

Deloitte.



A new frontier
in artificial intelligence
Implications of Generative AI for businesses

Deloitte AI Institute

Executive summary

The year 2022 was a watershed year for artificial intelligence (AI), with the release of several consumer-facing applications like ChatGPT, DALL.E, and Lensa. The common theme the use of Generative AI—a paradigm shift in the world of AI. While current generations of AI use pattern detection or rule-following to help analyze data and make predictions, the advent of transformer architectures has unlocked a new field: Generative Artificial Intelligence. **Generative AI can mimic the human creative process by creating novel data similar to the kind it was trained on, elevating AI from enabler to (potentially) co-passenger.** In fact, Gartner estimates that more than 10% of all data will be AI-generated by as early as 2025,¹ heralding a new age, the Age of With™.

Although early traction has been through consumer releases, which could be era-defining, **Generative AI also has the potential to add contextual awareness and human-like decision-making to enterprise workflows, and could radically change how we do business.** We may be only just beginning to see the impact of solutions like Google's Contact Center AI (CCAI), which is designed to help enable natural language customer service interactions,² and industry-specific solutions like BioNeMo from NVIDIA, which can accelerate pharmaceutical drug discovery.³ As such, Generative AI has attracted interest from traditional (e.g., Venture Capital (VC), Mergers & Acquisitions (M&A)) and emerging (e.g., ecosystem partnerships) sources. In 2022 alone, venture capital firms invested more than \$2B,⁴ and technology leaders made significant investments, such as Microsoft's \$10B stake in OpenAI⁵ and Google's \$300M stake in Anthropic.⁶

The far-reaching impacts and potential value when deploying Generative AI are accelerating experimental, consumer, and soon, enterprise use cases. And **even though much media coverage has focused on consumer use cases, the opportunities are widespread—and some are already here.** Still, questions remain about how individuals and enterprises could use Generative AI to deliver efficiency gains, product improvements, new experiences, or operational change. Similarly, we are only beginning to see how Generative AI could be commercialized and how to build sustainable business models.

Even so, Generative AI is in its infancy and not without risk.

Some of the most important risks to address relate to privacy and security, managing bias, transparency and traceability of results, IP ownership, and equal access, especially for those at greater risk of job displacement. As such, participants should balance commercialization, regulation, ethics, co-creation, and even philosophy, as well as expand the group of stakeholder thinkers and contributors beyond technologists and enthusiasts.

Ultimately, Generative AI could create a more profound relationship between humans and technology, even more than the cloud, the smartphone, and the internet did before.

Various analysts estimate the market for Generative AI at \$200B by 2032.⁷ This represents ~20% of total AI spend, up from ~5% today.⁸ Said another way, **the market will likely double every two years for the next decade. Numbers aside, we believe the economic impact could be far greater. To help understand the potential, this paper is equal parts primer and provocateur, adding structure to a rapidly changing marketplace.** We start with a brief explainer of the foundational elements, delve into enterprise and consumer use cases, shift focus to how players across the market can build sustainable business models, and wrap up with some considerations and bold predictions for the future of Generative AI.

Contents

Section I	Decoding the Generative AI magic trick	5
Section II	Consumer and enterprise use cases for Generative AI	9
Section III	Commerce and competition in Generative AI	17
Section IV	Adopting and commercializing Generative AI	27

About the Deloitte AI Institute

The Deloitte AI Institute helps organizations connect the different dimensions of a robust, highly dynamic and rapidly evolving AI ecosystem. The AI Institute leads conversations on applied AI innovation across industries, with cutting-edge insights, to promote human-machine collaboration in the "Age of With".

The Deloitte AI Institute aims to promote a dialogue and development of artificial intelligence, stimulate innovation, and examine challenges to AI implementation and ways to address them. The AI Institute collaborates with an ecosystem composed of academic research groups, start-ups, entrepreneurs, innovators, mature AI product leaders, and AI visionaries, to explore key areas of artificial intelligence including risks, policies, ethics, future of work and talent, and applied AI use cases. Combined with Deloitte's deep knowledge and experience in artificial intelligence applications, the Institute helps make sense of this complex ecosystem, and as a result, deliver impactful perspectives to help organizations succeed by making informed AI decisions.

No matter what stage of the AI journey you're in; whether you're a board member or a C-Suite leader driving strategy for your organization, or a hands on data scientist, bringing an AI strategy to life, the Deloitte AI institute can help you learn more about how enterprises across the world are leveraging AI for a competitive advantage. Visit us at the Deloitte AI Institute for a full body of our work, subscribe to our podcasts and newsletter, and join us at our meet ups and live events. Let's explore the future of AI together.

