LEK™

A Primer on Generative AI

Contents

About L.E.K. Consulting
We're L.E.K. Consulting, a global strategy consultancy working with business leaders to seize competitive advantage and amplify growth. Our insights are catalysts that reshape the trajectory of our clients' businesses, uncovering opportunities and empowering them to master their moments of truth. Since 1983, our worldwide practice — spanning the Americas, Asia-Pacific and Europe — has guided leaders across all industries from global corporations to emerging entrepreneurial businesses and private equity investors. Looking for more? Visit **www.lek.com**.

A new wave in artificial intelligence

Artificial intelligence — or AI — is technology that imitates human decision-making and problem-solving capabilities. It has touched almost all of us already, although we may not always realize it. AI increasingly interprets our online search requests, steering us toward specific products and services and pricing those offerings to the maximum we're personally willing to pay. Many businesses are using AI to detect fraud, manage inventory, prevent outages and more.
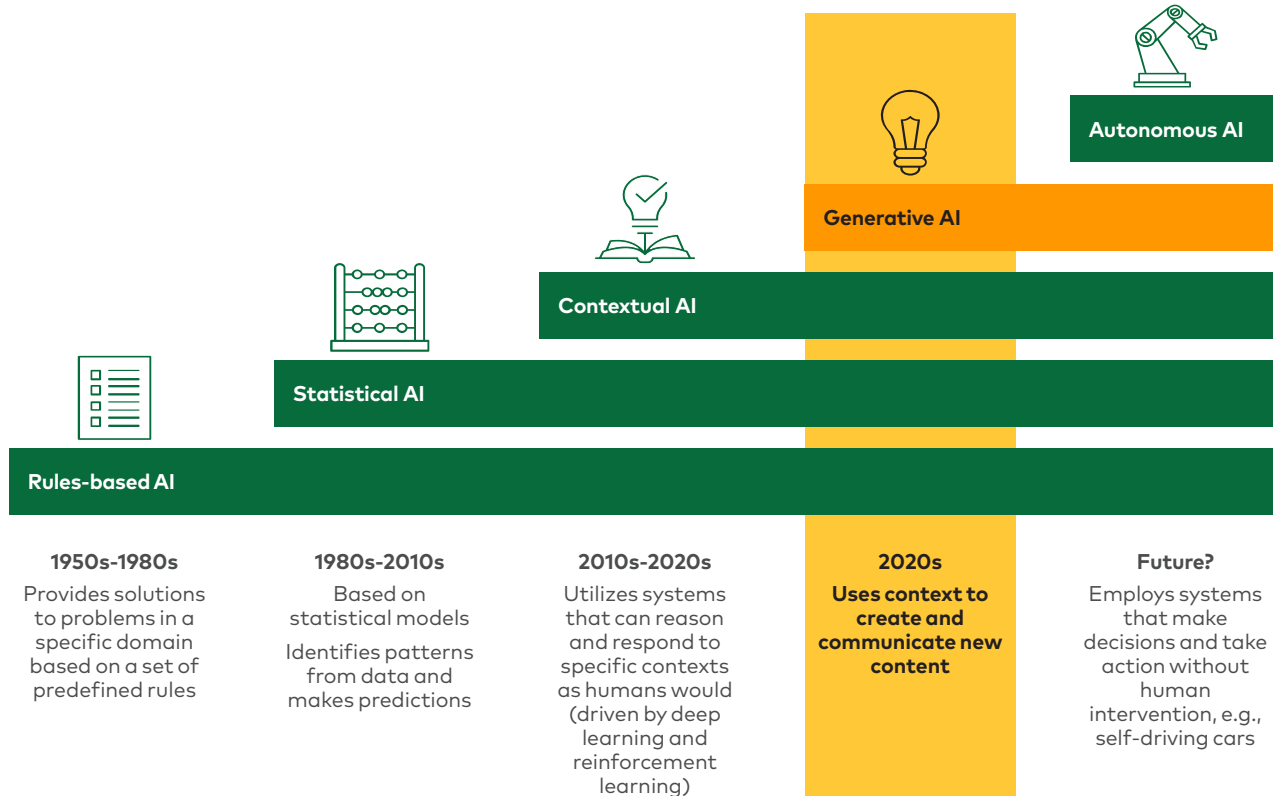
Today's AI applications are highly specific. They can be taught to reason and respond to specific contexts via deep learning algorithms operating on large data volumes, enabling decision-making at unprecedented scale and speed. This helps people do their jobs better and gives them more capacity to apply their expertise, experience and judgment. However, AI applications can't yet be employed autonomously to address a wide range of tasks as a human being can. AI designed to detect fraud can't also drive a car or write a song. And this journey — to so-called artificial general intelligence — is ongoing.

Generative AI is the latest significant step forward along this path. Generative AI can create content like a person does — whether visual, written or audio — and enable much more naturalistic human-to-computer interactions. Unlike previous AI technologies, which have typically been highly application specific, the models underpinning generative AI have broad applicability across a range of activities.

One way to think about how generative AI differs from older AI models is by comparing an online English-Chinese dictionary to a fluent speaker of both English and Chinese. Older AI models are more like the dictionary. They use statistical, rule-based machine translation techniques to translate phrases or words based on the learned patterns and relationships between words and phrases from bilingual collections of text and human-curated translations. The result is comparable to someone who can quickly make a calculated guess, but who may struggle with complex or context-dependent material.

Generative AI, on the other hand, is more like a fluent speaker. It can match the entire corpus of works in English to the entire corpus of words in other languages and infer the best way to translate a given piece. That brings us closer to a future where systems can make decisions without human intervention (see Figure 1).

