Reinking, Ernst; Becker, Marco

Working Paper

Opportunities for business use of today's AI models - Rapidly achievable personalization of Large Language Models (like ChatGPT) in times of Industry 5.0

IUCF Working Paper, No. 10/2023

Ernst Reinking, Marco Becker[1]

Opportunities for business use of today's AI models

Rapidly achievable personalization of
Large Language Models (like ChatGPT) in times of Industry 5.0

Abstract

The introduction of ChatGPT as one of the best-known Large Language Models not only opened a new chapter in artificial intelligence in the general perception – some authors even speak of an era of (business) informatics. It also heralds the fifth industrial revolution (Industry 5.0).

The aim of this working paper is not only to objectify the contradiction between hype and reality in the context of artificial intelligence, but also to show the opportunities and perspectives for the analysis of unstructured, internal company data. To this end, the authors have developed several prototypes based on their own research work, which form the basis of this working paper.

Zusammenfassung

Mit der Einführung von ChatGPT als eines der bekanntesten Large Language Models wurde in der allgemeinen Wahrnehmung nicht nur ein neues Kapitel der Künstlichen Intelligenz aufgeschlagen - manche Autoren sprechen sogar von einer Ära der (Wirtschafts-)Informatik. Sie läutet auch die fünfte industrielle Revolution (Industrie 5.0) ein.

Ziel dieses Working Papers ist es, nicht nur den Widerspruch zwischen Hype und Realität im Kontext der Künstlichen Intelligenz zu versachlichen, sondern auch die Chancen und Perspektiven für die Analyse unstrukturierter, unternehmensinterner Daten aufzuzeigen. Dazu haben die Autoren auf Basis eigener Forschungsarbeiten mehrere Prototypen entwickelt, auf denen die Grundaussagen dieses Working Papers basieren.

[1] **Dipl.-Ing. Ernst Reinking** is a Research Fellow at the NBS Northern Business School - University auf Applied Science in Hamburg, where he teaches and conducts research on business informatics, digital economy and artificial intelligence.
Prof. Dr. Marco Becker teaches and conducts research on controlling and financial management at NBS Northern Business School - University auf Applied Science in Hamburg and is deputy director of the Institute for Corporate Accounting, Controlling and Financial Management (IUCF). Furthermore, he is founder and partner of Marco Becker Management Consultants.

Starting point: Current hype around artificial intelligence and ChatGPT

November 28, 2022 marks the beginning of the fifth industrial revolution (Industry 5.0), ushering in the age of artificial intelligence (AI).[2] On that day, the company OpenAI released its application ChatGPT, a chat-based large language model (LLM), i.e. a complex artificial intelligence model based on deep neural networks and designed to solve complex natural language processing tasks. Such models are called "large" because they consist of a considerable number of parameters (GPT3 175 billion, GPT4 unpublished, estimated 1,000 billion), which gives them an impressive ability to generate and understand human language texts.[3]

The beginnings of AI date back to the 1950s, but after a very checkered history, it is only in recent years that it is experiencing a steady increase in importance, especially due to the availability of very powerful hardware and accompanied by the development of new algorithms.[4] Thus, GPT3 (on which ChatGPT is based) is also an evolutionary step. However, this was the first time that an AI language model was trained with an unimaginably large amount of data and a gigantic computing effort over years. Even the developers were initially completely surprised by the result and the performance. The system is based solely on searching through words stored in an n-dimensional vector space with the help of a few algorithms according to the principles of statistics and probability and finding the next word matching the question word by word and stringing them together as the answer. This goes so far that the model, although it is a pure text model, can also correctly solve simple questions about programming code or sometimes even simple mathematical tasks.[5]

The public provision of ChatGPT and the corresponding media accompaniment triggered an unprecedented IT hype. Within only 5 days, the number of users exceeded the 1 million mark and ChatGPT mutated into an omniscient tool in the public perception. An extreme competition between the big IT companies started, but also a very large and fast-growing open source community, in which many scientists can be found. In the meantime, there are a growing number of LLMs available in addition to ChatGPT/GPT. An immense media and also political echo accompanies this development with mostly unrealistic information.[6]

Main features of Large Language Models

In the course of the decades, very different manifestations of artificial intelligence with different characteristics have developed. The periodic table of artificial intelligence developed by BITKOM provides a summary overview.[7]

[2] Vgl. Becker/Daube et al. (2023).
[3] Vgl. Heaven (2023).
[4] Vgl. Taukki (2022): S. 6 ff.
[5] Vgl. Heaven (2023) and Zhang/Li (2021).
[6] Vgl. Heaven (2023) and Zhang/Li (2021).
[7] Vgl. Bitkom e.V. (2023).

