helping our customers make real-time decisions to reduce emissions, and mitigate climate risks with data and AI. For example, Google Cloud customers can reduce their cloud footprint with a feature called Active Assist, which uses machine learning to identify unused (and potentially wasteful) workloads that could reduce carbon emissions if removed.



Environmental footprint: *Leveraging AI to optimize our own operations, and working to reduce energy use and emissions from AI computing in our data centers.*

AI and machine learning workloads are quickly becoming larger and more capable, raising concerns about their energy use and their impact on the environment. With AI at an inflection point, predicting the future growth of energy use and emissions from AI compute in our data centers is challenging. Historically, research has shown that as AI/ML compute demand has gone up, the energy needed to power this technology has increased at a much slower rate than many forecasts predicted. We have used tested practices to reduce the carbon footprint of workloads by large margins; together these principles have reduced the energy of training a model by up to **100x and emissions by up to 1,000x**. We plan to continue applying these tested practices and to keep developing new ways to make AI computing more efficient. Google data centers are designed, built, and operated to maximize efficiency – even as computing demand grows.

We are committed to operating [carbon-free by 2030](#) and [replenishing 120% of the water we consume by 2030](#). We're also dedicated to raising our standard of water stewardship, improving water quality and security, and restoring the health of ecosystems in the communities in which we operate.

Shared Responsibility



As we continue to develop our AI platform, systems, and foundational models, our belief in shared fate and our experience in using these technologies guides us to invest in end-to-end governance tools, opinionated guidance, and best practices to help our customers keep their data and AI models safe.

At Google Cloud, we are committed to helping enterprises develop effective AI risk management strategies to be able to use the full potential of gen AI. While risk profiles are often complex, this is especially true for generative AI because of how intricate the models can be. Importantly, [risk management can vary](#) depending on how the organization has chosen to use AI – Is it developing its own AI applications, using AI applications developed by a third party (including those developed by Google Cloud), or a mix? How enterprise-ready those services are is also a factor.

We believe that [shared responsibility](#) is a core component of that effort and a critical concept in securing AI workloads on Google Cloud. This means that both Google and the customer play essential roles in safeguarding AI systems, as further discussed [here](#).

Looking through the lens of how a customer might participate in the AI ecosystem, we see [four basic scenarios](#) that require different risk management strategies: build it yourself, customize the model to your needs, integrate the model as-is, or consume the model out of the box. A key difference between these four scenarios is the level of direct control an organization has over the AI model, as compared to what is outsourced to an external provider, such as Google Cloud. Across all four scenarios, customers can rely on Google Cloud to uphold our strong [AI privacy commitments](#) and to [protect customers' data](#), enabling them to pursue data-rich use cases while [complying](#) with relevant regulations and laws.

Best Practices



Governance

Organizations can successfully implement AI by following [best practices](#) like identifying stakeholders, defining principles, utilizing frameworks, documenting policies, articulating use cases, leveraging data governance, collaborating with relevant departments, establishing escalation points, providing status visibility, and implementing an AI training program. These practices help organizations navigate AI implementation challenges and ensure responsible integration.

Acceptable Use

Organizations that want to use AI in a safe, secure, dependable, and robust way should devise their own “building code” for gen AI through an internal [Acceptable Use Policy](#) (AUP). It’s important to align the use of Gen AI to an organization’s overall goals and values, as well as the broader regional and industry requirements that may apply. An AUP can be an important, multifaceted guide in shaping any organization’s governance structure and its relationship to Gen AI because it ties into other organizational governance pillars, including broad-scale awareness campaigns, training, and ongoing monitoring for compliance.

Security

While gen AI does represent a new security world, it’s not the end of the old security world, either. Securing AI does not magically upend security best practices, and much of the wisdom that security teams have learned is still correct and applicable. We believe that many of the security principles and practices that apply to traditional systems also apply to AI systems. By understanding the [differences between securing a traditional enterprise software system and an AI system](#), organizations can develop a more comprehensive security strategy to protect

their AI systems from a variety of security threats. Now is also the time to [take steps](#) to prevent potential attacks from happening in the first place. When securing AI systems, it is important to think like an attacker. Consider known weaknesses and identify the ways that an attacker could exploit a system. Work with other teams in the organization – including data science, engineering, and security – to develop a comprehensive security.

Privacy and Data Governance

ML models learn from training data and make predictions on input data. Sometimes the training data, input data, or both can be quite sensitive. Although there may be benefits to building a model that operates on sensitive data, it's essential to consider the potential privacy implications in using sensitive data. This includes not only respecting the legal and regulatory requirements, but also considering social norms and typical individual expectations. It's essential to offer users transparency and control of their data.

Fortunately, the possibility that ML models reveal underlying data can be minimized by appropriately applying various techniques, some of which include:

- Identify whether your ML model can be trained without the use of sensitive data, e.g., by utilizing non-sensitive data collection or removing sensitive data from the training set.
- If it is essential to process sensitive training data, strive to minimize the use of such data. Handle any sensitive data with care: e.g., comply with required laws and standards, provide users with clear notice and give them any necessary controls over data use, follow best practices such as encryption in transit and rest, and adhere to privacy principles such as the ones found on the [Google Cloud Privacy Resource Center](#).
- Anonymize and aggregate incoming data using best practice data-scrubbing pipelines: e.g., consider removing personally identifiable information (PII) and outlier or metadata values that might allow de-anonymization, for example by using the [Cloud Data Loss Prevention](#) API to automatically discover and redact sensitive and identifying data.

Staying on Top of Gen AI Developments

We're often asked how to [stay on top of AI developments](#), both technological and regulatory, and how to empower teams with the knowledge, skills, and an understanding of the risks in using gen AI. When approaching how to enable your workforce for gen AI adoption, it's important to recognize it isn't just about technology, but about investing in your people. By demystifying gen AI, focusing on strategic skills building, and creating a culture that values continuous learning, your enterprise can unlock the full potential of gen AI as a transformative technology and prepare for the future of work. It is critical that both IT and business teams understand how gen AI works, how these risks materialize, and what to do about them.

We believe the best way to learn gen AI is to actually use the models – experiment with them, spend time with them, and apply them in your work. [Learning paths](#) are also available to provide customers with a wide range of upskilling offers for different roles and expertise levels on, for example, gen AI concepts, fundamentals of large language models, and responsible AI principles. In addition, our [recommended practices for AI](#) are a helpful guide to follow when designing, developing, testing and using AI systems with a focus on fairness, interpretability, privacy, safety and security, including relevant examples and documentation for the implementation of each.

Conclusion

We aim to be at the forefront of advancing AI through our deep research to develop more capable and useful AI. We're pursuing innovations that will help unlock scientific discoveries and tackle some of humanity's greatest challenges. From this research and development, we are bringing innovations into the world to assist people and benefit society everywhere through our infrastructure, tools, products, and services, and by enabling and working with others to benefit society.

To further the dialogue, we publish educational content, research and other forms of documentation to enable transparency and support our customers. These include [Responsible AI Guides](#) with best practices to assist customers in defining AI use cases and assessing their impact, and [AI research](#) on a range of topics including machine intelligence, natural language processing, and many others that maximize both scientific and real-world impact.