Figure 1
Long-run development of AI

| | | | 1950s-1980s | 1980s-2010s | 2010s-2020s | 2020s | Future? |

1950s-1980s
Provides solutions to problems in a specific domain based on a set of predefined rules

1980s-2010s
Based on statistical models
Identifies patterns from data and makes predictions

2010s-2020s
Utilizes systems that can reason and respond to specific contexts as humans would (driven by deep learning and reinforcement learning)

2020s
Uses context to create and communicate new content

Future?
Employs systems that make decisions and take action without human intervention, e.g., self-driving cars

Source: L.E.K. research and analysis

For businesses aiming to stay competitive in an evolving technological space, it's critical to understand the best way to use generative AI. All businesses will find opportunities with generative AI, and many face strategic questions of profound importance. In this special report, L.E.K. Consulting outlines what businesses need to know to navigate the challenges and opportunities ahead.

Moving beyond the hype

Generative AI burst into public view in late 2022 when Microsoft-backed startup OpenAI introduced ChatGPT, a tool that generated text and programming code in response to user instructions. Curious users flocked to the platform and discovered its incredible (and addictive) flexibility. Some developers tried to create novel solutions for unfamiliar problems. As word of ChatGPT's performance spread, interest in AI spiked (see Figure 2), triggering a broader reappraisal of AI among many businesses.

Figure 2
Growth in interest in AI

Google searches, indexed to 2013 levels
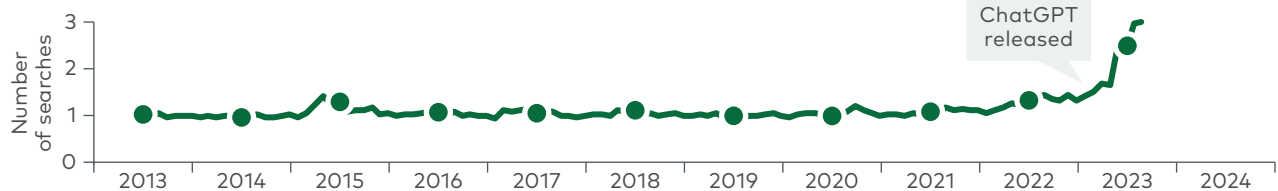
Source: Google Trends

Today, generative AI is nearing the top of the hype cycle, with pioneers continuing to explore the technology's possibilities. It will likely settle into a steady level of productivity two or three years from now, joining a raft of technologies (like deep learning and big data) from prior waves; still, despite all these high expectations, AI technology promises hugely significant benefits.

Skeptics may be forgiven for wondering whether generative AI isn't simply another in a series of technological false starts. From blockchain to Web 3.0 and autonomous vehicles, the road to a better future is littered with developments that failed to live up to their initial hype.

Although a dose of realism may be called for, it would be a mistake to dismiss generative AI as just another technological fad. The technology is far more mature than, say, autonomous vehicles. And the use cases for generative AI are far more tangible than those for blockchain (to name one example). Indeed, the most compelling comparisons to generative AI are breakthroughs in technology infrastructure — think smartphones, cloud computing, the internet or the graphical

L.E.K. Consulting

user interface — that are hard to ignore. It seems likely that generative AI will affect industries in ways similar to new technologies in the past. This means that, among other things, generative AI is poised to:
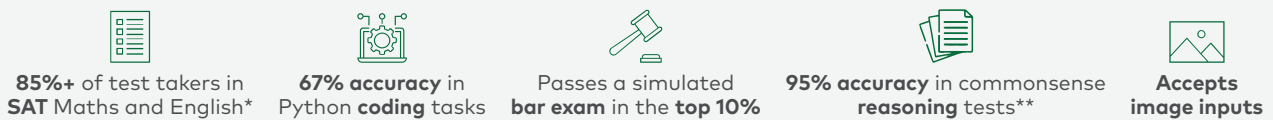
- Dramatically change and enhance how users interact with businesses, as search engines did
- Prompt a wave of development activity at the application layer (and new ecosystems) built on top of foundational technology, as did Apple's App Store and the dot-com boom that ended in the early 2000s
- Completely change market landscapes such that those directly exposed will need to rapidly pivot or face extinction, as we saw with Blockbuster's bankruptcy and Netflix's shift from DVDs to streaming
- Face the same legislative scrutiny as social media (which is already starting on both sides of the Atlantic), particularly in areas of ethics, privacy and data regulation

Getting better, fast

March 2023 saw the release of GPT-4, a more powerful version of ChatGPT. Users of GPT-4 have reported an impressive level of performance across a range of tasks (see Figure 3).

Figure 3
ChatGPT-4 performance outcomes

| **85%+** of test takers in **SAT** Maths and English* | **67% accuracy** in Python **coding** tasks | Passes a simulated **bar exam** in the **top 10%** | **95% accuracy** in commonsense **reasoning** tests** | **Accepts image inputs** |

*SAT Evidence-Based Reading and Writing
**Commonsense reasoning around everyday events
Source: Forbes; The Verge; OpenAI; Geekwire; World Economic Forum; The New York Times

That same month, Midjourney released version 5 of its image generation platform. In just 15 months, Midjourney had gone from producing images that looked like an expression of modernist art to ones that were virtually indistinguishable from a photograph (see Figure 4).