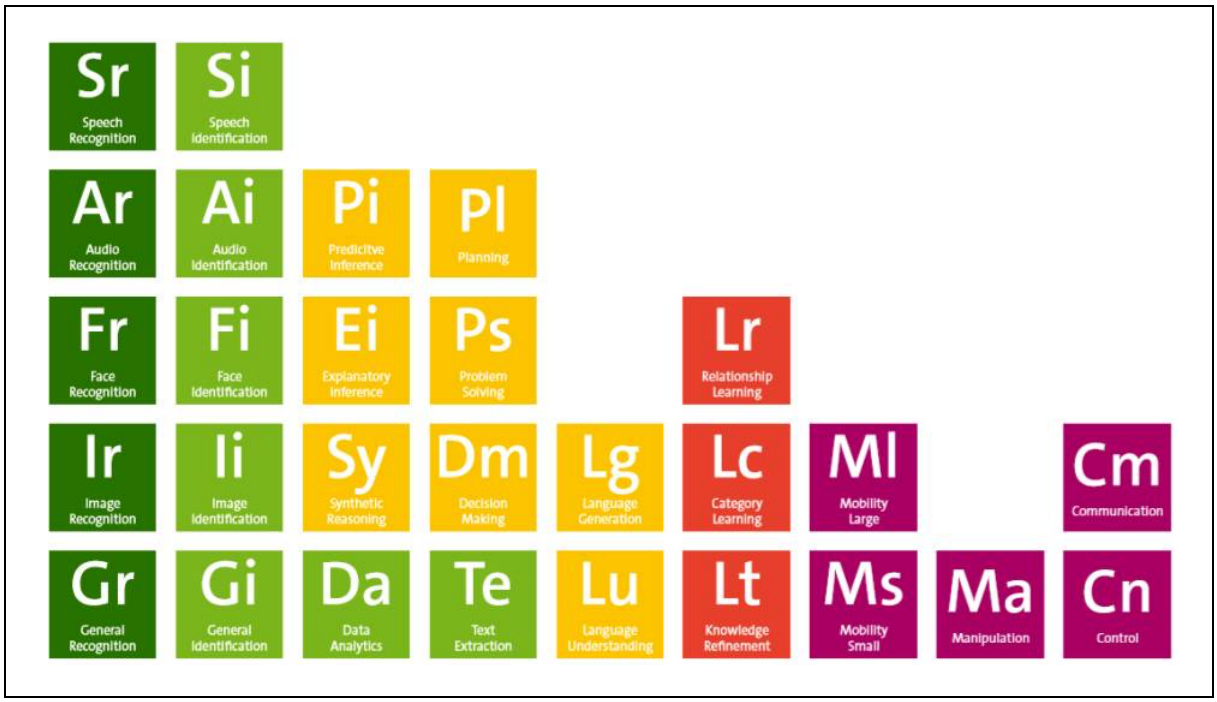


Figure 1: Periodic table of artificial intelligence [8]

As can be seen from the periodic table of artificial intelligence, LLMs represent a combination of several disciplines (including Te, Da, Te, Dm, Ps, Lu, Lg ...). The following key factors are of central importance for LLMs in this context.:[9]

- Text generation:
  LLMs can generate text that resembles human writing style. They can be used in various contexts, such as writing news articles, marketing materials, or even literary works.

- Language understanding:
  These models can interpret and respond to complex queries. They are able to grasp the meaning behind words and phrases, which is useful in areas such as customer service and virtual assistants.

- Sentiment Analysis:
  LLMs can detect the emotions behind a text and evaluate whether it is positive, negative, or neutral.

- Text translation:
  You can translate texts between different languages without losing the meaning.

- Knowledge extraction:
  LLMs can extract and synthesize information from large amounts of text, which can be used in research and analysis.

[8] Own illustration based on: Bitkom e.V. (2023).
[9] Vgl. DB Systel GmbH (2023) as well as its own considerations.

- Pretraining:
  The major LLMs are usually already trained with a very large general/"world" knowledge that users can leverage.

- Interfaces:
  LLMs have versatile program interfaces via which other programs can ask questions and receive answers. This allows them to become part of other program environments and thus also to be integrated into automations.

LLMs centrally map the interface between the AI engine and the user, as is done in ChatGPT, for example.[10]

Large Language Models in the Corporate World

LLMs have the potential to fundamentally change the business world by automating processes, improving the customer experience, and enabling data-driven decisions. LLM techniques and generative AI capabilities will also make their way into IT programs used today, such as ERP, Office, mail systems and others. Companies will recognize and take advantage of this advanced technology and use it as a competitive advantage.

