protiviti®
Global Business Consulting

Success With Generative AI Requires Balancing Risk With Reward

By Christine Livingston

Face the Future with Confidence®

Table of Contents

When ChatGPT launched in November 2022, it took just two months to garner a record 100 million users. A generative artificial intelligence (AI) platform, ChatGPT has captured broad market attention and spurred new thinking and conversations in the boardroom and C-suite.

Generative AI (GenAI), of which ChatGPT is just one of many models and applications, holds tremendous promise to transform business and operations. Leaders are eager to realise GenAI's enormous potential and are actively seeking opportunities to leverage this fascinating technology. At the same time, however, they need to identify and manage the risks associated with this technology, often by reinvigorating or establishing AI governance programs to ensure they are deploying AI responsibly and managing stakeholder expectations.

Opportunities and challenges

According to a survey conducted by the IBM Institute for Business Value, more than 60% of organisations plan to pilot or operate GenAI applications by 2024. Venture capital firms have jumped into the fray as well, investing more than $1.7 billion in GenAI solutions over the past three years. Clearly, leaders have taken notice of AI's immense market potential and are eager to capitalise on GenAI to create value and growth for their own businesses. What's more, GenAI is poised to radically change the way businesses operate. It is critical for business leaders to experiment and learn about the technology or they run the risk of their businesses being disrupted.

What technology leaders are saying

> *"Entire industries will reorient around it [generative AI]. Businesses will distinguish themselves by how well they use it."*
>
> – Bill Gates

> *"Use the tools, get a sense of the possibilities and participate in the conversation so that safe [artificial general intelligence] is as beneficial as possible."*
>
> – Sam Altman

> *"AI isn't a technology, and it isn't singular. It is, in fact, a complex system of many different technical components, systems, infrastructures and processes."*
>
> – Genevieve Bell

Morgan Stanley is calling AI a $6 trillion opportunity and notes that as AI tools and their ability to digitise untapped offline spending continue to evolve, industries such as advertising, e-commerce, travel, shared economy and public cloud are poised to benefit significantly.

Further, the firm has launched its own GenAI platform and rolled it out to 900 of its advisors as a pilot, with the intention of expanding it to more than 15,000 advisors.

While leaders are eager to advance GenAI-driven applications and opportunities in their organisations, they are struggling with three challenges:

- How to mitigate GenAI's risks — both those that already are well known and those that are just coming into view
- How to apply GenAI in areas and functions with the most potential to drive meaningful business value
- How to execute on and sustain those ideas

Morgan Stanley Co-President Andy Saperstein said at a presentation in May that he was "100% convinced" that any organisation that is not embracing AI tools "is going to be really left behind."

The risks — The five top risks that GenAI presents

Organisations need to appreciate the importance of adopting GenAI responsibly — and fortunately, most do. Despite the hype around GenAI and the rush in the market to achieve potential value, business leaders realise that this technology is nascent and requires an awareness of and preparation for the associated risks. Some of these risks are still emerging, and only time will tell how the legal and regulatory landscape will evolve to mitigate or alter these prospects.

At present, some of the top risks that GenAI presents include the following:

Veracity (hallucination) risk

Textual generative AI is optimised to predict the next word in a sequence, based on the massive amounts of information and examples of word sequences it has been trained on. It is not trained or optimised to evaluate truthfulness or to distinguish fact from fiction. Large language models (LLMs) are inherently designed to produce a well-presented response, but they do not have any concept of the veracity of that response, often leading these models to produce confident and convincing responses that are factually incorrect — a concept more recently termed hallucination. As a result, it is imperative that any text generated with AI is evaluated to test its truthfulness. Consider this real-life example: An attorney wrote a legal brief using ChatGPT, which produced detailed information citing a number of legal cases that had never occurred. The technology conjured them in a resoundingly persuasive brief — so persuasive that the attorney failed to consider whether the cases actually existed. According to the *New York Times*, hallucinations compound the challenges of fact checking. The publication warns that as chatbots become more reliable, users may become too trusting.

Authenticity risk

Authenticity risks are akin to hallucination risks and arise when it becomes impossible to determine whether something was fabricated by AI. One example is when deepfakes (AI-manipulated synthetic media) replace the likeness of one person with another in photo, video or audio files. Very little sample data is needed to construct a deepfake.

