

Demystifying generative AI

An insider's view: 4 ways Gen AI can impact the enterprise

Our down-to-earth overview can help your organization evaluate the benefits and challenges of large language models.

brillio



Generative AI's ability to create new text, images and audio makes it a far more disruptive technology than businesses have previously experienced.

Image was created using [Adobe Firefly \(Beta\)](#)



What are the implications of Gen AI for business?



In healthcare, Gen AI can create new compounds for drug molecules



In technology, Gen AI can optimize chip design



In quick-serve restaurants, Gen AI can generate receipts & kitchen orders that note personal preferences



Thanks to ChatGPT's skyrocketing popularity, 2023 is emerging as a breakout year for generative AI. The prototype chatbot has singlehandedly catapulted large language models out of data and research labs—and into enterprise business strategies.

And yet, generative AI's ability to create new text, images and audio makes it a far more disruptive technology than businesses have previously experienced. Microchips, cloud computing, 5G—all pale in comparison to generative AI's potential to change our world and the way we live and work in it.

For enterprises, the use cases are impressive, as they illustrate generative AI's ability to increase efficiency and productivity, reduce costs and introduce new areas of growth. But

the buzz obscures an important reality: Along with the wide-ranging applications comes confusion. Enterprises struggle with FOMO and uncertainty over how generative AI could impact their business.

We're here to demystify it.

Read on to gain a clear view of why generative AI is different for business, and how to balance the science behind it with the problem solving and engineering needed to

make it succeed. We also cover important issues such as infrastructure and Gen AI's tendency to produce "hallucinations" so you can analyze the caveats.

By fully understanding generative AI's unique mix of potential business impact and technology disruption, you'll be better positioned to make informed, grounded decisions that fit your organization's objectives.

1 | Why Gen AI is different for business

Until now, businesses have mainly used artificial intelligence to detect patterns and churn out analytics. Generative AI goes a giant step further. It not only observes and predicts but also converts existing data—in the form of text, images or audio—into entirely new forms.

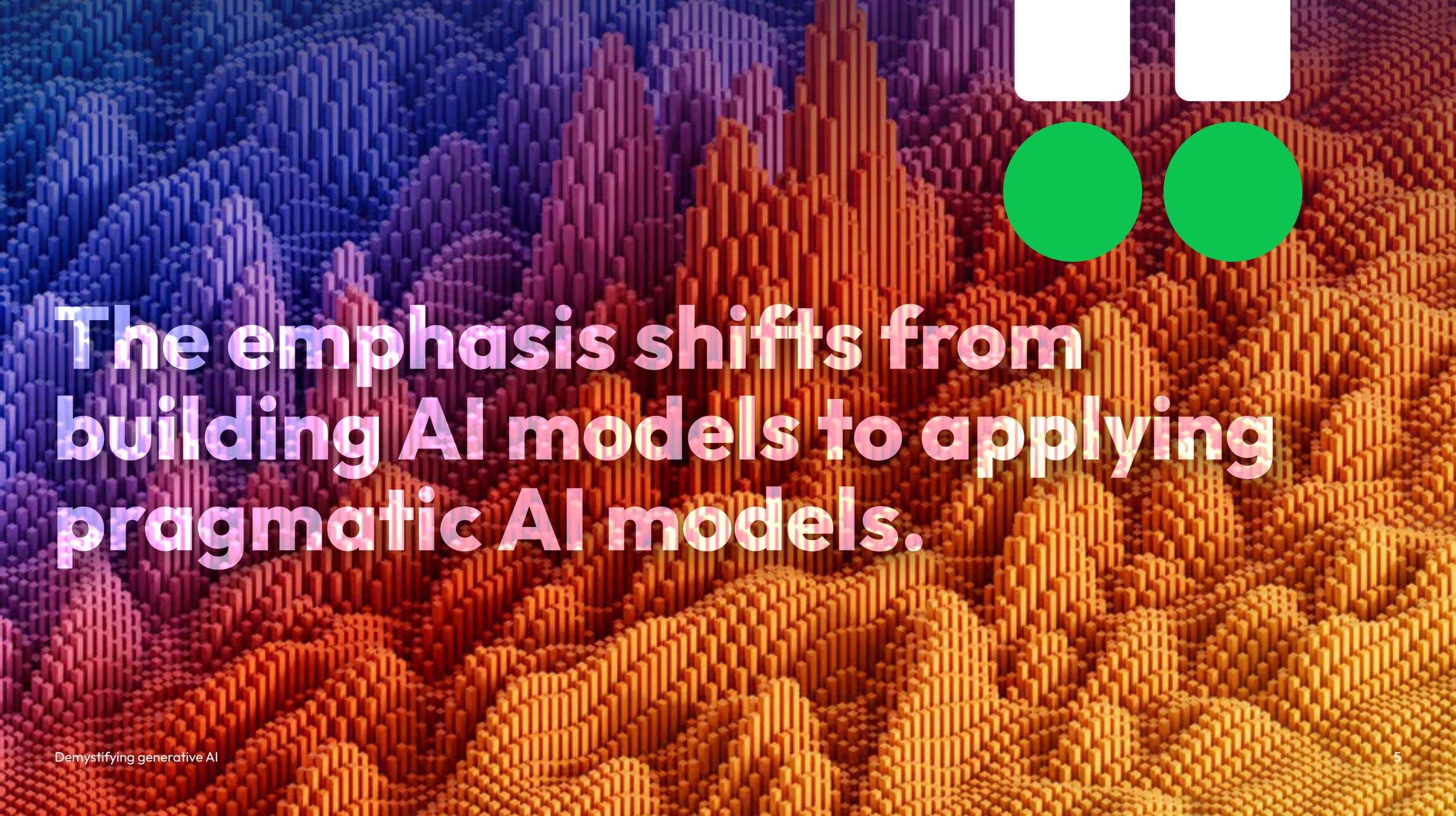
In other words, it doesn't just work with the data fed into its models—it creates new things from it. The implications for business are vast (see sidebar, page 7). In healthcare, generative AI can create new compounds for drug molecules. In technology, it can optimize component placement in semiconductor chip design. In quick-serve restaurants, speech-