A few application areas will be mentioned here as examples:[11]

- Customer interaction:
  LLMs can be used in chatbots and customer service systems. With their ability to understand and respond to natural language, they provide personalized and efficient interactions. Automating customer support leads to reduced costs and increased customer satisfaction.

- Data analysis and processing:
  Big data analytics is critical for many businesses. LLMs can analyze unstructured data, identify trends, and provide valuable insights into business processes. This drives data-driven decision making and strategy development.

- Personalization of marketing campaigns:
  The ability of LLMs to identify individual preferences and behavior patterns enables marketing teams to create personalized campaigns. This increases the engagement rate and ROI (return on investment) of marketing strategies.

- Automated translations:
  With LLMs, it is possible to translate texts quickly and accurately into different languages. This promotes the global expansion of companies and facilitates communication with international customers and partners.

[10] Vgl. DB Systel GmbH (2023) as well as its own considerations.
[11] Vgl. DB Systel GmbH (2023) and Kucharavy/Schillaci et al. (2023): S. 35 ff.

Large Language Models and their » knowledge«

„If the company knew what it knows...."[12]

Generating knowledge from internal company data is one of the key success factors of Industry 5.0. The difficulty here lies in the data format, since a large proportion of the relevant data (approx. 80% to 90%) is available in unstructured form. They are thus lost for conventional evaluations.[13]

Structured vs. unstructured data

They represent two basic categories of data used in data processing and analysis. They differ in many ways. To illustrate the differences, a comparison follows:

Structured data [14]

- Definition: Structured data is data that is organized in a fixed structure or schema. They follow a specific order and a fixed format.
- Examples: Database tables, Excel spreadsheets, CSV files.
- Format: Organized into rows and columns, with each column representing a specific type of data (e.g., integer, string).
- Analysis: Easy to analyze and query because they are organized in a fixed schema
- Storage: Often stored in relational databases.
- Size: Generally smaller than unstructured data because it is held in a fixed format.
- Applications: ERP systems, CRM systems, business intelligence.

[12] Davenport/Prusak (1999).
[13] Vgl. MIT SLOAN SCHOOF OF MANAGEMENT (2023).
[14] Vgl. LexisNexis (2023).

Unstructured data [15]

- Definition: Unstructured data does not have a specific structure or schema. It can exist in a variety of formats and types.
- Examples: Emails, social media posts, images, videos, text documents.
- Format: No fixed format. Can come in different forms and types.
- Analysis: More difficult to analyze because special tools and techniques are required to understand the data (e.g., text analysis, image recognition).
- Storage: Can be stored in different types of databases or file systems, including NoSQL databases.
- Size: Often more extensive than structured data, as it can come in many different forms and formats.
- Applications: Big Data analytics, artificial intelligence, machine learning, social media analytics.

LLMs are text models. They are exclusively suitable for processing unstructured text data. Until now, unstructured data in companies has generally only been processed or analyzed to a limited extent for technical reasons. An LLM promises significant added value and represents a long-sought solution for the analysis of unstructured data.[16]

Data basis of the LLM

More than 90% of the effort and thus the costs for the creation of an LLM are attributable to the data. This applies to data acquisition and cleaning as well as to the extremely computationally intensive and lengthy learning processes of the model. No reliable cost data is available for ChatGPT, but it can be assumed that the computing time required for the learning processes alone will be in the high two-digit millions. In addition, there is the preparation of the data, which has taken place over years with a great deal of human work.[17]

With the very valuable knowledge from the described learning processes, an LLM is available for use that has only a general knowledge, a „world knowledge“, albeit a very extensive one.[18] If operational questions are to be answered with the LLM, this alone is not sufficient; domain knowledge is also required:

[15] Vgl. LexisNexis (2023).
[16] Vgl. LexisNexis (2023) and Kucharavy/Schillaci et al. (2023): S. 4 ff.
[17] Vgl. DB Systel GmbH (2023).
[18] Vgl. DB Systel GmbH (2023).