Tan Wang, a third year Ph.D. student at Nanyang Technological University in Singapore, collaborated with Microsoft to create a model called Disentangled Control for Referring Human Dance Generation in Real World (DisCo). The model splits an image into three parts — the pose of the person in the shot and the background and foreground to be used in the video. Then, using information AI has trained on from TikTok dance videos, the model morphs the person into a series of poses to create individual frames. When compiled back into a video, these frames produce realistic footage of that person dancing. Wang said "...With these things, you can try to compose anything you want. If you want Elon Musk to dance, you can just use our [code]."

Deepfakes have become a major security risk, accentuated by the fact that voice authentication, already in use in financial services, can be the target of audio deepfakes.

Data privacy risk

Data privacy becomes an issue when insiders using public tools like ChatGPT inadvertently input proprietary information. It is critical for adopters of GenAI to read and review thoroughly the terms and conditions associated with any software and understand the potential impact that using that software can have on data privacy. Caution is advised when public tools are used as the prompts and queries may become part of the public data pool and available to all future users. In one case, employees at a software company asked ChatGPT to optimise proprietary code, making that code available to every other ChatGPT user. Major software and services vendors are attempting to address this issue by making changes to their terms of service. For example, videoconferencing service vendor Zoom reverses policy that allowed it to train AI on customer data (engadget.com).

Cybersecurity risk

Cybersecurity risks result from hackers bypassing an AI's restrictions on its use to develop illicit services like scripts to steal data or develop new malware. The Open Worldwide Application Security Project (OWASP) Foundation has identified the top 10 security risks for LLMs, which include poisoning of training data, disclosure of sensitive information and prompt injection. Understanding the cybersecurity implications of GenAI must be a priority for those considering using the technology.

Ownership risk

GenAI typically does not provide or cite its sources of information, and the models are continuously learning from both newly published information and information gleaned from users' interactions with the models. If a developer uses GenAI to help them write code, some of the generated code may have originated from a competitor's intellectual property if that competitor previously used the model to optimise its code. Neither the developer nor the competitor — nor anyone else, for that matter — could readily trace or otherwise establish original ownership. Garnering a better understanding of how trademarks are defined and the potential liability of infringements is something that cannot be ignored. The law firm Christian & Barton offers guidance for brand owners that illustrates some of the most critical considerations for creating content and highlights concerns such as inadvertently using the intellectual property of well-known brands to create confusingly similar trademarks. There are additional ownership risks as well. For example, one provider of stock images has sued the makers of a public AI text-to-image model for using the provider's proprietary images to train the maker's image generator. While some of the copyright legalities are still in flux, the U.S. Copyright Office is offering some guidance on how AI-developed artwork is protected under copyright and now states that when "a work's traditional elements of authorship were produced by a machine, the work lacks human authorship and the Office will not register it." While that amounts to little more than a partial and temporary solution, new rulings will potentially become foundational in determining future copyright decisions.

Governance — A governance framework should support scalability and flexibility

GenAI has fractured many existing governance frameworks for conventional AI.

GenAI has fractured many existing governance frameworks for conventional AI, as the ability to inform the algorithm based on user inputs and GenAI's ability to create new and novel outputs may not fit into existing governance rules. To develop effective policies for GenAI use, ensure GenAI's ethical application in the organisation and avoid the well-documented hazards of AI bias, policymakers must develop a framework that supports both scalability and flexibility, allowing consideration for each GenAI use case based on several key characteristics. Understanding these GenAI concepts properly will ensure that policies are granular enough to address a variety of AI usage scenarios and avoid the hazards of a one-size-fits-all approach to governance.

Four essential characteristics of a generative AI application

Early efforts to govern the use of GenAI within the organisation might have included a single policy encompassing everything from employees using ChatGPT to development and tuning of proprietary machine learning (ML) models. Policies are more effective, however, when they are specific to four characteristics of generative AI applications:

1. The value or function the application delivers
2. How it is trained
3. The data on which it is trained
4. The architecture on which it runs

These characteristics should inform the organisation's program to evaluate, build, govern and sustain GenAI solutions. Governance frameworks may be modelled on sets of responsible and ethical AI principles already defined by organisations including Microsoft[1], IBM, Google and risk management frameworks by standards organisations such as the National Institute of Standards and Technology (NIST). Consideration of ethics, governance, security, compliance and change management should form the foundation of any AI program.

