2025

Practical Guide to Generative AI for SAI Professionals

Sivré, Vincent

25/06/2025

FOREWORD

The emergence of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs) has transformed the field of natural language processing. These technologies offer powerful capabilities in text generation and language understanding, adding value to various processes across numerous fields. In this course, we explore the different ways Gen AI can enhance workflows, from data analysis to knowledge sharing and interactive applications. GenAI presents interesting use cases that auditors will discover as they familiarize themselves with this new tool. But it also presents risks that must be kept in mind.

This introductory course aims to mitigate such risks and demystify generative AI and LLMs for both technical and non-technical audiences across diverse Supreme Audit Institutions (SAI). It provides a comprehensive overview of foundational knowledge for understanding LLMs, core concepts in Gen AI, and practical applications of Gen AI in a variety of contexts.

The course, performed by Vincent Sivré, is intended for SAI professionals, equipping them with skills to enhance their work.

Vincent Sivré possesses a robust set of competencies that enable him to design and lead training sessions on the use of generative artificial intelligence (GenAI) for audiences such as auditors from higher control institutions.

As a senior auditor, he has a profound knowledge of public financial control, governance, and accountability frameworks. This background equips him to tailor AI training to the specific needs and challenges faced by auditors in public institutions.

Vincent Sivré's educational background includes prestigious institutions such as Sciences Po Paris, the École nationale d'administration (ENA), Sorbon University, École normale supérieure de Paris-Saclay. This academic and training experience underpins his ability to convey complex GenAI concepts clearly and effectively to professionals.

He emphasizes the importance of measuring and realizing productivity gains through AI projects, advocating for the use of analytical indicators and prioritizing improvements in operational efficiency. This practical orientation ensures that his training is not only theoretical but also actionable for auditors seeking to leverage AI in their control activities.

In summary, Vincent Sivré combines high-level expertise in AI and data, extensive knowledge of public sector auditing, and strong pedagogical skills, enabling him to effectively design and animate training programs on generative AI tailored to the needs of auditors in higher control institutions.

TABLE OF CONTENTS

SYNTHESYS 5

INTRODUCTION 8

1 UNDERSTANDING GENERATIVE ARTIFICIAL INTELLIGENCE 10

1.1 AI DESIGNS 10

1.1.1 A symbolic conception: man teaches the machine (between 1960 and 1990) 10

1.1.2 A connectionist or statistical conception: the machine learns on its own (1990 to the present day) 10

1.1.3 The design of artificial neural networks 10

1.2 THE ARCHITECTURE OF GENERATIVE ARTIFICIAL INTELLIGENCE (GenAI) 13

1.2.1 A new AI architecture 13

1.2.2 An architecture capable of predicting the next word to generate content 16

2 USE CASES AND RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE 18

2.1 EXAMPLES OF USE CASES 18

2.2 THE RISKS OF GenAI 19

2.2.1 The risks of “imperfection” 20

2.2.1.1 Biases 20

2.2.1.2 Hallucinations of GenAIs 21

2.2.2 Other risks 23

3 USING GENERATIVE ARTIFICIAL INTELLIGENCE IN YOUR PROFESSIONAL ACTIVITY 24

3.1 MAKE A REQUEST 24

3.1.1 General recommendations 24

3.1.2 Specific recommendations for auditors 26

3.1.2.1 Professional Secrecy and the Protection of Personal Data at the State Audit Office of Georgia 26

3.1.2.2 How to use GenAI while respecting professional secrecy and the GDPR? 27

3.2 EVALUATE, IMPROVE AND USE THE RESULT GENERATED 29

3.2.1 General recommendations 29

3.2.2 Specific recommendations for auditors 30

3.2.2.1 Transparency in the use of GenAI within financial jurisdictions 30

3.2.2.2 Transparency in the use of the GenAI towards those audited and the general public 31

3.3 TRAINING IN THE USE OF GenAI 31

4 APPENDICES 33
4.1 Mistakes to avoid when using ChapGPT 33
4.2 Prompt engineering techniques 34
4.3 Glossary 35

SYNTHESYS

This practical guide, structured into three main sections, is intended to assist audit teams in becoming familiar with generative artificial intelligence (GenAI), so that they can understand and integrate this technology into their professional practice.

Distinguishing Symbolic Artificial Intelligence from Generative Artificial Intelligence

Unlike symbolic AI, where humans teach the machine through predefined rules, generative AI enables machines to learn autonomously from examples using a deductive logic. Human intervention is thus limited to ensuring the quality of the machine's learning process and verifying that the task is properly performed.

GenAI is characterized by several key features that must always be kept in mind:

- It is **not** a reliable source of information: GenAI is neither an encyclopedia nor a search engine, but a system capable of generating content without any guarantee of accuracy.
- It does **not** understand the meaning of the words, images, or sounds it generates.
- Because it can produce coherent text, GenAI gives the **illusion** of understanding; this impression is misleading, and it is essential not to anthropomorphize the machine.
- **Humans assign meaning** to the output generated by GenAI; therefore, the responsibility for interpreting the generated result—whether text, image, or sound—lies with the user.

Risks associated with the use of generative Artificial Intelligence

While GenAI offers promising use cases that audit teams will discover as they familiarize themselves with the tool, it also entails significant risks:

- It is **neither** an encyclopedia **nor** a search engine.
- It can produce plausible, seemingly factual content that is actually incorrect—these are known as **hallucinations**.
- It exhibits **design biases** (eg, risk of discrimination in hiring) and may amplify **cognitive biases** in users (eg, intellectual complacency).
- It poses **confidentiality risks**, as all prompt data may be reused by the AI system's owner for training purposes. Even when claims are made to the contrary, there is no way to verify them independently.

The Importance of Evaluation and Transparency in the Use of GenAI

The use of generative artificial intelligence (GenAI) in public administrations, particularly in SAIs, enhances the efficiency and personalization of services. However, it also demands ethical and legal vigilance, especially concerning the protection of personal data and professional secrecy.

Audit teams must avoid transmitting sensitive information to unauthorized GenAI systems, such as publicly available tools, and instead prioritize internal solutions (like ChatJF in the French SAI[1]) or, alternatively, anonymize or pseudonymize data as needed.

Evaluating the results generated by GenAI systems is essential for identifying errors and avoiding liability. Hallucinations must be detected, and a critical mindset is necessary. It is recommended to use GenAI systems that disclose their sources, though this does not guarantee the absence of errors.

Training is vital to understand the capabilities and limitations of the tools, to optimize their use, and to assess their reliability. Financial jurisdictions must ensure transparency in their use of GenAI, through clear documentation and independent audits to ensure compliance with standards, particularly data protection regulations.

[1] The platform https://ia.ccomptes.fr/ provides access to a conversational agent powered by a large language model (LLM) based on artificial intelligence. It is accessible only via the JF network or through a VPN connection. The platform operates using open-source models made available by digital players such as Meta (LLaMA models) and Mistral (Mistral and Mixtral models). These public models come with certain limitations when compared to proprietary models like Google's Gemini or OpenAI's ChatGPT.

First, there are computational constraints: the JF platform does not possess the same processing power as that of Microsoft or Google. Second, public models are limited in context length, meaning they can process only a smaller amount of text at once.

It is important to note that the platform is neither certified nor authorised to handle classified information (i.e. secret or top secret). As such, users are strongly advised not to input confidential data.

INTRODUCTION

The emergence of generative artificial intelligence (GenAI) is transforming the way we produce and consume information. This new technology is already influencing many sectors—such as education, research, law, and the press—by opening new possibilities for efficiency and creativity, but also by raising unprecedented technical, ethical, and legal questions.

In particular, the popularization of GenAI in the form of publicly accessible conversational agents—especially ChatGPT —has introduced a form of automation into language-based tasks that were previously considered to require human intelligence. These models now enable any user to produce text, summarize or reformulate content, write software code, or even generate images, sound, or videos from simple textual instructions (prompts). In so doing, they expand the range of what individuals and institutions can accomplish.

Generative AI is therefore of particular interest to Supreme Audit Institutions, which are themselves primarily information-based organizations. However, the characteristics of these tools and the risks they present—especially with respect to data confidentiality and the reliability of results—require a cautious, informed, and well-supervised approach. In the public sector, where transparency, responsibility, and compliance with legal obligations are paramount, this necessity is even more pronounced.

This practical guide is intended for all audit team members and aims to:

- Introduce the key concepts behind generative AI;
- Outline its current capabilities and limitations;
- Describe the main risks associated with its use;
- Proposes an initial framework for secure and responsible experimentation;
- Provide practical examples to facilitate its integration into daily tasks.

This document is not exhaustive. It is designed to evolve over time, as tools develop, as legal frameworks adapt, and as new practices emerge. Above all, it invites all users—regardless of their level of technical proficiency—to develop a critical, reasoned, and ethical perspective on the use of this powerful yet still immature technology.

Article 3-1) of European Regulation No. 2024/1689 of 13 June 2024 on artificial intelligence defines "artificial intelligence system" (AI system), as "[…] *an automated system that is designed to operate at different levels of autonomy and can demonstrate adaptive capacity after deployment, and which, for explicit or implicit objectives, infers, from the inputs it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments* ".

A Goldman Sachs Survey shows that generative AI could automate an average of 25% of tasks across all sectors in the United States. This average share varies depending on the field of activity. In legal professions, the automation rate could reach 44%. Furthermore, the study shows that nearly 60% of legal jobs are likely to be completed by GenAI.

This report and this study announce profound transformations in the professions practiced within SAIs in the years to come.

Indeed, GenAI represents a considerable advance in AI in the field of linguistics and opens up possibilities that are still difficult to measure. Thanks to this new technology, the accomplishment of certain tasks by a machine becomes possible and, with it, productivity gains (example: management and drafting of letters, emails, meeting minutes, drafting of publications on social networks, newsletters, blogs, preparation of control plans, drafting of parts of reports or communications, production of summaries, research, assistance with reflection and the search for new ideas, etc.).

It therefore appears essential that SAI auditors take ownership of these new tools and familiarize themselves with them in order to measure what they can bring to them and, *ultimately*, deploy them in the various chambers and integrate them into their business processes.

Since GenAI presents benefits and risks, the use of these systems should be governed by recommendations and best practices. This is the objective of this guide.