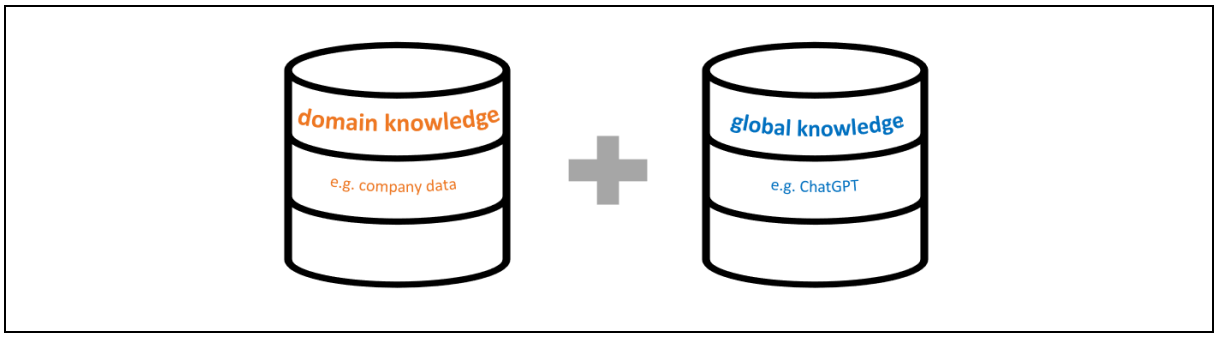


Figure 2: Domain knowledge and world knowledge [19]

Options for processing enterprise data in a pretrained LLM

Basically, there are two ways to add domain knowledge, fine-tuning the LLM and content injection.

Fine-Tuning

In fine-tuning, the pre-trained LLM model is further trained with additional specific data sets in an additional training phase. The procedure is thus a continuation of the original training and can only be performed by someone who also has the model in its entirety.[20]

In the case of GPT, OpenAI offers this option to customers for a fee. The customer has to deliver his data in a predefined format, OpenAI then carries out the elaborate training and subsequently holds its own GPT version ready for the customer. This procedure is associated with very high one-time and ongoing costs. If at all, it can only be considered for very long-lived data supplements, since the model itself is permanently changed.[21]

However, initial experience also shows that this procedure works only inadequately. This is due to the fact that the model has already been trained on a large amount of general text data and the domain-specific data is usually insufficient in quantity to overwrite what has already been learned.

Content-Injection

This procedure takes advantage of the fact that the question posed to the LLM is allowed to have a length that is quite comfortable for normal questions.[22]

The length is defined in so-called tokens. Tokenization is the process of breaking a text into smaller parts or "tokens" so that it can be processed by a model. Tokens vary in length and have an empirical average length of 3.5 letters. In the further course of processing, tokens are incidentally converted into numerical values.[23]

[19] Own illustration.
[20] Vgl. OpenAI (2023c) and Grosse/Hallgarten et al. (2023) and Wang/Li et al. (2019): S. 3.
[21] Vgl. OpenAI (2023c) and Grosse/Hallgarten et al. (2023).
[22] Vgl. Ferus (2022).
[23] Vgl. Ferus (2022).

The following table contains the different input lengths of selected models:

GPT-Modell	Technical Name	Max. Token	~ DIN A4 Pages*
GPT-4 (32k)	gpt-4-32k-0613	32.768	6,3
GPT-4 (8k)	gpt-4-0613	8.192	1,5
ChatGPT (16k)	gpt-3.5-turbo-16k-0613	16.384	3,0
ChatGPT (Standard – 4k)	gpt-3.5-turbo-0613	4.096	0,8

*) One DIN A4 standard page corresponds to 1,500 characters incl. spaces [24]

Figure 3: Maximum dialog length in selected LLMs [25]

With content injection, the question is constructed, for example, with upstream programs from

- a predefined text
- supplied content of own documents
- the actual question [26]

> *Answer the question at the end of the text using the following information. If you don't know the answer, just say that you don't know, don't try to make up an answer.*

> *3.3. Scope of work*
>
> *For term papers, a length of 15 pages and for bachelor theses, a length of 40 to 60 pages is defined as the rule. For a master's thesis, between 60 and 80 pages are to be planned. The length of the internship thesis is 20 to 30 pages...*

> Question: How long should a term paper be in the subject Business Informatics?

Fine-Tuning vs. Content-Injection

Fine-tuning is only considered for permanent model changes, the effort is very high, the effect remains low, because the amount of data for fine-tuning is usually much smaller than for pretraining.[27]

Content injection can be changed very dynamically from query to query, the effect is good, but the amount of content per query is very limited.

[24] Vgl. VG WORT (2023): S. 7.
[25] Own presentation based on: OpenAI (2023a) und OpenAI (2023b).
[26] Vgl. Ferus (2022).
[27] Vgl. OpenAI (2023c).

Advanced solution approach for content injection

Since the content injection method is currently by far the most sensible option for integrating your own data into an LLM, all considerations must deal with the limited amount of content per query.