and video-enabled kiosks can identify customers, retrieve their profile and offer personalized menu items.

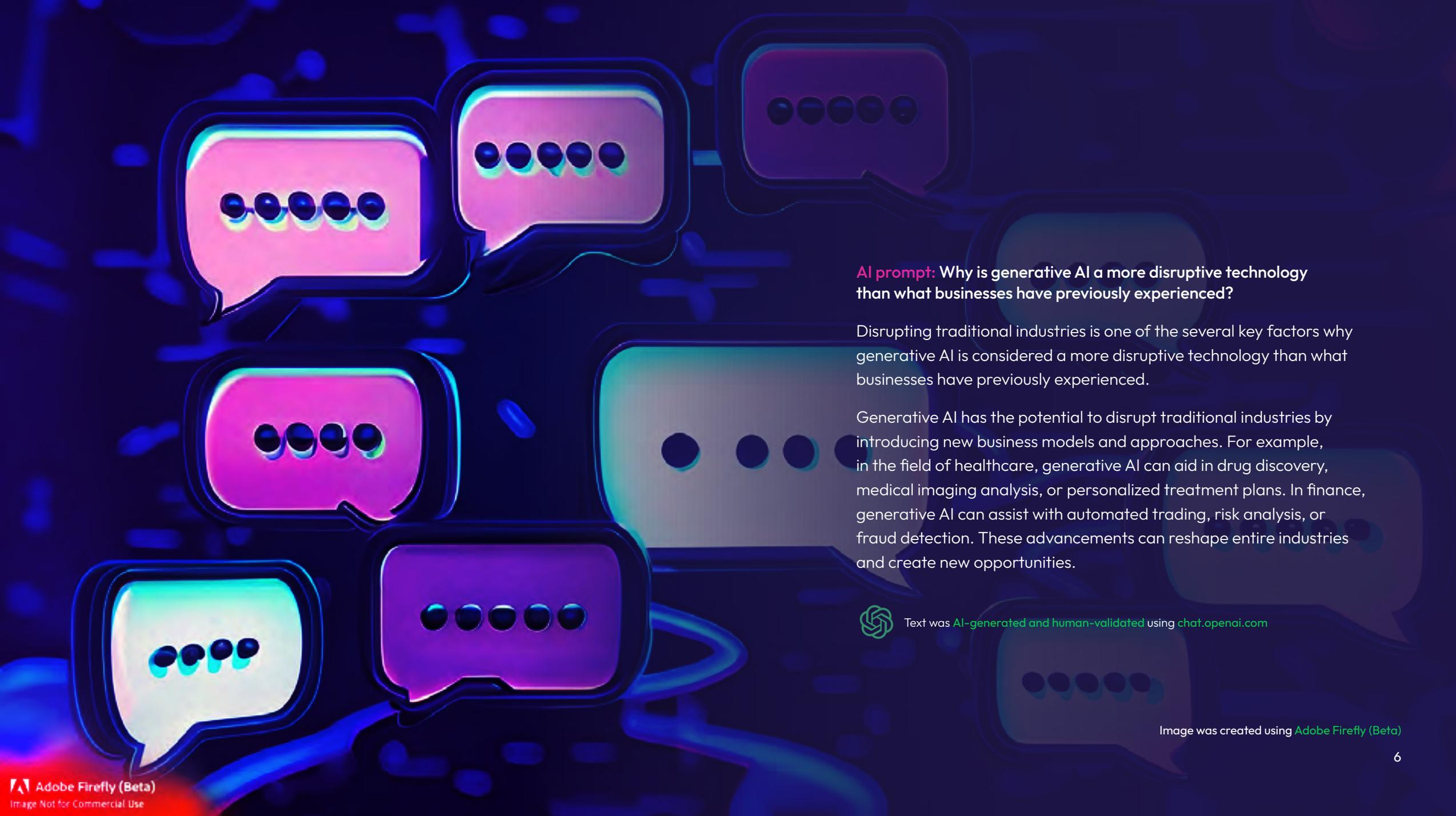
Large language models (LLM) are a type of generative AI trained to use internet data to produce an output. There are many LLM models available, from BLOOM, LaMDA and LLaMA, to MT-NLG, Alpaca and FLAN UL2. Some are open source, and others are available as paid versions.