www.deloitte.com/us/AIInstitute

SECTION I

Decoding the Generative AI magic trick

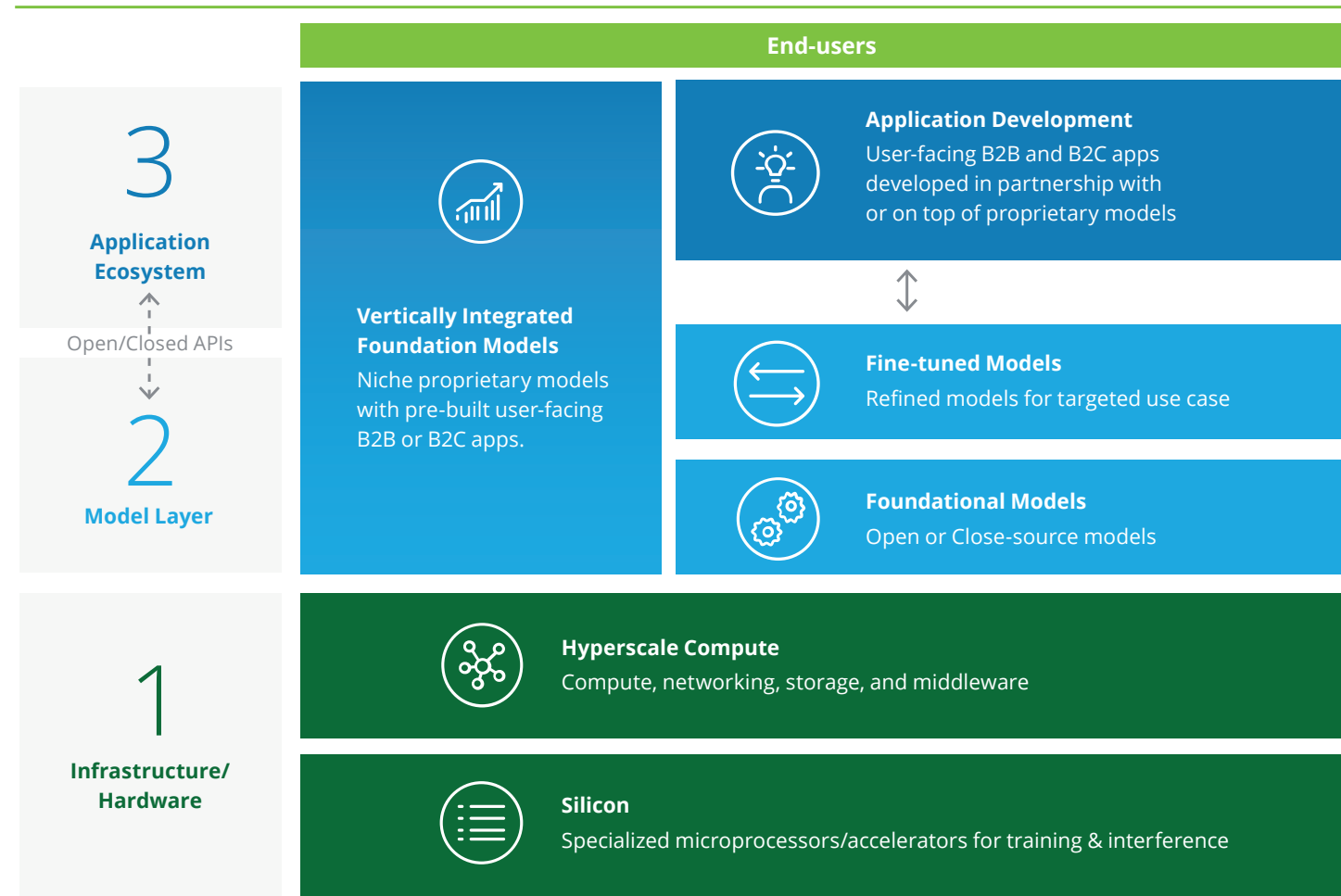
The lofty expectations for Generative AI depend on continued progress and innovation across an interconnected hardware, software, and data provider ecosystem.

The tech stack underlying Generative AI, however, is in some ways similar to others that came before. It consists of three layers: infrastructure, platform, and applications.

Infrastructure is generally accepted as the most established, stable, and commercialized layer. Incumbents offer compute, networking, and storage, including access to specialized silicon (microprocessors) like NVIDIA's GPUs and Google's TPUs optimized for AI workloads. Meanwhile, the application layer is evolving rapidly and consists of leveraging and extending foundation models, which is Generative AI's equivalent of a platform.