A very promising approach is to determine the parts of the document that are relevant for answering the question in an upstream AI model and to deliver only these fragments as content to the LLM within the permissible token length.

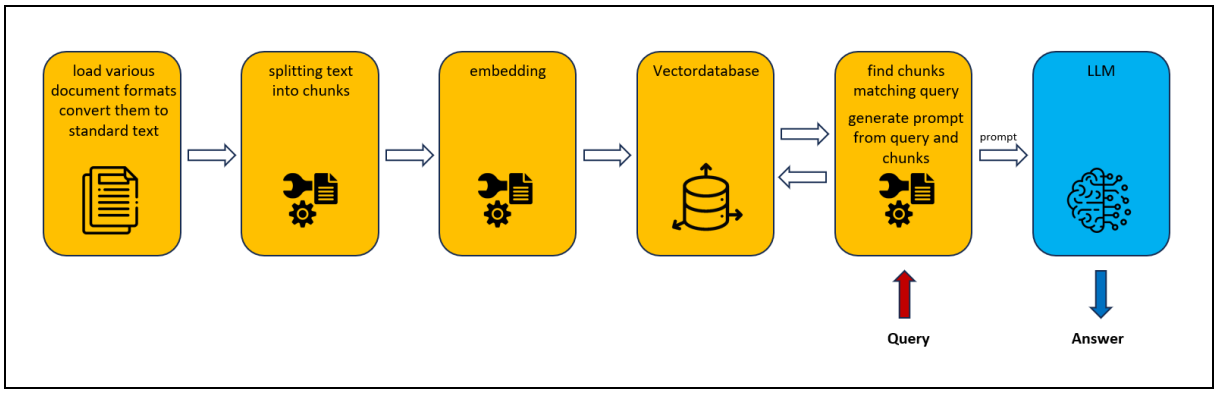


Figure 4: Exemplary structure of an upstream model [28]

The key steps of the upstream model are briefly characterized below:

1. Load and convert documents:
   The first step is to read in the various document formats with their unstructured data. They range from .pdf to word processing formats (.doc, .docx, .odt, .txt, .rtf, .md, ...), mail formats (.html, .eml, ...), web texts and formats that have to be processed with OCR text recognition. They are converted to pure ASCII and freed from all control characters.

2. Split texts:
   The text of a read-in document is split into parts, so-called chunks, which, depending on the setting, comprise a few hundred to a few thousand characters. The splitting is done in such a way that the chunks overlap by a likewise predefined number of characters. This is to ensure that the continuity of the content or the meaning is not lost or distorted during the technically necessary division.[29]

3. Embedding:
   Embedding is generally concerned with making the texts of chunks comparable to other texts or chunks. This is done by converting the texts into unique numerical vectors. For this purpose, so-called embedding models are used; for example, there is an embedding model published by OpenAI that generates the vectors for a 1536-dimensional space.[30]

[28] Own illustration.
[29] Vgl. Lee (2023).
[30] Vgl. Lee (2023).

4. Vector database:
   A vector database is a special type of database used to store, organize, and query vector data. A common application of vector databases is similarity search. One can run a query with a given vector and find the vectors in the database that are most similar to it. This is achieved through various similarity or distance metrics such as cosine similarity or Euclidean distance.

5. Find relevant chunks and generate prompt:
   In this step, the user's query is received and after prior preparation, a similarity search is started against the chunks stored in the vector database. The result, the chunk or chunks matching the query, are again converted to text and combined as content together with the query as described above to form a prompt for the LLM.[31]

Evaluation

Based on their own research, several prototypes were developed. With the help of these self-developed IT systems, the authors have demonstrated that the amount of content to be transmitted to the LLM can be minimized without loss of information. This means that an LLM can be used sensibly with the company's own – even internal – data.[32]

This approach still offers many possibilities for optimization and further development as well as adaptation to specific conditions and tasks. The further development of this approach is the focus of current research. For example, the first vector database models are now available in a size and performance that allow all of a company's unstructured data to be held in reserve and kept permanently up to date, and not just collected on demand.

Extensions

Chains

A technique known from programming for structuring processes and algorithms is the chain, a procedure in which tasks or operations are executed in a specified order. This order can be strictly sequential (sequential chain) or conditional, i.e. dependent on data ("if..then..else") (router chain). These procedures can be easily integrated into the model described above.

Sequential Chain:

As an example, let's consider the following flow:

- Request to the LLM on a specific topic
- Request to the LLM to translate the answer to the topic into another language
- Saving the translation on the hard disk

[31] Vgl. Lee (2023).