This practical guide, structured in three main parts, aims to help audit teams familiarize themselves with GenAI so that they understand and integrate this technology into their professional practice.

This guide begins with a detailed explanation of generative artificial intelligence and aims to enlighten control teams on its design and operation.

Next, this guide will discuss examples of possible uses of GenAI and the associated risks. These elements will help control teams understand what GenAI can be used for and highlight the challenges this technology presents them with.

Finally, the last part of this guide will propose good practices and vigilance rules to control teams to enable them to use, for their professional practice, the possibilities offered by these tools in a responsible and informed manner.

1 UNDERSTANDING GENERATIVE ARTIFICIAL INTELLIGENCE

Generative artificial intelligence (GenAI) represents a major turning point in the history of AI.

To fully understand its architecture and the risks it entails, it is first necessary to present the two conceptions of AI as they result from history.

1.1 AI DESIGNS

The history of AI, dating back to the 1940s and 1950s, shows two different conceptions of AI. The second led to the design of an artificial neural network.

1.1.1A symbolic conception: man teaches the machine (between 1960 and 1990)

It is based on hypothetico-deductive (and rationalist) reasoning: "if... then". AI assumes a set of instructions programmed in advance by humans to give rise to expert systems designed to carry out a specific task.

Example: this is how the vast majority of spell checkers we know today work in word processing software, man having taught the machine, by means of a set of instructions (expert system), the grammatical rules (the difficulty being to determine what a mistake is and how to classify them into spelling mistakes, grammar mistakes, style mistakes, etc.).

1.1.2A connectionist or statistical conception: the machine learns on its own (1990 to the present day)

It is based on artificial neural networks, the links between neurons, through their different layers, allowing a machine to learn on its own to perform certain tasks (example: recognizing a cat from a dog in an image). This is what we call "automatic learning" or "machine learning" or *deep learning*.

For example: face and object detection in your digital camera or smartphone, translation, voice recognition used by various digital assistants, not forgetting GenAI systems, which we reserve the right to study.

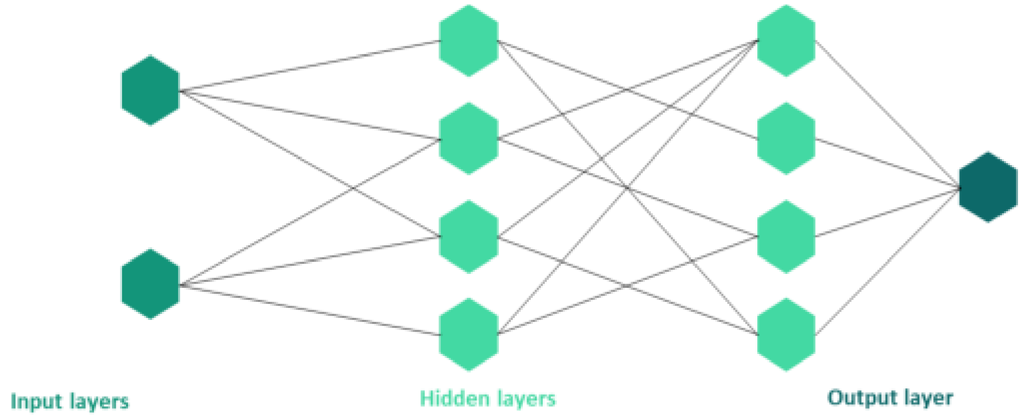
1.1.3The design of artificial neural networks

Artificial neural networks (ANNs) are models inspired by the functioning of biological neurons in the brain, composed of layers of interconnected nodes.

They consist of an input layer, hidden layers, and an output layer, where each node performs computations based on input signals and adjustable weights.

ANNs are fundamental to deep learning and artificial intelligence, enabling machines to learn and improve from training data.

Schéma n° 1 : Artificial neural networks



Source: National Bar Council, Practical Guide - Use of Generative Artificial Intelligence Systems - September 2024.

In the diagram above, the neural network architecture is composed of an input layer of two neurons, two hidden layers of four neurons each, and an output layer composed of one neuron.

A neural network mimics the human brain: the input data passes through several layers of "neurons" before providing a result, in our example distinguishing a dog from a cat.

First, it is necessary to train the machine.

During this first phase, the machine learns to identify statistical links (i.e. similarities) between the data by training on a large data set, the learning data. This is statistical learning.

More precisely, each neuron contains parameters that allow the machine to detect, by means of statistical calculations, recurring patterns (statistical links) in the training data.

When this training data is labeled, the machine knows the expected result: for example, for a machine to distinguish a dog from a cat, it is trained on data containing a label specifying whether it is a photo of a dog or a cat. Thus, the machine can, by means of an error backpropagation mechanism, correct the parameters of the neurons itself in the event of an error. This type of learning is called "supervised".

To learn and obtain good results, the machine is trained on a large number of examples (thousands, even millions of images), which requires two things: data produced in large quantities, which has been made possible by the digitalization of our societies ; computing power, supported in particular by the emergence of the cloud.

In some cases, humans perfect machine learning using a system that closely resembles a reward system: if the answer is correct, the reward is given; if not, the reward is delayed. This is called "reinforcement learning."

In a second step, the machine applies what it learned during the training phase to new data to perform a task.

NAMELY

ChatGPT 3.5, developed by OpenAI , was trained on 45 petabytes, or 45 million gigabytes, of texts sourced from the WebText database (excerpts from web pages), Wikipedia, books and scientific publications [2]. To give an idea of the size, your computer's storage capacity rarely exceeds a few terabytes, or a few thousand gigabytes.

It should be noted that the deeper a network is, that is to say the more hidden layers it contains, the more the machine is able to abstract, that is to say to recognize, with a very high probability (97 or 98%), the characteristics of a dog and a cat on new data which are, by definition, distinct from the training data.

IN SUMMARY

Unlike symbolic AI, in which humans teach machines, in neural network (connectionist) AI, the machine learns on its own from examples using deductive logic. Thus, human intervention is limited to ensuring the quality of the machine learning and that the expected task is performed correctly.

This is how GenAI works.

[2] How was ChatGPT formed? - ChatGPT info (chatgpt-info.fr).

1.2 THE ARCHITECTURE OF GENERATIVE ARTIFICIAL INTELLIGENCE (GenAI)

GenAI is a variety of AI that relies on neural networks and machine learning.

While AI has become very efficient, particularly in image processing (examples: detection of faces, animals or objects), GenAI allows considerable progress to be made in the linguistic field.

Unlike an AI trained to distinguish a dog from a cat, GenAI is capable of generating content (text, image, audio, video) in response to a user's query. All of this progress has been made possible by a new AI architecture, i.e., a new learning model.

1.2.1A new AI architecture

A distinctive feature of GenAI is its learning model, the Large Language Model (LLM). This type of model is also called a foundation model or a general-purpose model because of its wide variety of use cases.

GenAI was born out of a major technological breakthrough: transformers, discovered by a team of Google researchers in 2017 [3].

Attention Is All You Need: A Turning Point in AI

"Attention Is All You Need" is a 2017 research paper that introduced the Transformer architecture, revolutionizing machine translation models.

The model relies solely on attention mechanisms, eliminating the need for recursion and convolutions, which improves parallelization and reduces training time.

This approach has become fundamental to modern language models, influencing diverse applications such as multimodal generation and querying.

These transformers have the advantage of being free from supervised learning which relies on the use of labeled data (data labeling being done manually, labeled data sets are fewer in number, as they are very expensive). With transformers, it is possible to train the machine with unlabeled data which is available in much greater quantity. This so-called "unsupervised" training allows the machine to learn sentence structures to bring out linguistic models. Then, this initial learning is refined by means of a supervised training phase, using labeled data and a reinforcement learning phase, based on a reward system for all correct answers.

The major breakthrough for transformers is "attention" (or self-attention), according to which the meaning of a word is understood only in its context. Transformers are able to focus attention on the most relevant word in a sentence based on the words that precede it. They are then able

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, & I. Polosukhin, " Attention is all you need ", 2017.

to understand that the same word can have a different meaning depending on the structure of the sentence, whether in a simple sentence, but also, this is an important breakthrough, in a complex sentence[4]:

1. Word: "bank"

- Simple Sentences:

She sat on the river bank and read a book.

→ “bank” = land alongside a river.

He deposited the cheque at the bank.

→ “bank” = financial institution.

Here, self-attention helps the model focus on surrounding words like river or cheque to infer the correct sense of bank.

- Complex Sentence:

After fishing all morning by the river bank, he stopped at the bank to withdraw some cash.

→ “bank” is used with two meanings in one sentence. The model uses different attention contexts for each occurrence.

2. Word: "light"

- Simple Sentences:

The box is very light.

→ “light” = not heavy.

She turned off the light before bed.

→ “light” = illumination.

- Complex Sentence:

Though the package was light and easy to carry, it still contained a powerful flashlight that emitted a very bright light.

→ The model distinguishes "light" as weight in the first clause and illumination in the second, guided by context like carry and bright.

3. Word: "bear"

- Simple Sentences:

He saw a bear in the forest.

[4] A complex sentence is a sentence that contains one independent clause and at least one dependent clause. The independent clause expresses a complete thought, while the dependent clause adds additional information but cannot stand alone. Understanding complex sentences helps in forming better sentence structures, improving clarity in writing, and expressing relationships between ideas effectively.

→ “bear” = animal.

I can’t bear this noise any longer.

→ “bear” = tolerate.

- Complex Sentence:

She couldn’t bear to look at the bear as it lumbered past the tent.

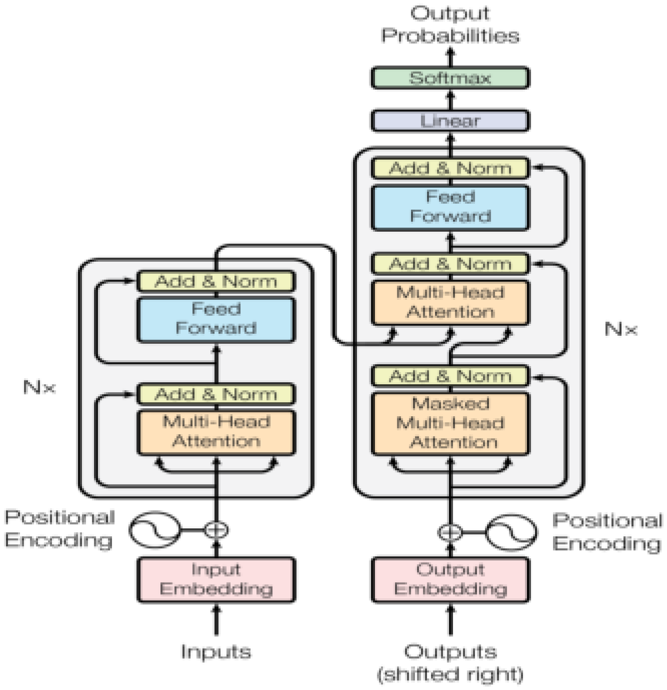
→ The first "bear" = tolerate; the second = animal. Contextual cues like look and lumbered past guide attention.