Generative AI Tech Stack



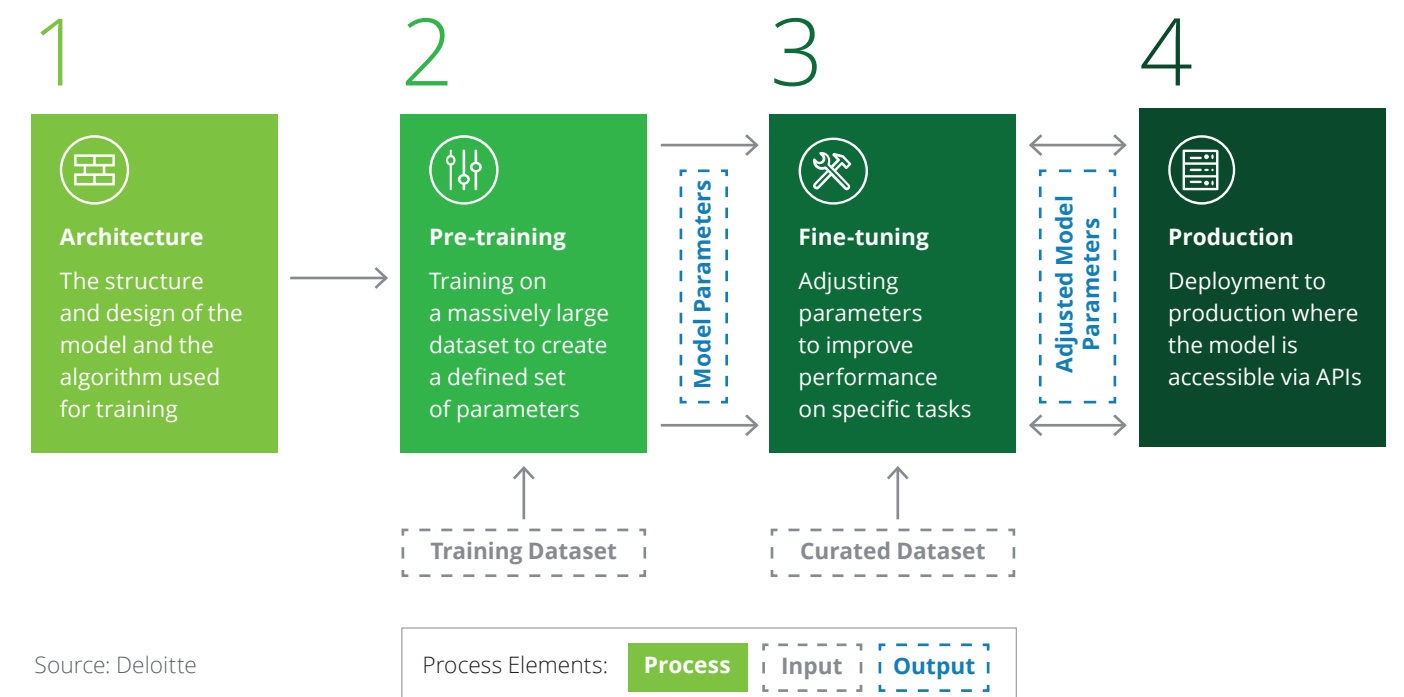
Source: Deloitte

Foundation Models, however, are what differentiate the Generative AI tech stack from AI that came before. At its core, a Foundation Model, a term coined by Stanford University's Center for Research on Foundation Models, is a machine learning (ML) model pre-trained on a broad dataset that can be adapted to solve a range of problems.⁹ Just as Microsoft's Win32 offers APIs for developers to access base-level hardware and OS functions, and NVIDIA's CUDA allows graphic-intensive applications like game engines simplified access to GPU resources, the model layer is designed to connect ambitious application developers to optimized hardware to help accelerate the adoption of and democratize Generative AI.

These models are often available to developers via closed and open APIs, where developers can fine-tune models with additional training data to improve context, relevance, and performance to specific use cases, all while optimizing delivery costs.

Foundation models are typically developed in four stages, which are illustrated below.

Development of Foundation Models



Source: Deloitte

While this framework is applicable across AI architectures, state-of-the-art Foundation Models today (e.g., GPT-3, Stable Diffusion, Megatron-Turing) are based on a neural network architecture called transformers, invented by a team at Google Brain in 2017.¹⁰ Transformers represent a step change in ML performance and differ from prior architectures in their ability to assign context, track relationships, and predict outcomes. The most mature Foundation Models today are in the text domain, primarily driven by vast quantities of available training data, which accelerated the development of Large Language Models (LLMs), a type of Generative AI foundational model. LLMs are trained to generate text by predicting the next word in a sequence or missing words within a paragraph.

Moreover, Generative AI can create artifacts across various modes—code, images, video, audio, and 3D models. **This could both disrupt and drive step changes in productivity across a range of capabilities, from copywriting to research and software engineering.** For example, in advertising, Generative AI could create original copy, product descriptions, and images in seconds. It can generate synthetic X-ray images in healthcare, helping physician diagnostic training.

Indeed, Generative AI could transform how businesses operate and interact with customers and may even redefine an "employee" as we know it. This transformation is already underway in some consumer and enterprise spaces.



SECTION II

Consumer and enterprise use cases for Generative AI

In 2022, OpenAI's DALL·E 2 captured the world's attention with its text-to-image capabilities.¹¹ The model creates images from simple prompts, from something as direct as "a lion in a jungle" to something more comical like "two lions playing basketball in the style of Picasso."

Ever since, Generative AI has occupied the news cycle, punctuated by other launches like ChatGPT and previews like MusicLM. No wonder we've seen broad-market consumer use cases, like Bing's internet search powered by OpenAI's ChatGPT.¹² These are emblematic of a Cambrian explosion in consumer apps, touching everything from search to therapy.

To help contextualize this explosion, we group consumer use cases—those that individuals invoke in their personal lives—into four broad categories based on the utility provided:

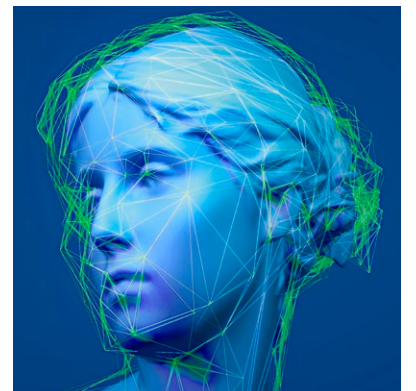
Efficiency | Optimizing tasks like planning, research, and product discovery

Instruction | Providing personalized guidance or learning content

Creation | Generating or enhancing content, replicating the creative process

Entertainment | Building games, virtual personas, and other entertainment

This is just an early view of the market; there will likely be overlapping categories as work evolves. Moreover, new, category-defining use cases are expected to emerge as future generations of AI (e.g., those that enable multi-model engagement or run entirely on-device) mature.

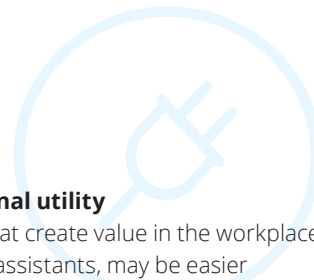


The pace of change can make predictions challenging, but as of early 2023, we expect consumer use cases with the following aspects as having staying power:



Speed to market

Consumer awareness, increasingly through social media, could lower acquisition costs, allowing companies to piggyback on coverage, work out product kinks, and scale efficiently with an active and contributory user base.



Occupational utility

Products that create value in the workplace, like writing assistants, may be easier to fit into a sustainable business model, as opposed to products attached to a “hype cycle,” like social media filters.



Seamless integrations

Solutions that integrate into platforms could be discovered through existing workflows, driving more “sticky” adoption. Grammarly was early to market with this on PCs and, more recently, OpenAI with Bing.

Consumer use cases can also be indicators of the possibilities in the enterprise. However, unlike consumers, enterprises require advanced features, proven ROI, customization, organizational content, security, and technical support. In today’s formative era of Generative AI, the most popular enterprise use cases—invoked to drive internal or B2B outcomes—will be general purpose or applicable across industries or functions (“horizontal”). However, like technologies that came before, there are often more sustainable value-creation opportunities in industry-specific enterprise use cases (“vertical”).

These efficiencies may even redefine job expectations, making prompt engineering (i.e., asking AI the right questions) a differentiating skill set.

