

AI-Fueled Productivity: Generative AI Opens New Era of Efficiency Across Industries

Author: Cliff Edwards

A watershed moment on Nov. 22, 2022, was mostly virtual, yet it shook the foundations of nearly every industry on the planet.

On that day, OpenAI released ChatGPT, the most advanced artificial intelligence chatbot ever developed. This set off demand for generative AI applications that help businesses become more efficient, from providing consumers with answers to their questions to accelerating the work of researchers as they seek scientific breakthroughs, and much, much more.

Businesses that previously dabbled in AI are now rushing to adopt and deploy the latest applications. Generative AI — the ability of algorithms to create new text, images, sounds, animations, 3D models and even computer code — is moving at warp speed, transforming the way people work and play.

By employing large language models (LLMs) to handle queries, the technology can dramatically reduce the time people devote to manual tasks like searching for and compiling information.

The stakes are high. AI could contribute more than \$15 trillion to the global economy by 2030, according to PwC . And the impact of AI adoption could be greater than the inventions of the internet, mobile broadband and the smartphone — combined.

The engine driving generative AI is accelerated computing . It uses GPUs, DPUs and networking along with CPUs to accelerate applications across science, analytics, engineering, as well as consumer and enterprise use cases.

Early adopters across industries — from drug discovery , financial services , retail and telecommunications to energy , higher education and the public sector — are combining accelerated computing with generative AI to transform business operations, service offerings and productivity.

Today, radiologists use AI to detect abnormalities in medical images, doctors use it to scan electronic health records to uncover patient insights, and researchers use it to accelerate the discovery of novel drugs.

Traditional drug discovery is a resource-intensive process that can require the synthesis of over 5,000 chemical compounds and yields an average success rate of just 10%. And it takes more than a decade for most new drug candidates to reach the market.

Researchers are now using generative AI models to read a protein's amino acid sequence and accurately predict the structure of target proteins in seconds, rather than weeks or months.

Using NVIDIA BioNeMo models, Amgen, a global leader in biotechnology , has slashed the time it takes to customize models for molecule screening and optimization from three months to just a few weeks. This type of trainable foundation model enables scientists to create variants for research into specific diseases, allowing them to develop target treatments for rare conditions.

Whether predicting protein structures or securely training algorithms on large real-world and synthetic datasets, generative AI and accelerated computing are opening new areas of research that can help mitigate the spread of disease, enable personalized medical treatments and boost patient survival rates.

According to a recent NVIDIA survey , the top AI use cases in the financial services industry are customer services and deep analytics, where natural language processing and LLMs are used to better

respond to customer inquiries and uncover investment insights. Another common application is in recommender systems that power personalized banking experiences, marketing optimization and investment guidance.

Advanced AI applications have the potential to help the industry better prevent fraud and transform every aspect of banking, from portfolio planning and risk management to compliance and automation.

Eighty percent of business-relevant information is in an unstructured format — primarily text — which makes it a prime candidate for generative AI. Bloomberg News produces 5,000 stories a day related to the financial and investment community. These stories represent a vast trove of unstructured market data that can be used to make timely investment decisions.

NVIDIA, Deutsche Bank, Bloomberg and others are creating LLMs trained on domain-specific and proprietary data to power finance applications.

Financial Transformers, or “FinFormers,” can learn context and understand the meaning of unstructured financial data. They can power Q&A; chatbots, summarize and translate financial texts, provide early warning signs of counterparty risk, quickly retrieve data and identify data-quality issues.

These generative AI tools rely on frameworks that can integrate proprietary data into model training and fine-tuning, integrate data curation to prevent bias and use guardrails to keep conversations finance-specific.

Expect fintech startups and large international banks to expand their use of LLMs and generative AI to develop sophisticated virtual assistants to serve internal and external stakeholders, create hyper-personalized customer content, automate document summarization to reduce manual work, and analyze terabytes of public and private data to generate investment insights.

With 60% of all shopping journeys starting online and consumers more connected and knowledgeable than ever, AI has become a vital tool to help retailers match shifting expectations and differentiate from a rising tide of competition.

Retailers are using AI to improve customer experiences, power dynamic pricing, create customer segmentation, design personalized recommendations and perform visual search.

Generative AI can support customers and employees at every step through the buyer journey.

With AI models trained on specific brand and product data, they can generate robust product descriptions that improve search engine optimization rankings and help shoppers find the exact product they’re looking for. For example, generative AI can use metatags containing product attributes to generate more comprehensive product descriptions that include various terms like “low sugar” or “gluten free.”

AI virtual assistants can check enterprise resource planning systems and generate customer service messages to inform shoppers about which items are available and when orders will ship, and even assist customers with order change requests.

Fashable, a member of NVIDIA Inception’s global network of technology startups, is using generative AI to create virtual clothing designs, eliminating the need for physical fabric during product development. With the models trained on both proprietary and market data, this reduces the environmental impact of fashion design and helps retailers design clothes according to current market trends and tastes.

Expect retailers to use AI to capture and retain customer attention, deliver superior shopping experiences, and drive revenue by matching shoppers with the right products at the right time.

In an NVIDIA survey covering the telecommunications industry, 95% of respondents reported that they were engaged with AI, while two-thirds believed that AI would be important to their company’s future success.

Whether improving customer service, streamlining network operations and design, supporting field technicians or creating new monetization opportunities, generative AI has the potential to reinvent the telecom industry.

Telcos can train diagnostic AI models with proprietary data on network equipment and services, performance, ticket issues, site surveys and more. These models can accelerate troubleshooting of technical performance issues, recommend network designs, check network configurations for compliance, predict equipment failures, and identify and respond to security threats.