By far, the best known LLM right now is GPT-4 (fourth-generation Generative Pretrained Transformer) from OpenAI, the San Francisco-based company that released ChatGPT.

LLMs underscore another fundamental difference for generative AI: Instead of building new models from scratch, developers fine-tune and test pre-existing models. The emphasis shifts from building AI models to applying pragmatic AI models. As a result, the skill sets required shift from mathematics and the ability to build an algorithm and test it, to more deeply understanding the human element of a use case. It's a significant change for development teams.



The emphasis shifts from building AI models to applying pragmatic AI models.



AI prompt: Why is generative AI a more disruptive technology than what businesses have previously experienced?

Disrupting traditional industries is one of the several key factors why generative AI is considered a more disruptive technology than what businesses have previously experienced.

Generative AI has the potential to disrupt traditional industries by introducing new business models and approaches. For example, in the field of healthcare, generative AI can aid in drug discovery, medical imaging analysis, or personalized treatment plans. In finance, generative AI can assist with automated trading, risk analysis, or fraud detection. These advancements can reshape entire industries and create new opportunities.



Text was AI-generated and human-validated using chat.openai.com

Image was created using [Adobe Firefly \(Beta\)](https://firefly.adobe.com)

An array of business use cases for Gen AI



Banking

Algorithm training. Create synthetic data to train the machine learning algorithms used in KYC. Generative AI can also create more accurate natural language processing models for virtual assistants.

Interpret loan applications. Assess small business loan applications that contain nonnumeric data, such as business plans.

Real-time customer analysis. Speed commercial banking tasks, such as answering questions in real time about a customer's financial performance in complex scenarios.



Healthcare and life sciences

Drug discovery and development. Generate new compounds for drug molecules and optimize drug candidates.

Accelerate clinical trials. Construct summarizations, Q&As, translation and knowledge graphs out of massive volumes of unstructured data.

Medical insurance assistant. Interact with customers and provide assistance on inquiries regarding health plans and other issues.

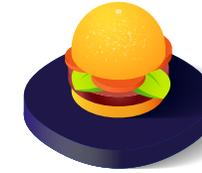


Technology

Chip design. Use AI to optimize component placement in semiconductor chip design.