Why It Matters for Transformers ?

The breakthrough with attention is that the model doesn’t process each word in isolation. Instead, it dynamically weighs which parts of the sentence are most relevant for interpreting each word — even across long and syntactically complex structures. This allows models to handle ambiguity, nested clauses, and idiomatic usage with surprising accuracy.

Thus, GenAI learns words by considering the relative importance of words in their context, and does so over billions of examples. It learns sentence structure to build a statistical model capable of predicting the next word. This is how GenAI is able to produce new content (text, image, audio, video) distinct from its training data.

Schéma n° 2 : Transformative architecture



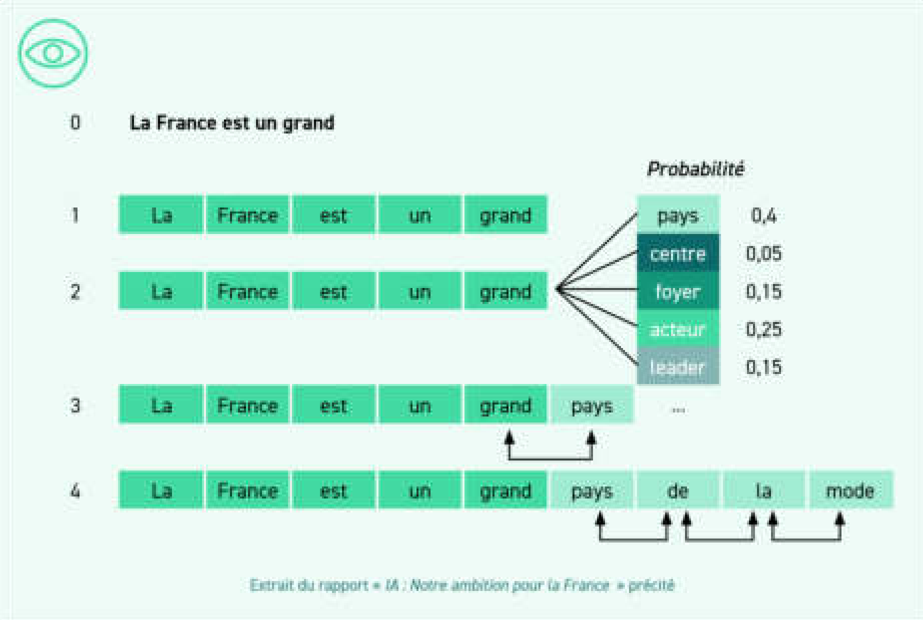
Source: Excerpt from the article, “Attention is all you need ”, Vaswani & alii, 2017.

1.2.2An architecture capable of predicting the next word to generate content

The GenAI produces (as output) a new text which, while being similar to the training data (as input), is not identical: the creation of the GenAI is the result of the statistical model born from its training.

GenAI is a next-word prediction system: the next word is the most likely word.

Schéma n° 3 : Transformative architecture


0 La France est un grand
Probabilité
1 La France est un grand
pays 0,4
centre 0,05
2 La France est un grand
foyer 0,15
acteur 0,25
leader 0,15
3 La France est un grand pays ...
4 La France est un grand pays de la mode
Extrait du rapport « IA : Notre ambition pour la France » précité

Source: Excerpt from the report, « Une ambition pour la France », March 2024[5].

While GenAI is similar to your smartphone's next-word prediction system when you type a message, it is infinitely more sophisticated because it is able to understand the importance of words in a sentence.

Technically, the transformer used by the GenAI operates according to a decomposition-recomposition process:

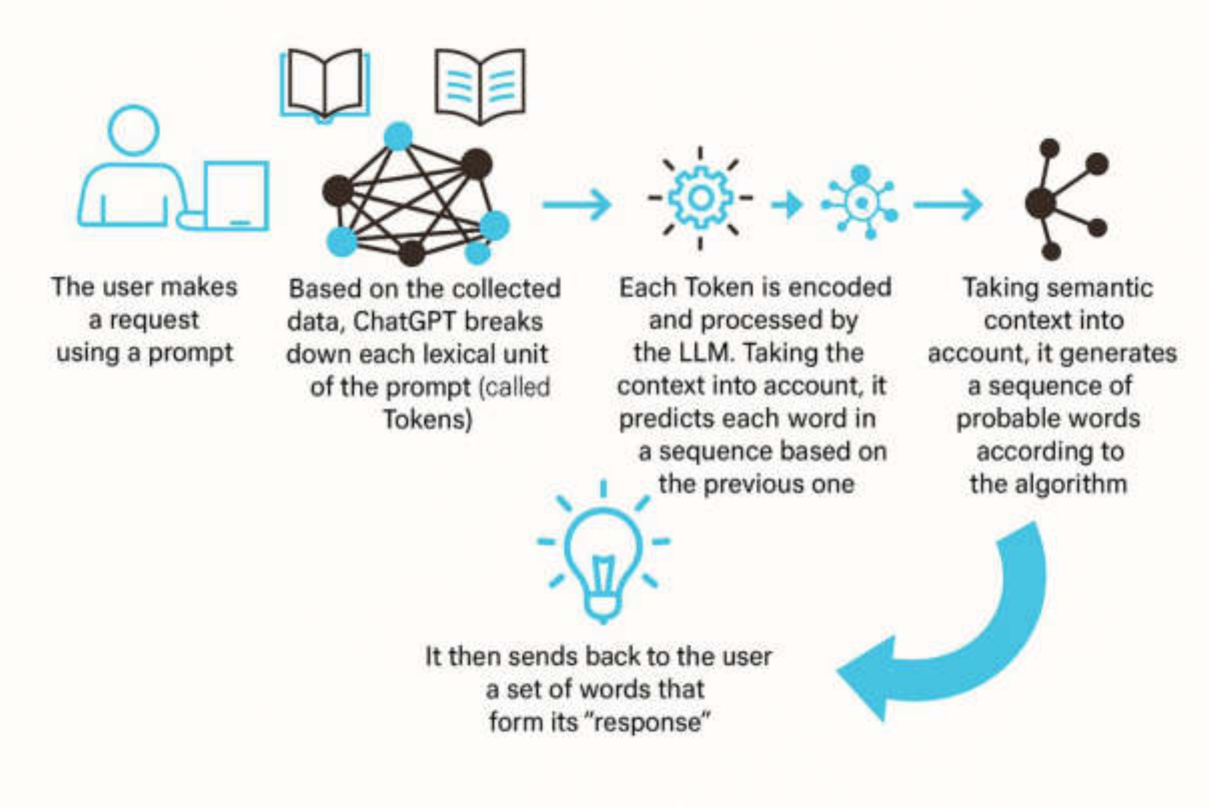
- encoding : the transformer breaks down the sentences of a text into *tokens* ;

[5] France | is | a | great → [model predicts next word with probabilities:]

Word	Probability
country	0.4
center	0.05
hub	0.15
actor	0.25
leader	0.15

- decoding : the tokens are then recomposed by the decoder which generates a new text (the model identifies, after analysis of the tokens, the different probabilities of following tokens and chooses the most probable lexeme.

Schéma n° 4 : The encoding-decoding process


The user makes a request using a prompt
Based on the collected data, ChatGPT breaks down each lexical unit of the prompt (called Tokens)
Each Token is encoded and processed by the LLM. Taking the context into account, it predicts each word in a sequence based on the previous one
Taking semantic context into account, it generates a sequence of probable words according to the algorithm
It then sends back to the user a set of words that form its "response"

Source: University of Bordeaux, mission to support teaching and innovation.

Some applications, such as Microsoft's Copilot or OpenAI 's ChatGPT , allow the user to configure, within certain limits, the model's probability level in generating the text:

- more creative: the system retains, in the choice of the next word, lower probabilities to have greater chances of creating new ideas;

- more precise: the system retains the greatest probability in the choice of the next word, for increased relevance;

- balanced: the system is set to achieve a balance between creation and precision.

It is important not to confuse the tokens that concern the way in which the sentence is digitally decomposed, according to its structure to be readable and processed by the model, with the parameters of the neural network that concern the way in which the model has learned to generate content (text, image, sound, video).

Finally, let us point out that GenAI is essentially a multimodal system that can transform tokens into text, images, sound or video.

IN SUMMARY

The GenAI is characterized by the following features, which should always be kept in mind when using it:

- GenAI is not a reliable source of information: GenAI is not an encyclopedia or a search engine, but a system for generating content without any assurance of its reliability;

- the GenAI does not understand the meaning of the words, images and sounds it generates;

- GenAI, because it is capable of generating text, gives the impression that it understands the meaning of the text it generates. This impression is misleading, and care must be taken not to humanize the machine;

- it is the human who gives meaning to the result generated by the GenAI. Thus, the meaning of the result generated by the GenAI, whether it is text, an image or a sound, is the responsibility of the human.

2 USE CASES AND RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE

GenAI offers benefits in terms of efficiency and productivity. This is the question of GenAI use cases. Since the technology is new, we will provide examples of use cases, without claiming to be exhaustive. GenAI also presents risks that you should be aware of when using it. These risks are well known, mainly concerning bias and hallucinations.

2.1 EXAMPLES OF USE CASES

GenAI is based on a foundation model (example: ChatGPT , Gemini, Claude, Mistral, etc.), that is, a general-purpose model characterized by an absence of specialization.

Since GenAI is new, it is necessary to familiarize yourself with it to determine the use cases suited to your needs.