Figure 4
Midjourney image evolution

V1   V5
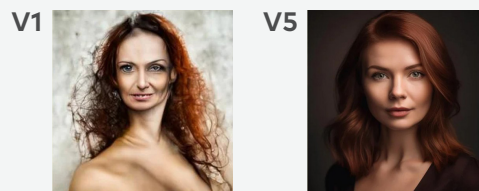
Source: Alex Valaitis, "Same prompt, but 1 year apart in @midjourney," Twitter, March 29, 2023. https://twitter.com/alex_valaitis/status/1641108801016197122

Of all these technologies, cloud computing may be the closest analog. The infrastructure-as-a-service platforms we now recognize as cloud computing first hit the market in 2006. Now they store 60% of the world's corporate data. Generative AI is expected to mature in a way similar to cloud computing with:

- Seemingly infinite capacity that users can tap at a moment's notice
- Pay-as-you-go subscriptions that let organizations of any size use the services without having to invest in equipment
- Low and declining prices, thanks to increases in scale
- An end state wherein most users won't necessarily recognize where or how the underlying AI is empowering the tools and solutions they're interacting with

It's also likely that a limited number of platforms will be clear leaders in the generative AI foundation model market, just as three providers (Amazon Web Services, Microsoft Azure and Google) serve 65% of the cloud computing market today.
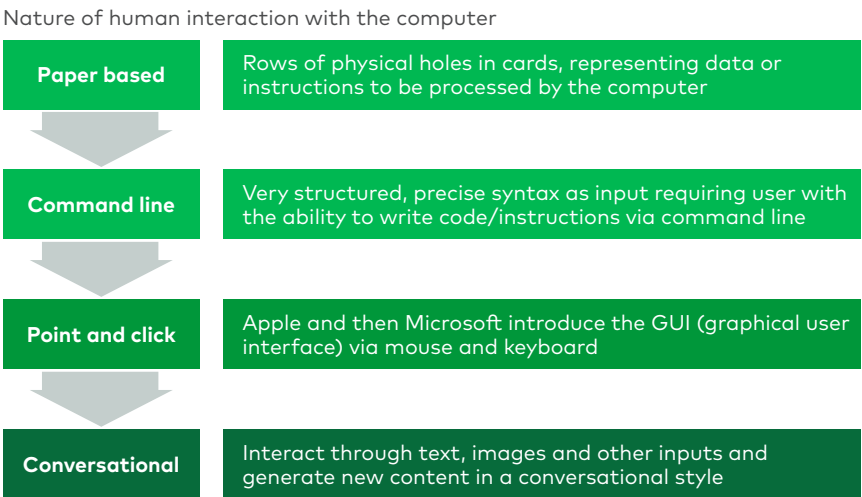
A look under the hood

Generative AI greatly simplifies the way computers interact with people — they now speak our language. Foundational models can absorb information (from people or documents) and create realistic, original outputs in response.

If the foundational model is a large language model, the output is text. Examples include ChatGPT-4 or Google Bard, which are trained on an extensive text corpus and can understand context, generating relevant responses and simulating conversation. If the foundational model is a diffusion model, the response is an image. Midjourney, Stable Diffusion and OpenAI's DALL-E are all examples of generative AI tools that produce digital images. Each is trained on hundreds of millions of images with text captions scraped from the internet.

By the same token, generative AI doesn't need you to use prescribed menus, toolbars or buttons to communicate with it. Instead, you can just tell it what you want to do. This is a transformative change for computer interfaces and the next major step in making technology more accessible (see Figure 5).

Figure 5
The evolution of computer interfaces

Nature of human interaction with the computer

|	
Paper based	Rows of physical holes in cards, representing data or instructions to be processed by the computer
Command line	Very structured, precise syntax as input requiring user with the ability to write code/instructions via command line
Point and click	Apple and then Microsoft introduce the GUI (graphical user interface) via mouse and keyboard
Conversational	Interact through text, images and other inputs and generate new content in a conversational style

Source: L.E.K. research and analysis

Companies are already building generative AI into their technology stack. Microsoft added ChatGPT to its Bing search engine, and an AI assistant feature called Copilot is now available to Microsoft 365 users. Google's Bard chatbot, released in March 2023, can converse in 43 languages.[1] Zoom rolled out several AI-powered features for its namesake videoconferencing platform. Salesforce integrated Einstein, a set of AI technologies, with its customer relationship management platform. This combination of generative AI with existing technologies is where the most impactful advancements are likely to be made, exposing billions of users to the power of generative AI.

But it won't be just big tech integrating generative AI applications. The barriers to building new applications on top of existing foundation models are low, and developer tools — think Singapore-based LangChain — are continually improving. As a result, the use of generative AI within applications will proliferate, whether in ecommerce like Shopify or social media platforms like Quora. Even enterprise-specific solutions are becoming available, with businesses like Santa Clara-based startup Vectara helping corporations implement their own generative AI-enabled internal search (see Figure 6).

