The Generative AI Playbook

We told you. Generative AI is here to stay.

We released our previous white paper (”How generative AI will disrupt Belgian businesses”) in January 2023. Our promise came true: this technology has definitely disrupted our business landscape and beyond. It’s time to revisit the technology and its implications, as much has happened since then.

To start this white paper on a **more personal note**, the technology has definitely disrupted my life in a way I could have never imagined. Since January, I have shared my knowledge about generative AI on television (VRT Laat, Play4, Kanaal Z), on the radio (Studio Brussel, Radio 1), and in the paper (De Tijd, De Morgen, VRT NWS). We have given countless keynotes and workshops for our clients (okay, we counted: 55 to date). On average, we provided two keynotes/masterclasses each week. It has been a ride, to say the least, especially since the landscape is evolving extremely rapidly. Preparing a keynote a week in advance meant it would be outdated by the time we had to give it. We edited our slides almost daily as news would pop up like mushrooms.

So, now, we have decided to write a successor for our previous white paper, by combining all the knowledge we have gained since releasing the first. We go beyond the basic notions of generative AI and dive deeper into the technical advancements, discuss new guidelines, and offer strategies for successful implementation.

As we release this white paper in September 2023, note that this information will probably be outdated in the blink of an eye — especially in this rapidly evolving landscape. Nevertheless, we hope to provide a guideline that will help you make more informed decisions when implementing this technology into your organization. Any questions? Just ask.

Enjoy the read!

Michiel Vandendriessche
Co-founder Raccoons

In our previous white paper, we discussed the potential of generative AI since the technology was new to the broader public. We examined its influence on Belgian businesses and explained creativity was no longer solely the domain of the human mind, but it is now shared with artificial intelligence. By offering basic definitions of these AI models and showcasing diverse applications for, e.g., marketing, HR, and even media, we aimed to underline the technology's role in the future.

Today, as expected, the ever-evolving landscape of generative AI has experienced some significant shifts. The introduction of ChatGPT has been one of the most important developments of the previous year, and many updates in the field have followed. These **new developments, more advanced algorithms, and a broader understanding of the opportunities and challenges** have triggered us to write a successor for our first white paper. Our promise? More depth, more insights, more guidance.

We are revisiting the topic, diving deeper, and offering a fresh perspective. After giving countless keynotes and experimenting with the technologies, we are ready to go beyond the hype and offer actual value. We could call this white paper a v2.0, an update, but it’s more than that. It’s a **comprehensive guide** — a playbook — tailored for CEOs, CTOs, managers, forward-thinking leaders. Anyone with a remote interest in incorporating the technology in their business or life.

We go beyond the basic notions of generative AI and dive deeper into the technical advancements, discuss new guidelines, and offer strategies for successful implementation. We end the white paper with a glimpse of the future and some recommendations.

1. Evolution of Generative AI

Since January, the landscape of generative AI has been anything but static. It has moved at a remarkable speed, and it seems like its developers took no vacation days. The rapid progression of the technology is fueled by **breakthroughs** that have expanded the capabilities of generative systems.

But first, let's recap. As we have mentioned in our previous white paper, generative AI is a subset of artificial intelligence that is focused on the ability of machines to generate new and unique outputs based on a set of input data. This can include creating **images, videos, music, and text**.