Example of different GenAI use cases:

- check the spelling, grammar and style of a text, and suggest improvements;

- compare texts to highlight their common points and differences (examples: contract, clause, conclusion, etc.);

- write minutes of meetings or telephone calls;

- highlight the actions to be taken following a meeting;
- make summaries;
- make presentations in slide format;
- help with managing and writing emails (examples: softening the tone of a message, summarizing a particularly long conversation, prioritizing emails, etc.);
- help with teamwork on a project (example: summary of team members' emails every morning, etc.);
- translate texts into different languages;
- carry out or assist in research;
- write communication content (examples: newsletters or websites, publications on social networks, etc.);
- write draft emails or email templates, letters or letter templates, etc.;
- help with thinking and organizing work to write a control plan;
- help with thinking and finding new ideas;
- create images (examples: room communication, etc.);
- create videos (examples: room communication, etc.);
- help overcome writer's block.

POINT OF ATTENTION

The case of research by a GenAI.

Be careful not to mistake the GenAI for what it is not: a search engine. Indeed, GenAIs, when they are not integrated with search engines, do not necessarily have access to up-to-date information. Some of them are blocked in 2022 or 2023.

The case of image and video creation

In this case, we recommend using a GenAI specialized in image or video generation and checking whether this GenAI requires special commands to use its query system.

2.2 THE RISKS OF GenAI

Generative artificial intelligence offers significant benefits for financial jurisdictions, but it also carries risks that need to be well understood in order to manage them. Several types of risks can be identified.

2.2.1The risks of “imperfection”

This risk of imperfection (inadequacy, faultiness, deficiency, etc.) is not necessarily detected by simply reading the generated text.

Indeed, the generated texts are apparently plausible, coherent and therefore truthful – generative AI is, moreover, assertive – which helps to create trust in the user's mind, while they are false, either because the GenAI has invented a fact or because it has distorted the understanding of an existing fact.

The risk lies in the gap between the plausibility of what is generated and its veracity.

A study published in the journal Science Advances found that false information is more easily believed when written by GenAI models. The use of these tools may have worked against some lawyers who have argued against case law "hallucinated" by GenAIs. For example, a New York lawyer cited in his brief, in a personal injury liability case against an airline, 17 court decisions that did not exist: the lawyer had misused ChatGPT by using it as a search engine (which it is not) and by not verifying the results produced.

These risks are the biases and hallucinations of GenAI.

2.2.1.1 Biases

Biases can be distinguished into two categories:

2.2.1.1.1 Design biases

Design biases are those that are integrated directly into the GenAI statistical model during its training.

These biases appear when the data on which the GenAI is trained is likely to contain cultural stereotypes or discrimination against gender and minorities. These biases reflect a problem of representativeness of the database on which the GenAI is trained.

The Future of Privacy Forum, a think tank, has identified four main types of harm, or unintended consequences, that GenAI bias can cause:

– loss of opportunity: example: discrimination in hiring of certain groups (case: an algorithm tended to exclude women from job interviews);

– economic loss: example: an African American had his credit limit significantly reduced, from $10,800 to $3,800, because he shopped at places where customers statistically had poor credit repayment histories;

– social harm: example: the Facebook news feed offers the content that we are most likely to see, and to read ideas that confirm our own beliefs about the world, whether they are biased or not;

– loss of freedom: example: certain states in the United States, which authorize the use of a predictive algorithm for recidivism, are likely to lead to court decisions made partly on the basis of social or ethnic origin criteria.

In addition to these harms, there is the macro-level loss of sovereignty and influence (of continental law, for example), which can result from cultural biases.

These biases, integrated from training, can then influence the responses generated by the GenAI, introducing errors and risks of amplification and aggravation of discrimination.

2.2.1.1.2 The user's personal cognitive biases

Personal cognitive biases of the user also pose a risk.

We would like to point out once again that the GenAI is not a search engine. Indeed, GenAIs, when not integrated with search engines, do not necessarily have access to up-to-date information. Some of these are blocked in 2022 and 2023.

Bias involves the user formulating a query based on their own prejudices/expectations. For example, if the lawyer formulates a question that suggests guilt/innocence, the tool could provide an answer that reinforces this assumption rather than producing an objective result. The other risk in using GenAIs lies in intellectual laziness, not looking beyond the generated result.

To limit the risks associated with user-induced bias, it is important to train the members of each control team in the art of querying (see below, Use of the query and training).

2.2.1.2 Hallucinations of GenAIs

Hallucinations of GenAI also pose a significant risk to users of these systems.

In practice, hallucinations are erroneous results or imaginary facts that have no basis in the training data. It is the probabilistic system of predicting the next word that explains these hallucinations, and the GenAI could invent, for example, quotes or even case law.

Examples of hallucination, that is to say of a plausible content, apparently true, but which is in reality false:

◇ **1. Fabricated Academic Citations**

Prompt: “Give me sources for the theory of cognitive dissonance.”

Hallucinated Response:

Festinger, L. (1957). *The Dissonance Theory of Cognitive Dynamics*. Journal of Behavioral Psychology, 12(4), 321–334.

Why It’s a Hallucination:

While Leon Festinger *did* publish on cognitive dissonance, this specific journal and article title do not exist. The citation is fabricated but sounds highly academic and convincing.

◇ **2. Nonexistent Legal Cases**

Prompt: “What are some case laws that deal with AI copyright?”

Hallucinated Response:

In *Smith v. OpenAI Corp.*, 2022, the court held that AI-generated images could not be copyrighted because they lacked human authorship.

Why It's a Hallucination:

No such case exists as of June 2025. The case name and conclusion are plausible but completely made up.

◇ 3. Incorrect Medical Facts

Prompt: "Can you give a treatment for Lyme disease?"

Hallucinated Response:

Recent studies suggest that turmeric-based supplements can fully cure Lyme disease without antibiotics.

Why It's a Hallucination:

While turmeric has some anti-inflammatory properties, there is *no scientific evidence* that it cures Lyme disease. Standard treatment is antibiotics. This is a harmful medical hallucination.

◇ 4. Invented Historical Events

Prompt: "Tell me about early feminist movements in Asia."

Hallucinated Response:

In 1892, the "Bangkok Women's Liberation Front" led a major suffragette protest that resulted in voting rights for Thai women.

Why It's a Hallucination:

There's no historical record of such a movement or event at that time. It combines plausible elements (Thailand, feminism, suffragette era) but fabricates the details.

◇ 5. Misattributed Quotes

Prompt: "Give a famous quote by Albert Einstein about education."

Hallucinated Response:

"Education is not the learning of facts, but the training of the mind to think." — Albert Einstein

Why It's a Hallucination:

This quote is often attributed to Einstein, but there is no verifiable source linking it directly to him. It's a common misattribution.

2.2.2Other risks

Other risks may also be stated:

- risks linked to malicious use, such as the creation of content allowing a malicious person to carry out phishing, cyberattacks or create any form of illicit content;

- risks linked to the revelation of confidential information present in the training data;

- systemic risks, such as the dissemination of false information which, if disseminated en masse, can constitute a danger for democracy (examples: creation of false statements during election periods, creation of disinformation articles to manipulate public opinion, etc.);

- confidentiality risks linked to the reuse of data transmitted by the owner of the GenAI model: when the user transmits data to a GenAI, there is a risk of reuse of this data by the owner of the GenAI model, calling on the user of these models to be particularly vigilant when formulating their request, in order to ensure compliance with the protection of personal data and professional secrecy, and more generally the confidentiality of the data transmitted;

- risks to the protection of personal data: to function effectively, the GenAI requires massive amounts of personal data; this collection can lead to risks to the privacy of the persons concerned if the data is not correctly used and pseudonymized (see point on pseudonymization); in addition, risks to personal data may also appear if the GenAI is the target of cyberattacks (see above) or in the event of bias present in the training data (see above);

- risks for copyright protection: GenAI allows the generation of complex content, such as texts, photos, videos, music, often based on existing protected works; this use raises important copyright issues, such as questions relating to copyright infringement, because GenAI can use protected works without authorization, leading in particular to risks of counterfeiting or reproduction, or raises questions relating to the authorship of the works created by GenAI; the origin of the training data used lacks transparency, which leads to difficulties, in particular in terms of copyright.

IN SUMMARY

The GenAI presents interesting use cases that control teams will discover as they familiarize themselves with this new tool.

But it also presents risks that must be kept in mind:

- GenAI is not an encyclopedia nor is it a search engine;

- it can produce plausible content, apparently true, but which is in reality false: these are hallucinations;

- it includes design biases of which one must be aware (example: risk of discrimination, particularly in hiring) and cognitive biases of the user (example: risk of intellectual laziness);

- it presents a risk of loss of confidentiality, because all the query data is reused by the owner of the GenAI to train his model and when he claims that this is not the case, there is nothing to ensure it.

3 USING GENERATIVE ARTIFICIAL INTELLIGENCE IN YOUR PROFESSIONAL ACTIVITY

The use of GenAI in public administrations is profoundly transforming the sector, both in terms of efficiency, service personalization, and data management. However, these advances require in-depth ethical reflection and adequate regulation to ensure they benefit all citizens while respecting democratic principles and fundamental rights.

In terms of use within financial jurisdictions, we will focus our developments on GenAIs capable of generating text, a use case of interest primarily to control teams.

For financial jurisdictions, the use of a GenAI system requires increased vigilance to ensure its effectiveness and relevance, in compliance with legal and professional requirements, such as the protection of personal data and professional secrecy.

The first rule that applies to financial jurisdictions is to guarantee compliance with professional secrecy in all circumstances: control teams must never communicate data covered by professional secrecy to generative artificial intelligence, under penalty of sanctions (see below).

The objective of this first guide is to propose good practices and rules of vigilance in the use of this new tool, in the formulation of queries and in the exploitation of their results.

Terminological precision: GenAI is a learning model that requires an application for its use, the most common being the conversational robot or chatbot. In the remainder of this guide, this application will be referred to as what it is for a control team: a tool.

3.1 MAKE A REQUEST

3.1.1General recommendations

Here are some general guidelines for formulating your requests:

Define your objectives , that is, the expected result based on the characteristics of the GenAI system you are using and the result you expect:

- Generating text is not the same as generating an image or video. For example, Midjourney , which converts text to an image, assumes that every query begins with "/imagine";
- master the capabilities and limitations of language models to exploit them to the maximum: systems can give different results, because they have not been trained with the same training data, the same supervised or reinforcement learning or have different architectures:
- for example, generating communication content or an email template can be done with a general GenAI (trained on non-legal data);
- But to generate legal content, we recommend using a specialized GenAI trained on legal data so that training is done on diverse and representative datasets, thus ensuring comprehensive and balanced coverage of legal topics.