Ultimately, horizontal use cases will create a commercial foundation for more specialized applications. Enterprises must start deploying these early to help build capabilities and a knowledge base, making the value case for vertical applications over time.

Today, some enterprises are already driving tangible returns from investments in horizontal use cases.

We’ve seen research teams summarize third-party information, product managers write requirements documentation, social media marketers refine copy, and customer service teams create case summaries and suggested resolutions. **However, tangible ROI could depend on proprietary and serviceable data, secure model partitioning, talented product leaders and ML engineers, enabling MLOps tooling,** and new commercial and operating models. These are investments that enterprises should evaluate, whether they see themselves as early adopters, fast followers, or late entrants.

Potential targets of horizontal use cases are well-established automation centers, offer a substantial volume of training data (e.g., knowledge base, support chat logs), and are the focus of cost optimization and productivity improvement efforts. For example, creative marketing tasks like writing advertising copy, blogs, or social media captioning can take hours or days for humans to author. In contrast, Generative AI can complete workable drafts in minutes, requiring only editing from humans.

A Sampling of Consumer Use Cases Available Today

	Efficiency	Instruction	Creation	Entertainment
High Complexity ↑	Creating a health & wellness plan	Conversing with virtual companions	Generating & editing video files	Creating original games
	Discovering new products	Creating personalized financial plans	Creating interior design mockups	Chatting with pop culture figures
	Conducting research with citations	Teaching new languages	Curating outfits & fashion ideas	Rendering 3D environments
Low Complexity ↓	Curating content	Synthesizing research papers	Modifying & editing design files	Remixing or sampling music
	Answering general questions	Guiding & informing personal writing	Creating art & editing images	Generating original fictional short stories
Sample vendors				
	Synthesis.ai Consensus	Grammarly Lingostar.ai	Luminar AI Lensa	Jasper Scenario

Source: Deloitte

Sampling of Vertical and Horizontal Enterprise Use Cases

	Consumer & Retail	Life Sciences & Health Care	Banking & Fin. Services	Technology	Media & Telecom	Industrial & Manufacturing	Government & Public Sector
VERTICAL	Personalized Conversational Retail Experience ● ●	AR/VR Content Generation for Digital Therapy ● ● ● ●	Fraud Simulation & Pattern Detection ● ●	Personalized AR/VR Experience Generation ● ● ● ●	Original Games Creation ● ● ● ● ● ●	Geological Assessment for Oil Exploration ● ● ● ● ●	Academic 24/7 Office Hours Virtual Assistant ● ● ● ● ●
	Customized Product Design & Recommendation ● ● ● ●	Predictive & Virtual Patient Triage ● ●	Tax and Compliance Audit & Scenario Testing ● ●	Automated Product & Hardware Design ● ● ● ●	Trailer & Summary Generation ● ● ● ●	Generative Simulation & Safety Testing ● ●	Infrastructure Mapping & Planning ● ● ● ●
	Product Details & Photography Generation ● ● ●	3D Images of Anatomy for Education ● ●	Retail Banking Transaction Support ● ●	Personalized & Automated UI/UX Design ● ● ● ●	Script/Score Design & Subtitle Generation ● ●	3D Env. Rendering: Well Sites, Pipelines, etc. ●	Disaster Recovery Simulation ● ●
	Fashion Outfit Curation ● ● ●	Healthy & Wellness Plan Creation ● ● ● ●	Personalized Virtual Financial Advisor ● ●	Product Testing & Feedback Generation ● ● ●	Personalized News & Content Generation ● ● ● ●	Automated Tech. Equipment Training ● ● ● ● ●	Fraud, Waste & Abuse Prevention Reports ●
	Personal Art Creation & Edits ●	Drug Discovery Through Molecule Simulation ●	Financial Reporting Analysis & Insight Gen. ●	Software Sales, CX & Retention Support ● ●	Original Fictional Short Stories Generation ●	Generative Automation for Smart Factories ● ● ●	Research w/ Citations & Explaners ●
HORIZONTAL	Personalized Conversational Retail Experience ●	Self-serve HR & IT Functions ● ●	End-to-end Automated Customer Service ● ●	Customer Feedback Sentiment Classification ● ●	Automated Code Debugging & Issue Resolution ● ● ●	Dialogue Generation for Virtual Assistants ●	
	Enterprise Search & Knowledge Mgmt. ● ● ●	3D Environment Rendering: Metaverse ●	Marketing/Sales Content Generation ● ●	Accessibility Support (text-to-speech & speech-to-text) ● ●	Autonomous Code Generation & Completions ● ● ●	Personalized Targeted Ads across platforms ● ● ● ● ●	

Generative AI Modality

- Text
- Image
- Audio
- Video
- 3d Model
- Code
- Others

Source: Deloitte

In contrast, vertical use cases target industry-specific workflows that require domain knowledge, context, and expertise.

For these, foundation models may need to be fine-tuned or may even require new special-purpose models. For instance, Generative AI can be used to create a customized portfolio of securities based on risk-reward descriptions or recommend personalized treatment plans based on a patient’s medical history and symptoms. However, delivering performant vertical use cases requires a nuanced understanding of the field. In software, for example, Generative AI can design composable blocks of code based on simple prompts, which requires tacit knowledge of efficient coding, coding languages, and an understanding of technical jargon.

Enterprise buyers have unique purchase decisions relative to consumers, as **model performance (speed, relevance, breadth of sources) is not expected to exclusively drive vendor selection.** on early opinions from both advocates and naysayers, frequently cited criteria to adopt Generative AI are:

Ease of use | Integrations into systems and workflows via out-of-the-box connections and low/no code tooling, reducing expensive IT resources and enabling frontline users.