Figure 6
Generative AI technology stack
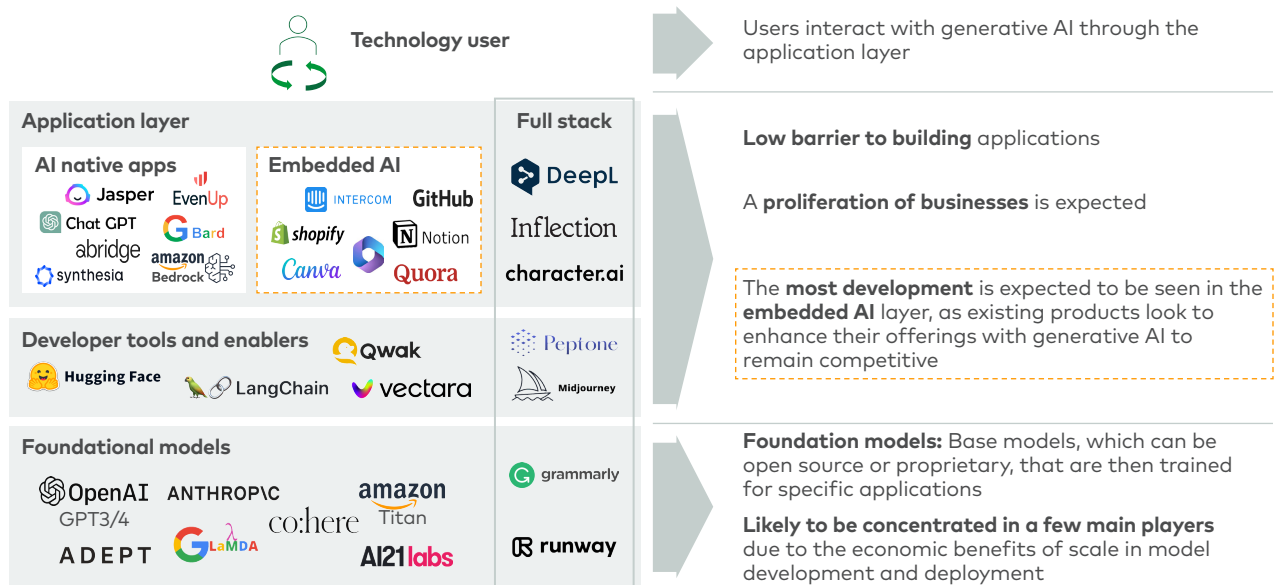
Source: Bessemer Venture Partners; Sequoia

In contrast to applications, however, the market for foundation models will likely concentrate on a few main players due to the economic benefits of scale in model development and deployment. It's too early to say which models might win. Current ones have different strengths and weaknesses. Meanwhile, proprietary models like Google's Pathways Language Model (PaLM) and OpenAI's GPT-4 — as well as open source models like Meta's Llama 2 — are competing for adoption.

Although text and coding applications built on large language models were first to achieve broad recognition, Midjourney and Stable Diffusion launched their platforms (and attracted attention from mainstream media) before ChatGPT hit the market. Now these and other fast-developing image generators are edging into the spotlight. Over time, generative AI capabilities will appear for video, gaming and 3D as well, although these may not be ready for prime time until 2030 or so (see Figure 7).

Figure 7
Illustrative view of generative AI capability by domain

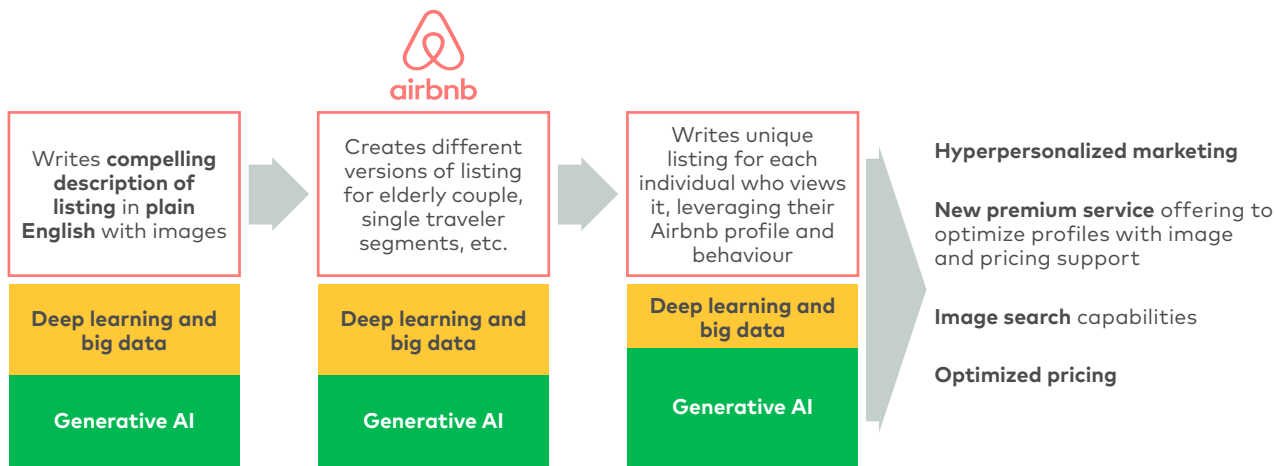|	Pre-2020	2020	2022	2023?	2025?	2030?
Text	Spam detection Translation Basic Q&A	Basic copywriting First drafts	Longer form Second drafts	Vertical fine-tuning improves (scientific papers, etc.); models' context improves	Final drafts better than human average Models are trainable	Final drafts better than professional writers
Code	One-line auto-complete	Multiline generation	Longer form Better accuracy	More languages More verticals	Text to product (draft)	Text to product (final) better than full-time developers
Images			Art Logos Photography	Mock-ups (product design, architecture, etc.)	Final drafts (product design, architecture, etc.)	Final drafts better than professional artists, designers, photographers
Video/3D/ gaming			First attempts at 3D/video models	Basic/first draft videos and 3D files	Second drafts	AI Roblox Video games and movies are personalized dreams

Large model availability:
● First attempts ● Almost there ● Ready for prime time

Source: Sequoia Capital; L.E.K. research and analysis

Generative AI's input and output domains will probably converge as well. That will create major business opportunities, including real-time, hyperpersonalization of content for customers (see Figure 8).