Be clear and concise:

- Get straight to the point by starting with a clear verb or term to describe the task to be accomplished (examples: “explain”, “compare”, “summarize”, “justify”, “research”, etc.); the longer the query text and the more complex the concepts, the greater the risk of receiving a poor quality response;
- keep in mind that GenAI systems do not handle conflicting notions in the same query well;
- If your query is complex and cannot be simplified, you can use punctuation to clarify it.

Be specific when:

- when requesting a summary, specify the key point(s) that the summary should cover;
- when requesting text generation, it is recommended to specify relevant themes, keywords or details;
- When asking a question, it is recommended to be as specific as possible: Example: "What are the most effective methods for reducing carbon emissions and mitigating climate change, and how have these strategies been implemented in different parts of the world?"

Include context in the query:

- contextualization is a situational setting of the GenAI; if you request the summary of an article, providing the author, the title of the article, the publication date can help improve performance;
- specify the capacity of the requester (example: if the need is to write a draft questionnaire containing external audit vocabulary, it will then be necessary to specify in the request that the author is an external auditor);
- specify historical elements (example: if you are asking for help with a control plan, specify the elements already carried out).

Specify the format of the text generated as a result :

- if you want bullet points, please specify;
- a short text of a few sentences;
- a text without technical words for better popularization;
- a text in a formal or more casual style.

Arbitrate between the level of creativity and freedom of the model:

- if you are looking for new ideas, give the model more creativity;
- the balance between constraints (a query that is too constrained does not produce a satisfactory result) and freedom: to find this balance, it is recommended to: start with more general queries that will be refined by successive iterations, encourage other interpretations of the same fact, which gives freedom to the GenAI, for example by reformulating the query.

It is worth noting that the first skill created by the emergence of GenAI is that of "Prompt Engineer ", a rapidly growing skill, that is, the ability to design and create queries to lead GenAI to perform, as efficiently as possible, a specific task. This requires a combination of technical and creative skills, and a good understanding of GenAI systems.

This is a crucial element of the added value of the magistrate or auditor. It is essential that they train to become the best "Prompt Engineers ." It is likely that GenAI tools will integrate prompt aids in the future, such as shortcuts to the most used prompts. In any case, members of the audit team will need to master the art of prompting to assess whether these shortcuts are relevant to their needs.

3.1.2 Specific recommendations for auditors

In a GenAI system, the chatbot is the interface to the LLM, a foundation model hosted in a cloud. Users access it via the internet.

Any request sent to an GenAI then results in the transmission of data to the company that operates the GenAI system, which, in the majority of cases, is American. Indeed, the global GenAI and cloud markets are dominated by American companies.

This transmission is accompanied by the retention and processing of this data by the company, the transmitted data sometimes being used to improve the models.

These companies are characterized by the opacity of the processing of the data transmitted to them.

However, SAI auditors are subject to strict obligations regarding confidentiality and data protection, under the provisions of GDPR, national legislation (the financial Juridictions code in France), professional standards and the Code of Ethics.

This transmission of information to a third party raises two difficulties: the first relates to the confidentiality of the data transmitted, which calls into question the professional secrecy of the auditors ; the second relates to the protection of personal data (GDPR). To resolve these difficulties, the auditors can use the tools made available to them by the digital department (for example ChatJF at the French Court of Accounts) or, alternatively or as a subsidiary measure, the anonymization or pseudonymization of personal data.

3.1.2.1 Professional Secrecy and the Protection of Personal Data at the State Audit Office of Georgia

In the execution of its constitutional mandate, the State Audit Office of Georgia (SAOG) operates as the supreme body of state financial and economic control. Its activities are governed by the Law of Georgia on the State Audit Office, which ensures its independence and delineates its responsibilities. According to Article 3 of this law[6], the SAOG is independent in its activities and is bound only by the law, prohibiting any interference or control over its operations unless expressly provided by law.

The SAOG is entrusted with the responsibility of conducting audits to promote the legal, efficient, and effective use of public funds. In fulfilling this role, it must access various forms of information, including those protected by confidentiality provisions. Article 5 of the Law on the State Audit Office mandates the SAOG to ensure that personal, state, official, and

[6] Law of Georgia on State Audit Office | სსიპ ”საქართველოს საკანონმდებლო მაცნე”.

commercial secrets are protected in accordance with Georgian legislation. Furthermore, audit results must remain confidential until the preparation of an audit report.

The processing of personal data by the SAOG is subject to the oversight of the Personal Data Protection Service of Georgia[7]. This body is responsible for monitoring compliance with data protection legislation and ensuring the lawfulness of data processing activities. The Service conducts audits, handles complaints, provides consultations, and informs the public about data protection matters.

In the context of internal investigations, the SAOG is authorised to obtain personally identifiable information on individual taxpayers by court order and pursuant to the rules set out in the Tax Code of Georgia. Audit staff are obliged to comply with confidentiality requirements concerning personal, public, corporate, and commercial information as stipulated by Georgian legislation.

Conversely, when engaging with third parties, the SAOG must ensure that it does not disclose any protected confidential information. Article 25 of the Law on the State Audit Office stipulates that while the SAOG may publish audit reports and other information on its activities, it must not disclose legally protected secret information related to the auditee, except in cases envisaged by Georgian legislation.

It is imperative that auditors refrain from transmitting any data relating to audited entities or individuals to generative artificial intelligence platforms, except for those internal tools expressly approved and provided by the SAOG's digital services department. Disclosing such information to external GenAI systems constitutes a potential breach of professional secrecy and may subject the auditor to disciplinary or legal sanctions.

3.1.2.2 How to use GenAI while respecting professional secrecy and the GDPR?

In practice, how do you ask the GenAI to make a summary, write a paragraph of a report, or check the spelling and grammar of a text?

In order to use the resources of generative artificial intelligence while protecting their data, auditors are advised to use local application (as ChatJF application at the French Court of Accounts).

An alternative approach is to practice anonymization or pseudonymization of data.

Anonymization makes data irreversibly unidentifiable. It removes the personal nature of the data, which excludes it from the scope of the GDPR.

[7] ACTIVITIES

Anonymization of personal data

A personal data anonymization process aims to make it impossible to identify individuals within datasets. It is therefore an irreversible process. Once this anonymization is complete, the data is no longer considered personal data and the requirements of the GDPR no longer apply.

By default, we recommend that you never consider raw datasets as anonymous. An anonymous dataset must necessarily result from an anonymization process that will eliminate any possibility of re-identification of individuals, whether by:

- individualization: it is not possible to isolate part or all of the records relating to an individual;
- correlation: the dataset does not allow two records relating to the same person or group of people to be linked;
- inference: it is impossible to deduce the value of an attribute of a person from information internal or external to the dataset.

These data processing operations usually involve a loss of quality in the resulting dataset. Opinion G29 on anonymization techniques describes the main anonymization techniques used today, as well as examples of datasets wrongly considered anonymous. It is important to note that there is no universal solution for anonymizing personal data. The decision to anonymize or not the data, as well as the selection of an anonymization technique, must be made on a case-by-case basis depending on the context of use and need (nature of the data, usefulness of the data, risks for individuals, etc.).

Pseudonymization is defined by the GDPR as a process of making personal data unattributable to a specific individual without additional information**.** [8]It replaces identifying data, such as a name, with non-identifying data (e.g., a number). This process is reversible, as it is possible to trace a person's identity using separate, protected information.

From a legal perspective, pseudonymized data remains personal data, subject to the GDPR, while anonymized data is exempt from this regulation, as it no longer allows an individual to be identified.

Pseudonymization of personal data

Pseudonymization represents a compromise between the retention of raw data and the production of anonymized datasets.

It refers to the processing of personal data in such a way that the data relating to a natural person can no longer be attributed without the use of additional information. The GDPR insists that this additional information must be kept separately and be subject to technical and organizational measures to prevent re-identification of the data subject. Unlike anonymization, pseudonymization is a reversible process.

[8] According to Article 4(5) of the GDPR, “pseudonymisation” is “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”.

pseudonymization process consists of replacing directly identifying data (surname, first name, etc.) in a dataset with indirectly identifying data (alias, number in a ranking, etc.) in order to reduce their sensitivity. They can result from a cryptographic hash of individual data, such as their IP address, user ID, or email address.

Data resulting from pseudonymization is considered personal data and therefore remains subject to the obligations of the GDPR. However, the European regulation encourages the use of pseudonymization in the processing of personal data. Furthermore, the GDPR considers that pseudonymization helps reduce risks for data subjects and contributes to compliance with the regulation.

To date, pseudonymization does not apply to all tasks that a GenAI might perform.

For example, pseudonymization cannot be applied to the production, live, of a report of an interview with a person being checked. If confidential information is exchanged during this interview, you will transmit this confidential information to the GenAI, which will prepare a report.

Therefore, it is appropriate to make appropriate use of the GenAI.

IN SUMMARY

The use of generative artificial intelligence (GenAI) in public administrations, particularly in SAI, improves the efficiency and personalization of services. However, this requires ethical and legal vigilance, especially regarding the protection of personal data and respect for professional secrecy. Control teams should avoid transmitting sensitive information to unauthorized GenAIs, such as consumer tools, and favor specialized secure solutions, anonymizing or pseudonymizing data if necessary.

3.2 EVALUATE, IMPROVE AND USE THE RESULT GENERATED

3.2.1General recommendations

It is essential to rigorously evaluate the results produced by generative artificial intelligence (GenAI) systems, in order to detect any errors or hallucinations, otherwise you may be held liable.

To do this, you must first identify hallucinations, i.e., errors produced by the system. You thus become both a verifier and a validator of the results generated by the GenAI.

It is essential to take a step back and maintain a critical mindset, because although machine-generated content may seem perfectly coherent, it may turn out to be erroneous. For example, the following examples should be kept in mind:

A New York lawyer cited 17 non-existent court decisions in his brief in an airline injury liability case;

In another case (Mark Walters v. OpenIA), ChatGPT "hallucinated" the meaning of a court decision by designating as guilty a person who was not.