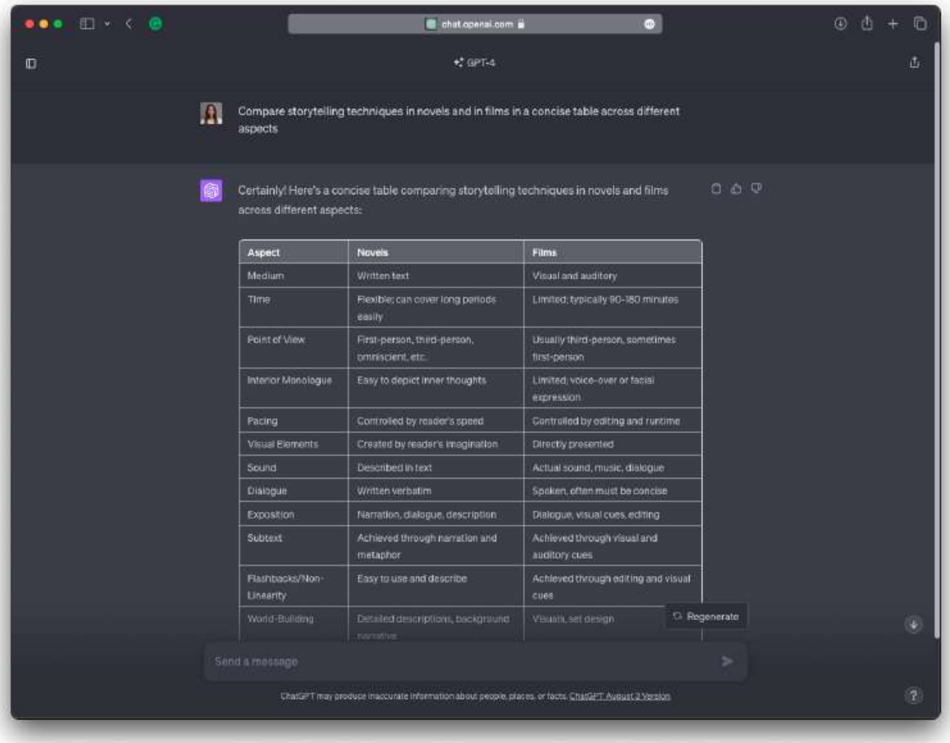
To structure this chapter, we will dive into the evolution of these four areas: text, images, audio/music, and video. We will discuss the most significant improvements and important tools that have propelled the technology even more.

1.1 Text

In the domain of Natural Language Processing (NLP), generative models are used to perform a wide range of tasks such as:

- **Text generation**: Generating new text that is similar to a given input text, such as writing a story or an article.
- **Text summarization**: Generating a short summary of a given text, such as a news article or a book.
- **Dialogue systems**: Generating responses for chatbots or virtual assistants that can carry out a conversation with a human.

And many more. Thanks to ChatGPT, we have seen remarkable results for all of the above, and as with every technology, continuous refinement is key. Whereas the first version of ChatGPT was based on GPT-3.5, we now can use GPT-4, which is more creative, reliable, and able to handle much more nuanced instructions than GPT-3.5. For instance, GPT-4's ability to reason over text is mind-blowing. You can actually ask it questions about any given text and it will analyze, interpret, and understand immediately.



Compare storytelling techniques in novels and in films in a concise table across different aspects

Certainly! Here's a concise table comparing storytelling techniques in novels and films across different aspects:

Aspect	Novels	Films
Medium	Written text	Visual and auditory
Time	Flexible; can cover long periods easily	Limited; typically 90-180 minutes
Point of View	First-person, third-person, omniscient, etc.	Usually third-person, sometimes first-person
Interior Monologue	Easy to depict inner thoughts	Limited; voice-over or facial expression
Pacing	Controlled by reader's speed	Controlled by editing and runtime
Visual Elements	Created by reader's imagination	Directly presented
Sound	Described in text	Actual sound, music, dialogue
Dialogue	Written verbatim	Spoken, often must be concise
Exposition	Narration, dialogue, description	Dialogue, visual cues, editing
Subtext	Achieved through narration and metaphor	Achieved through visual and auditory cues
Flashbacks/Non-Linearity	Easy to use and describe	Achieved through editing and visual cues
World-Building	Detailed descriptions, background narrative	Visuals, set design

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT August 3 Version

In addition to OpenAI's improvements and new models, other big tech players have come out with their reaction to OpenAI's domination. Microsoft began rolling out Bing Chat in February 2023, powered by the Microsoft Prometheus model, which has been built on top of OpenAI's GPT-4 foundational LLM. Moreover, Google refined its PaLM (2) model to launch Bing-counterpart

Bard and power its Search Generative Experience (SGE) and Google Duet. Moreover, Meta has open-sourced LLaMA, and many other open source models have seen the daylight. We are expecting many more alterations and upgrades over the next few months.