Procedural content generation. Produce game content such as levels, maps and quests, based on predefined rules and criteria.

Game analysis. Analyze player data, such as game-play patterns and preferences, to provide personalized experiences and help developers increase player engagement and retention.



Quick-service restaurants

Self service. Use speech- and video-enabled kiosks to identify customers and retrieve their profile and previous transactions to offer personalized menu items.

Automated bill generation. Generate receipts and kitchen orders that note personal preferences.

Personalization. Create unique combos for customers to deliver customized experiences.

2 | The challenge: Balancing the science with problem solving

Given all the attention being paid to ChatGPT, it's no surprise that 67% of senior IT leaders plan to prioritize generative AI over the next 18 months, according to [a new Salesforce survey](#). The survey also underscores the divide among leaders' opinions of the technology: While 33% say generative IT is their top priority, an equal number say it's over-hyped.

Despite the anticipation—or perhaps because of it—leaders struggle with how to approach LLMs. In the race to capitalize on ChatGPT's potential, they devote substantial resources to evaluating the language models and

algorithms—and shortchange the assessment of generative AI's ability to solve current challenges and meet their business objectives.

That is, they over-index on the science of the technology and overlook the problem solving and engineering that's equally necessary for successful generative AI initiatives.

For example, it's not uncommon for business use cases to rest on overly general statements such as, “how can we use ChatGPT to improve CX”? Or, “how can we use it to resolve service requests faster”?

Businesses gain far more useful insights when they invest the time in evaluating which customer touchpoints would benefit most from applying generative AI, and where it would produce meaningful business outcomes. For example, generative AI-powered chatbots can learn to respond swiftly to complex customer queries, but they're also useful for automating customer surveys or improving segmentation.

67%

of senior IT leaders plan to prioritize generative AI over the next 18 months.

33%

say generative IT is their top priority; an equal number say it's over-hyped.

Source: Salesforce

Demystifying generative AI





5 caveats of Gen AI

ChatGPT's current ubiquity makes generative AI seem like a ready-set-go technology. The reality is far more nuanced. Generative AI is an emerging technology that requires business leaders to proceed with an abundance of caution.

Not only are technologists still learning about it, but many practical and ethical issues also remain unanswered:

- 1 Data ownership.** All information uploaded to ChatGPT becomes part of its model.
To ensure sensitive data is not exposed, more than a few enterprises have blocked access to ChatGPT.
- 2 Accuracy of results.** Like people, generative AI can produce inaccurate information in response to users' questions. Worse, there's no built-in mechanism to signal inaccuracies to the user, or to challenge the result being provided.
- 3 Tagging of inappropriate content.** The built-in filters of current models are not effective at catching inappropriate content, such as profanity.
- 4 Systemic biases.** Gen AI systems draw from massive amounts of data that can include unwanted biases. For organizations, a key task will be adapting the technology to incorporate their ethics and values.
- 5 Intellectual property claims and ownership.** For results generated from public data, plagiarism is a potential risk that can lead to questions of legal claims and ownership.



