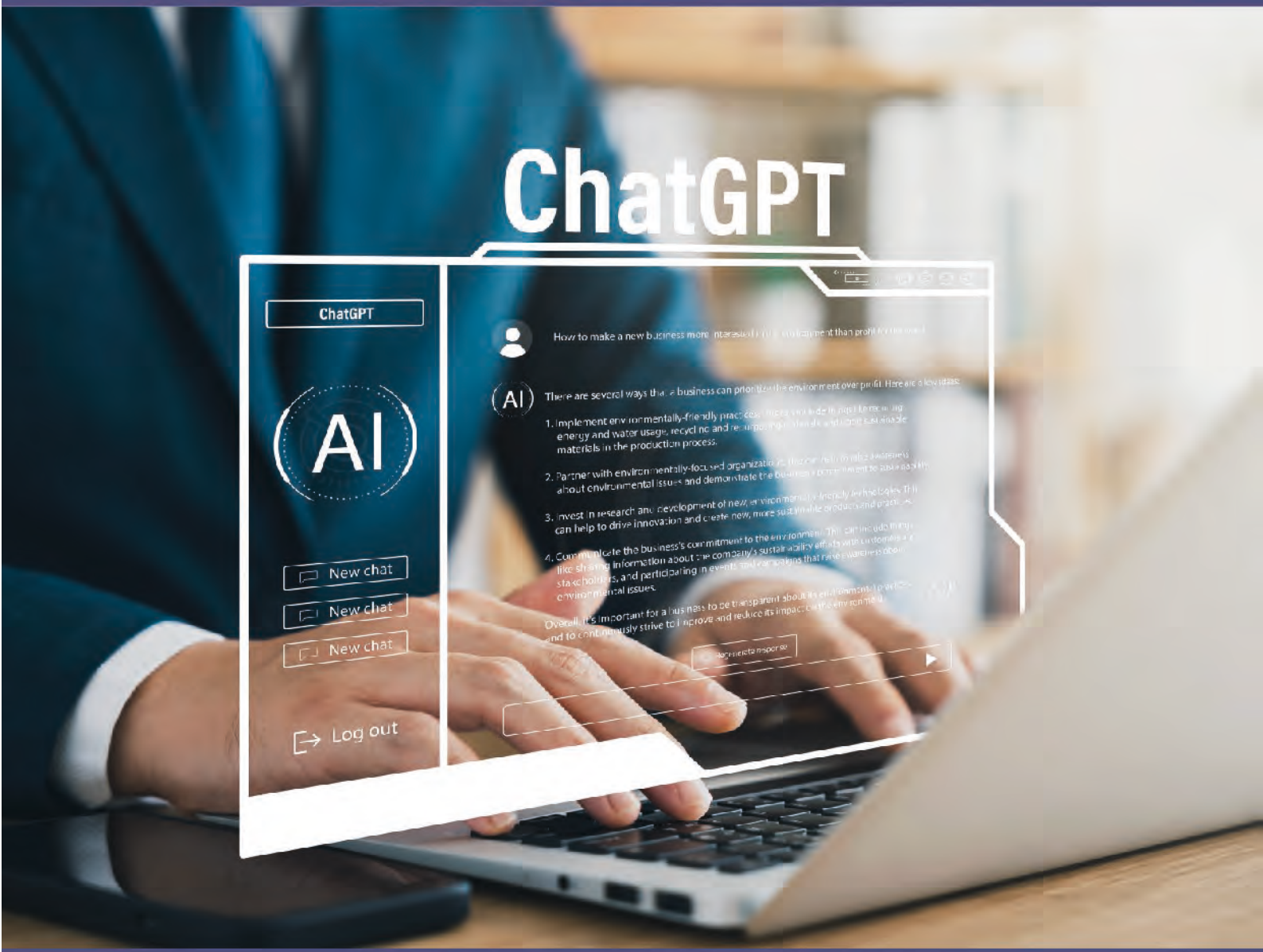
Swiss Data Science Center

# Generative AI: Opportunities and Risks

A Briefing Paper by the Swiss Data Science Center, an initiative of EPFL and ETH Zurich




With applications such as ChatGPT gaining traction, many questions are arising about Generative AI, the technology behind those tools sometimes described as disruptive. How does it work, how can it support organizations today and what are the risks and opportunities associated with its use?