The advancements in text generation over the past months are **not just about improved writing capabilities** but about **deeper understanding and contextual relevance**. With giants like Microsoft, Google, and Meta pushing boundaries, it's evident that generative AI in the realm of text is an integral tool in businesses and beyond.

1.2 Images

Regarding images, the advancements made over the past years have been nothing short of revolutionary. From merely enhancing existing images to creating realistic visual content, the realm of image generation has been expanding rapidly. Key tasks of generative models include:

- Image synthesis: Generating new images that are similar to a given input image, such as creating a new picture of a person's face or a landscape.
- Super-resolution: Generating high-resolution images from low-resolution images, such as increasing the resolution of an image taken with a low-quality camera.
- Image-to-image translation: Translating an image from one domain to another, such as converting a picture of a day scene to a night scene or a sketch to a photograph.
- Text-to-image generation: Generating high-resolution images from textual inputs (prompts).

We have seen a rise in text-to-image models starting in 2022 — and once again, OpenAI initially was a dominant player with its DALL-E (2). However, competitors quickly approached and exceeded the quality of the generated images. For instance, Midjourney's quality is mind-blowing, and other open source models like Stable Diffusion (StabilityAI) have revolutionized the field.

To prove that this technology actually speeds up workflows, especially for creatives, Adobe has released Firefly. Now, you can easily alter pictures by selecting a part of your image and spelling out exactly what you want the AI model to do. The results? Stunning — all while shaving off hours of manual work.

Midjourney

1.3 Audio and music

Audio and music generation now show capabilities that were thought to be years away at the start of 2023. Let's look at what we're capable of today. In the domain of audio, generative models are used to perform a wide range of tasks, such as:

- **Voice cloning**: Generate a text-to-speech engine based on someone's voice
- **Music generation**: Creating a new piece of music from scratch, such as a new classical piece or pop song
- **Music style transfer**: Applying the style of one piece of music to another, such as converting a rock song to a classical piece

Last year, you probably would have hired a voice actor when you needed to clone someone's voice to make a compelling deepfake. Now, you can easily upload 30 seconds of their voice in ElevenLabs and receive an amazing voice clone in mere seconds in more than 30 languages (even Dutch!).

Moreover, the possibility of crafting **entirely new compositions from scratch** — be it a soulful classical symphony or a catchy pop tune — has become more accessible. You can now generate instrumental snippets through multiple apps, and we ourselves have dabbled in music style transfer to test out the possibilities. As a result, we have successfully covered songs with the voices of world-class pop stars (TikTok and YouTube are full of them by now), and we expect this to get easier and easier — which raises a few concerns in light of authenticity, rights, and the essence of human creativity.

1.4 Video

The only domain that isn't quite as evolved as the others is text-to-video generation. Many audio deepfakes of celebrities and even songs have seen the daylight and we expect this to boom even more once video generation What we can expect of generative models in this domain:

- **Text-to-video generation**: Generating video from textual inputs (prompts).
- **Image-to-video generation**: Adding motion to a single image or filling in the in-between motion to two images.
- **Video synthesis**: Creating variations of videos based on the original.

There are already a few results that show we are on the right track regarding video generation. They are becoming more precise, realistic, and controllable. However, it should be noted that the results thus far are not as spectacular as those of, for instance, image generation. Meta's Make-A-Video is one example; however, the best example you can find on the market today is Runway Gen-2. They are certainly at the forefront of video generation today, as they offer mind-blowing video-to-video and text-to-video generation.

TL;DR: Since January 2023, generative AI has evolved at a remarkable speed when it comes to text, image, and music generation. Even the first steps toward video generation have been made.