1. The value or function the application delivers

AI applications offer a variety of opportunities to create value, each of which carries its own considerations for governance and policy. However, it also is critically important to understand the intended use of the GenAI application and to align each GenAI application to practical and relevant guidelines. For example, using GenAI to craft a witty poem to send with your Mother's Day flowers holds vastly different benefits and risks than providing personalised investment advice.

Identifying the expected business value of any AI solution is key, and that solution must be deployed in a simple, intuitive way to achieve its full potential. The following is a summary of selected functions and applications of GenAI:

- **Decision-support** applications enable people and systems to make more informed decisions by curating and synthesising large amounts of relevant data and presenting that data at the right time with appropriate context.
- **Customer experience** applications help with customer service needs by answering queries and speeding response times, providing personalised experiences, and simplifying interactions for customers' convenience.
- **Knowledge management** applications unlock enterprise data, making the knowledge of its workers and enterprises readily accessible through a variety of mediums and at any point in time, reducing the need to rely on an individual's personal knowledge to source information.
- **Process efficiency and process automation** applications infuse intelligence into automation technologies such as robotic process automation or business process automation, scaling and speeding processes end to end.

[1] Read also: "Microsoft announces new Copilot Copyright Commitment for customers," Brad Smith and Hossein Nowbar, Microsoft, September 7, 2023.

2. How it is trained

There are several techniques that can be used to train AI, including the following:

- **Supervised learning:** Trainers provide AI with labelled information from which it learns: this is white, this is black; this is an apple, this is a pear; and so on. With this method, humans exert the highest control over AI training.
- **Unsupervised learning:** AI infers its own categories and classifications from data without being provided explicit labels. With this method, AI learns representations of the input that can be used to explore, analyse or generate new data.
- **Reinforcement learning:** An AI model is rewarded when it accomplishes a defined objective or optimisation. In this scenario, it is important to understand what the model is optimised, or reinforced, for.

3. The data on which it is trained

Data might be sourced from the public domain, subscription data feeds, enterprise-owned data and industry-specific data. Understanding where the data came from, if it was labelled, who labelled it, and/or how it was labelled is important. For public domain information, a general knowledge of how data was vetted (or not) for inclusion should be obtained.

4. The architecture on which it runs

The architecture on which AI runs fundamentally informs the application's security and privacy. Architecture decisions are central to management decision-making, as well as to choices related to ethics, responsibility and governance.

Forthcoming regulations will have a measurable impact on how GenAI applications are developed, deployed and used.

- **The public domain** includes AI tools anyone can use. Every user's input feeds the ever-evolving model, and these interactions guide how public-domain AIs will behave in the future. For businesses, public-cloud generative-AI tools like ChatGPT, DALL-E and Bard can boost productivity. They're risky, however, because they are beyond the control of the enterprise. Some organisations dictate that staff should use these tools only under certain conditions; others instruct their staff not to use these tools at all.
- **"Enterprise cloud"** describes AI in a hosted-cloud environment, which often affords additional data and application security and privacy according to the agreed-upon hosting terms and conditions.
- **"On-premises"** describes AI that is hosted on servers owned and operated by the enterprise. While this offers the greatest privacy and security, it also comes at the highest cost, including hardware and in-house operational expertise.

The evolving regulatory landscape

Leaders will likely see the regulatory landscape change over the coming months. Forthcoming regulations will have a measurable impact on how GenAI applications are developed, deployed and used. Organisations will need to anticipate that changes are coming and should prepare themselves by understanding the characteristics of their AI applications. Leaders will need that information so they can make the best-informed decisions when regulations change.

As just one example, the European Union is developing the Artificial Intelligence (AI) Act, which will strengthen existing regulations on AI development and use. The act will focus on "strengthening rules around data quality, transparency, human oversight and accountability [and] address ethical questions and implementation challenges in various sectors ranging from healthcare and education to finance and energy," according to the World Economic Forum.