The Swiss Data Science Center provides insights into this technology and analyzes how it can generate value within an organization.

# Generative AI: Opportunities and Risks

By the Swiss Data Science Center, an initiative of EPFL and ETH Zurich

Generative AI has the potential to bring about significant changes in various fields. AI algorithms can be used to generate new works of art, music, and writing. In the business world, generative AI can be used to automate tedious tasks, such as data entry and report writing, freeing up valuable time for more meaningful work. Additionally, the ability of generative AI to create new and unique outputs could lead to the development of new products and services, driving economic growth.

However, with any new technology come concerns and challenges. The use of generative AI-based tools and applications raises questions and has implications that should be addressed proactively in order to assess their relevance to business requirements.

## What is Generative AI?

Generative AI is an expression referring to types of artificial intelligence (AI) that can be used to create new text, images, video, audio, code, and other forms of media. Recently, this technology has received a surge of public attention thanks to applications such as:

- ChatGPT, a large language model developed by OpenAI. It is capable of generating human-like text based on a prompt provided to it.
- DALL-E, a generative model developed by OpenAI that can generate unique images from textual descriptions.

The enabling technology behind state-of-the-art generative models is a machine learning paradigm called *transformers* and more specifically its recent evolution named *generative pretrained transformers* (GPT). More in detail:

- **Transformers** are deep learning approaches based on “attention”, i.e., the capability of a neural network to focus on specific areas of a text in order to learn about its meaning. They are “generative” AI models as they are usually optimized to solve text generation tasks such as finding the upcoming or missing word in a sequence. Introduced in 2017 by Google (BERT being a well-known example), they have since become the standard approach to textual understanding and generation.

- **Generative Pre-trained Transformers** (GPTs) are a form of generative models developed since 2018 by OpenAI, with similar approaches proposed by other tech companies. GPTs are based on a deep neural network architecture capable of learning complex features about human language thanks to its exposure to billions of words from different origins (Wikipedia, Reddit, books, web pages, etc.) during training. **The latest GPT models - such as GPT-3, launched in 2020 - excel at generating plausible content and are sometimes said to have "learned to learn"**. This refers to their ability to perform tasks like translation or summarization in line with the user's intentions and without being specifically optimized to solve them. This behavior can be triggered by "prompts", i.e., textual cues to the desired behavior. Given the input: "Translate English to French. I love you → je t'aime. I hate you → _", the prompt "*I hate you"* is all that's needed to hint at the intended output.

## ChatGPT

Released in November 2022, ChatGPT is an interactive system developed by OpenAI based on a specialized version of the GPT-3 model called InstructGPT[1].

- **What makes ChatGPT disruptive** – a dialogue interface capable of maintaining the conversation context for a "natural" feel and – perhaps more importantly – a direct interface to a type of logic that so far had been only implicitly available behind applications such as Google search and Meta's Facebook platform. In other words, ChatGPT offered many users the first opportunity to directly assess and interact with a state-of-the-art AI approach to content generation.
- **What it's best at** – summarizing, simplifying, and bootstrapping text, providing hints or stubs to inspire human creation.
- **Limitations** – despite being much more effective than its predecessors at creating truthful content, it is not exempt from "hallucination", i.e., the production of plausible but incorrect content.

## Linking Text to Images

Generative AI models have been successfully used also for generating images based on textual descriptions. Examples of recently released models include **Stable diffusion**, a latent text-to-image diffusion model capable of generating photo-realistic images given any text input, and **DALL-E**, a generative model that can generate

[1] https://openai.com/blog/instruction-following/

unique images from textual descriptions, i.e., prompts. Such models make it possible to design products based on textual specifications and produce creative artwork.