2. Deep Dive: Technical Advancements

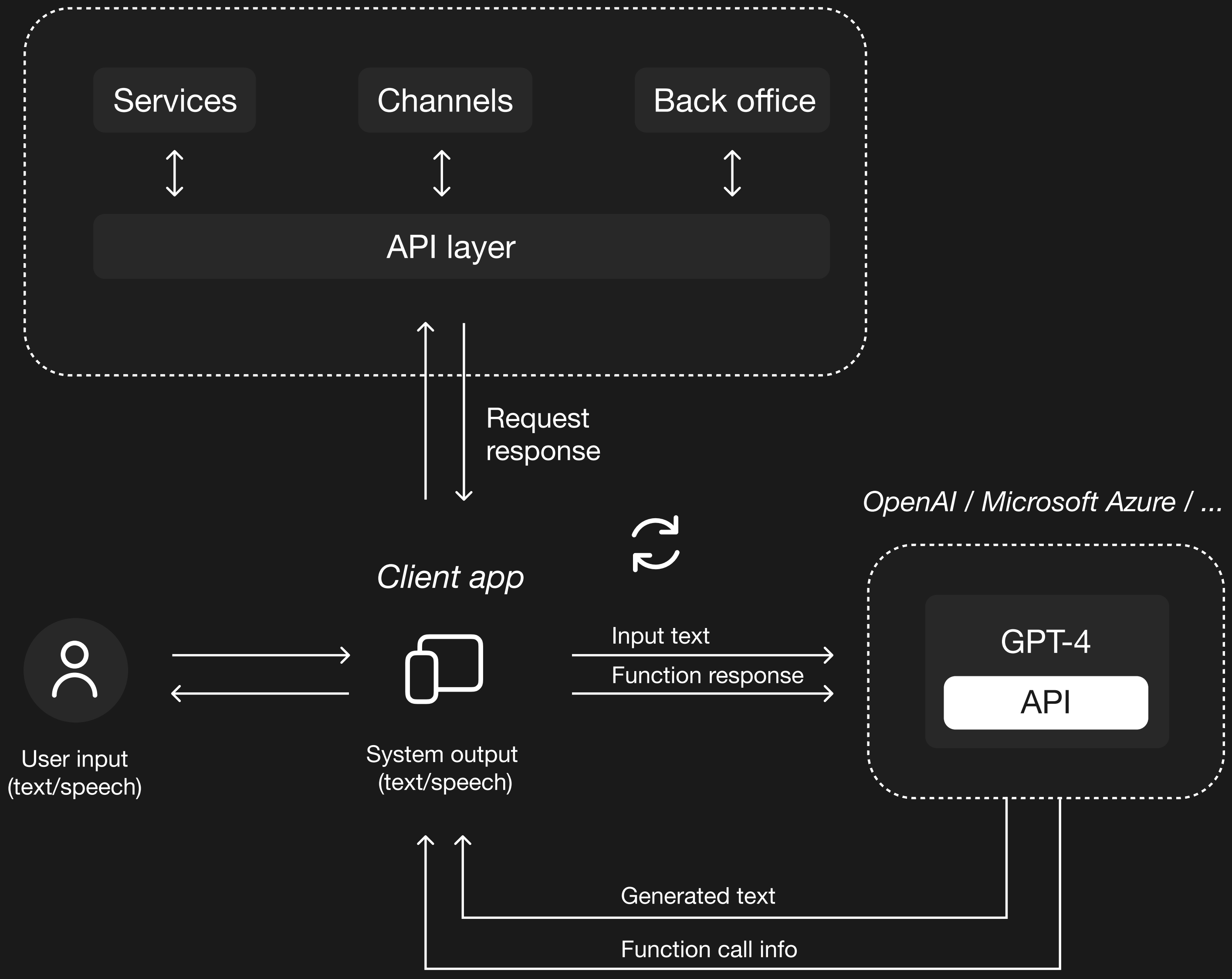
In the first chapter, we discussed the evolution of generative AI as a whole. In this chapter, we will dive into the field that (according to us) has the most notable advancements for business applications: text generation, and more specifically GTP. We will discuss three technical advancements that have had an enormous influence on how we can implement the technology to efficiently solve our clients' challenges: **function calling, knowledge mining, and fine-tuning**.