3 | Don't overlook infrastructure

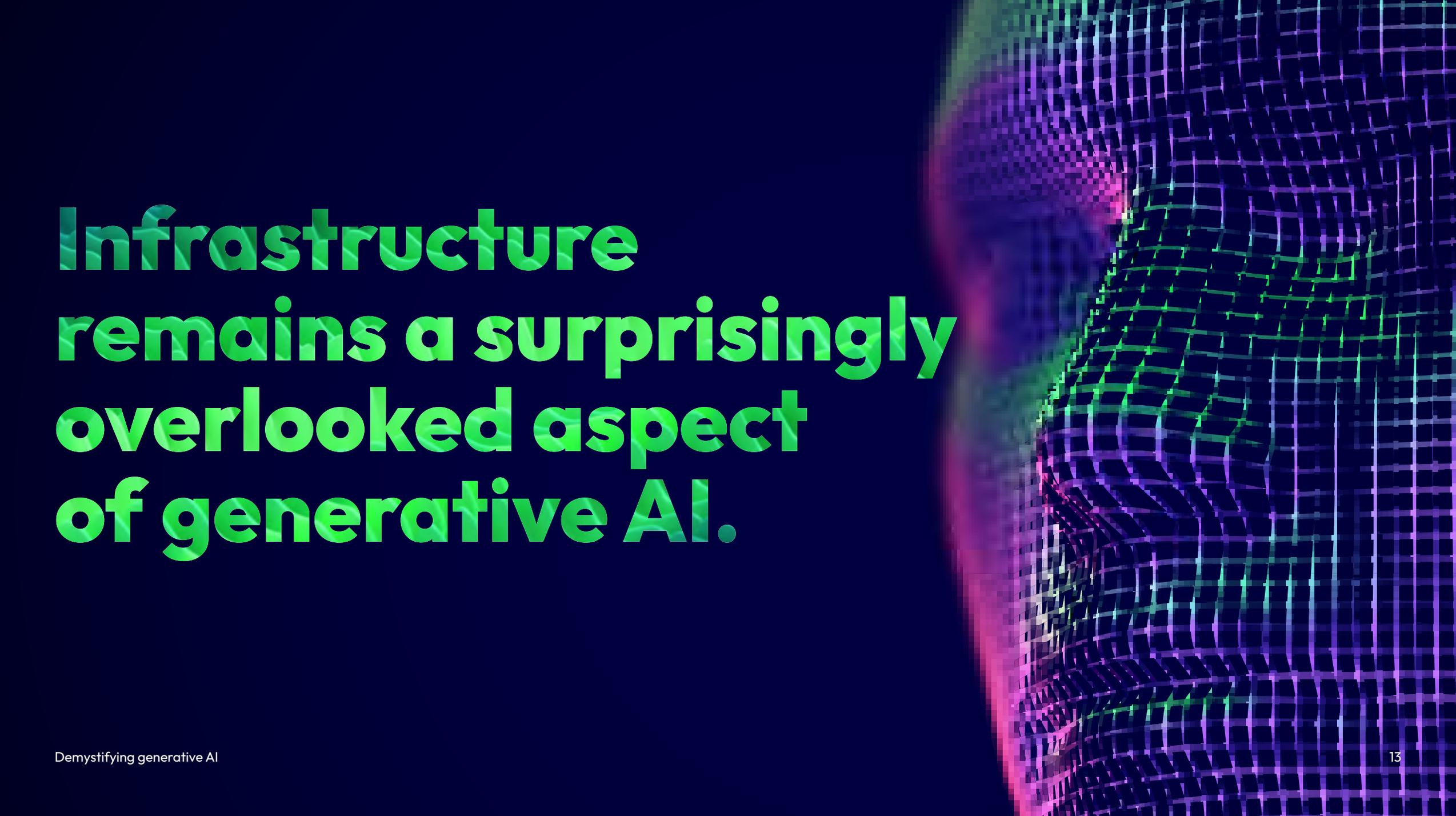
Infrastructure remains a surprisingly overlooked aspect of generative AI. Maintaining LLMs requires businesses to be in the right place on the AI and data maturity curve and to use the design technique that's most appropriate for their existing architecture.

For example, an organization at Level 0 on the maturity curve won't have an adequate infrastructure for building and scaling AI due to a low inventory of AI/ML models. Creating a proof of concept will require a wide range of tasks, from setting up the infrastructure from scratch and identifying the right data sets and models, to fixing data quality issues and implementing metrics to track ROI.

An enterprise that's further up the maturity curve may find its model inventory over-indexed on classic ML and structured data; it might have the right architecture for data but not for ML Ops. In this case, the company's generative AI efforts would need to include a focus on the AI scaling and maintenance infrastructure.

Organizations with the highest level of maturity, reflecting a strong infrastructure for data and ML Ops, have the foundation in place to focus on building a generative AI model to achieve defined outcomes.

In addition, because LLMs are resource-intensive, and can lead to high latency and high costs, building and scaling them requires an architecture that's been designed for optimization and efficiency. Optimization approaches include gradient accumulation and checkpoint, AdamW, mixed precision training and DeepSpeed ZeRO. Efficient inference techniques include low-rank parameterization, model quantization, hardware acceleration and model parallelism.