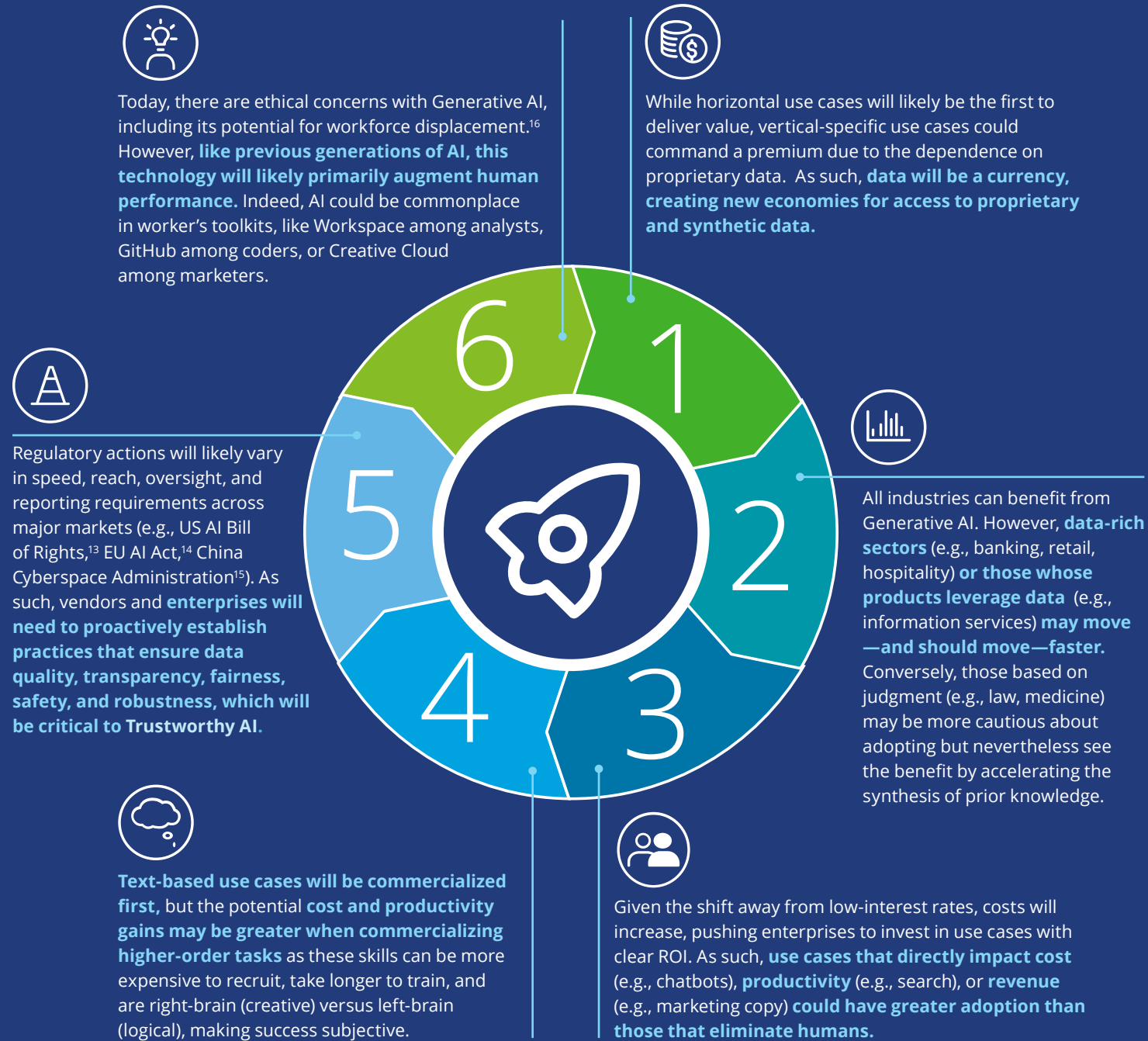
Security and privacy | Compliance with data security standards (e.g., SOC 2, HIPAA, GDPR) and role/persona-level access control over confidential data.

Robust ecosystems | Broad set of development and service partners to extend, customize, and co-develop specialized data sets, use cases, and applications.

Transparency and explainability | Understanding how model outputs and responses are derived and the ability to perform root cause analysis on inaccurate results.

Flexibility and customization | Ability to create parameters, train on proprietary data, and customize embeddings while maintaining privacy and ownership of data and tuning.

Even as new use cases emerge at an accelerating pace, we believe the market will unfold in six ways:



Despite its promise, myriad challenges should be overcome before Generative AI can be deployed at scale. We discuss these in more detail, but there is also the question of commercial viability. In other words, for all the fascinating possibilities and use cases for Generative AI, it still needs to be determined how vendors will build a sustainable business model.





SECTION III

Commerce and competition in Generative AI

The battle for value capture will be fought on multiple fronts, and each layer of the stack will have its competitive dynamics driven by things like scale, data access, brand, and a captive customer base.

However, we see two primary competitor archetypes: pure-play providers operating within a single layer—infrastructure, model, and application - and integrated providers that play in multiple layers. As with incumbent technology, we expect consumer pricing to be simple (e.g., per user, per month) and enterprise pricing to be more complex (e.g., per call, per hour, revenue share). However, **pricing simplicity, predictability, and value will be important to scaling within the enterprise beyond early adopters or edge use cases.**

To begin, the infrastructure layer, which is the most mature of the Generative AI technology stack, is where hyperscalers dominate the market. The business model here is proven: provide scalable compute with transparent, consumption-based pricing. To help make Generative AI workloads “sticky,” hyperscalers have entered commitments with model providers to guarantee future workloads, including Azure with OpenAI,¹⁷ Google with Anthropic,¹⁸ and AWS with Stability.ai,¹⁹ alongside their proprietary models.



While the cloud service providers (CSP) deliver abstracted services, there is another enabling layer within infrastructure that is rapidly evolving: silicon.