Figure 8
Example of a generative feedback loop with large language modeling

Writes **compelling description of listing** in **plain English** with images	Creates different versions of listing for elderly couple, single traveler segments, etc.	Writes unique listing for each individual who views it, leveraging their Airbnb profile and behaviour
Deep learning and big data	**Deep learning and big data**	**Deep learning and big data**
Generative AI	**Generative AI**	**Generative AI**

Hyperpersonalized marketing

New premium service offering to optimize profiles with image and pricing support

Image search capabilities

Optimized pricing

Source: L.E.K. research and analysis

Technical and strategic growing pains

Generative AI still has a number of growing pains to get through. Here are some common ones, along with the approaches providers are taking to deal with them.

Fidelity and realism. A model may struggle to produce images or videos that are consistently realistic or of high fidelity. More training and feedback filters are being deployed to help generative AI models improve the visual quality and realism of generated media.

Bias. Existing models learn from existing datasets, which can often reflect biases present in the data and perpetuate stereotypes or discrimination. Model developers are working to mitigate these issues with more representative training data and techniques.

Control and manipulation. Fine-grained control, such as specifying certain attributes or styles within a picture or video, can be challenging today. Look for models that allow for user-specified conditions, attributes and interactive feedback, including style transfer from reference images.

Copyright infringement. Models can inadvertently generate content that infringes on copyrighted material, particularly if they're trained on licensed datasets. Better data selection with content filtering and moderation systems are helping to address this issue.

Integration. Most models are specific to one domain. To address this, developers are working on large multimodal models that work across multiple domains, like text and images.

These technical and legal challenges will continue to evolve, and new ones may come up. Even so, businesses face important strategic decisions today. Chief Marketing Technologist touches on some of these in its article "Exploring the 2nd order effects of generative AI in marketing and martech." They include the growing need for AI governance, the disposability of easy-to-create software and the risk of buyers tuning

out even the most personalized communications amid an avalanche of new content.[2] As generative AI develops, the second-order impacts and combinations with other technologies are likely where some of the most important strategic questions will arise for businesses (see Figure 9).

Figure 9
AI's zero-, first- and second-order effects

	Time		

AI trigger	Zero-order effects	First-order effects	Second-order effects
AI generates content	**Content** is **easier, faster and cheaper to create** and available to more people	The **quantity of content** in the world grows exponentially Personalized content is fully generated Spam grows exponentially	Buyers further shut out, "pushed" marketing and sales content **Trusted sources** are even **more valuable**
AI generates code	**Software** is **easier, faster and cheaper to build** and available to more people	The **quantity of software** in the world grows exponentially Many **software** programs are built "**on demand**"	More **software becomes disposable** — easier to re-create from scratch than maintain More **software becomes ambient** — we think less about explicit "apps"
AI absorbs data and content to answer questions	Asking an **AI agent** is the easiest, fastest way to **answer questions** from large amounts of content and data	Traditional **search engines are displaced** The long tail of content and data is now widely accessible	**Proprietary** data becomes **best AI moat** **Analytics** is massively **democratized**
AI executes digital tasks autonomously	Digital "busy work" can be delegated to AI agents	The **quantity of automation** in the world grows **exponentially** Massive **boost to productivity**	API services become a first-class marketing channel **Bot commerce** grows **exponentially** Challenges of "big ops" — orchestration and governance

Note: API=application programming interface
Source: ChiefMartec
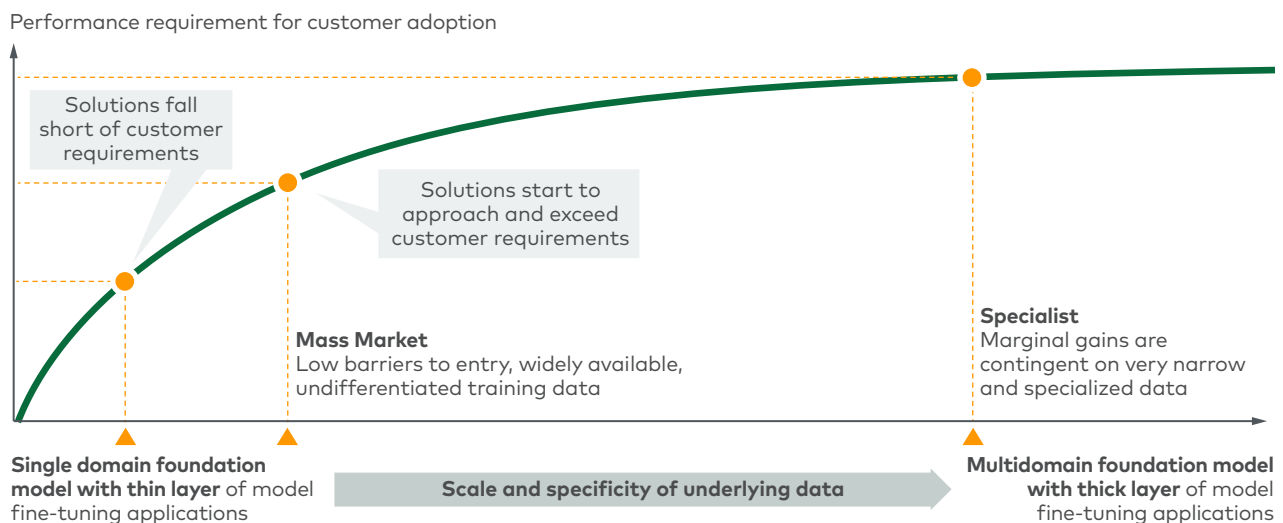
The impact on businesses

Generative AI foundation models are likely to become important platforms for consumers and businesses. We anticipate four broad routes to generative AI adoption:

- Widely available public services like ChatGPT or Google Bard
- Integrated office productivity tools like Salesforce (Einstein) or Microsoft Office 365 (Copilot)
- Purpose-built, industry-specific offerings like Google Med-PaLM 2 for medicine or BloombergGPT for finance
- Proprietary solutions like Cowbell Cyber's MooGPT for underwriting or Allen & Overy's Harvey for legal work

The last two routes will require greater accuracy than the mass-market uses we've seen so far. It's one thing for generative AI to produce an email or product pitch that's off the mark. It's quite another if it misinterprets a medical image or miscalculates the performance of a composite material. Specialist capabilities like these will rely on marginal gains achieved by domain convergence, fine-tuning layered applications and training foundation models with highly specific data (see Figure 10).