[32] When processing internal company data, the general regulations on data protection (in particular GDPR and BDSG) must be observed. In addition, suitable protective measures for intelligence protection (IP) and compliance regulations must be established. The IP problem in particular could be solved by installing an LLM locally.

In a chain, all steps are automatically executed in the correct order and consistently. Resources other than the LLM can also be included, provided that the program itself has access to them.

To give another example: It is also possible to combine several steps into one subsequent step:

- Request to the LLM on topic 1
- Request to the LLM for topic 2
- Request to the LLM on topic 3
- Request to the LLM: "Summarize answers to 1 to 3".

Router Chain:

A router chain is similar to a sequential chain, but instead of each operation being executed in sequence, a routing decision determines which operation is executed next. Here is another example:

- Step 1: Query to LLM on a topic as sentiment analysis (result positive, neutral or negative).
- Step 2: (only executed if answer to step 1 = positive)
- Step 3: (will only be executed if answer to step1 is not positive)

Agents

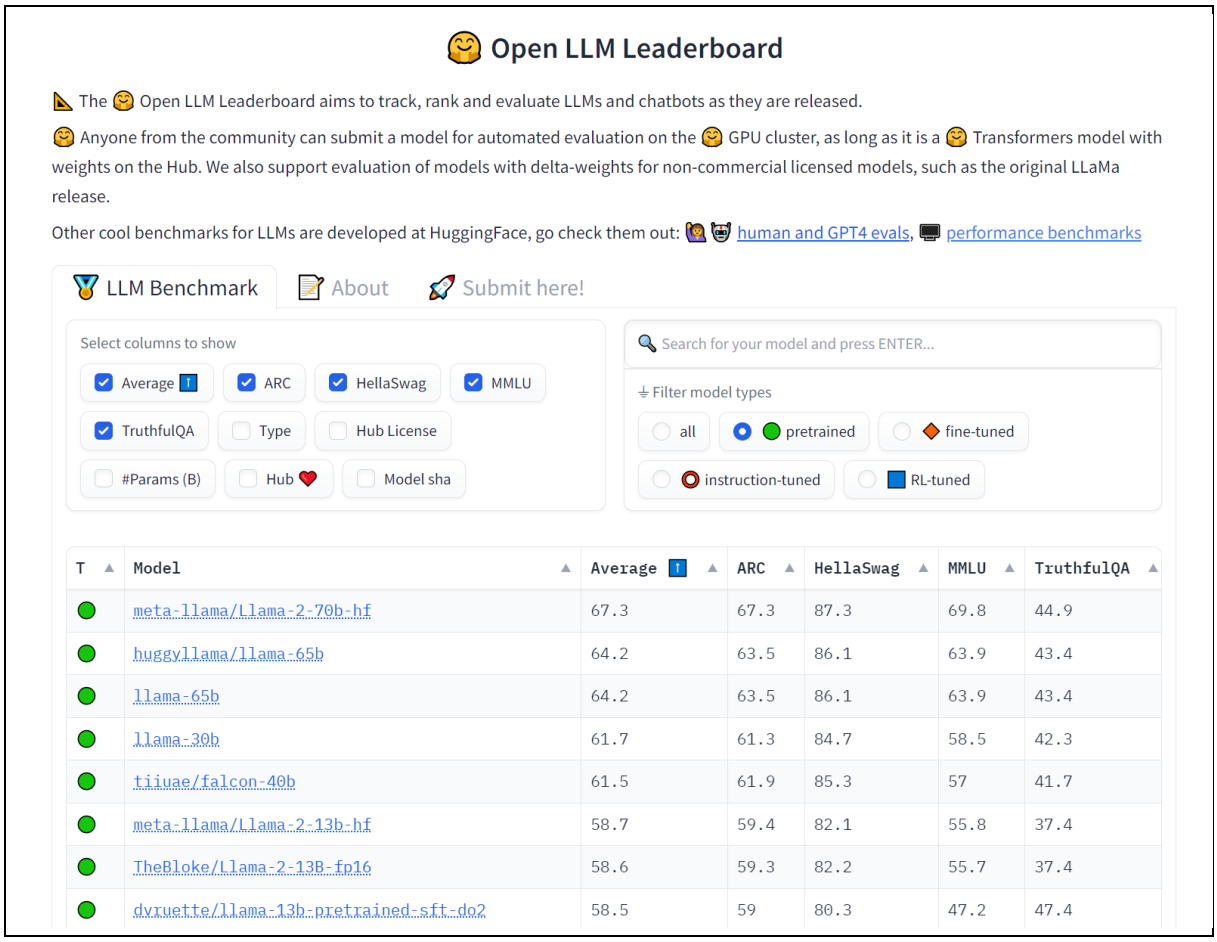
- An agent in artificial intelligence (AI) is an autonomous entity that performs actions in an environment to achieve a specific goal.
- In the context of Sequential Chains or Router Chains, an agent can take on the role of an entity that navigates through the chain and performs actions. For example, an agent in a router chain might receive the output of a previous operation, make a decision about which operation to execute next, and then send the appropriate data to that operation.
- Agents can vary in complexity. Simple agents follow set rules or algorithms to make decisions. Goal-oriented agents based on machine learning or deep learning are much more complex and are able to learn from experience and adapt their strategies over time. Agents can also interact with each other and learn from each other in a multi-agent environment.
- The model described above can be supplemented by simple agents without much effort. The challenge is to set sufficient limits to such an autonomously acting agent so that unpredictable behavior cannot occur.
- Agents that make their strategies based on feedback from an LLM are fundamentally error-prone because their results are based on probabilities. The use of these agendas can currently only be justified under human control.
- Agents are an active area of research. There are ongoing efforts to develop methods and techniques to address the challenges mentioned above. In addition, regulations and standards for the safe, fair, and responsible use of AI agents are being discussed.