We recommend using GenAI systems that list their sources so you can verify their accuracy. However, be warned: this does not protect against hallucinations.

It's also important to remember that GenAI systems are not search engines and may not have been trained on recent data. For example, some GenAIs are powered by information from 2022-2023.

The conditions of use of the various services also provide for notable exclusions of liability:

Copilot (free, general public version): “The Online Services are provided for entertainment purposes. They are not error-free, may not function as intended, and may generate incorrect information. You should not rely on the Online Services or use them for advice of any kind. You assume the risks associated with using the Online Services.”;

ChatGPT (free, general public version): “Given the probabilistic nature of machine learning, use of our Services may, in certain situations, result in Output Data that does not accurately reflect real-life people, places, or events.”

Finally, to improve the quality of the results obtained, iteration is essential. By proceeding with successive adjustments, you will gradually be able to refine the desired result. The more you use these tools, the more familiar you will become with their features and the more able you will be to optimize your performance.

3.2.2Specific recommendations for auditors

3.2.2.1 Transparency in the use of GenAI within financial jurisdictions

The use of generative artificial intelligence within a chamber should be carried out in complete transparency. Within a chamber, the presiding officer must ensure that the terms and purposes of using GenAI tools are clearly communicated to all of its staff, particularly with regard to their inherent risks and confidentiality risks. This includes training staff on the capabilities and limitations of GenAI (see below), as well as information on the security and confidentiality practices to be observed in this use. In other words, this communication must allow each member of the control team to know when and how GenAI tools can be used within the chamber and to understand the implications of their use.

In practice, this information can be communicated within an IT charter or an internal user guide, and during dedicated training sessions.

3.2.2.2 Transparency in the use of the GenAI towards those audited and the general public

SAIs must ensure transparency in the use of generative AI by making it public and specifying the tools and criteria used. Reports must detail the role of AI and include a clear summary for the public. While AI can analyze data, the final decision must be made by human agents, with prior verification of the generated results. Independent auditing and full documentation of processes must be carried out to ensure compliance. The use of AI must comply with personal data protection standards (GDPR), with sensitive information anonymized. Practices must be reassessed regularly, and annual reports must be published on the effectiveness of AI in controls. Finally, stakeholder consultation and public accessibility of results are essential to ensure transparency and trust.

3.3 TRAINING IN THE USE OF GenAI

Given the impact of artificial intelligence in our society, it is essential to train all audit teams in the use of GenAI tools.

This training, in that it allows you to understand and fully exploit the capabilities of these tools and to know how to appreciate their limits, presents several strategic advantages:

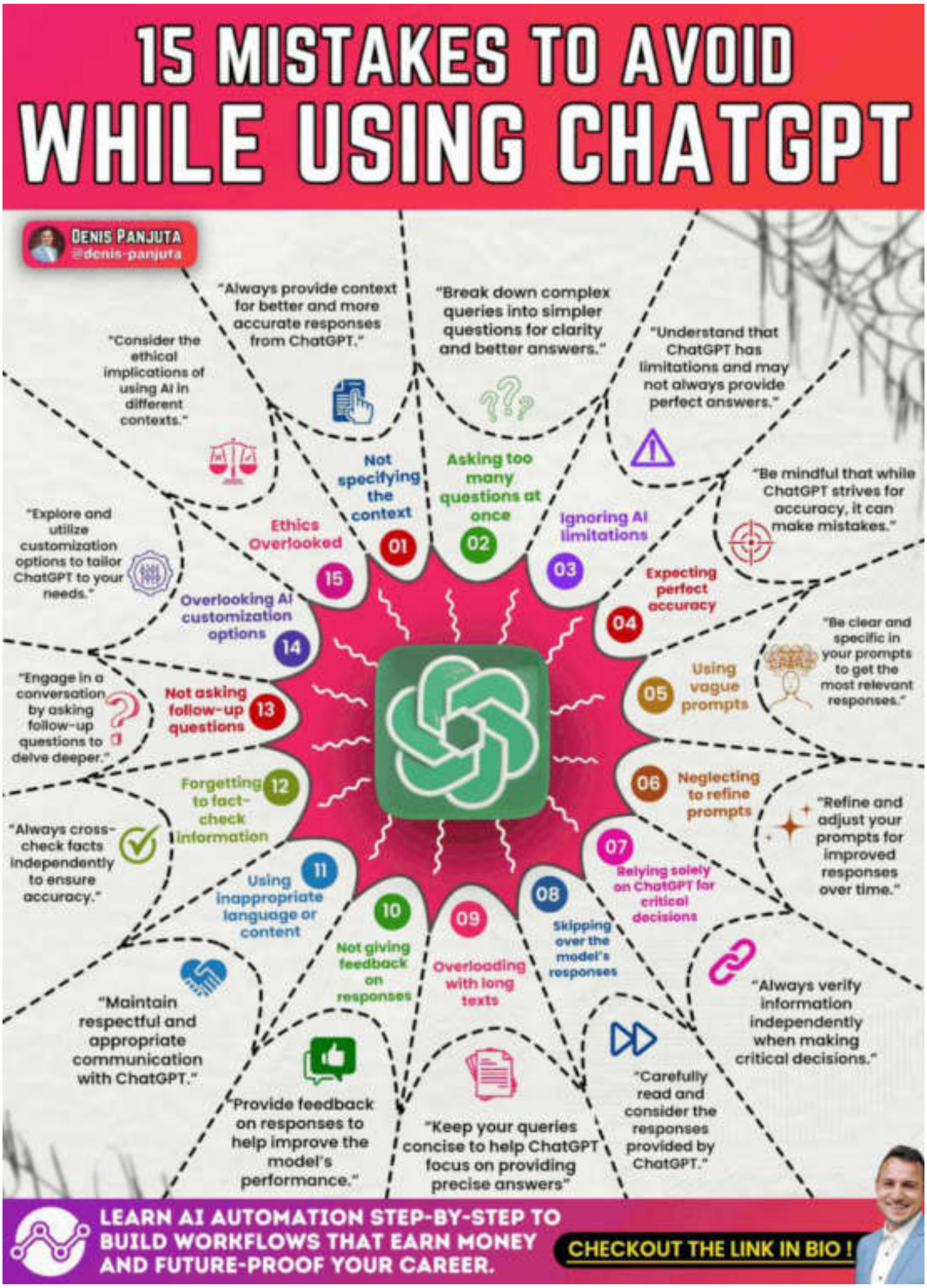
- Firstly, training can enable each auditor, and more generally each SAI, to gain in productivity; in fact, mastery of GenAI tools can enable SAIs to optimize practices by transforming time-consuming tasks (examples: legal research, document analysis, synthesis, etc.) into a faster and more efficient process;
- secondly, the training enables each member of the control teams to strengthen their ability to use and evaluate these tools in an informed and responsible manner. Training in “prompt engineering” or similar training appears essential in that it enables each member of the control teams to be able to formulate clearer and more relevant requests, thus improving the quality of the responses produced by the tool and optimizing the usefulness of these tools for financial jurisdictions. To another extent, this training also provides each member of the control teams with in-depth knowledge of the functionalities and limitations of the tool, thus enabling them to maintain a critical mind with regard to the result provided (see above), to identify any biases/hallucinations of the machine, and to ensure the reliability of the information obtained (role of verifier/validator of each member of the control teams with regard to the result provided by the tool);
- Finally, it is important to remember that all agents of financial jurisdictions must be trained in the use of these tools; in fact, each agent must be able to understand when and how to use the GenAI in carrying out their tasks, and this, in strict compliance with their obligations (professional, ethical, legal).

IN SUMMARY

Evaluating the results generated by generative artificial intelligence (GenAI) systems is essential to detect errors and avoid liability. Hallucinations must be identified, and critical thinking is required. It is recommended to use GenAI systems that disclose their sources, but this does not guarantee the absence of errors. Training is essential to understand the capabilities and limitations of the tools, optimize their use, and assess their reliability. Financial jurisdictions must ensure transparency in the use of GenAI, with clear documentation and independent audits to ensure compliance with standards, particularly regarding data protection.

4 APPENDICES

4.1 Mistakes to avoid when using ChapGPT



15 MISTAKES TO AVOID WHILE USING CHATGPT

DENIS PANJUTA
@denis-panjuta

01 Not specifying the context
"Always provide context for better and more accurate responses from ChatGPT."

02 Asking too many questions at once
"Break down complex queries into simpler questions for clarity and better answers."

03 Ignoring AI limitations
"Understand that ChatGPT has limitations and may not always provide perfect answers."

04 Expecting perfect accuracy
"Be mindful that while ChatGPT strives for accuracy, it can make mistakes."

05 Using vague prompts
"Be clear and specific in your prompts to get the most relevant responses."

06 Neglecting to refine prompts
"Refine and adjust your prompts for improved responses over time."

07 Relying solely on ChatGPT for critical decisions
"Always verify information independently when making critical decisions."