2.1 Function calling

Meet function calling, a groundbreaking feature that allows AI models to interact with external tools and APIs in a structured manner.

Simply said, we can get the model to generate text and execute specific tasks **by calling predefined functions**. For example, it can convert a user query like "What's the weather like in Brussels?" into a function call that fetches real-time weather data.

How it works? The GPT model can now intelligently decide which function should be called at what time. And it goes even further: the model automatically provides all the arguments and parameters that should be passed to the function call. The only thing developers need to do for it to work is **describe the purpose of the function and its arguments in natural language**. When you pass the result of a function call back to the GPT model, it can use this result to provide an answer to the question or decide to call another function, and so on.



Function calling

While this may sound technical, the implications for your business are profound. Here are some ways function calling can revolutionize your operations:

- **Automated decision-making**: Imagine a chatbot that can automatically schedule meetings, pull relevant sales data, or initiate customer service protocols. Function calling enables the model to interact with your existing software tools, making automation more seamless.
- **Data-driven insights**: Ever wondered who your top customers are this month? Function calling can convert such natural language queries into database function calls, providing you with real-time insights to make informed decisions.
- **Enhanced user experiences**: Thanks to function calling, your users can now enjoy more contextual and real-time responses. For instance, a customer service bot could now check the status of a user's order in real time by calling a function that interacts with your inventory management system.

This list is just the tip of the iceberg. The possibilities are endless, limited only by your imagination and business needs. By partnering with experts in generative AI, you can quickly integrate function calling into your existing systems, reducing the time-to-market significantly. We would argue function calling is not just a technical update — it's a **business enabler**. It opens up unseen possibilities for automation, user experience, and data-driven decision-making. Businesses can get ahead of the pack by understanding its potential and integrating it effectively.

2.2 Knowledge mining

While function calling enables AI models to interact with external tools and APIs in a structured manner for task automation and real-time data retrieval, **knowledge mining** is another way to incorporate facts into your large language models.

Knowledge mining aims to enrich the model's output with **domain-specific information**, making it insightful for your business. What’s important to remember is that on their own, large language models are not meant to be used for facts, as they do not have any specific business knowledge. It is meant to generate a reply in natural language only. With knowledge mining, you can now add those facts into your answers.

Simply put, knowledge mining is when you feed the model specific facts or data relevant to your business (structured and unstructured data), using the model's language generation capabilities to **produce responses that are linguistically coherent and factually aligned with your business context**.