Here, NVIDIA is a leader with their Ampere and Hopper series GPUs purpose-built for training and inference workloads, respectively, coupled with their Selene supercomputing clusters that speed up training time.²⁰ Similarly, AMD's CDNA2 Architecture is purpose-built for exascale computing on machine learning applications, advancing competition in the high-performance computing market.²¹



Infrastructure Layer

Offering	Description	Examples	Primary Customer			Primary Monetization	
			Enterprise	Developer	Consumer	Model	Metric
Cloud Service Provider	Hyperscale and purpose-built compute, storage, and networking	Amazon Baidu Google Microsoft	Yes	Yes	No	Consumption	Per minute By CPU/GPU type
Generative AI Service Providers	Specialized services to accelerate deployment (e.g., security, monitoring, testing, model isolation)	Amazon Co:here Google	Yes	Yes	No	Consumption	Per hour Per generation Per embedding
Chip Provider	Purpose-built semiconductors, including GPUs and CPUs	AMD NVIDIA	Yes	No	No	One-time lease	Per component

Next is the model layer, where the market is evolving fast. This area can be resource intensive; model builders must continually revisit architectures (e.g., parameters, embeddings) to maintain performance. They have to attract and retain AI talent (i.e., architects, engineers, and data scientists) to design the frameworks, guardrails, and learning mechanisms to ensure the robustness and reliability of models. Finally, Generative AI **workloads can run up large bills due to their compute-heavy nature and need for specialized silicon.**²² No wonder we've seen players start to recoup the investment by charging fees or integrating into monetized products (e.g., GPT-3.5 into Edge, LaMDA into Google Search).

Another less-considered path to monetization could be developing and licensing model architectures or development platforms.

In other industries, like semiconductors, ARM (CPU) and Qualcomm (wireless networking) create large, stable business models built on licensing fees.



Model Layer

Offering	Description	Examples	Primary Customer			Primary Monetization	
			Enterprise	Developer	Consumer	Model	Metric
Closed-source Model Providers	Hosted and managed models built on a vast data corpus	Co:here Google OpenAI	Yes	Yes	No	Consumption	Per token Per API call
Open-source Model Providers	Foundation models maintained by communities	Meta Stability.ai	No	Yes	No	Monetized via fine-tuned models or model hubs	
Fine-tuned Model Providers	Use case-specific versions of foundation models	Co:here C3.ai	Yes	No	No	Consumption	Per token Per API call
Model Hubs	Marketplace, community or hosting services for models	Github Hugging Face Replicate	Yes	Yes	No	Subscription Consumption Rev. share	Per month Per hour
Model Service Providers	Proprietary architectures, synthetic data, weights, and embeddings	Co:here MostlyAI RealAI	No	Yes	No	One-time Subscription License	Per embedding Per month Per user

Finally, the application layer serves as the gateway between models and end users.

Today's apps are typically monetized through subscriptions and recurring transactions, a model that will likely persist, albeit with modifications suited to Generative AI.



Application Layer

Offering	Description	Examples	Primary Customer			Primary Monetization	
			Enterprise	Developer	Consumer	Model	Metric
Platforms	SDKs, frameworks, and tools to build and distribute apps	Google Hugging Face Microsoft	Yes	Yes	No	License Rev. share	Per user
Standalone application	Full-feature solutions to modify workflows	Boomy Canva Lensa	Yes	No	Yes	Subscription Consumption One-time	Per user Per month Per service
Plugins	Extensions and features to supplement tasks and workflows	AI Art Grammarly Jasper	Yes	No	Yes	Subscription Consumption	Per user Per month

Competition within the application layer could unfold within several markets. However, given the wide range of applications and use cases that may emerge, we should look at “micro-markets.” Broadly, today’s real and predicted enterprise use cases fall into five categories where competitive lines could be drawn:



Accelerate

Improve productivity by speeding up outcomes. These do not eliminate human intervention but provide high-quality inputs upon which to build.



Personalize

Create intimacy and personalization, which previously would have taken significant effort. Here, models can leverage personal data to tailor content.



Automate

Deliver business and technical workflows and, in certain instances, replace humans. Vendors often demo these due to the immediate cost-saving potential.



Create

Push the boundaries of intellectual property development, leveraging prompts (a new art form unto itself) to generate novel content like images, video, text, and media.



Simulate

Create environments in which workflows, experiments, and experiences can be simulated before being pushed into production, saving time, cost, and physical resources.

Sampling of Enterprise Micro-Markets

	Accelerate	Personalize	Automate	Create	Simulate
APPLICATIONS	Email outreach	Social media marketing	Calendar mgmt./ Admin assistant	Image/logo creation	3D modeling
	Note taking	Gaming environment design	Keynote speaker notes	Advertising copy	Marketing campaigns
	Content marketing	Physical goods design	Support chatbots	Short-form video generation	Medical testing (R&D)
	Advertising video editing	NLP-based email/ app. responses	Content summarization	Product ideation & PRD authoring	Chemical interactions
	Code completion	Personal assistant	Basic code generation & documentation	Music scoring	Disaster response management
MODELS	Anthropic Co:here OpenAI GPT-3	Facebook OPT GATO Microsoft X-CLIP	BigScience BLOOM OpenAI Codex Tabnine	OpenAI DALL.E 2 Soundify Stable Diffusion	Cradle DreamFusion NVIDIA GET3D

Source: Deloitte

This may have implications for the model and infrastructure layers. The vendors lower in the stack could remain relevant by creating purpose-built infrastructure, models, and services that enable innovation in micro-markets.

A second archetype, in contrast to pure-play providers who monetize through first- and third-party channels, are **vertically integrated or multi-layer players. These players lead with bundled pricing, proprietary data, special-purpose clouds, or cross-domain expertise to gain a competitive advantage.**

We see integration happening in two ways. First, companies like Anthropic and Midjourney have released applications for specific use cases. Lower in the stack, companies like NVIDIA have released specialized models, including BioNeMo, a pharmaceutical pipeline development accelerator that is optimized to run on NVIDIA GPUs.