For additional information on regulations, policy and the latest news related to GenAI, please visit Protiviti's Transformation and Innovation Insights web page.

Devising a GenAI strategy

Formulating a GenAI strategy should begin with a small, cross-functional team to pilot the technology. Representation should be broad, but the team also should be small enough to support rapid decision-making. Sponsors should ensure the team is educated at an appropriate depth level according to their expected contributions around key AI concepts.

Most enterprises have yet to develop or acquire significant expertise in developing GenAI programs or applications. Leaders may want to rely on external experts who can guide efforts and transfer knowledge about effective GenAI governance and use. They'll want partners who can accelerate progress in developing AI programs, governance and strategy, often bringing accelerators and domain expertise to accelerate time to value (TTV).

Getting started: Six steps to AI enablement

Organisations can enable GenAI via a six-step process. Each step can operate independently once its predecessors' key outputs are developed, offering a flexible approach.

Step 1. Formulate the vision and team

The most successful deployments of AI orient the team around a shared vision and purpose, such as improving the customer experience or reimagining operational processes. Once the vision is defined, organisations should:

- Align an executive sponsor and line-of-business sponsors for each functional area of operation.
- Develop cross-functional representation for technology, risk management, privacy, security and change management to support the sponsors.
- Afford sponsors and stakeholder groups with both autonomy and authority to evaluate ideas and experiment with GenAI, with increasing levels of oversight and governance accordingly as applications approach production.

Step 2. Identify opportunities

This step begins with an exploration of how GenAI is being used in the organisation's own industry and beyond. This is the "think big" opportunity — where teams can be challenged to disrupt or be disrupted. Ideate on the possibilities and aim to outline 10 to 20 practical opportunities that are aligned with the stated vision and have the potential to drive meaningful business value.

Despite the challenges and uncertainties connected to GenAI, business leaders should be eager to seize the opportunities offered or else risk their businesses being disrupted.

Step 3. Prioritise ideas

Once the team has defined potential ideas, each use case should be evaluated for feasibility and complexity while also considering the availability of technologies to support desired results. Team members should:

- Analyse technology requirements, data integrations and operational readiness considerations for each application idea.
- Look for areas where technical capabilities may be reusable and avoid creating siloed applications
- Establish high-level anticipated value of each idea, using the benefit-complexity analysis to recommend a pilot application.
- Develop specific value measures expected from the selected pilot GenAI use case, with an adaptive eye for future value that may not have been originally anticipated.
- Outline an indicative road map to optimise investments and visualise correlations and reusability across use cases.

Step 4. Initialise the pilot

This step includes familiar software development preparatory steps such as sprint planning, as well as considerations unique to AI, such as:

- Selecting and procuring AI-enabled technologies
- Building or configuring the model
- Procuring data sources and pipelines
- Experimenting with various data science methods
- Conducting practical AI training with the team

Step 5. Develop the pilot application

In this step, the team builds the GenAI-enabled application, exploring and tuning the model(s) iteratively. To inspire confidence in the pilot and identify opportunities to tune capabilities further, the team may develop a demonstration model of the pilot application.

This results in a proof-of-concept, alpha version of the application that will demonstrate its value and may be made available for testing with select stakeholders.

As the team establishes metrics for performance and value delivery, it can add application features to monitor attainments in these areas. Automating metrics and monitoring enables data-driven decisions about subsequent enhancements and possible future AI initiatives.

Step 6. Adapt and reflect

After delivering the pilot's alpha version, the team evaluates both the development process and the application itself, and uses its findings to improve the pilot application, as well as its methods. Also during this step, the team revisits the high-level road map for deploying AI use cases and the technical capability inventory that has been developed.

In closing — act now to achieve rewards while managing risk

Despite the challenges and uncertainties of GenAI, business leaders should be eager to seize the opportunities offered or else risk their businesses being disrupted. With 60% of organisations planning to pilot or operate GenAI applications by 2024, the time is obviously now to begin to experiment and learn; organisations that don't take these steps risk being left behind.