08 Skipping over the model's responses
"Carefully read and consider the responses provided by ChatGPT."

09 Overloading with long texts
"Keep your queries concise to help ChatGPT focus on providing precise answers"

10 Not giving feedback on responses
"Provide feedback on responses to help improve the model's performance."

11 Using inappropriate language or content
"Maintain respectful and appropriate communication with ChatGPT."

12 Forgetting to fact-check information
"Always cross-check facts independently to ensure accuracy."

13 Not asking follow-up questions
"Engage in a conversation by asking follow-up questions to delve deeper."

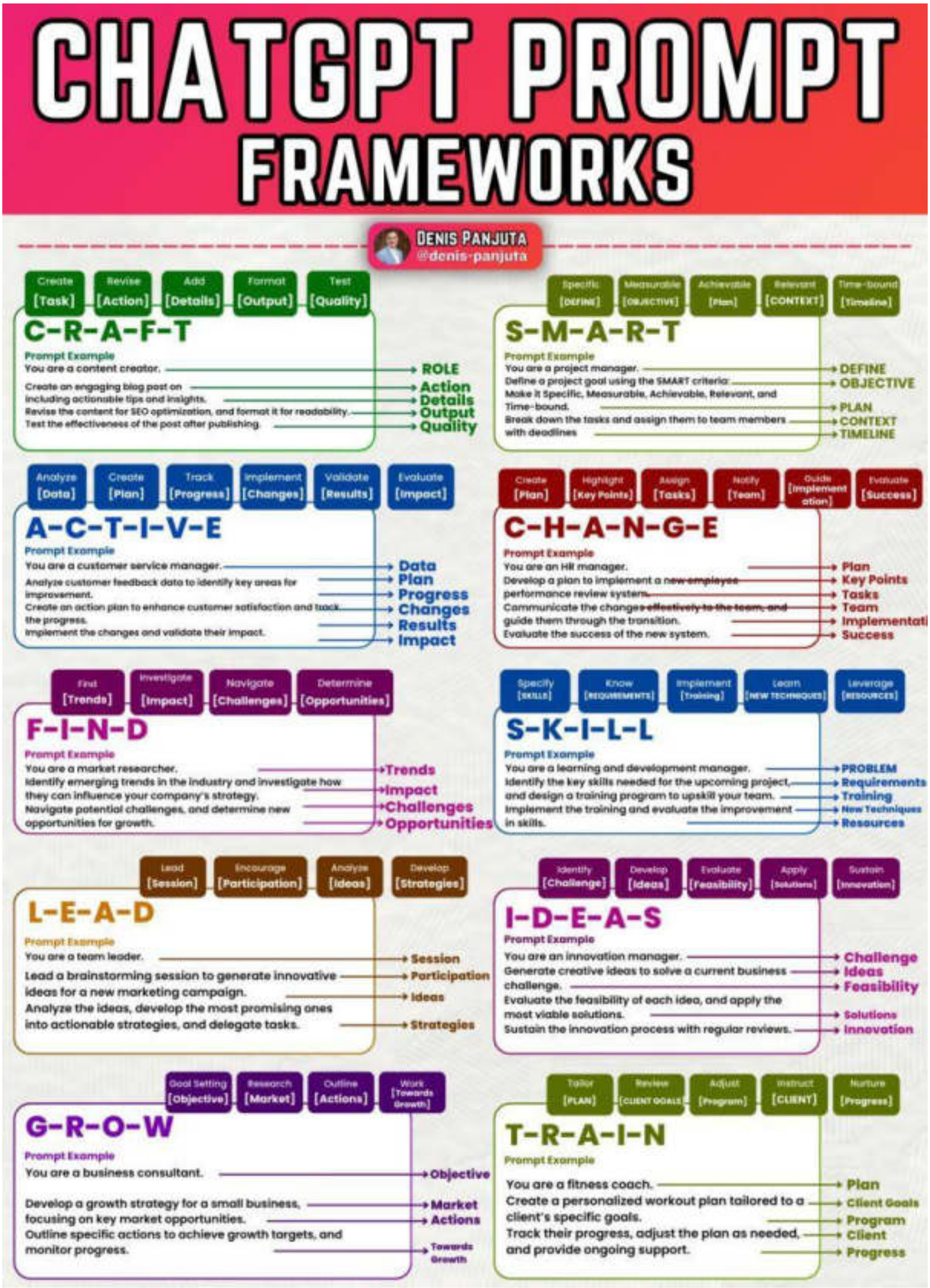
14 Overlooking AI customization options
"Explore and utilize customization options to tailor ChatGPT to your needs."

15 Ethics Overlooked
"Consider the ethical implications of using AI in different contexts."

LEARN AI AUTOMATION STEP-BY-STEP TO BUILD WORKFLOWS THAT EARN MONEY AND FUTURE-PROOF YOUR CAREER.

CHECKOUT THE LINK IN BIO !

4.2 Prompt engineering techniques



CHATGPT PROMPT FRAMEWORKS

DENIS PANJUTA
@denis-panjuta

Create [Task] · Revise [Action] · Add [Details] · Format [Output] · Test [Quality]

C-R-A-F-T

Prompt Example

You are a content creator. → ROLE

Create an engaging blog post on → Action

including actionable tips and insights. → Details

Revise the content for SEO optimization, and format it for readability. → Output

Test the effectiveness of the post after publishing. → Quality

Specific [DEFINE] · Measurable [OBJECTIVE] · Achievable [Plan] · Relevant [CONTEXT] · Time-bound [Timeline]

S-M-A-R-T

Prompt Example

You are a project manager. → DEFINE

Define a project goal using the SMART criteria: → OBJECTIVE

Make it Specific, Measurable, Achievable, Relevant, and Time-bound. → PLAN

Break down the tasks and assign them to team members → CONTEXT

with deadlines → TIMELINE

Analyze [Data] · Create [Plan] · Track [Progress] · Implement [Changes] · Validate [Results] · Evaluate [Impact]

A-C-T-I-V-E

Prompt Example

You are a customer service manager. → Data

Analyze customer feedback data to identify key areas for → Plan

improvement. → Progress

Create an action plan to enhance customer satisfaction and track → Changes

the progress. → Results

Implement the changes and validate their impact. → Impact

Create [Plan] · Highlight [Key Points] · Assign [Tasks] · Notify [Team] · Guide [Implementation] · Evaluate [Success]

C-H-A-N-G-E

Prompt Example

You are an HR manager. → Plan

Develop a plan to implement a new employee → Key Points

performance review system. → Tasks

Communicate the changes effectively to the team, and → Team

guide them through the transition. → Implementati

Evaluate the success of the new system. → Success

Find [Trends] · Investigate [Impact] · Navigate [Challenges] · Determine [Opportunities]

F-I-N-D

Prompt Example

You are a market researcher. → Trends

Identify emerging trends in the industry and investigate how → Impact

they can influence your company's strategy.

Navigate potential challenges, and determine new → Challenges

opportunities for growth. → Opportunities

Specify [SKILLS] · Know [REQUIREMENTS] · Implement [Training] · Learn [NEW TECHNIQUES] · Leverage [RESOURCES]

S-K-I-L-L

Prompt Example

You are a learning and development manager. → PROBLEM

Identify the key skills needed for the upcoming project, → Requirements

and design a training program to upskill your team. → Training

Implement the training and evaluate the improvement → New Techniques

in skills. → Resources

Lead [Session] · Encourage [Participation] · Analyze [Ideas] · Develop [Strategies]

L-E-A-D

Prompt Example

You are a team leader. → Session

Lead a brainstorming session to generate innovative → Participation

ideas for a new marketing campaign. → Ideas

Analyze the ideas, develop the most promising ones

into actionable strategies, and delegate tasks. → Strategies

Identify [Challenge] · Develop [Ideas] · Evaluate [Feasibility] · Apply [Solutions] · Sustain [Innovation]

I-D-E-A-S

Prompt Example

You are an innovation manager. → Challenge

Generate creative ideas to solve a current business → Ideas

challenge. → Feasibility

Evaluate the feasibility of each idea, and apply the

most viable solutions. → Solutions

Sustain the innovation process with regular reviews. → Innovation

Goal Setting [Objective] · Research [Market] · Outline [Actions] · Work [Towards Growth]

G-R-O-W

Prompt Example

You are a business consultant. → Objective

Develop a growth strategy for a small business, → Market

focusing on key market opportunities. → Actions

Outline specific actions to achieve growth targets, and

monitor progress. → Towards Growth

Tailor [PLAN] · Review [CLIENT GOALS] · Adjust [Program] · Instruct [CLIENT] · Nurture [Progress]

T-R-A-I-N

Prompt Example

You are a fitness coach. → Plan

Create a personalized workout plan tailored to a → Client Goals

client's specific goals. → Program

Track their progress, adjust the plan as needed, → Client

and provide ongoing support. → Progress

4.3 Glossary

The following glossary has been established by Fan Yang, Jake Goldenfein, and Kathy Nickels, in collaboration with the ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) and the Office of the Victorian Information Commissioner (OVIC).

GenAI Concepts - ADM+S Centre

This page is also available in PDF format.

Generative AI or GenAI

Generative AI or GenAI are both short for generative artificial intelligence. These are software systems that create content as text, images, music, audio and videos based on a user's 'prompts'.

Types of GenAI models in the market

Prompt and prompt engineering

A prompt is an instruction, query, or command that a user enters into a **GenAI** interface to request a response from the system.

Read more

Machine learning

Machine learning (sometimes seen as ML) is a set of techniques for creating algorithms so that computational systems can learn from data.

A machine learning algorithm is a set of rules or processes that helps an AI system to perform specific tasks, such as finding patterns in data or making predictions based on inputs. In this way, the model's behaviour reflects the data or the learning experience.

Higher quality data helps algorithms improve their accuracy in various tasks, such as recognising faces in photos, predicting the weather or recommending products to buy.

Read more

Large language models (LLMs)

Large language models (LLMs) are data transformation systems. They are trained with large numbers of parameters, which are numerical values that developers adjust to shape the inputs and outputs of AI models. When a user inputs a prompt, the model generates text content in response.

Read more

Knowledge cut-off date

The knowledge cut-off date of a **GenAI** model is the date when the training data for a specific **LLM** was last updated. It defines the limitations of the model's understanding and knowledge.

Read more

Chatbot

Chatbots are popular applications of **GenAI** and **LLMs**. A chatbot is a computer program that interacts with humans through natural language conversations. Some chatbots use LLMs to generate content according to user inputs.

Chatbot functionality is often embedded in customer services (e.g. banking, tax, logistics, ecommerce). Chatbots might be **fine-tuned** on an organisation's private datasets to answer specific queries.

Read more

General purpose AI (GPAI)

General purpose AI is a new paradigm where very large AI models become competent at a wide range of tasks, often without substantial modification or **fine-tuning**. General-purpose AI systems are often **LLMs**.

Read more

Foundation model

Foundation models, sometimes called **general-purpose AI systems**, provide a basis for future modification and **fine-tuning** for specific tasks. They are trained on vast amounts of data at scale, including images, text, audio, video and other data types like 3D models, with less emphasis on data quality so that they can be adapted to a wide range of downstream tasks.

Read more

Frontier model

Frontier models are larger than foundation models, with more parameters. They are potentially much more capable than existing models and raise additional **safety** risks particularly when developed for critical applications such as health, social welfare, defense and military

Read more

Transformer architecture

Transformer architecture makes **LLMs** possible. Put simply, transformers convert text. This type of architecture was proposed in 2017 by eight authors from Google in the paper **Attention is all you need**.

Read more

Transfer learning

Transfer learning is a model's ability to apply information about one situation to another. In this way, the model builds on its internal knowledge.