Integrated Players

Offering	Description	Examples	Primary Customer			Primary Monetization	
			Enterprise	Developer	Consumer	Model	Metric
Model and application	Applications built on proprietary, first-party models	Anthropic Co:here Midjourney OpenAI	Yes	No	Yes	Subscription Consumption	Per month Per user Per service Per download
Model and infrastructure	Fully managed infrastructure and model-as-a-service	Google NVIDIA	Yes	Yes	No	Consumption	Per hour Per API call Per embedding
Silicon and infrastructure	Purpose-built horizontal and vertical clouds for ML workloads	Amazon Azure Google NVIDIA	Yes	Yes	No	Consumption	Per minute By CPU/GPU type
End-to-end	Applications built on first-party models and clouds	None yet	Yes	No	Yes	Consumption Subscription	Per user Per month Per hour

Competitive dynamics are unfolding across both technical and commercial dimensions.

On the technical front, newer, more sophisticated silicon, datasets, and models are emerging, with some models now likely to exceed one trillion parameters.²³ On the commercial side, things are in flux as consumer solutions gain traction (e.g., the Pro version of ChatGPT). In the enterprise, solutions have yet to be commercialized at scale and may continue to be hamstrung by computing costs and risk-averse adoption. As such, we offer a few considerations that organizations should evaluate when thinking of going from product to business.

- As innovations in architecture deliver diminishing returns, performance could depend on the volume and quality of training data, HITL training, and guardrails. As such, **competitive advantage will be driven by access to proprietary datasets and scarce talent.**
- **While critical to the advancement of the field, general-purpose models may not capture the lion's share of value.** Instead, companies that build vertical use cases and industry-focused solutions could have the most potent impact within the enterprise.
- **Hyperscalers will compete against a new generation of CSPs** like Coreweave, targeting fungible AI workloads with custom hardware and cheaper pricing. As such, incumbents should innovate and rely on Independent Software Vendor (ISV) and System Integrator (SI) ecosystems to help retain a competitive advantage.

- **Ecosystem orchestrators and system integrators** will continue to play an important role in assisting enterprises in preparing their data, prioritizing use cases, complying with local regulations, stitching together a panoply of models and applications, and derisking adoption.

- **Integrated plays could have a clearer path to ROI**, primarily due to the high cost-of-compute, which may be abstracted through solutions that combine infrastructure, model, and app. As supply blockages ease and new GPUs emerge, however, ISVs may regain value.

Even though Generative AI will likely herald a new age of productivity, some elements could parallel previous waves of technological transformation, which can serve as a blueprint for the market.



SECTION IV

Adopting and commercializing
Generative AI

Generative AI could transform business models, processes, and value dynamics and change how individuals work, learn, and interact. As with other disruptive technologies, this is likely to transpire slowly at first and then rapidly.

Take software development as an example. By some estimates, less than 1% of people know how to code.²⁴ Yet, **software is integral to many businesses and business models today. Generative AI, if harnessed strategically, can democratize coding and reduce the gap between ideas and revenue** by synthesizing product requirements, converting prompts to code, auditing code to find and address bugs, suggesting code optimizations, and proactively provisioning environments optimized for test and run use cases.

Similarly, Generative AI can optimize the end-to-end customer acquisition funnel. If you are in sales and marketing, consider demand generation, where LLMs could author marketing copy across channels and run digital marketing campaigns. Gartner estimates that 30% of outbound marketing will be synthetically generated by 2025.²⁵ Further down the funnel, Generative AI could gather account intelligence, create a first-call presentation, suggest a talk track to account executives, and document and track

outcomes and actions. Finally, Generative AI could proactively suggest pricing and discounting, author a contract, and update customer and CRM records. This would allow marketers and sellers to focus on higher-value activities, such as developing relationships and applying pricing judgment.

We've discussed other ways that adopters can leverage Generative AI across industries (see section 2), from market research to note taking and improving customer support interactions. Further, there are sectorized use cases like customized financial planning for wealth managers, medical diagnoses in health care, generating new worlds and experiences in media and entertainment, and outfit curation for retailers. In fact, **the benefits that adopters can expect to achieve may be significant**; we've offered some early thoughts below, indexed to the idea of enterprise micro-markets (see section 3).



	Benefit	Accelerate	Personalize	Automate	Create	Simulate
FINANCIAL	Expand TAM ¹		●		●	
	Generate revenue		●		●	
	Reduce investment	●		●	●	●
	Utilize assets	●	●	●		●
OPERATION	Move faster	●		●	●	●
	Reduce labor input	●	●	●	●	
	Reduce non-labor input	●				●
EXPERIENTIAL	Improve experience		●		●	●
	Grow skills		●			●
	Build relationships				●	
	Spur innovation	●		●	●	●

● = Primary Benefit,
¹ Total Addressable Market

Source: Deloitte



With that in mind, for enterprises with a commercial interest in Generative AI, we believe there are five “no-regret” moves to start considering today.

If I am a Technology Provider...	If I am an Enterprise Buyer/User...
<p>1 Take a platform approach to monetization that includes models, out-of-the-box use cases, and low/no code tooling, all while building a web of ecosystem players.</p>	<p>Educate the executive leadership team on the potential and risks of Generative AI to ensure a shared understanding and alignment on a path forward.</p>
<p>2 Build solutions that equally serve the customers and developers, given the critical role the latter will play in scaling through fine-tuning apps and extensions.</p>	<p>Identify and prioritize a set of use cases, starting with horizontal ones, especially in areas within the organization that haven't benefited from automation.</p>
<p>3 Develop a roadmap of verticalized solutions, as these will flip the model from loss leader to margin capture.</p>	<p>Lay out a clear technology strategy, including data engineering and pipelines, MLOps tools, and AI-ready talent.</p>
<p>4 Invest early in cross-modal and multi-modal solutions, as the “second and third generations” of Generative AI are on the horizon.</p>	<p>Determine sources of competitive advantage, especially proprietary data, and begin curating these for the coming wave of Generative AI use cases.</p>
<p>5 Allocate funds to support opportunistic M&A, including acquihires and IP-driven investments, especially if valuations remain muted.</p>	<p>Proactively engage your ecosystem of advisors and partners to create a first-mover advantage, gain favorable pricing, and experiment with new solutions.</p>

Generative AI does present risks, and progress and adoption may slow if these are not considered and mitigated when scaling.

