# CONTRIBUTION TO AN AI STRATEGY FOR GERMANY

Version 1.4 from November 2, 2023

Artificial Intelligence (AI) combines several sub-disciplines of computer science with the intention of making machines more intelligent. One of these sub-disciplines deals with large language models (LLMs). LLMs process natural language. Other models, on the other hand, deal with images, movies, music, protein folds, etc. In domains where machines previously failed to deliver adequate results in the past, LLMs are becoming increasingly reliable.

This reliability comes from the fact that models are trained with huge amounts of data and, based on their programming, independently define several hundred billion parameters for their intended use in some cases. The purposes of LLMs are, for example, summarizing texts or answering questions.[1] A prominent example of an LLM application is ChatGPT. Not only do LLMs require an extensive database and need to be programmed, but developing them also requires specialized computing capacities to train them. In 2020, training a comprehensive LLM cost around USD 10 million,[2] which is a relatively high market entry barrier for new market participants.

As directed by LEAM, Germany urgently needs to catch up in regards to these kinds of generic LLMs.[3] Established companies, like OpenAI, Anthropic, Microsoft, Google as well as Baidu and Huawei, are developing at a very fast pace[4] and investing huge sums of money in this field. Imitating this approach would not directly differentiate German companies. Even though digital sovereignty is often mentioned when it comes to Europe's strategic autonomy in the digital field, this has proved to be an insufficient competitive feature for European companies looking to assert themselves in the digital market, e.g., in the case of establishing European HyperScale technology. Therefore, for Germany to have any chance of success, the answer lies in a different business and development model.

Several aspects are crucial for the competitiveness of LLMs: On the one hand, the quantity and content of the data used to train the models define the value of the models. German industry has high-quality, specialized data in various fields which can be used to train its own LLMs, including data from manufacturing and automation processes. Furthermore, in terms of making public data available for research and development, there is still great potential in Germany (and the EU). On the other hand, it is essential that sufficient computing capacity is accessible for training the models. There is a strategic need for investment here because the available computing capacity in Germany is limited. The capacities cannot simply be bought in since NVIDIA's graphics processors, among other things, must currently be procured primarily from the USA. The traceability, transparency, and trustworthiness of the models will also play an important role in the future. One opportunity arises from the fact that these criteria are currently often at odds with the business models of established AI companies.

**When developing AI, Germany should focus on its competencies and consider the needs and assets of local technology-leading industries, society, and science. Only then, and only with swift action, can decisive, competitive advantages emerge despite the current backlog.**

[1] https://www.cloudcomputing-insider.de/was-ist-ein-large-language-model-llm-a-9b7bdd0c3766b5a9c0ee1e0c909790a3/
[2] Sharir, Peleg, and Shoham (2020): The Cost of Training NLP Models, URL: https://arxiv.org/pdf/2004.08900.pdf
[3] The Large European AI Models (LEAM) initiative calls for a high-performance computing center to be established specifically for the development of AI applications and the research thereof in the field of artificial intelligence (see https://ki-verband.de/leam-grosse-ki-modelle-in-europa/).
[4] Ratnaparkhi (May 30, 2023): Evolution of NLP. URL: https://medium.com/@Ratnaparkhi/evolution-of-nlp-unleashing-the-potential-of-large-language-models-and-prompts-fee7ba02f72b

To achieve this goal, SPRIND proposes four parallel strands of action:

1. **Enable SPRIND Challenges for the development of application-specific AI** for the concrete use and further development of AI in the verticals.
2. **Enable a SPRIND Challenge for the development of data pools** for the provision and curation of high-quality, unique data pools from business, administration, research, and society.
3. **Promote open source LLMs** analogous to the Sovereign Tech Fund model to boost local, future champions.
4. **Provide computing capacity** with parallel **development of specialized hardware**.

**This document describes a mission-oriented approach that can be quickly put into action. SPRIND can be used as a tool for implementing Challenges and providing project funding. SPRIND can be commissioned in-house.**

## 1. Enabling SPRIND Challenges for the development of application-specific AI

SPRIND Challenges have proven to be a powerful tool for developing innovations in complex subject areas. Five SPRIND Challenges are currently in progress.[5] Challenges address actors from science, business, and civil society and thereby highlight all aspects of a field. The process is quick, requiring only a few weeks between the end of the submission deadline and the allocation of funds. Challenges are efficient because they make use of pre-commercial contracting without grant or project funding logic, which involves costly controlling, and without excluding smaller actors a priori. With the multi-stage competition format, SPRIND Challenges also generate a great deal of momentum.