# How AI-Generated Content Supports Businesses Today

It is important to note that GPT-type architectures have been a (discreet) part of our daily lives since their introduction. For instance, Google search has been using transformer-based models to represent user queries and web page content since the introduction of BERT. Its MUM model, released in 2021, was already 1000 times more powerful and is now being followed by LamDA and Bard[2]. The latest developments offered by GPT-3's successors GPT-3.5[3], InstructGPT, and ChatGPT offer further opportunities to generate and synthesize text with high relevance to business requirements. Let us look at some applications:

1. **Foreign language learning and content creation**
   The language learning company Duolingo is known to provide French grammar corrections and automate the generation of English tests with GPT-3[4]. The Swiss media creator Coteries trained its own GPT-type model, Cedille[5], to provide its users with high-performing text generation for creating automatic email replies and high-quality translations to French and German, amongst other applications.
2. **Workplace productivity**
   The recent agreement between Microsoft and OpenAI has hinted at the integration of GPT-like functionalities within its proposed products and services. One concrete example consists in the premium features recently announced for the Teams communication platform[6]. These include intelligent recaps, the automatic creation of markers and tasks, and the annotation of salient moments. This kind of feature is expected to augment workplace productivity.

[2] https://blog.google/technology/ai/bard-google-ai-search-updates/
[3] https://platform.openai.com/docs/model-index-for-researchers
[4] https://blog.duolingo.com/test-creation-machine-learning/
[5] https://cedille.ai/
[6] https://www.microsoft.com/en-us/microsoft-365/blog/2023/02/01/microsoft-teams-premium-cut-costs-and-add-ai-powered-productivity/

3. **Code generation**
   Text generation is not limited to natural language. As a matter of fact, several recent demonstrations of ChatGPT have focused on its ability to generate programming code based on a natural language prompt as well as to explain its functionalities in plain English. While such capabilities require validation by human experts, they are currently part of commercial and open-source solutions. For example, Copilot is known to use Codex – a GPT-like model for programming code – to assist engineers with code writing[7].

**Use Case:** Automated Data Entry and Cleaning
**Problem:** Data entry and cleaning are time-consuming and error-prone tasks in a BioPharma/CDMO organization.
**Solution:** Use ChatGPT to automate data entry and cleaning tasks.
**Example Output:** Given a dataset with inconsistent formatting and missing values, ChatGPT could clean and standardize the data as follows: "Patient ID, Age, Gender, Diagnosis, Treatment, Outcome".

More generally, generative AI technologies may bring significant benefits to companies in terms of increased productivity, creativity, and competitiveness. These include:

1. **Automation of routine tasks:** Generative AI has the potential to automate routine tasks, freeing up time for employees to focus on more meaningful work and increasing productivity.
2. **Improved decision-making:** Generative AI can be used to analyze large amounts of data, providing companies with insights on various aspects of the business such as market trends, customer behavior, and product performance, and helping them make more informed decisions.
3. **Development of new products and services:** Generative AI can be used to generate unique outputs, leading to the development of innovative products and services and driving customer satisfaction and loyalty.

**Use Case:** Marketing and Branding
**Problem:** a company wants to create unique and visually appealing images for marketing and branding purposes.
**Solution:** Use DALL-E to generate images based on specific descriptions.
**Example Output:** Given a description of "A pharmacist holding a bottle of medicine with a backdrop of a sunrise over a mountain range", DALL-E could generate the following image:

4. **Cost savings:** by automating routine tasks and reducing the need for human labor, companies can reduce their costs and improve their bottom line.
5. **Competitive advantage:** companies that adopt and leverage generative AI will have a competitive advantage as they will be able to operate more efficiently and effectively.

[7] https://github.com/features/copilot

Generative AI technologies are poised to play a major role in driving growth and innovation in the coming years. Let's look at some specific applications.