Figure 10
Generative AI readiness for commercial impact

Performance requirement for customer adoption

Solutions fall short of customer requirements

Solutions start to approach and exceed customer requirements

Mass Market
Low barriers to entry, widely available, undifferentiated training data

Specialist
Marginal gains are contingent on very narrow and specialized data

Single domain foundation model with thin layer of model fine-tuning applications

Scale and specificity of underlying data

Multidomain foundation model with thick layer of model fine-tuning applications

Source: L.E.K. research and analysis

Software development offers us a glimpse of generative AI in action. Developers are using the technology to complete unfinished code, test completed code, detect bugs and suggest new code in response to natural language prompts. Developers can also do code synthesis and translation using generative AI tools.

For example, Accenture developers found that CodeWhisperer, Amazon's generative AI coding platform, reduced their efforts by 30%. Goldman Sachs announced that AI was at least as good as humans are at classifying documents, and it produced usable code up to 40% of the time.

In the future, generative AI will likely be able to write, execute and iterate most code to the same standard as humans. This will improve agility and cost-effectiveness for many businesses while freeing developers and data scientists for more complex, higher-value tasks.

Given its broad applicability, generative AI is a call to action for businesses to explore and understand the many distinct ways they can use the technology for opportunity and competitive advantage.

Generative AI opportunities for businesses

Potential opportunity	Example
New revenue streams	Unlock new products or services, including access to AI systems and new, unique content
Customer acquisition and retention improvement	Enhance communications with potential and existing customers through greater targeting, personalization and response time
Product and service enhancement	Enhance the core business product or service through design generation, allowing greater customization, personalization or improvements in quality control
Cost reduction	Reduce cost through increased productivity and/or automation within the business or within the supply chain
Process transformation and reorganizations	Optimize business systems and processes to integrate seamlessly with generative AI solutions
Process automation	Gain cost efficiencies through automation of routine, repetitive back-office tasks
Competitive advantage building	Gain a sustainable competitive advantage through the ownership of proprietary algorithms, models or data

The impact on jobs

Indications are that generative AI is already having an impact on skilled, knowledge-based work. Just how big that impact will be — and how quickly it will happen — is up for debate. But businesses will need to think about how certain roles — including those in office, administrative, legal, architecture and engineering fields — will meet their objectives in the age of generative AI (see Figure 11).