**SPRIND therefore is proposing a series of Challenges that address promising competencies of German science, business, and civil society with the aim of applying AI in these fields and driving forward developments in the competition format. This list can be extended at any time and can be adapted to the partners' focus of interest, for example:**

- **AI Engineer** (e.g., supporting development in mechanical engineering, robotics)
- **AI Scientist** (e.g., generating data, deriving and testing hypotheses in pharmaceutical research, molecular biology, or materials research)
- **AI Lawyer** (e.g., streamlining legislation, proposal writing and processing, contracting, and taxation)
- **AI Doctor** (e.g., supporting anamnesis, pathogenesis, salutogenesis)
- **AI Public Officer** (e.g., optimizing and monitoring administrational work processes)
- **AI Reflector** (e.g., verifying and testing LLMs or the authenticity and identity of documents)
- **AI Developer** (e.g., automated software development, quality assurance)
- **AI Teacher** (e.g., rethinking education in the age of AI)

## 2. Enabling a SPRIND Challenge for the development of data pools

Data determines the quality of the models and their performance in the pertinent cases of application. Despite several initiatives by the federal government, it has not yet been possible to provide a sufficiently large amount of relevant data. The main reasons for this are actors' unwillingness to share data and also a general uncertainty regarding the General Data Protection Regulation. There is also a lack of targeted support for data-oriented projects and the associated acquisition of skills in handling data.

[5] https://www.sprind.org/de/challenges/

In many of their systems, German companies have large volumes of data and various data classifications. Most companies are not capable of preparing or even merging their own data. Because of this, companies often know little about the potential that lies in their data. Another SPRIND Challenge could focus on enabling companies to recognize and leverage this potential.

A SPRIND Challenge could accelerate the development of data pools in Germany by activating existing data and bringing together different sources.[6] To develop the greatest possible innovative power, the Challenge would define certain framework conditions, like that the data must be openly accessible despite a certain degree of exclusivity and that a business model be developed for subsequent monetization.

Established companies, like Google, are already highly advanced when it comes to developing their horizontal data pools. However, considering the in-depth data of German industrial companies, it does not make sense to imitate those other companies. Instead, it would be possible to quickly establish vertical data pools across several stages of a value chain if there is a high level of willingness to cooperate. Plus, it also enables significantly greater innovation potential than with horizontal data pools. Therefore, the aim of the Challenge could be a kind of vertical crawler for creating data pools. Instead of focusing on the finished data pool, the initial objective could be to develop a crawler demonstrator as an instrument for establishing that pool. This demonstrator could then be subjected to more extensive practical tests in the second stage of the Challenge. Both the developers of a crawler and the data providers can take part in the Challenge.

## 3. Promoting Open Source LLMs

Commercialization quickly began after the success of previously open models by companies like OpenAI. Thus far, this has culminated in Microsoft investing 10 billion US dollars in OpenAI, which began as an open source, non-profit.[7] This investment marked the end of this company's open source ambitions.[8] The AI open source community rapidly developed as a result. It began to develop its own models to stop this centralization of power and to enable LLMs for general public use.[9] Consequently, the number of open source LLMs has significantly increased recently.[10] This increase suggests that it will be easier for new market players to offer services based on open source LLMs because a major hurdle has been removed.[11] Market entry with open source LLMs does not require investment in the development and training of entirely new, proprietary LLMs. Instead, open source LLMs can be used as a basis and, for example, adapted for specific areas of application. This fine-tuning is inexpensive and LLMs can be optimized much faster compared to an LLM that is developed centrally and proprietary.[12] Finally, these open source LLMs can also run on private computers, which means they can be used assuredly in numerous scenarios. Compared to cloud solutions, this makes them relatively protected and safe from the unintentional outflow of data.

