Leveraging ChatGPT & Large Language Models (LLMs) in your organization

accilium 2023

A new pivot point in Artificial Intelligence (AI) has been reached

Artificial Intelligence

Machine Learning (ML)

Deep Learning (DL)

Natural Language Processing (NLP)

Large Language Models (LLMs)
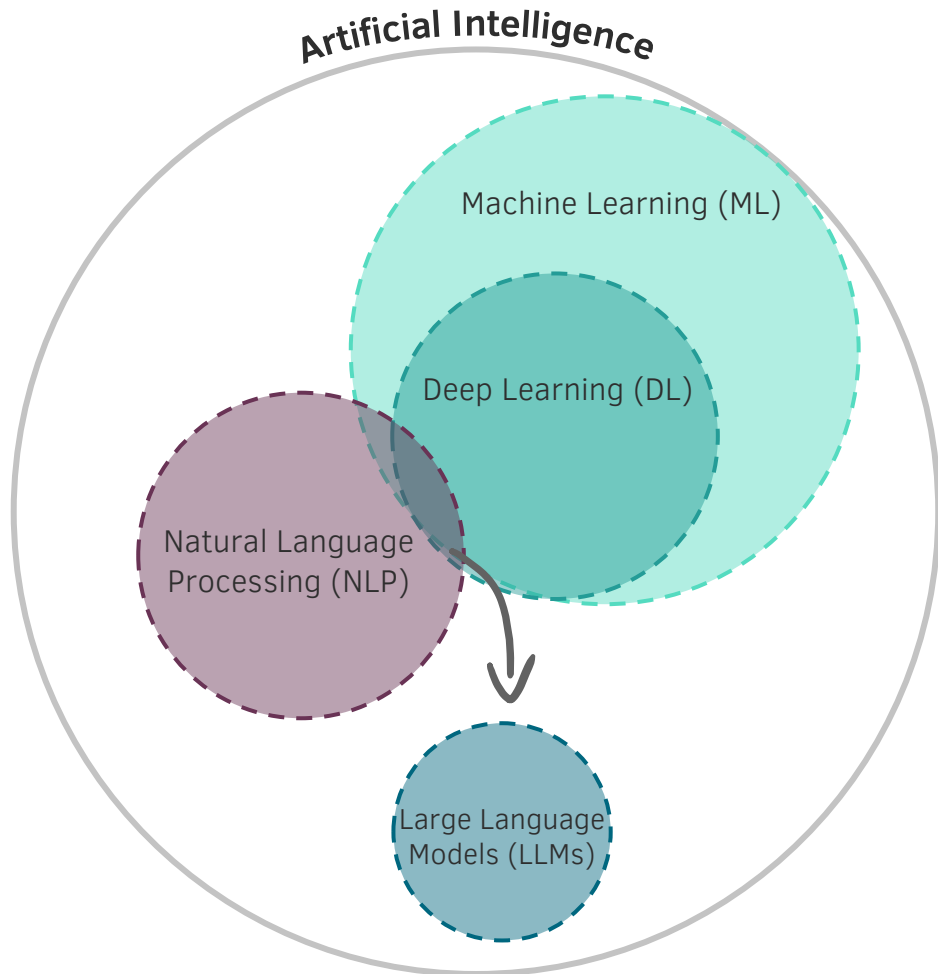
› Never in history an application has reached 100 Million users in just 2 months time[1]. This immense popularity has been reached for a good reason. It brings the power of artificial intelligence using Lange Language Models (LLM) into the hands of users around the world in a broad and easy way like never before.

› Answers to questions are not "looked up" anymore to create a response. They are not programmed and come with a human like contextual understanding using information of previous questions asked.

› Business leaders identified the significance of this new technical capability, realizing how transformable it will be to their company's business, science and society itself. Almost every job will be affected by this technology (knowledge, communication, processes)

› The question will not be anymore, if your company will use LLM's for their good, it's about the how and who will be able to generate the most value out of it.

[1]: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

accilium

How does a Large Language Model (LLM) work?