LLM and vector databases

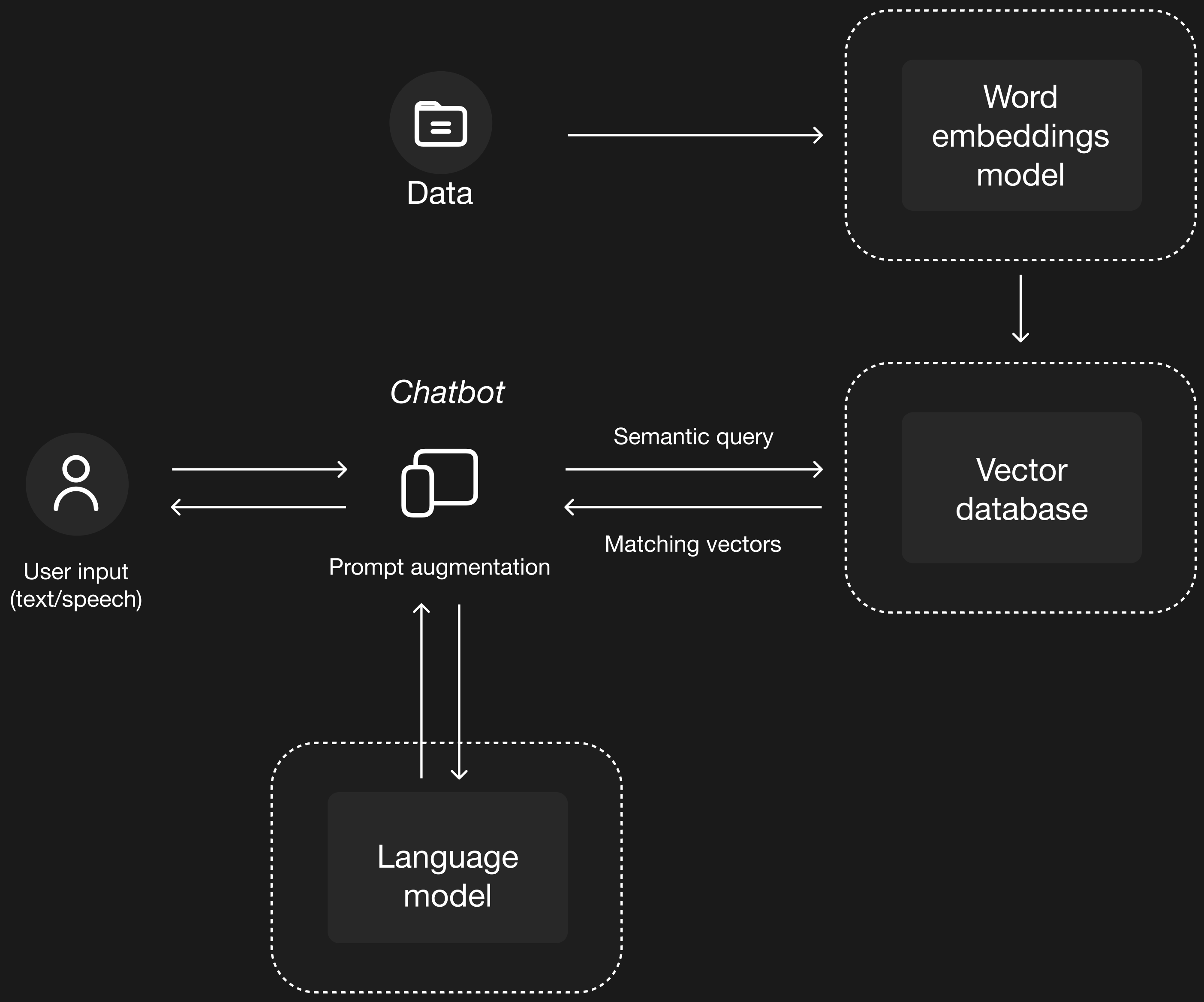
The combination of large language models — like the GPT series — and vector databases offers a powerful tool for businesses to extract valuable insights from a sea of data. The explanation gets technical, but let's scratch the surface and see how it works.

Usually, users interact with a conversational agent powered by asking the LLM a question. This input is then processed by what's known as an "embedding model" which transforms the text into a vector embedding. This vector can be described as a compact representation of the user's query, capturing its essence in a way that can be easily compared to other pieces of information.

Once this vector is generated, it is **matched against a vector database**. These databases are specialized storage systems that contain vectors representing a wide range of domain or business-specific information. The system can **identify the most relevant information** and articulate an appropriate response by comparing the user's vector to the database vectors. The answer is generated and sent to the user based on this matching process.

One of the key advantages of using vector databases is **contextual accuracy** — by feeding the model with domain-specific facts, you ensure that the generated text is grammatically correct, contextually accurate, and relevant. Moreover, this method is much more efficient when retrieving information, as the responses are generated in real time, reducing the response time.

By understanding the collaborative power of LLMs and vector databases, businesses can leverage these advanced technologies to gain a competitive edge in today's data-driven landscape.



Knowledge mining

2.3 Fine-tuning

The last technical advancement we will discuss is fine-tuning, a feature we have eagerly awaited. This feature enables the customization of the GPT-3.5 Turbo model (at the time of writing; will be added to GPT-4 in the future) to suit specific use cases better, enhancing its performance. This is especially useful for businesses that require a more tailored approach to AI-powered solutions.