**SPRIND proposes to support open source LLMs and other types of models. Funding could, for example, correspond to the Sovereign Tech Fund.[13] Plus, targeted support should be given to local companies that commercialize the models, e.g., the highly successful open source enterprise model that companies like SUSE Linux, RedHat, and others have been using for decades. The projects initiated by Gaia-X, such as Catana-X and the upcoming Manufacturing-X, as well as the other**

[6] Projects that have already been started can also be accelerated (like the Idealist Project from the Helmholtz Zentrum München).
[7] https://www.golem.de/sonstiges/zustimmung/auswahl.html?from=https%3A%2F%2Fwww.golem.de%2Fnews%2Fopenai-microsoft-investiert-in-chat-gpt-unternehmen-2301-171383.html&referer=https%3A%2F%2Fwww.google.com%2F
[8] https://de.wikipedia.org/wiki/OpenAI
[9] A list of current Open LLMs can be found at https://github.com/eugeneyan/open-llms
[10] https://medium.datadriveninvestor.com/list-of-open-source-large-language-models-llms-4eac551bda2e
[11] https://cobusgreyling.medium.com/large-language-models-are-being-open-sourced-537dcd9c2714
[12] https://bdtechtalks.com/2023/05/08/open-source-llms-moats/
[13] https://sovereigntechfund.de/de/

**lighthouse projects initiated by the BMWK, could also experience a significant AI boost (for example, like the TEAM-X and Health-X dataLOFT in the health sector).**

4. Providing AI-Specific Hardware

Current computing power available for AI activities is considered too insufficient, and the existing, powerful systems are predominantly used by research institutions.[14] SPRIND therefore recognizes three important needs in the short and medium terms:

First, SMEs, industry, and digital society should be empowered to use AI by having special AI hardware developed for them. It is important to make hardware generally available, to promote knowledge about the necessity of using AI, to provide human and financial resources, and to encourage expertise and security awareness among users.

Second, so-called middleware should be developed, which provides appropriate software interfaces and services for complex applications.

Third, advanced, promising, and rapidly available hardware solutions for high-performance computing should be supported.

From SPRIND's standpoint, these necessities can be concretely met as follows:

a) Expand existing supercomputer centers with AI-specific, "classical" hardware, possibly in collaboration with the European Union.
b) Support hardware development projects that foreseeably promise the greatest success and whose results will quickly be available for an application. From the work in the last three years and from SPRIND's New Computing Concepts Challenge, the SPRIND environment has already enabled several approaches.[15]
c) Develop a software solution with all the necessary software components (full-stack development) to make the underlying hardware easier to use.
d) Enable AI-supported chip and software development, e.g., to optimize FPGA designs for use in AI applications. Reference cases should be created here to demonstrate the benefits of AI-assisted development.

In the medium term, Germany should develop its own AI-specific hardware, both for previously unoccupied niche applications and for those that provide an advantage in terms of efficiency by a factor of at least 10 compared with the current technology. At the same time, algorithms and software should be developed to match the new hardware, maximizing the benefits. SPRIND has two examples from its current work:

- **Memristors and capacitors** are among the most important components of future hardware for AI.[16] That is why an ecosystem for developing memristors should be established and developed. SPRIND's subsidiary Memlog GmbH can provide support in this regard. Also, the teams in SPRIND's New Computing Concepts Challenge can help initiate an appropriate ecosystem. This could be done, for example, with the special compilers and arithmetic logic units for processors; these are already being developed and adapted to memristors.
- Accelerate projects from SPRIND's New Computing Concepts Challenge to **create alternative hardware** for accelerating AI and increasing efficiency with the goal of achieving a factor of at least 10 in each case.

[14] https://background.tagesspiegel.de/digitalisierung/wo-stecken-die-ki-milliarden
[15] https://www.sprind.org/de/challenges/newcomputing
[16] https://www.enas.fraunhofer.de/de/geschaeftsfelder/micro_and_nanoelectronics/Beyond-CMOS-und_HF-Bauelemente/Memristoren_fuer_die_Rechner_von_morgen.html

## 5. Summary

To keep up with the rapid pace of developing technologies, we suggest that these four plausible strands of action be started simultaneously and as soon as possible:

1. Enable Challenges for the development of application-specific AI
2. Enable a Challenge for the development of data pools
3. Promote Open Source LLMs
4. Provide computing capacity with parallel development of specialized hardware