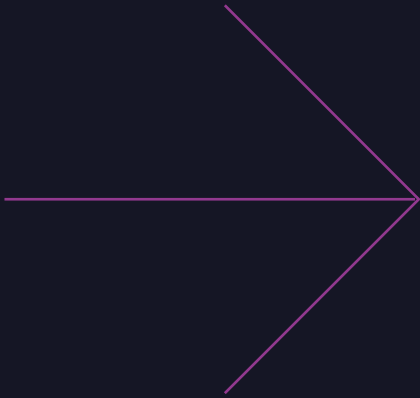


# Gener(AI)ting the future



**Arthur Mensch**  
CEO  
Mistral AI



# The CEO Corner

*in discussion with*



**Aiman Ezzat**  
CEO  
Capgemini







**Arthur Mensch**  
CEO, Mistral AI

Arthur Mensch is a French entrepreneur and scientist.

In 2023, Arthur Mensch, along with Guillaume Lample and Timothée Lacroix, founded Mistral AI with the mission of making generative AI ubiquitous and pioneering a new approach to AI - one that is more open, portable, independent, and accessible to all.

After more than 10 years of academic work focused on the possibilities of machine learning in the field of brain imaging and on optimization of machine learning, he joined DeepMind Paris in 2020 as a researcher, where he spent three years and played a key role in the development and deployment of flagship projects in generative AI.



**Aiman Ezzat**  
CEO, Capgemini

With more than 20 years' experience at Capgemini, Aiman Ezzat has a deep knowledge of the Group's main businesses. He has worked in many countries, notably the UK and the US, where he lived for more than 15 years.

Aiman was appointed CEO in May 2020. Prior to that, from 2018 to 2020, he served as the Group's COO and, from 2012 to 2018, as CFO. Aiman is also on the Board of Directors of Air Liquide and is a member of the Business Council and the European Round Table (ERT) for Industry.

## What inspired you to form a new player [in Mistral] in the generative AI (Gen AI) space – and why in Europe?

— **Arthur:** My co-founders and I have been working in the Gen AI space for over 10 years, previously in large US-based organizations. When development accelerated at end-2022, it gave us an opportunity to create some very strong models in a short period of time. We secured funding, assembled a dedicated team and the GPUs [graphic processing units] required to train the LLMs [large language models], and were ready to go.

Why Europe? Europe is a great place to start a company. The education systems in France, Poland, or the UK, for example, are great for training AI scientists. We brought in recent PhDs from Paris; we were able to get the most important thing to get started – the team. As the only player in Europe in the field of conventional language models, we had some strong geographical business opportunities.

“

**We use both an open-source model and a portable platform for model deployment.”**

Arthur Mensch



### What do you see as the advantages of the open-source gen AI model?

— **Arthur:** We use both an open-source model and a portable platform for model deployment. Even our commercial models are licensed. This allows users to customize the models to their needs.

It offers portability and comfort. With a model that you can deploy on any platform, on a private cloud or on-premise or on dedicated instances on the cloud, you can use the technology where your data is. So, this adapts to the data-governance constraints of the enterprises, and our customers very much appreciate this flexibility.

— **Aiman:** There are clear advantages to both open and closed approaches. Openness boosts innovation and drives collaboration. Open models also allow everyone to scrutinize the model for potential sources of bias, demystifying the “black box” nature of AI models. There are also challenges. Customizing open models for a particular industry or organization is tricky, but using open models out of the box can lead to suboptimal performance. Fine-tuning any foundation model, open-source or proprietary, is a time-consuming, resource-intensive process that requires significant financial investment. Hence, it is important for enterprises to assess ROI carefully before pushing out the Gen AI boat.

**"Open models also allow everyone to scrutinize the model for potential sources of bias."**

Aiman Ezzat



## Do you see organizations using a generalized Gen AI model going forward or many different specialized models?

— **Arthur:** We see the field moving in these two directions simultaneously. A strong generalized model gives a good platform for testing solutions. But this can be a slow and costly process, offering poor ROI for specific tasks. You want your LLM to offer an intelligent, dynamic solution for a specific issue, whether that's parsing the logs of an IT system or parsing the conversation between a customer and a customer agent.

From a scientific point of view, smaller models can solve specific issues, but they must be finely tuned. We want to bring solutions to market that develop the smallest possible model to solve a specific defined task, which will allow for low-latency applications.

Smaller models also mean the applications are less costly to run and, more importantly, if you have a model that is 100 times smaller, you can call it 100 times more for the same cost, bringing a little more intelligence to your application with each call. We call this "compressed knowledge". We specialize models in order to make differentiated applications that go fast, that call LLM often and that are cost-controlled.



**If you have a model that is 100 times smaller, you can call it 100 times more for the same cost, bringing a little more intelligence to your application with each call."**

Arthur Mensch

— **Aiman:** There's a very clear market for both generalized and specialized models. A generalized model can serve those use cases that don't require extensive customization. These are "low-hanging fruit" that rapidly demonstrate the power of Gen AI.

Developing and training specialized models for some basic use cases might even be counterproductive in terms of cost and sustainability. That said, there are use cases that benefit from specialized models, for instance, in terms of performance characteristics or in detecting and responding to specific nuances of the industry or use case. Any use case that requires high performance or deep domain expertise will likely continue to go down the path of specialized models. At the same time, specialized models potentially require significant resources in terms of maintenance and regular updates, so organizations might prefer a generalized model for use cases with less stringent requirements. I see a future where both types of models coexist harmoniously.

“

Any use case that requires high performance or deep domain expertise will likely continue to go down the path of specialized models.”