Development status of Large Language Models

In the past, significant developments were published in scientific papers. ChatGPT is the first exception here. Nevertheless, all major IT companies have very similar starting conditions in the competition that is now developing. Thus, LLM developments take place in most of the well-known software companies. Among the commercial providers, OpenAI (with Microsoft as licensee) and Google with several models

each, Meta with several smaller models also as open source, as well as products of start-ups founded by former OpenAI and Google developers respectively (e.g. Anthropic) are to be mentioned primarily today.

At the same time, a powerful open source community is developing, involving renowned research institutions and others. Meta was the first company to make its AI model available to the AI community. In the meantime, numerous variants have already developed. The currently most powerful open-source LLM Falcon-40B, which has just been released, now comes very close to the commercial top group with its performance values. The community website huggingface.co publishes a leaderboard of open-source LLMs. Here, 101 pre-trained models were listed and downloadable at the time of the query.[33]



T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA
🟢	meta-llama/Llama-2-70b-hf	67.3	67.3	87.3	69.8	44.9
🟢	huggyllama/llama-65b	64.2	63.5	86.1	63.9	43.4
🟢	llama-65b	64.2	63.5	86.1	63.9	43.4
🟢	llama-30b	61.7	61.3	84.7	58.5	42.3
🟢	tiiuae/falcon-40b	61.5	61.9	85.3	57	41.7
🟢	meta-llama/Llama-2-13b-hf	58.7	59.4	82.1	55.8	37.4
🟢	TheBloke/Llama-2-13B-fp16	58.6	59.3	82.2	55.7	37.4
🟢	dvruette/llama-13b-pretrained-sft-do2	58.5	59	80.3	47.2	47.4

Figure 5: Open LLM Leaderboard[34]

Quantization of models is a very active field of research. In computer science, single precision floating point numbers are stored in a fixed length data field of 32 bits. For a simple model with 1 billion parameters, a (main) memory requirement of 4 GB is needed. If the number format is reduced to 8bit, the memory requirement is reduced to 1 GB. At the same time, computation time and teach-in time improve to a similar extent. Thus LLMs can be operated also on common servers.

[33] HUGGINGFACE (2023).
[34] HUGGINGFACE (2023).

By the change of the number format the accuracy of the represented number decreases. However, in the application of LLMs, this affects the performance much less than expected. In the list of open source models, several identical models can be found, each with different data formats.

Bibliography

Becker, Marco; Daube, Carl Heinz; Ernst, Reinking (2023): *Industrie 5.0*, IUCF Working Paper Nr. 4/2023, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg.

Bitkom e.V. (2023): *Periodensystem der Künstlichen Intelligenz*, Online-Quelle: https://periodensystem-ki.de/Mit-Legosteinen-die-Kuenstliche-Intelligenz-bauen, letzter Zugriff am: 14.08.2023.

Davenport, Thomas; Prusak, Laurence (1999): *Wenn Ihr Unternehmen wüßte, was es alles weiß... Das Praxisbuch zum Wissensmanagement*, Landsberg/Lech, Moderne Industrie.