Generative AI applications on handheld devices can support field technicians by scanning equipment and generating virtual tutorials to guide them through repairs. Virtual guides can then be enhanced with augmented reality, enabling technicians to analyze equipment in a 3D immersive environment or call on a remote expert for support.

New revenue opportunities will also open for telcos. With large edge infrastructure and access to vast datasets, telcos around the world are now offering generative AI as a service to enterprise and government customers.

As generative AI advances, expect telecommunications providers to use the technology to optimize network performance, improve customer support, detect security intrusions and enhance maintenance operations.

In the energy industry, AI is powering predictive maintenance and asset optimization, smart grid management, renewable energy forecasting, grid security and more.

To meet growing data needs across aging infrastructure and new government compliance regulations, energy operators are looking to generative AI.

In the U.S., electric utility companies spend billions of dollars every year to inspect, maintain and upgrade power generation and transmission infrastructure.

Until recently, using vision AI to support inspection required algorithms to be trained on thousands of manually collected and tagged photos of grid assets, with training data constantly updated for new components. Now, generative AI can do the heavy lifting.

With a small set of image training data, algorithms can generate thousands of physically accurate images to train computer vision models that help field technicians identify grid equipment corrosion, breakage, obstructions and even detect wildfires. This type of proactive maintenance enhances grid reliability and resiliency by reducing downtime, while diminishing the need to dispatch teams to the field.

Generative AI can also reduce the need for manual research and analysis. According to McKinsey, employees spend up to 1.8 hours per day searching for information — nearly 20% of the work week. To increase productivity, energy companies can train LLMs on proprietary data, including meeting notes, SAP records, emails, field best practices and public data such as standard material data sheets.

With this type of knowledge repository connected to an AI chatbot, engineers and data scientists can get instant answers to highly technical questions. For example, a maintenance engineer troubleshooting pitch control issues on a turbine's hydraulic system could ask a bot: "How should I adjust the hydraulic pressure or flow to rectify pitch control issues on a model turbine from company X?" A properly trained model would deliver specific instructions to the user, who wouldn't have to look through a bulky manual to find answers.

With AI applications for new system design, customer service and automation, expect generative AI to enhance safety and energy efficiency, as well as reduce operational expenses in the energy industry.

From intelligent tutoring systems to automated essay grading, AI has been employed in education for decades. As universities use AI to improve teacher and student experiences, they're increasingly dedicating resources to build AI-focused research initiatives.

For example, researchers at the University of Florida have access to one of the world's fastest supercomputers in academia. They've used it to develop GatorTron — a natural language processing model that enables computers to read and interpret medical language in clinical notes that are stored in electronic health records. With a model that understands medical context, AI developers can create numerous medical applications, such as speech-to-text apps that support doctors with automated medical charting.

In Europe, an industry-university collaboration involving the Technical University of Munich is demonstrating that LLMs trained on genomics data can generalize across a plethora of genomic tasks, unlike previous approaches that required specialized models. The genomics LLM is expected to help scientists understand the dynamics of how DNA is translated into RNA and proteins, unlocking new clinical applications that will benefit drug discovery and health.

To conduct this type of groundbreaking research and attract the most motivated students and qualified academic professionals, higher education institutes should consider a whole-university approach to pool budget, plan AI initiatives, and distribute AI resources and benefits across disciplines.

Today, the biggest opportunity for AI in the public sector is helping public servants to perform their jobs more efficiently and save resources.

The U.S. federal government employs over 2 million civilian employees — two-thirds of whom work in professional and administrative jobs.

These administrative roles often involve time-consuming manual tasks, including drafting, editing and summarizing documents, updating databases, recording expenditures for auditing and compliance, and responding to citizen inquiries.

To control costs and bring greater efficiency to routine job functions, government agencies can use generative AI.

Generative AI's ability to summarize documents has great potential to boost the productivity of policymakers and staffers, civil servants, procurement officers and contractors. Consider a 756-page report recently released by the National Security Commission on Artificial Intelligence. With reports and legislation often spanning hundreds of pages of dense academic or legal text, AI-powered summaries generated in seconds can quickly break down complex content into plain language, saving the human resources otherwise needed to complete the task.

AI virtual assistants and chatbots powered by LLMs can instantly deliver relevant information to people online, taking the burden off of overstretched staff who work phone banks at agencies like the Treasury Department, IRS and DMV.

With simple text inputs, AI content generation can help public servants create and distribute publications, email correspondence, reports, press releases and public service announcements.

The analytical capabilities of AI can also help process documents to speed the delivery of vital services provided by organizations like Medicare, Medicaid, Veterans Affairs, USPS and the State Department.

Generative AI could be a pivotal tool to help government bodies work within budget constraints, deliver government services more quickly and achieve positive public sentiment.

Across every field, organizations are transforming employee productivity, improving products and delivering higher-quality services with generative AI.

To put generative AI into practice, businesses need expansive amounts of data, deep AI expertise and sufficient compute power to deploy and maintain models quickly. Enterprises can fast-track adoption with the NeMo generative AI framework, part of NVIDIA AI Enterprise software, running on DGX Cloud . NVIDIA's pretrained foundation models offer a simplified approach to building and running customized generative AI solutions for unique business use cases.

Learn more about powerful generative AI tools to help your business increase productivity, automate tasks, and unlock new opportunities for employees and customers.

Original URL: <https://blogs.nvidia.com/blog/2023/07/13/generative-ai-for-industries/>