› Large Language Models use neural network algorithms which are inspired by how a human's brain is working. A deep neural network consists of layers of interconnected nodes called neurons. Each neuron receives input data, applies computations, and produces an output.

› Those networks learn by adjusting the weights of connections based on training data. Neural networks excel at generalizing patterns from training data to make predictions on new data. Humans, on the other hand, can understand complex contextual information, make nuanced judgments, and adapt to novel situations.

› Neural networks often struggle with explaining their decision-making process as well as when it comes processing data which was not part of the trainings set (new data, not available for computational processing).

› Due to the extreme high complexity and cost to build and operate a LLM (e.g., GPT-3 requires 175 billion calculations for each new question being asked) a thorough planning is essential for your organization.

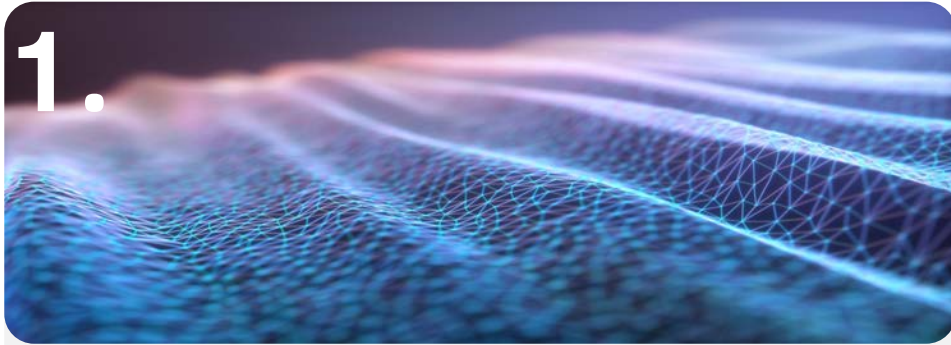Input Layer   Hidden Layers   Output Layer

← neuron
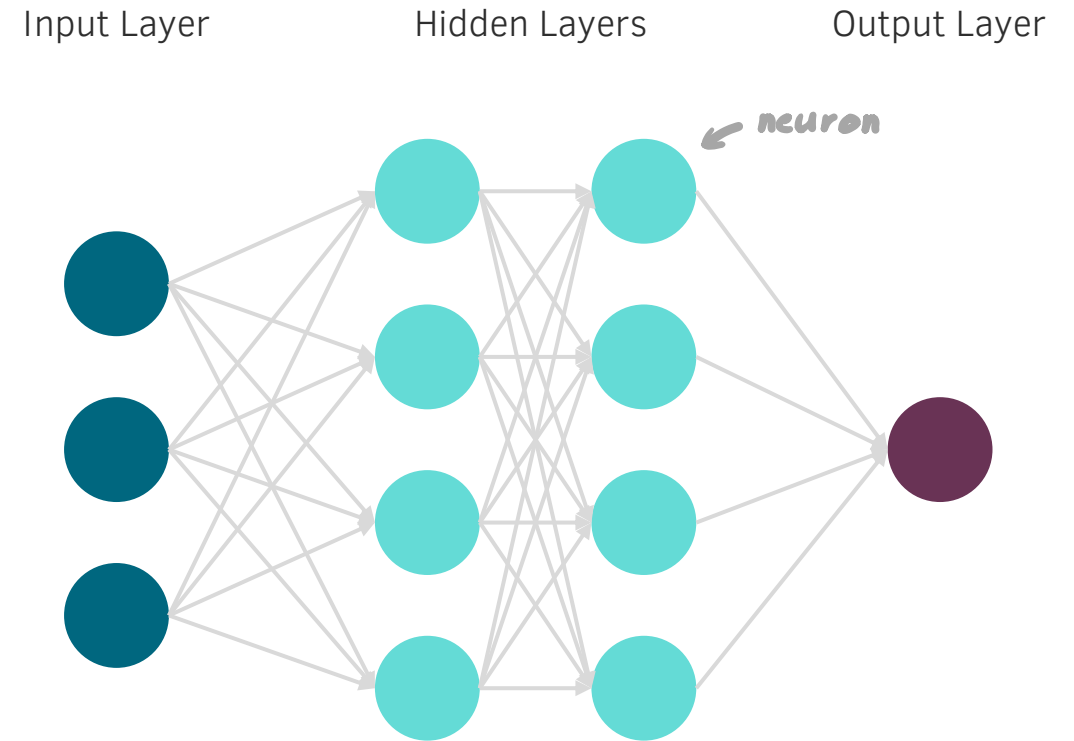
Note: GPT-3 has about 175 billion parameters; its training dataset is about 500 billion tokens, compared to 3 billion tokens for Wikipedia[1]. GPT-4 has an unknown number of parameters