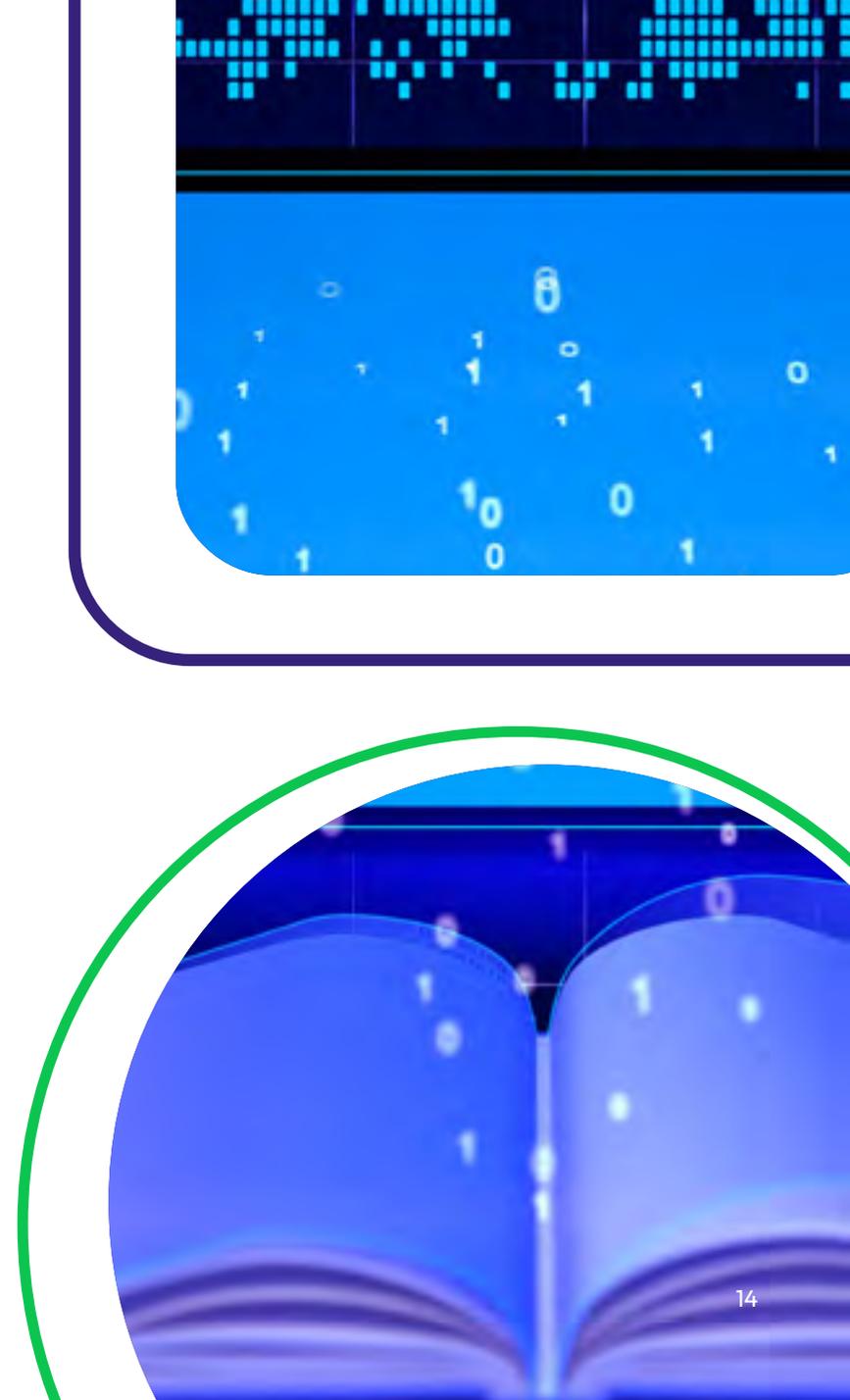
**Infrastructure
remains a surprisingly
overlooked aspect
of generative AI.**

4 | What about LLM hallucinations?

Still frustrating to developers is generative AI's ability to produce false statements known as hallucinations. ChatGPT has been known to cite **fake newspaper articles**, invent **research that doesn't exist** and even **profess its love** for users.

Although earlier versions of GPT struggled with hallucinations, GPT 4 has made significant improvements in mitigating this problem. However, the challenge of dealing with falsehoods continues to persist. [This paragraph was **AI-generated, human-validated content**. Generated using chat.openai.com.]