So, what's this fine-tuning all about, then? It's pretty simple. You can now **adapt a pre-trained language model to perform better on specific tasks or within certain contexts**. Essentially, you train the model further on a specific dataset relevant to your business. Note that fine-tuning is not meant to give the model more facts about your business! Knowledge mining is a better architecture to achieve this.

A use case could be your custom tone of voice, for instance. If your company has a unique brand voice, you can now fine-tune the model to make the output more consistent with your brand's tone. Another use case would be learning specific jargon. This means that if you use domain-specific formats or terminology, you could easily train the model to use these words in writing reports.

To conclude, this feature ensures we are no longer confined to the general abilities of a pre-trained model. Organizations can now change GPT to fit their unique requirements. Of course, developers could already get to work with open-source models like LLaMA 2, which can be fine-tuned as well.

We believe fine-tuning is **most effective when combined with other techniques**: good prompt engineering, function calling, and even knowledge mining. GPT can now be used as a real solution for various business needs.

Conclusion

As the landscape of generative artificial intelligence keeps evolving, this information will probably be outdated in a few months — or even weeks. Nevertheless, we believe these advancements will be the basis of applications businesses will need and should develop in the coming years. Each feature (function calling, knowledge mining, and fine-tuning) offers distinct benefits, yet they all share a common thread: the potential to revolutionize how businesses interact with AI.

While fine-tuning is all about customization (i.e., tailoring the model to specific business needs), knowledge mining is about context (i.e., enriching the model's output with domain-specific information through vector databases), and function calling is about extending capabilities (i.e., interacting with external services). While each of these advancements can be impactful on itself, their true power lies in their potential for integration.

Function calling, knowledge mining, and fine-tuning are new features that enable businesses to use large language models in a more effective way.

3. Ethical Landscape

As generative AI technologies become increasingly integrated into business processes, the ethical considerations surrounding their use also arise. In this chapter, we will delve into the ethical landscape of generative AI. With every new technology, concerns enter the picture, and so do calls for guidelines. We will navigate this complex terrain and summarize where we stand today.

Challenges that come with generative AI include issues related to privacy, ownership, and copyright disputes. Moreover, artificial intelligence has always had the risk of exacerbating existing biases in society, and for generative AI, this is no different. For instance, if an AI model is trained on historical data that is inherently biased — think of the Amazon HR scandal — it could make biased decisions. The rapid advances in the technology have made these risks more real than ever — they are not just theoretical problems anymore. The technology is already being used for unethical purposes, think of generating disinformation.

3.1 Guidelines

The EU AI Act

The European Union has taken a pioneering role in regulating AI through its proposed AI Act. This comprehensive law categorizes AI applications into three risk categories: unacceptable risk, high-risk applications, and mostly unregulated applications. For example, high-risk applications like CV-scanning tools that rank job applicants are subject to specific legal requirements. This act could potentially become a global standard, but will be first rolled out in Europe.

ACM's Statement of Principles

Professional organizations like the ACM (Association for Computing Machinery) have also issued guidelines to help navigate the ethical complexities of AI. These principles focus on a holistic approach to technology, considering not just technical performance but also the broader societal implications. They advocate for transparency, inclusivity, and multi-stakeholder conversations to address the ethical challenges posed by AI.

3.2 Best practices for businesses

For businesses looking to adopt generative AI, as a first step, it's crucial to align the technology with the organization's objectives and values. This involves having internal conversations about the potential unintended consequences of adopting new technology. This way, you can stay one step ahead and catch issues even before they arise. Moreover, value-alignment is extremely important for the adoption of the technology by your employees.

We believe organizations should take a multi-faceted approach, considering everyone who could be potentially impacted. Moreover, anyone who has artificial intelligence on their mind should already get to know the implications of the EU AI Act, as this act will come into effect in a few years. Preparation is key!

Conclusion

As generative AI becomes more deeply integrated into our lives and businesses, the ethical implications become increasingly significant. By knowing the emerging guidelines and industry standards and by engaging in open, multi-stakeholder conversations, businesses can responsibly use generative AI in their operations. Moreover, when implementing a technology, be mindful of which supplier you choose. Some companies have worked tirelessly on mitigating the challenges and problems that come with generative AI. So, do not just choose a technology based on its functionality, but always think of the ethical aspects! By doing this, you mitigate risks and ensure that the technology truly benefits our society.