Figure 11
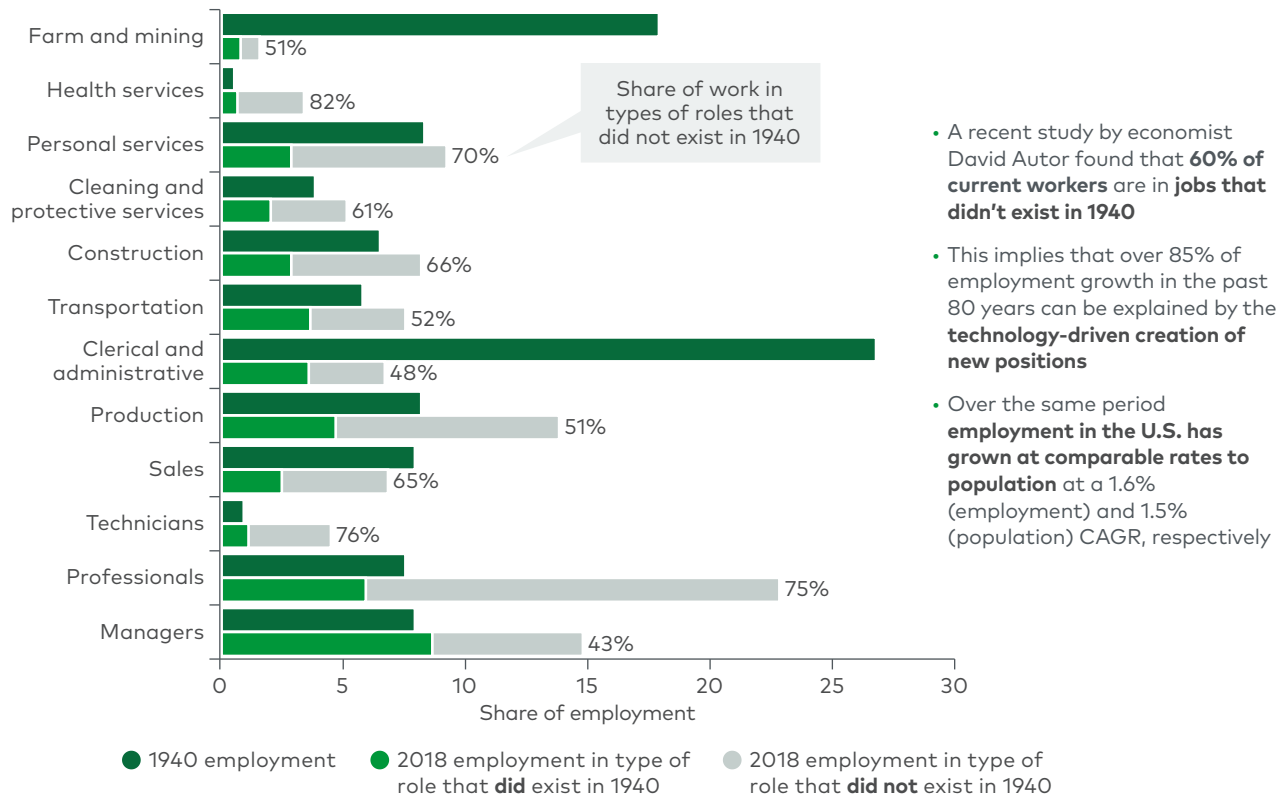Share of industry employment exposed* to automation by generative AI**
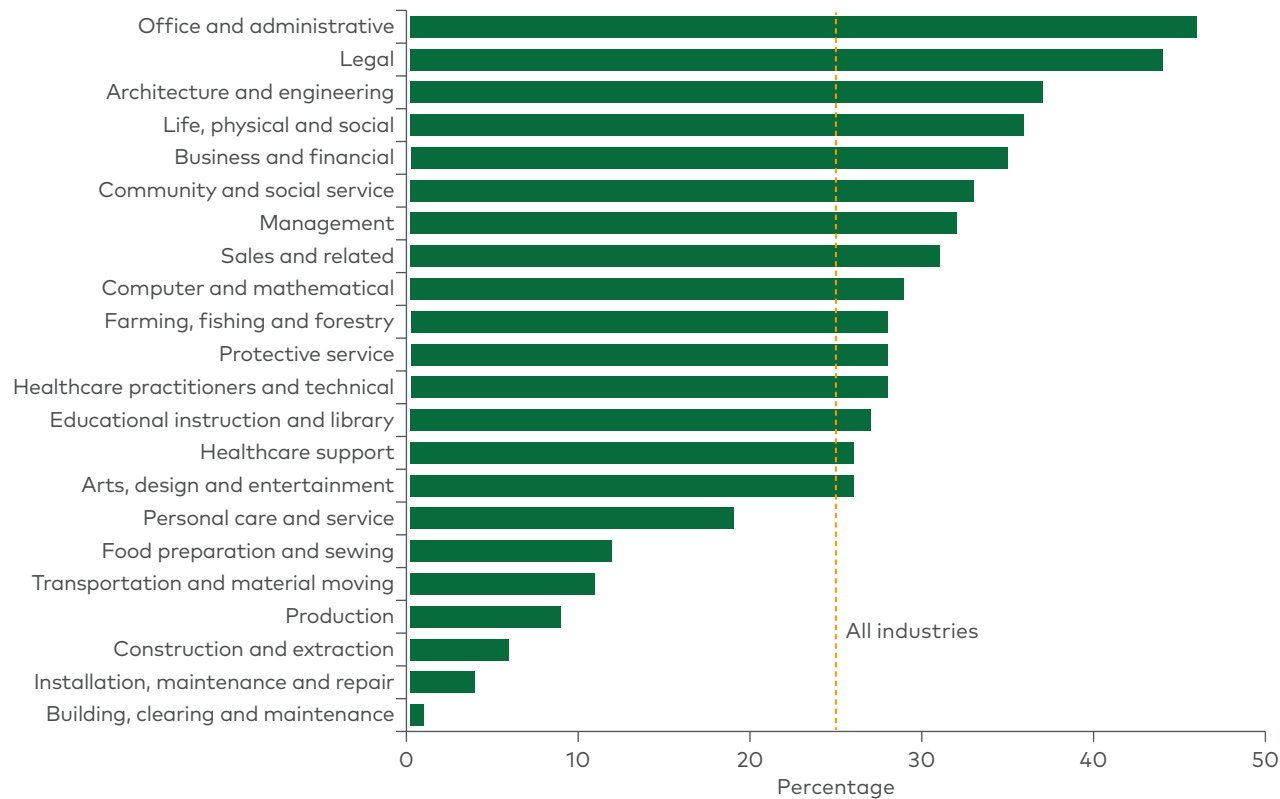
*Exposure is defined as the percentage of current work tasks that generative AI could automate
**Research by Goldmans Sachs: The Potentially Large Effects of Artificial Intelligence on Economic Growth
Source: Bloomberg; Reuters; Goldmans Sachs

That's not to say the overall number of jobs will decline. According to a study from the Massachusetts Institute of Technology, most of today's U.S. workers are in jobs

that didn't exist in 1940.[3] Put another way, history tells us that technological change creates new jobs requiring new skills (see Figure 12).

Figure 12
US employment across major occupations in 2018 and whether that type of role existed in 1940

Share of work in types of roles that did not exist in 1940

- A recent study by economist David Autor found that **60% of current workers** are in **jobs that didn't exist in 1940**

- This implies that over 85% of employment growth in the past 80 years can be explained by the **technology-driven creation of new positions**

- Over the same period **employment in the U.S. has grown at comparable rates to population** at a 1.6% (employment) and 1.5% (population) CAGR, respectively

● 1940 employment   ● 2018 employment in type of role that **did** exist in 1940   ● 2018 employment in type of role that **did not** exist in 1940

Note: CAGR=compound annual growth rate
Source: MIT; U.S. Census Bureau; Bureau of Labor Statistics

The implication is that more than 85% of employment growth in the past 80 years can be explained by the technology-driven creation of new positions. Over the same period in the U.S., the compound annual growth rate of employment has kept pace with population growth (1.6% versus 1.5%, respectively).

Where to go from here

Innovation begets innovation. Businesses are already boosting the performance and functionality of new and existing products with generative AI capabilities. They're also using generative AI to accelerate their development cycles.