In **GenAI**, products must adapt and work well in different situations. It's like teaching a model skills in one area and then letting the model use those skills in other areas too. In this way, models can learn from one thing and apply it to many different tasks, making them versatile and useful.

Read more

Open-source and closed-source LLMs

Open-source LLMs have publicly accessible source code and underlying architecture, allowing developers, deployers, researchers and enterprises to use, modify and distribute them freely or subject to limited restrictions.

Closed-source LLMs have proprietary underlying source code and architecture. They are accessible only under specific terms defined by their developers.

Open-source LLMs and closed-source LLMs can be accessed and deployed via application programming interfaces (APIs), which are sets of rules or protocols that allow two software programs to communicate with each other to exchange functionality. AI developers, deployers or users can integrate data, services and functionalities from other applications through provided APIs, rather than developing them from scratch. The following table presents different features of open-source LLMs and closed-source LLMs.

Comparison of Open-Sourced LLMs and Close-Sourced LLMs

Token

A token is the smallest unit of data used by GenAI systems.

For text-based models like GPT, a token can be a word, part of a word, punctuation marks, spaces or other elements of the text. For image-generating AI models like DALL-E, a token is a pixel of the image. For audio-based AI models like MusicLM, a token might represent a short sound segment.

Tokens allow AI models to understand, memorise and generate meaningful responses. They play a vital role in memory capacity, determining how much information the AI model can recall.

Read more

Reinforcement
learning from human feedback

(RLHF)

Reinforcement learning from human feedback (RLHF) resembles the human learning process. When we learn a skill our teacher/instructor says things like ‘well done’ if we do something right or ‘let’s try it again and differently’ to improve our skill.

Read more

Diffusion models

Diffusion models are used for AI image generation. They work by destroying training data (by adding visual noise) and then learning to recover the data by reversing the noise.

Read more

Inference

AI inference is the process of applying trained **machine learning** models to new, unseen data to derive meaningful predictions or decisions. When users give a GenAI system a prompt, the computational system used to produce the output is called inference. The energy required for inference is much lower than training a model but is still significant and is a large part of the cost of using a GenAI system.

Read more

Datasets

LLMs are trained from different types of datasets. The material in datasets may be protected by intellectual property or information privacy laws. Organisations like **Common Crawl** scrape the public internet periodically and create gigantic datasets that can be used for model training.

Read more

Data licensing

Entities that control large amounts of text, visual, musical, code or video content may license it to AI firms for a fee. Datasets may also come with licensing stipulations that dictate what others can do with that dataset, for instance whether they can share it, modify it or use it for commercial purposes.

Read more

Developer or Dev

AI developers belong to a broader category of programmers and engineers. They write code and algorithms to enable machines to perform tasks that normally require human intelligence. They build models in-house by training on a mix of public and private data for various applications, from chatbots and virtual assistants to self-driving cars.

In emerging AI regulations, developers are responsible for creating AI software products and complying with regulations.

Read more

Data labelling and annotation

Data labelling and annotation describe the process of tagging or labelling text, images, audio and video data for AI training.

Comparison of data annotation and labelling

Supply chains

An AI supply chain refers to the process of creating, sourcing and integrating the components needed to **develop** and **deploy** AI systems or products.

View infographic

Read more

AI libraries

AI libraries include pre-written codes and algorithms for common AI tasks, such as data preprocessing, model training and evaluation.

Read more

Machine learning environments

Machine learning environments are where machine learning models are built, trained and deployed. Provided by the big tech corporations like Amazon and Microsoft, these environments contain fully managed infrastructure and tools to enable reproducible, auditable and portable machine learning workflows across different environments. Deployers can run training scripts or host service deployments, often referred to as computer targets, within these environments.

Read more

Fine-tuning

Fine-tuning happens during model development and deployment. It involves modifying a trained AI model with a smaller, targeted fine-tuning dataset. Fine-tuning maintains the original capabilities of a pre-trained model while adapting it to suit more specialised use cases.

Read more

Deployer or ML/ops

While **AI developers** create the AI system, AI deployers make them available for use in realword applications. In some cases, the distinction between a developer and a deployer is minimal. However, deployers may be the organisations that make AI tools available to others.

Read more

User

Users are the individuals or organisations that interact with AI technology through text, voice, images or other inputs. Users can also be anyone whose rights and activities are influenced by AI system outputs.

Read more

Risks

AI regulation is developing as risk-based regulation. This approach imposes obligations depending on the risk level of the system. Most regulatory proposals have three risk categories: low risk, medium risk and high risk.

The measurement and mitigation of AI-generated risks requires **human oversight**, **AI auditing**, **transparency** and **explainability**.

Read more

Privacy and data protection

The wide use of AI systems presents challenges to privacy. The datasets used to train **foundation models** and other machine learning systems often include personal information without consent. There have been cases where chatbots reproduce personal information from training data in response to prompts. Privacy and data protection laws impose obligations on entities that collect and process **personal information**.

Read more

Hallucination

Hallucination refers to AI models making up facts to fit a **prompt's** intent. When an **LLM** processes a prompt, it searches for statistically appropriate words, not necessarily the most accurate answer. An AI system does not 'understand' anything, it only recognises the **most statistically likely answer**. That means an answer might sound convincing but have no basis in fact. This creates significant risks for organisations that rely on chatbots to give advice about products or services because that advice might not be accurate.

Read more

Safety

In AI regulation, AI safety means that AI should be designed, developed and used in a way that is human-centric, trustworthy and responsible (see **The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023**). The main concern of AI safety is to make sure that AI systems are developed, deployed and used in ways that align with the norms and values of the broader public or specific user groups.

Read more

Transparency

AI transparency is a fundamental regulatory and ethical requirement; however, transparency has different meanings for different stakeholders (e.g. **developers**, **deployers**, users, regulators). Transparency represents how well users, regulators and other stakeholders understand how the system operates and how it was built. This is important for ensuring trust in AI outputs. Transparency is related to the **openness** and **explainability** of AI systems.

Read more

Copyright

Copyright law typically applies to original works of authorship, such as literary works, artistic works and computer programs. Copyright material has been included in AI training datasets without permission or rights clearance.

Read more

Data governance and security

Data governance refers to the system of rules for management and control of data.

Data governance requirements may involve compliance with laws, standards or ethical guidelines to ensure data within an organisation is collected, stored and used appropriately.

Within an organisation, data governance responsibilities include ensuring data quality (i.e. data is accurate) as well as data security (i.e. data is appropriately encrypted and protected).

This may mean additional care when using organisational data to fine-tune a GenAI system or when using organisational or sensitive data as part of a prompt.

Data life cycle management refers to the processes for managing data from creation to disposal, including data archiving and removal.

Read more

AI standards

International AI standards serve as a global governance mechanism to help achieve AI policy goals. Standards organisations create standards through stakeholder input and cover various technical and regulatory dimensions of AI systems. Compliance with standards may be required for AI products to be market-worthy; procurers may also require compliance with specific standards. In the EU system, if developers comply with standards, then they likely comply with their obligations under the **EU AI Act**.

Read more

Guardrails

AI guardrails are predefined **standards**, limitations and operational protocols to prevent AI systems from making decisions or taking actions that could lead to harmful or unintended consequences.

Guardrails are often sets of algorithms or rules that filter and edit inputs (i.e. prompts) and outputs of GenAI systems to ensure that outputs comply with legal and safety requirements. These guardrails are designed to prevent outputs from breaching copyright, produce political misinformation, create biased or discriminatory information or generate hate speech. This has proved very difficult to achieve in practice but remains an important aspect of AI system design.

Read more

Regulatory sandbox

Regulatory sandboxes are software testing environments where businesses can conduct limited testing of innovations and test their regulatory status.

This allows businesses to receive feedback from experts, including regulators, investors, innovators and other stakeholders, regarding the potential and viability of the innovation.

Read more

Human oversight

Human oversight is the requirement that human actors oversee the output of AI systems to ensure that systems create accurate and accountable results. There are different forms and degrees of human oversight, depending on the context and purpose of the AI system. Human oversight is usually required in high-risk systems.

Read more

AI auditing

AI auditing involves **humans** – normally researchers, programmers and regulators – looking closely at AI systems to evaluate risk and ensure AI systems act fairly and safely, and comply with relevant laws, regulations and ethical standards.

Different AI auditing methods have different advantages and limits:

- Technology-oriented audits focus on the properties and capabilities of AI systems.
- Process-oriented audits focus on technology providers’ governance structures and quality management mechanisms.

Some auditing methods may be simple compliance checkboxes; others may be comprehensive assessments of how AI systems might affect users and other stakeholders.

Read more

Explainable AI (XAI)

Explainable AI or XAI is a set of tools, techniques and algorithms designed to produce highquality interpretable, intuitive, human-understandable explanations of AI decisions. Many emerging AI regulations require some degree of explanation for higher risk system outputs.

The goal is not to prioritise complete explainability over performance or vice versa. Organisations should disclose the limits of transparency in the system to find a balanced approach that considers the risks and benefits of each AI application, while also considering human and environmental implication.

Read more

Post-market surveillance/monitoring

Post-market surveillance refers to monitoring the ongoing performance and **safety** of an AI product or service after it has been released to the market.

Read more

Floating-point operations per second (FLOPS)

Floating-point operations per second (FLOPS) in the context of computing and artificial intelligence are often used to measure the processing power or performance of hardware devices like CPUs (central processing units) and GPUs (graphic processing units).

Sometimes **frontier models** are defined by FLOPS, and some regulatory proposals have stricter compliance obligations for systems that surpass FLOPS thresholds.

Higher FLOPS values indicate higher computational power, more complex AI models, faster data processing, better graphic performance and higher energy consumption

Read more

Ecological

Training, deploying and using AI systems contribute to the global CO2 emissions. Typically, more powerful AI models require more energy. The servers that power AI models also generate considerable heat and are often water-cooled. The amount of water needed to train an AI model is immense. A team of researchers **disclosed** that "training GPT-3 in Microsoft's state-of-the-art U.S. data centers can directly evaporate 700,000 liters of clean freshwater, but such information has been kept a secret." And running GPT-3 inference for 10-50 queries **evaporate** 500 mililitres of water depending on when and where the model is hosted.

LLMs are among the biggest machine learning models, spanning up to hundreds of millions of parameters, requiring millions of GPU (graphic processing units) hours to train and emitting carbon in the process.