DB Systel GmbH (2023): *Large Language Models - Was hat ChatGPT dazu zu sagen?, in: Digital Foresight Januar 2023*, Online-Quelle: https://www.dbsystel.de/resource/blob/10372378/bc7c882c57e69eeb2cee72812a139be4/Download-Digital-Trend-Impuls-ChatGPT-data.pdf, letzter Zugriff am: 14.08.2023.

Ferus, Jacob (2022): *Improving ChatGBT With Prompt Injection*,Online-Quelle: https://levelup.gitconnected.com/improving-chatgpt-with-prompt-injection-b0c0c27b7df7, letzter Zugriff am: 14.08.2023.

Grosse, Tobias; Hallgarten, Philipp; Zwaller, Lukas; Hoellig, Christoph; Asgharzadeh, Pouyan (2023): *ChatGPT & enterprise knowledge: "How can I create a chatbot for my business unit?"*, Online-Quelle: https://medium.com/next-level-german-engineering/chatgpt-enterprise-knowledge-how-can-i-create-a-chatbot-for-my-business-unit-4380f7b3d4c0, letzter Zugriff am: 14.08.2023.

Heaven, Will Douglas (2023): *Generative KI: Die Geschichte hinter ChatGPT - Der Durchbruch von OpenAI kam gefühlt über Nacht – aber dahinter steht jahrzehntelange Forschung, in: MIT Technologie Review*, Online-Quelle: https://www.heise.de/hintergrund/Generative-KI-Die-Geschichte-hinter-ChatGPT-8243968.html, letzter Zugriff am: 14.08.2023.

HUGGINGFACE (2023): *Open LLM Leaderboard*, Online-Quelle: https://huggingface.co/spaces/gsaivinay /open_llm_leaderboard, letzter Zugriff am: 14.08.2023.

Kucharavy, Andrei; Schillaci, Zachary; Maréchal, Loic; Wünsch, Maxime; Dolamic, Ljiljana; Sabonnadiere, Remi; Davic, Dimitri Percia; Mermoud, Alain; Lenders, Vincent (2023): *Fundamentals of Generative Large Language Models*, Online-Quelle: https://arxiv.org/pdf/2303.12132.pdf, letzter Zugriff am: 14.08.2023.

Lee, Ernesto (2023): *Chunking Strategies for LLM Applications*, Online-Quelle: https://drlee.io/chunking-strategies-for-llm-applications-7a37d56e2b15, letzter Zugriff am: 14.08.2023.

LexisNexis (2023): *Datenanalyse: Wertvolle Informationen aus Daten - Mit Datenanalysen gewinnen Sie Erkenntnisse aus Daten. Hier erfahren Sie alles rund um die Vorteile, Methoden und Anwendungsgebiete der Datenanalyse.*, Online-Quelle: https://www.lexisnexis.de/begriffserklaerungen/data-integration/methoden-der-datenanalyse, letzter Zugriff am: 14.08.2023.

MIT SLOAN SCHOOF OF MANAGEMENT (2023): *Tapping the power of untructured data*, Online-Quelle: https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data, letzter Zugriff am: 14.08.2023.

OpenAI (2023a): *Models - GPT-3-5*, Online-Quelle: https://platform.openai.com/docs/models/gpt-3-5, letzter Zugriff am: 14.08.2023.

OpenAI (2023b): *Models - GPT-4*, Online-Quelle: https://platform.openai.com/docs/models/gpt-4, letzter Zugriff am: 14.08.2023.

OpenAI (2023c): *Fine-tuning - Learn how to customize a model for your application*, Online-Quelle: https://platform.openai.com/docs/guides/fine-tuning, letzter Zugriff am: 14.08.2023.

Taukki, Tom (2022): *Grundlagen der Künstilchen Intelligenz - Eine nichttechnische Einführung*, Delaware, Springer Nature.

VG WORT (2023): *Merkblatt zur VG WORT für Urheber und Verlage*, Online-Quelle: https://www.vgwort.de/fileadmin/vg-wort/pdf/dokumente/Merkblaetter/Auszahlungen/Merkblatt_Allgemeines_VG_Wort_2022.pdf, letzter Zugriff am: 14.08.2023.

Wang, Chenguang; Li, Mu; Smola, Alexander J. (2019): *LanguageModelswithTransformers*, Online-Quelle: https://arxiv.org/pdf/1904.09408.pdf, letzter Zugriff am: 14.08.2023.

Zhang, Min; Li, Juntao (2021): *A commentary of GPT-3, in: MIT Technology Review 2021*, Online-Quelle: https://www.sciencedirect.com/science/article/pii/S2667325821002193, letzter Zugriff am: 14.08.2023.