[1]: https://arxiv.org/pdf/2005.14165.pdf

INTERNAL
accilium

Two ways your enterprise can benefit from LLM's

1.

Consume: Connect to existing models

Companies can consume open-source LLM applications through APIs (e.g., gpt-3.5-turbo model by OpenAI).

Tailoring can be done only at a small degree through prompt engineering techniques (how to write questions to get the best answers).

2.

Customize: Operate and run open-source model on your own

Use existing algorithms (e.g., text-davinci-003) enriched with company internal data to benefit from customized results. This requires a more sophisticated (more expensive) implementation approach and a longer implementation cycle.

Both approaches have their benefits and drawbacks. You will need to take a close look at your business needs to decide which approach fits best

INTERNAL

accilium

Consume: Connect to existing models (Model-as-a-Service)

1.
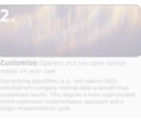
Description

Companies (e.g., OpenAI, AWS, Google Cloud Platform) offer APIs making it possible to directly connect to existing neural network models within your applications or programs. Based on the number of questions asked you will be charged a responding amount.
An integration within the Microsoft 365 application (Word, Excel, PowerPoint, Outlook etc.) will soon be available with Microsoft Copilot.
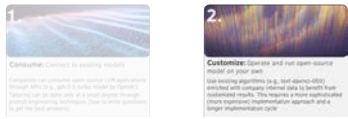
Benefits ✅

› Quick implementation of functionality
› Access to the most advanced and latest models currently existing
› Real-time usage

Limitations ⚠️

› Only publicly available information available to users
› Data privacy (data is shared with and processed by external provider)
› Potential high costs (charged per number of words generated)

INTERNAL

accilium

Customize: Operate and run open-source model on your own

2.

Description

With self managed open-source models organizations can use existing LLM that have been proved to be valuable results based on the defined use cases. Those models can run on-premise as well as within the cloud considering the potential extensive computational power required to interact with end-users in real-time.

Benefits ✅

› Models including company internal data (knowledge database, emails, contracts etc.)
› Potentially lower costs (for specific use cases customized models can be trained)
› Independence of third-party providers (regulatory, ability to explain results etc.)

Limitations ⚠️

› High internal expertise required due to high complexity
› Time to market (preprocessing, cleansing, training, model evaluation, deployment and operation required)
› Models have a much narrower field of input data (not as comprehensive answers as user is used to)

INTERNAL

accilium

5 step approach to benefit from LLM's

Given the rapid pace of innovation in this field, the most successful companies will be those which can apply the right approach and the right model for their specific use cases

1.

Start with Business engagement

2.

Identify areas of application

3.

Plan & decide on approach

4.

Implement & train users

5.

Level-up your AI capability

Top executive buy-in is required for an enterprise-wide implementation

1.

Start with Business engagement

› The potentially most transformational technology of the last years requires a well thought through approach from the beginning.

› Involving all necessary parties right from the beginning like (executives, business divisions, HR, legal, works council etc.) will be beneficial to get everybody affected involved early in the process

› Determining the business objectives and non-objectives for the usage of LLM's ensures a compliance with the company's values

› Implementing highly complex AI models is no easy task and comes with a high degree of uncertainty. Hence an adequate expectation management when it comes to time and cost of implementation has been beneficial

› Change and Communication during the implementation as well as for the years ahead will be one of the most crucial success factors when introducing LLM's within the organization.