GenAI's potential use cases and long-term benefits continue to be defined, as does a rapidly evolving regulatory landscape. Leveraging what GenAI can deliver requires that leaders develop ethical and responsible governance frameworks for their GenAI programs and start their generative AI journeys with innovative pilots. Those journeys help educate and inform business leaders of the potential offered while also creating institutional knowledge that development teams can lean on to advance additional use cases. It all comes down to finding the right balance between GenAI risk and reward.

About Protiviti

Protiviti (www.protiviti.com) is a global consulting firm that delivers deep expertise, objective insights, a tailored approach and unparalleled collaboration to help leaders confidently face the future. Protiviti and our independent and locally owned Member Firms provide clients with consulting and managed solutions in finance, technology, operations, data, analytics, digital, legal, HR, governance, risk and internal audit through our network of more than 85 offices in over 25 countries.

Named to the 2023 *Fortune* 100 Best Companies to Work For® list, Protiviti has served more than 80 percent of *Fortune* 100 and nearly 80 percent of *Fortune* 500 companies. The firm also works with smaller, growing companies, including those looking to go public, as well as with government agencies. Protiviti is a wholly owned subsidiary of Robert Half (NYSE: RHI). Founded in 1948, Robert Half is a member of the S&P 500 index.

About the Author

Managing Director Christine Livingston is responsible for artificial intelligence/machine learning and innovation solutions at Protiviti. With over a decade of experience in AI/ML deployment, she has delivered hundreds of successful AI solutions, including many first-in-class AI-enabled applications. She has helped several Fortune 500 companies develop practical strategies for enterprise adoption of new and emerging technology, including the creation of AI-enabled technology roadmaps. She focuses on identifying emerging technology opportunities, developing innovation strategies and incorporating AI/ML capabilities into enterprise solutions.

Contact Christine at Christine.Livingston@protiviti.com.

THE AMERICAS

UNITED STATES
Alexandria, VA
Atlanta, GA
Austin, TX
Baltimore, MD
Boston, MA
Charlotte, NC
Chicago, IL
Cincinnati, OH
Cleveland, OH
Columbus, OH
Dallas, TX
Denver, CO
Ft. Lauderdale, FL
Houston, TX
Indianapolis, IN
Irvine, CA
Kansas City, KS
Los Angeles, CA
Milwaukee, WI
Minneapolis, MN
Nashville, TN
New York, NY
Orlando, FL
Philadelphia, PA
Phoenix, AZ
Pittsburgh, PA
Portland, OR
Richmond, VA
Sacramento, CA
Salt Lake City, UT
San Francisco, CA
San Jose, CA
Seattle, WA
Stamford, CT
St. Louis, MO
Tampa, FL
Washington, D.C.
Winchester, VA
Woodbridge, NJ

ARGENTINA*
Buenos Aires

BRAZIL*
Belo Horizonte*
Rio de Janeiro
São Paulo

CANADA
Toronto

CHILE*
Santiago

COLOMBIA*
Bogota

MEXICO*
Mexico City

PERU*
Lima

VENEZUELA*
Caracas

EUROPE, MIDDLE EAST & AFRICA

BULGARIA
Sofia

FRANCE
Paris

GERMANY
Berlin
Dusseldorf
Frankfurt
Munich

ITALY
Milan
Rome
Turin

THE NETHERLANDS
Amsterdam

SWITZERLAND
Zurich

UNITED KINGDOM
Birmingham
Bristol
Leeds
London
Manchester
Milton Keynes
Swindon

BAHRAIN*
Manama

KUWAIT*
Kuwait City

OMAN*
Muscat

QATAR*
Doha

SAUDI ARABIA*
Riyadh

UNITED ARAB EMIRATES*
Abu Dhabi
Dubai

EGYPT*
Cairo

SOUTH AFRICA *
Durban
Johannesburg

ASIA-PACIFIC

AUSTRALIA
Brisbane
Canberra
Melbourne
Sydney

CHINA
Beijing
Hong Kong
Shanghai
Shenzhen

INDIA*
Bengaluru
Chennai
Hyderabad
Kolkata
Mumbai
New Delhi

JAPAN
Osaka
Tokyo

SINGAPORE
Singapore

*MEMBER FIRM

protiviti®