---

***"Organizations have a responsibility to educate their staff and particularly decision makers to the functionalities of generative AI. This will enable a critical, fact-based approach to the decision of integrating such functionalities within their business processes",*** *says*
*Silvia Quarteroni, Head of Innovation at the SDSC.*

---

# Applications Across the Organization

The use of generative AI tools may prove effective in maintaining commercial competitiveness and generating value within several organizational functions:

1. In **product design**, to understand/summarize product briefs, analyze state-of-the-art products, and generate textual and visual product candidates.
2. In **customer care**, to analyze feedback and direct it to the relevant organizational units and/or respond in a semi-automated way.
3. In **sales/marketing**, to generate promotional content (ideally, re-elaborated by human experts), launch ideas deriving from the processing of a large body of text, improve customer targeting through insights on customer behavior and preferences, increase customer engagement through the use of AI-generated customer testimonials and product reviews.
4. In **HR**, to identify pain points and suggest career paths for employees, as well as to create content optimizing the operational flows of the unit.
5. In **IT**, to accelerate the generation of prompts and code, suggest architectures, and handle operational tickets. Several Swiss organizations are already benefiting from GPT-like models for bootstrapping code generation or translating code to different programming languages.

In the context of a collaboration between the SDSC and a company offering HR services, GPT-3 is used to generate textual representations referring to skills and job descriptions with the purpose of matching talents to workplaces.

6. In **R&D and education**, to leverage large amounts of data for fact-based question-answering tasks, initiate essay generation, correct grammar, and identify prior/complementary avenues of research and development.

In the context of the Monitoring Patterns of Violence project, the SDSC is helping the International Committee of the Red Cross evaluate its humanitarian impact by using a generative AI approach to identifying occurrences of violent acts and related entities (victims, perpetrators) in the areas where the NGO operates. The prompt-based technique is used with different transformers-based models to this end.

## Limitations and caveats

As the name suggests, generative AI tools produce synthetic data. What's more, such data is produced by "mimicking" the distribution of words in a way that is optimized for plausibility but not validated by an external reference such as a knowledge base. Finally, the deep neural network infrastructure underlying even the most recent and powerful generative AI tools such as ChatGPT has only been exposed to a limited amount of text, not more recent than mid-2021. As a consequence, the truthfulness of AI-generated content is far from guaranteed. Generative AI's behavior of producing plausible but false content is referred to as "hallucination", a phenomenon making direct integration into "mission-critical" tasks extremely risky at the current stage.

Another characteristic of GPT-like models trained by aligning with human intention is their abilities - both powerful and limiting - to "change their minds", i.e., to review assertions based on the user's feedback during an interaction. This has resulted in many examples of users influencing ChatGPT to state falsehoods as a result of their "persuasive" behavior.

Finally, and perhaps most importantly, one business limitation of generative AI technologies lies in the access to such technological opportunities via external APIs served through the cloud (domestic or foreign). In other words, **submitting text to a cloud API exposes an organization to the processing of its confidential information by a third party**, incurring privacy risks and potential breaches of its policies towards customers or internal users. This is why many large companies currently block access to such APIs for their employees.

# Outlook

- ChatGPT's successors are coming soon, and so are improved features of generative AI models. The current limitations of GPT-like models and their resulting APIs will likely be at least in part addressed by upcoming approaches.
- Organizations have a responsibility to educate their staff and particularly decision-makers about the functionalities of generative AI. This will enable a critical, fact-based approach to the decision to integrate such functionalities within their business processing.
- Educational institutions such as universities should play a key role in this process and ensure that they convey technological innovation in an effective and timely fashion to stakeholders in civil society.
- Currently, the direct integration of ChatGPT-like APIs as-is within production processes poses few technical challenges but several potential risks. In other words, for mission-critical applications, it is not recommended to use the direct output of ChatGPT. The correct usage of this technology is for initial content generation.

---

*To summarize with a quote from ChatGPT,* ***"Generative AI tools [...] can help improve productivity, generate new ideas, and reduce the time required for tasks. However, technical knowledge and relevant skills are required to effectively implement and adopt these tools."***

---

*About the SDSC*

*The Swiss Data Science Center (SDSC) is a joint venture between EPFL and ETH Zurich. Its mission is to accelerate the adoption of data science and machine learning techniques within academic disciplines of the ETH Domain, the Swiss academic community at large, and the industrial sector. In particular, it addresses the gap between those who create data, those who develop data analytics and systems, and those who could potentially extract value from it. The center is composed of a multi-disciplinary team of data and computer scientists and experts in select domains with offices in Zürich, Lausanne, and Villigen.*

*For further information, contact: communications@datascience.ch*