First, models should be continually trained to improve performance, which leads to concerns about exposure to sensitive data, privacy, and security. Next, outcomes can only be as good as the quality of training. Therefore, any data biases (e.g., in representation or sampling) often appear in outputs. Other challenges include determining IP ownership of outcomes, high compute cost, and the need for expensive human-in-the-loop (HITL) reinforcement learning. **Everyone involved in the development, consumption, discussion, and regulation of Generative AI should strive to manage the following identified risks:**

Erosion of trust | Malicious—hallucination, deepfakes, phishing, and prompt injection²⁶—and ambivalent actors—not citing data sources—can expose the attack surface and erode customer trust.

Security and risk | Companies should stay ahead of a rapidly evolving regulatory landscape while maintaining confidentiality of data, embeddings, and tuning with inherently “multi-tenant” models.

Bias and discrimination | Generative AI is prone to mimicking biases and propagating discriminatory behavior if implemented without guardrails and continuous monitoring.

Data privacy and IP obscurity | Models will be trained on a corpus of proprietary, often private data, requiring regulatory compliance, node isolation, and source traceability.

Costs | Costs of a query/prompt using Generative AI can cost up to ten times²⁷ that of an index-based query. While these costs will likely come down over time, the economics should be factored into internal business cases and customer pricing to drive adoption.

Long-term worker displacement | Today, the highest ROI use cases will augment workflows and drive productivity; however, as models advance, there may be a risk of job displacement without proper upskilling and workforce planning.

Ultimately, Generative AI could create a more profound relationship between humans and technology, even more than the cloud, the smartphone, and the internet did before. **If you're willing to go along for the ride as a "true believer," we offer four predictions,** from those likely to unfold in the next 1–2 years to those farther afield:

1

Generative AI will change the future of work. AI agents will become an indispensable utility, and widespread adoption among employees will be the new norm and accelerate the Age of With™. Those who fail to adopt may be left behind in the workplace.

2

The race is not only for data but also trust. As Generative AI moves into the enterprise, it will be subject to intense scrutiny. Adoption, therefore, hinges on the ability to conform to expectations—both intuitive and factual—and earn trust.

3

Hyper-personalization will become a driver of growth. Businesses will leverage the ability to analyze large amounts of customer data to create dynamic, real-time, and tailored experiences, products, services, and communication.

4

LLMs are among the first forms of AI to be "general purpose," albeit text oriented. And while we are afield from multi-model, ubiquitous, cross-domain AI, the seeds have been planted. Could we now be in the first days of Artificial General Intelligence (AGI)?

Beyond this, it can be hard to imagine where Generative AI will take us, including the impact on the future of work, trust, and human-machine interaction. **However, time and again, new technology has allowed humans to conquer greater pursuits, and similarly, Generative AI will drive an unprecedented era of human potential.** Individuals could eventually be free from mundane, repetitive work, potentially allowing humanity to live in novel and unimagined ways.

Endnotes

- 1 [Gartner: Insights for Generative AI](#)
- 2 [ZDNet: Google makes Contact Center AI generally available](#)
- 3 [VentureBeat: Nvidia boosts generative AI for biology with BioNeMo](#)
- 4 [Financial Times: Investors seek to profit from groundbreaking 'generative AI' start-ups](#)
- 5 [Bloomberg: Microsoft Invests \\$10 Billion in ChatGPT Maker OpenAI](#)
- 6 [Financial Times: Google invests \\$300mn in artificial intelligence start-up Anthropic](#)
- 7 [Global NewsWire: Global Generative AI Market Size Value Cap Expected to Grow USD 200.73 Billion By 2032](#)
- 8 [Precedence Research: Artificial Intelligence Software Market Size & Forecast](#)
- 9 [Stanford's Human-Centered Artificial Intelligence: HELM—Center for Research on Foundational Models](#)
- 10 [Google: Attention is All You Need, Vaswani et al.](#)
- 11 [Tech Crunch: Now anyone can build apps that use DALL-E 2 to generate images](#)
- 12 [New York Times: Bing \(Yes, Bing\) Just Made Search Interesting Again](#)
- 13 [The White House: Blueprint for an AI Bill of Rights](#)
- 14 [European Council: Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights](#)
- 15 [Tech Crunch: China's generative AI rules set boundaries and punishments for misuse](#)
- 16 [The Atlantic: How ChatGPT Will Destabilize White-Collar Work](#)
- 17 [Engadget: Microsoft's OpenAI supercomputer has 285,000 CPU cores, 10,000 GPUs](#)
- 18 [Anthropic: Anthropic Partners with Google Cloud](#)
- 19 [AWS: Stability AI builds foundation models on Amazon SageMaker](#)
- 20 [Forbes: NVIDIA Provides More Details On Selene Supercomputer](#)
- 21 [Tech Insights: AMD CDNA2 Targets Supercomputers](#)
- 22 [InfoWorld: The cost and sustainability of generative AI](#)
- 23 [VentureBeat: Google trained a trillion-parameter AI language model](#)
- 24 [Evans Data Corporation: Worldwide Professional Developer Population of 24 Million Projected to Grow amid Shifting Geographical Concentrations](#)
- 25 [Gartner: Beyond ChatGPT: The Future of Generative AI for Enterprises](#)
- 26 [Business Insider: Breaking ChatGPT: The AI's alter ego DAN reveals why the internet is so drawn to making the chatbot violate its own rules](#)
- 27 [Reuters: For tech giants, AI like Bing and Bard poses billiondollar search problem](#)



This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.