# WHAT ARE FOUNDATION MODELS?

*With the arrival of ChatGPT, generative AI models are the talk of the town in AI.. OpenAI's ChatGPT is built upon advanced foundation models such as GPT-3.5 and GPT-4, which is also used by other companies, as for example in the Microsoft Copilot application. A foundation model is an AI model that is, as the name implies, the foundation or 'base' for other companies to build on and thus use it to build or train other AI applications.*

*In this brAInfood, you will learn more about what foundation models are and what they are used for. We explain this with some examples of applications of foundation models. Finally, we also delve into possible benefits and state some considerations about the way they are designed and used.*

## WHAT IS A FOUNDATION MODEL?

A foundation model is an AI model that is fed with a **huge amount and variety of data**. This model can be trained only once and then refined to **perform all kinds of different tasks**, such as generating text, images or audio.

Some foundation models **only take input in a particular modality**, such as text, while other models can take on **multiple modalities**, e.g. text, image, video, etc., and even **generate multiple types of output**. This output could be, for example, generating images, summarising text or answering questions.

What makes foundation models unique is that they can be **standalone systems**, but they can also be **used as 'bases' for other AI applications**. The organisation creating the foundation model opens up 'the product' to other organisations, often for a fee, to build on it and develop a new AI application.

## APPLICATIONS OF FOUNDATION MODELS

Two well-known foundation models are large language models (LLMs) and generative AI:

- **LLMs** are based in the field of Natural Language Processing (NLP). They are trained with billions of words of text and can generate natural language based on text prediction. These models calculate the probability of the use of a character, word or string based on the preceding or surrounding lingual context. LLMs fall under a broader umbrella of 'generative AI'.
- **Generative AI** can generate different types of content (text, images, video or audio) based on user input such as text prompts. Note that not all generative AI is based on foundation models as some applications are designed for a specific purpose, and not to be reused in new contexts.

Foundation models are often used for **internal purposes**, such as a portal where, at an employee's request, a chatbot shares necessary information about human resources, or an assistant summarising long legal texts for a lawyer.

## BENEFITS

Thanks to foundation models, **new applications can be rolled out faster** because the basis of the application already exists. In this way, duplication of work is avoided and new applications can go to market faster. It is no longer necessary for developers to first fully immerse themselves in the world of machine learning and artificial intelligence applications, because thanks to foundation models, such applications can be made **more easier to deploy without the need for in-depth knowledge about building these applications.** Developers thus have more time to focus on training the model on specific data sets with labeled examples. Of course, this does not mean that training these models does not require technical and/or programming knowledge.

## REGULATION

AI models that display significant generality, can competently perform a wide range of distinct tasks and can be integrated into a variety of downstream systems, are regulated under the EU AI act under the term "General-purpose AI models" or "GPAI models".

Providers of all GPAI models are obligated to draft technical documentation, provide information to providers who will integrate the GPAI-model into their AI-system, create a policy for their compliance with copyright law and publicize a summary of the content used for training. These obligations do not apply to open-source GPAI models with systemic risks.

Providers of GPAI-models with systemic risk must test for, assess and mitigate systemic risks, report serious incidents to specific authorities and ensure adequate cybersecurity of the model.

GPAI models have systemic risk if they have high impact capabilities or if they are designated as such by the Commission based on criteria set out in the AI act.

## CONSIDERATIONS

There are some considerations to developing and using foundation models or AI in general, such as:

- **Power imbalance**: because developers can start from an existing foundation model and work on it, this naturally gives a certain amount of power to the providers of those foundation models because they determine how they function. The users of foundation models should always critically analyse the model.
- **Lack of contextual understanding**: foundation models have no understanding of the output they generate. For instance, they lack the ability to understand and account for nuances or underlying emotions of a situation (such as sarcasm).
- **Hallucinations**: these models lack "common sense", the ability to understand the nuances of human experience and the real world. This can lead to a phenomenon known as "hallucinations", where the model generates plausible but factually incorrect information. For example: a model summarises a news article, but makes up quotes or details that are not in the original source.
- **Bias - garbage in, garbage out**: a term that states that a model only works as well as the data used to train it. If a foundation model is trained on data that is biased and/or negative, it may produce discriminatory, hateful and/or other forms of harmful content. A face recognition model trained only on examples of white individuals will have difficulty (correctly) recognising individuals with a different skin colour.
- **Security vulnerabilities**: building on the third consideration, there are several scenarios that can occur when training models. Because these models are trained with an awful lot of data, there is a chance that when importing new training data, harmful, sensitive or confidential information could be accidentally or deliberately inserted.
- **Difficulties in monitoring and/or understanding system behaviour**: foundation models are highly complex, so it is not always clear for users to understand their decision-making processes and outcomes. This lack of transparency is further exacerbated by the fact that the models are often not open source. The use of the models is (relatively) free, but that does not immediately mean that the source code is freely accessible for review or modification
- **(Environmental) cost**: developing, implementing and maintaining foundation models requires a high cost, both monetarily and in terms of energy consumption. Although no official figures have been released, the energy consumption for training ChatGPT is estimated at more than 1 Gigawatt-hour (GWh), comparable to the energy consumption of 120 households for a year.
- **Regulation**: Because foundation models are difficult, if not 'readable', it is difficult for legislators to regulate and legislate the use and functioning of these technologies. The EU has addressed this with a dual regulatory framework. The AI act distinguishes between generic models and systemic models. Generic models are subject to transparency obligations (technical documentation and respecting copyright legislation). Systemic models have, in addition to the obligations imposed on generic models, to conduct evaluations, assess and mitigate systemic risks, report incidents and ensure cybersecurity.

*Knowledge Centre Data & Society (March 2024). What are foundation models? brAInfood of the Knowledge Centre Data & Society. Brussels: Knowledge Centre Data & Society.*



Knowledge Centre Data & Society

Artificial Intelligence Flanders