Aiman Ezzat



### What are the most innovative use cases that you are seeing in Gen AI?

— **Arthur:** In financial services, for instance, Mistral has built models that extract information from financial reports and summarize it for bankers to analyze. This harnesses the power of generative AI to process a large amount of text and detect weak signals, which is very much the core business of banks. The other successful deployment is in customer services.

— **Aiman:** We have been working on several innovative cases using Gen AI across industries. In life sciences for instance, we have developed with generative AI a solution to design new drug molecules. This method significantly boosts the process of generating new structures, offering researchers a potent tool for designing molecules aimed at specific biological targets. It illustrates AI's transformative potential in accelerating and refining drug discovery, particularly in the preliminary phases.



## Given the energy required to create and train the large models, what are the sustainability implications for Gen AI?

— **Arthur:** Most of the compute and energy resources required to run these systems are used at inference time rather than at training time. So you train for a couple of months, and when the models are deployed on many, many GPUs, then the large energy consumption is more linked to the usage than to the training itself. There are trade-offs between the amount you spend on training and the compression that you can achieve. If you invest more in training, you can make smaller models, achieving the same performance as a larger model with less compute. These smaller models consume less energy to deploy at inference time.

At Mistral, we focus on compressing knowledge and making models that are smaller than those the competition produces. Limiting carbon emissions is a cause that is very dear to our heart and the reason why we deployed our solutions in Europe. In Sweden, in particular, renewables compose a high proportion of energy consumption.

“

**Most of the compute and energy resources required to run these systems are used at inference time, rather than at training time.”**

Arthur Mensch

— **Aiman:** Our research shows that more than three-quarters of organizations are conscious of environmental concerns around Gen AI. As a leader in the eco-digital revolution, we at Capgemini recognize the need to weigh the immense potential of Gen AI against its cost to the planet and society. We are committed to taking a “sustainable by design” approach to developing Gen AI solutions that harness cutting-edge data, AI, and climate tech to maximize business outcomes in a sustainable manner.

Mitigation strategies include optimizing the amount of data required to train the models, working on smaller, task-specific energy-efficient models that employ more efficient training and operating algorithms, and powering the AI infrastructure with renewable energy as well as using more energy-efficient datacenters. We also promote transparency in AI development and operation by monitoring and disclosing the energy consumption and carbon footprints of Gen AI models. Our Gen AI lifecycle analysis tools help organizations to mitigate environmental impact.

## How should large organizations address ethical considerations and potential bias in deployment of AI models?

— **Arthur:** When an organization is making an AI-driven product, it has to consider the decisions and outputs the system will make. So, these decisions and these outputs should be constrained to respect the company's role. What it means is that before deployment of a new AI product, the first thing to think about is how do you evaluate success. How do you ensure that the model is behaving as it should and not producing unwanted outputs? And is it able to provide a nuanced but unbiased answer to complex questions?

Owing to our open approach, the customer can make their own editorial choices from these evaluations.

— **Aiman:** Large organizations should be conscious of a variety of risks: Inherited risk, intellectual property, correctness, data leakage, and user privacy.

Organizations should establish employee guidelines for the safe use of Gen AI and validating outputs to eliminate bias. At Capgemini, we have applied a governance model to ensure this.

"The rare talent that we recommend every organization look for is the software engineer who can also do data science."

Arthur Mensch



## How are you bridging the AI talent demand-supply gap?

— **Arthur:** It has been a challenge to get the best AI scientists. We recommend hiring very strong data scientists who can undertake software development. Since we are making the tools and the foundation for the model itself, training the model is not a necessity within the enterprise setting. To make the most interesting products, clients must understand how to use the platform.

So, the capacity for doing this is really adjacent to what we used to call data science a decade ago. It's the ability to run experiments, to evaluate certain systems, to see what is failing, and to see how to try and improve it. This scientific mindset, running experiments on a computer and measuring success, which is really the data scientist's job. The changes with the data scientist's job today is that the software requirement is stronger because, if you want to make an interesting application, you also need to dive deep into the way you assemble the software, connect it to the LLM, the LLM to the database, and an LLM to tools. Having a system mindset is necessary to create successful applications. The rare talent that we recommend every organization look for is the software engineer who can also do data science.

— **Aiman:** We are investing over €2 billion over three years in Gen AI and have already trained over 120,000 team members on generative AI tools thanks to our Gen AI Campus. We have also launched a dedicated platform to industrialize our custom generative AI projects. We will also focus on obtaining certifications and building centers of excellence, as well as specific go-to-market skills. Ultimately, Gen AI training will be a key requirement in all of our development and training curricula.



### How do you see gen AI driving transformation in large organizations?

— **Arthur:** The first step is to take a model – a Mistral model, for instance – and connect it to the enterprise context. The enterprise context is located across different databases or SaaS [Software-as-a-Service] systems. You can then generate assistants with access to the enterprise context, to help every employee navigate the enterprise processes and organization. That's typically what our customers do first. They create a knowledge management tool or general assistant for employees.

— **Aiman:** Driving transformation with generative AI goes beyond the technology. Success depends on a broad strategic vision that covers everything from applying it to the right use cases, potentially adapting internal processes, to optimizing customer-facing operations. In addition, the value of generative AI depends on two key foundations: the data and the human elements. Leaders need to have the right data foundations in place to ensure they are realising the full potential of Gen AI. Equally important is training employees to not only use AI effectively but also to trust it, which is key to adoption.



---

**Arthur Mensch**  
CEO, Mistral AI

---

**"The first step is to take a model – a Mistral model, for instance – and connect it to the enterprise context."**



---

**Aiman Ezzat**  
CEO, Capgemini

---

**"There's a very clear market for both generalized and specialized models."**