Due to the quick rise of generative AI, many eyebrows were raised. What does this mean from an ethical viewpoint? To counteract these concerns, guidelines have been set that European companies will have to adhere to.

4. Strategies for Successful Implementation

As generative AI technologies continue to evolve, the question for many people is not just "What can AI do?" but "How can we implement AI successfully?" This chapter offers a step-by-step guide to starting to build a generative AI strategy that is tailored to meet the needs of your organization.

Step 1: Define objectives and scope

The first step in any successful AI implementation strategy is to clearly define what you hope to achieve. Whether it's automating customer service, enhancing data analytics, or improving product recommendations, having a clear objective will guide the rest of the process.

💡 **Key consideration**: Make sure that your objectives align with the broader goals and values of your organization.

Step 2: Assess existing infrastructure

Before diving into AI implementation, take stock of your existing IT infrastructure. This will help you identify what upgrades or changes are necessary for smooth integration. Evaluate the scalability and flexibility of your current systems to ensure they can accommodate the AI tools you plan to implement.

💡 **Key consideration**: Rely on experts to help you make this assessment.

Step 3: Choose the right tools

Selecting the right AI tools and technology partners is crucial. Look for solutions that not only meet your technical requirements but also have a proven track record and ethical standards.

💡 **Key consideration**: Consider the long-term viability of your chosen tools and partners. Make sure they are committed to continuous improvement and ethical AI practices.

Step 4: Skill development

The success of your AI implementation will largely depend on the skills and expertise of your team. Even if you outsource the actual development of the AI application, you should still make sure anyone impacted is on board and understands the effect AI can and will have on the organization.

💡 **Key consideration**: Don't just focus on your tech team. Make sure that everyone, from executives to sales, have a basic understanding of what AI can and cannot do.

Step 5: Pilot testing

Before full-scale implementation, conduct a pilot test to identify any issues or bottlenecks. This will give you an opportunity to make adjustments before rolling out the technology across your organization.

💡 **Key consideration**: Use the pilot phase to test the impact on your end users.

Step 6: Monitor, evaluate, and iterate

Once the AI tools are implemented, continuous monitoring is essential. Use KPIs aligned with your initial objectives to evaluate performance. Be prepared to make iterative changes to optimize both efficiency and ethical considerations.

💡 **Key consideration**: Keep an eye on how AI impacts not just business metrics but also employee satisfaction and customer experience.

Implementing generative AI in your organization is not just a technical endeavor but a strategic one that requires careful planning and skilled experts. By following this step-by-step guide, you can navigate the complexities of AI implementation, ensuring operational success.

5. Future Outlook

When it comes to our future, we think it's crucial to look ahead and consider where generative AI is headed. Generative AI is rapidly evolving and we can expect to see AI models that are more context-aware, capable of more nuanced interactions, and even more tightly integrated with other existing and emerging technologies.

What you can do now? Well, in a fast-paced technological landscape, **timing is everything**. Early adoption of generative AI technologies not only provides a **competitive edge but also allows for a more seamless transition as the technology evolves**. Organizations that invest in AI now will be better positioned to leverage future advancements — ensuring long-term success.

The future of generative AI is not just promising; it's imminent. The technology is advancing at an unprecedented rate, offering unparalleled opportunities for innovation and growth. The urgency to act is not a matter of keeping up with the competition; it's about staying ahead of the pack. The time to leverage the transformative power of generative AI is now.

And good news: you're in the right place if you need a strategic, innovative partner. Schedule a call with us today, and let’s see what generative AI can do for your business.

Get a free consultation call:

Book your free consultation with Michiel Vandendriessche via hello@raccoons.be.

About Raccoons

We chase our curiosities, build what's next and advance human potential. We lead and explore new ways of thinking about how modern technologies and services can be used to craft innovative experiences and products.