For enterprises that work with LLMs, that means it's critical to develop a strategy that minimizes hallucinations by evaluating and addressing them mathematically. For example, businesses can conduct a fact-checking evaluation to compare generated outputs with a knowledge base or set of trusted sources to ensure accuracy. Businesses can also use machine-translation metrics such as BLEU (bilingual evaluation understudy) and METEOR (metric for evaluation of translation with explicit ordering) to evaluate the accuracy of AI-generated text.



The background features a vibrant, multi-colored bokeh effect in shades of blue, green, and purple. Overlaid on this is a 3D grid of vertical bars that vary in height, creating a sense of depth and digital structure. The text is positioned on the left side of the image, overlaid on the bokeh and grid.

**The buzz is out
there—and so is
the potential.**

Gen AI: beyond the buzz

Before your organization can capitalize on generative AI's unique potential for business impact and disruption, there needs to be a clear understanding of its complexities.

That means taking a measured approach that sorts out the misconceptions and myths and evaluates the application for your organization's goals.

The buzz is out there—and so is the potential. By fully understanding the problem solving, engineering and science behind generative AI, you'll be better prepared to make informed, grounded decisions and maximize the ROI of this powerful technology.



About the author



Muthumari S

Global Head of Data & AI Studio, Brillio

Muthu is responsible for strategizing and executing data and AI solutions for Brillio's clients. Using her unique blend of technical expertise, business acumen and creativity, she solves business problems at scale. She also runs Brillio's AI Labs and Research Council, which incubates new technologies such as generative AI and AI TRiSM, as well as the use of AI in drug discovery.

Muthu was named to *CDO Magazine's* list of **Global Power Data Women 2022** and its 40 Under 40 Data Leaders in **2022**. Last year, she was one of **five global finalists** for the 2022 VentureBeat "Women in AI – Responsibility and Ethics of AI" Award. *Analytics India Magazine* named her to its list of **Top 10 Data Scientists 2019** and 40 under 40 Data Scientists 2021.

Muthu has delivered 20+ thought leadership sessions across the globe and is also, an active speaker in WiDS (Women in Data Science).

In 2016, she was a key contributor to the patent on video analytics, which targeted making sense of the high volumes of unstructured data created by different forms of media devices and digital content.

Muthu can be reached at muthumari.s@brillio.com or [linkedin.com/in/muthumari-s/](https://www.linkedin.com/in/muthumari-s/).



LEARN MORE

Stay tuned for the next installment in our *Demystifying generative AI* series, where we offer an insider's view on what businesses need to know about AI governance.

For more information visit the [Data & AI](#) section of our website or contact us at info@brillio.com or visit www.brillio.com.

ABOUT BRILLIO

Brillio is the partner of choice for many Fortune 1000 companies seeking to turn disruption into a competitive advantage through innovative digital adoption. Backed by Bain Capital, Brillio is one of the fastest growing digital technology service providers. We help clients harness the transformative potential of the four superpowers of technology: cloud computing, Internet of Things (IoT), artificial intelligence (AI) and mobility. Born digital in 2014, we apply our expertise in customer experience solutions, data analytics and AI, digital infrastructure and security, and platform and product engineering to help clients quickly innovate for growth, create digital products, build service platforms and drive smarter, data-driven performance. With 17 locations across the U.S., UK, Romania, Canada, Mexico and India, our growing global workforce of over 6,000 Brillians blends the latest technology and design thinking with digital fluency to solve complex business problems and drive competitive differentiation for our clients. Brillio was certified by Great Place to Work in 2021 and 2022. Learn more at www.Brillio.com.

ABOUT BRILLIO RESEARCH

Brillio Research develops original thought leadership that addresses emerging business-technology trends, delivering guidance and insights for getting ahead in today's vital digital economy. We actively collaborate with industry, technical and academic experts to drive critical insights that help your business avoid hidden obstacles and achieve a competitive edge. Brillio's thought leadership provides practical ways to manage your immediate business goals and pursue longer term digital strategies that are essential for growth. To learn more, visit Brillio.com/insights, or contact April Vadnais, Head of Global Thought Leadership Programs, at April.Vadnais@brillio.com.