› A concrete communication plan will support addressing potential fears and resistance early in the process

Critical success factors

› Executive buy-in from the beginning

› Involving all necessary parties

› Change and Communication will be key

› Business objectives and non-objectives

Embrace a humble start, ascend to remarkable success

2.

Identify areas of application

1. **Performance boost areas:**
   LLM's support in a broad range of business activities. Nevertheless, they work best in the following areas:

   › **Coding:** Productivity can be significantly improved by generative AI. Tasks like code generation, optimization, understanding of code, error handling and documentation

   › **Creating:** Support creative tasks like writing/summarizing texts, creating presentations, research support, new idea generation, personalized content

   › **Advising:** customer support, sales enablement, data analysis, research base advising (legal, HR, tax etc.)

   › **Automating**: business process automation will be at a transformative new level with generative AI code creation and automation platforms

2. **Application areas:**
   Which of my current job portfolios have the highest potential to be supported by a generative AI tool (check against current level of automation, data literacy, human feedback)

3. **Value case definition:**
   Calculate value cases to prioritize focus areas in the transformation journey. Define job description of the future

Critical success factors

› Deep understanding of the capabilities and limitations of LLM's

› Identify areas with high potential and data readiness for support

› Structured approach to evaluate prioritized business areas

› Value case calculation to efficiently allocate resources

accilium

Bringing a horse to water is only half the battle; you still need to make it drink

3.

Plan & decide on approach

› 2 areas (technology and people) will be the most crucial one to define a future ready set-up and need to be aligned with each other.

› **Technology** (tool/model & governance perspective):

  › Decide which approach (consume or customize) will fit your identified use cases best

  › Have a clear understanding of the costs and calculated benefits when it comes to selecting a company wide approach (required data, data availability, internal knowledge building, available tech stack and architecture, data platform requirements etc.)

  › Plan organizational changes (data governance, (ethical) guidelines etc.)

› **People** (change & communication/training):

  › Clear picture of the transformational requirements for executives, focus groups and entire workforce

  › Plan detailed steps to support the cultural transformation. Support forerunners to become more efficient, while enabling users to not leave anybody behind

  › Take a close look at potential job-, task-, and skill redesign within your organization

Critical success factors

› Clear picture of current capabilities (technological & organizational/ cultural)

› Start focused & become big. A rolling approach might be most suitable for majority of companies

› Generative AI is here to stay – think big as capabilities will rapidly increase

› Organizational change & training initiatives are crucial success factors

INTERNAL

accilium

Generative AI-tool implementations should rather be seen as marathons not sprints

4.

Implement & train users

› This step will require the biggest investments.

› Based on the approach selected and the complexity of the solutions to be implemented it significantly vary in terms of time and cost

› Companies which have already a high data maturity level will significantly benefit from already taken investments from past years.

› Internal competence development (technical and organizational) will require dedicated investments and newly defined roles to embed the change

› High computational power required by such models require a modern tech stack also considering environmental impact of such tools

› Setting up an organizational model to be ready for further technical developments within the generative AI area

› Building upon existing initiatives for digital literacy and reskilling programs will support filling even faster growing the talent gap

Critical success factors

› Parallel implementation applying "consume" & "customized" solutions

› Experiment with open-source APIs in protected areas & use knowledge for custom developments

› Setting up a realistic time-horizon to not lose supporters and users along the way

› Permanent upskilling due to newly emerging technical capabilities

Continuously innovate and improve your data maturity

5.

Level-up your AI capability

› Progress and innovation has never been as slow as it is now.

› AI companies around the world are rapidly increasing and is soon expected to cover 1 out of 4 dollars invested by Venture Capital (VC) companies[3].

› Newly emerging platforms and solutions will tremendously change the way we are working

› Customers expectation to their user experience will be benchmarked against the ones who achieve the most personalized and fluent experience

› Companies will be required to continuously improve their internal capabilities; technologically and people wise to stay relevant in the competition.