This introduces the greater challenge for executives: anticipating the strategic impact on their businesses and industries. It could ultimately make a product's or service's value proposition obsolete. Companies could fail to move quickly enough, assuming the hype will pass. They may focus on the wrong strategies or hire the wrong kind of AI talent. Competitors may gain advantage by using the technology in a better way.

There's no one-size-fits-all approach to tackling these opportunities and challenges. But executives can begin to size up the strategic impact of generative AI by answering the following questions about their organization:

- How significant is content development to our products or services?
- What level of automation exists in the business today?
- How standardized or customized are our products, services, customer interactions and communications?
- How important are software development cycles to our competitiveness?
- What level of data sophistication informs decisions in our company?

Executives must also then assess how quickly or extensively these impacts will occur. Key considerations include:

- How regulated is our industry?
- How risk averse is our industry?
- How precise do our decisions need to be?
- Is the content that we produce largely text-based or is it image-based?
- How much of the information used in our industry is readily available rather than privileged (e.g., confidential or proprietary)?

Generative AI is a step change technology on many fronts. It's revolutionizing the creation of most forms of content, advancing the scale and speed of decision-making, and greatly simplifying interactions between humans and other technologies. Although it's still the early days for this latest iteration of artificial intelligence, it's

not too soon for businesses to think about how they can use generative AI to boost productivity and create new competitive advantages in a fast-changing marketplace.

For more information, please contact **technology@lek.com.**

A glossary of AI-related terms

Term	Definition
Application layer	An interface for users to interact with the underlying model
Application programming interface (API)	A protocol for programs/applications to interact
Artificial general intelligence (AGI)	The point at which machines can comprehend, learn and perform intellectual tasks at least as well as humans can
Artificial intelligence (AI)	The simulation of human intelligence processes by machines
Big data	Data that's too voluminous or messy for traditional data-processing applications to readily handle
Deep learning	A subset of machine learning based on multiple layers of neural networks
Diffusion model	A type of algorithm used to create data that closely resembles the data the algorithm is trained on; these models learn by adding noise to images and removing it, thus generating new and diverse high-resolution images that are reminiscent of the original data
Domain	A type of input a model is specialized for (e.g., text, image, video)
Explainability	The concept that a machine learning model and its output can be explained in a way that humans understand and trust
Foundation model	A model, trained on a large amount of unlabeled data, that can be adapted to many applications
Full stack	The entire range of technologies that go into building a software application
Generative AI	A branch of artificial intelligence that creates new output across multiple domains (e.g., text, images, code) by learning patterns from vast amounts of data and mimicking the underlying structure
Generative pre-trained transformer (GPT)	A family of LLMs and a prominent framework for generative AI (the most notable GPT foundation models have been from OpenAI)
Graphical user interface (GUI)	A form of user interface that allows users to interact with electronic devices through visual components
Large language model (LLM)	A type of AI algorithm that uses deep learning techniques and extremely large datasets to understand and generate new text-based content
Large multimodal model (LMM)	An AI system capable of understanding and generating information from multiple data domains (e.g., text and image)
Machine learning	A subset of AI broadly defined as the process of using mathematical models to help a computer learn without direct instruction
Natural language processing (NLP)	Machine learning technology that gives computers the ability to interpret, manipulate and comprehend human language
Neural networks	An AI method that teaches computers to process data in a way that is inspired by the human brain
Statistical AI	An AI method that uses principles to derive useful information from data
Structured data	Data that has a standard format for efficient access by software and humans (typically tabular with rows and columns that clearly define attributes)

Endnotes

[1]Sam Schechner, "Google's Bard AI Chatbot Adds More Languages to Take On ChatGPT," Wall Street Journal, July 13, 2023. https://www.wsj.com/articles/googles-bard-ai-chatbot-adds-more-languages-to-take-on-chatgpt-a2acfc5b

[2]Scott Brinker, "Exploring the 2nd order effects of generative AI in marketing and martech," Chief Marketing Technologist. https://chiefmartec.com/2023/04/exploring-the-2nd-order-effects-of-generative-ai-in-marketing-and-martech/

[3]David Autor, David Mindell and Elisabeth Reynolds, "The Work of the Future: Building Better Jobs in an Age of Intelligent Machines," Massachusetts Institute of Technology, 2020. https://workofthefuture.mit.edu/wp-content/uploads/2021/01/2020-Final-Report4.pdf

About the Authors

Philip Meier
PARTNER, LONDON
Philip Meier is a Partner based in L.E.K. Consulting's London office. Philip leads the Analytics practice in Europe and works across multiple sectors. He has experience helping clients assess and understand changes in software, data and analytics and deploy analytics to enhance propositions and business performance. He has deep expertise in technology development and commercialisation, with a particular focus on technologies that enable the energy transition.

Alan Lewis
MANAGING DIRECTOR, BOSTON
Alan Lewis is a Managing Director based in L.E.K. Consulting's Boston office and a member of the firm's consumer sector team. Alan is also a leader on the global Advanced Analytics team. He is an expert in growth strategy and has helped many companies within the travel sector and beyond to define and execute on ways to grow and create value.

Stuart Robertson
PARTNER, LONDON
Stuart Robertson is a Partner based in L.E.K. Consulting's London office and the firm's Disruptive Analytics Lead. Stuart focuses primarily on the transport industry, covering operators and equipment and infrastructure suppliers. With particular expertise in the interaction between the public and private sectors, he has deep experience in transport infrastructure, pricing and revenue management, transport technologies, consumer engagement, and litigation and dispute resolution.