› The responsible use of AI and environmental impact of these tools will become even more important for customers

Critical success factors

› Set up internal capabilities to pivot towards an even more innovative learning culture

› Continuously scan for emerging solutions to your business problems

[3]: A sharp increase in AI-related venture capitalist investments could transform global economies and shape the future of artificial intelligence - OECD.AI

INTERNAL

accilium

What are use cases of ChatGPT ?...

Boost your software engineer's satisfaction by >60% and cut development time by 50%[1]

Description

› Writing code is the most important aspect of a software engineer's work.

› It's a time-consuming and labor-intensive process that involves substantial trail and error as well as error investigation, understanding of code purpose, structuring of code lines and documentation work within the code.

› Most companies face a shortage of skilled personnel especially for very specified programming languages leading to implementation backlogs and delays.

› AI-based coding support tools can save time for exactly those tasks, with a direct integration into the development environment.

› Cost vary based on the software provider depending on generated characters, but also fixed-fee subscriptions range from € 10 - € 30.

How to get started

› Start within a small functional team of experts, experimenting with different tools available on the market

› Check and tools performance based on accuracy and for intellectual property and security topics.

[1]: Research: quantifying GitHub Copilot's impact on developer productivity and happiness | The GitHub Blog

INTERNAL

accilium

Cut your personnel's onboarding time with LLM's

Description
› Use your internal knowledge base (documentation, PDF, pictures, internal blogs, intranet etc.) and other documentation to train a generative AI tool.
› Enrich your pre-trained model from publicly available data with company internal information to provide your personnel a 24/7 personal assistant independent of any language.
› Significantly reduce onboarding time and cut search time for administrative topics in an easy-to-use way.
› No question-answer coding required due to contextual understanding of the LLM.

How to get started
› Through planning required due to the increased complexity and investment required of this use case
› Potentially tech-stack setup required with data cleansing effort (based on the internal data availability and quality)
› Model training and quality check as well as user training to unlock full potential of AI tool

INTERNAL

accilium

Support your customer service with real time answers

Description

› Users will demand accurate, personalized and human-like answers to their inquires (will become industry standard such as hard-coded chatbots are today)

› Multilingual customer support will enable a globally unified user experience

› Customer feedback and chat history will be directly integrated into the service experience suggesting personalized offers for the customer to raise satisfaction and utilize cross-selling opportunities

› Conversational tools can be trained to recognize and respond to common customer complaints, such as issues with product quality, shipping delays, or billing errors and automatically offer potential solutions

How to get started

› Significant investment required as the described use-case requires extensive availability of cross-functional data and business input

› Detailed market screening of potential out-of-the-box solutions, fulfilling the level of complexity of inquires

› Expansion stage of internal assistant to train and gather feedback first internally, before direct customer interaction

accilium

Use AI as your personal assistant and become a more efficient yourself

Description

› The native integration of large language models into M365 apps (Word, Excel, PowerPoint, Outlook, Microsoft Teams etc.) through Microsoft Copilot[1] will rapidly change the way of working.

› The Work Trend Index 2023 lists "AI delegation (prompts)"[2] among the top new skills for employees. Not just writing efficient prompts will be an essential skill also where and how to use these tools will be essential.

› All our consultants use LLM's regularly to reduce time for administrative work or tasks that can easily be done by machines to focus on creative processes

› Compliance policies and Intellectual Property (IP) matters are essential to be addressed; but efficiency gains and benefits to employee's satisfaction level should leading drivers

How to get started

› Define top-down the strategic goals and values concerning its use

› Bring all stakeholders onto one table to get everybody on board

› Evaluate and develop short-term upskilling (use cases, prompt engineering) and long-term reskilling plans to be future ready and guide your people along the path

[1]: Introducing the Microsoft 365 Copilot Early Access Program and the 2023 Work Trend Index – Stories

accilium

Alex Rinner

Manager
+43 676 553 68 83
alex.rinner@accilium.com

David Schneiderbauer

Associate Manager
+43 676 903 54 59
david.schneiderbauer@accilium.com