Business plan for a DIN SPEC project according to the PAS procedure on

# ”Knowledge Graphs for Language Models and Language Models for Knowledge Graphs: Hybrid Applications of symbolic and subsymbolic AI”

Status:

**For developing the DIN SPEC after adoption on 2024-09-18**

Recipients of this business plan are requested to name all patent rights known to them to be relevant to the project and to make available all supporting documents.

Berlin, 2024-09-19 (Version 2)

## Table of contents

# 1 Status/version of the business plan

For developing the DIN SPEC after adoption on 18.09.2024
No Changes to the previous version 01

# 2 Initiator and other consortium members

- **Initiator:**

| Person/Organization | Short description |
|---|---|
| Prof. Dr. Georg Rehm | Prof. Dr. Georg Rehm works as a Principal Researcher in the Speech and Language Technology Lab at the German Research Center for Artificial Intelligence (DFKI), in Berlin. He is also an adjunct professor at Humboldt-Universität zu Berlin and deputy spokesperson of the DFKI Lab Berlin. Georg Rehm was appointed honorary professor for outstanding achievements in research and education at Humboldt-Universität zu Berlin where he is affiliated with the Institut für deutsche Sprache und Linguistik. In October 2018, Georg Rehm was awarded the honorary appointment as a DFKI Research Fellow for outstanding scientific achievements and special accomplishments in technology transfer. |

- **Other potential participants:**

  This DIN SPEC will be developed in a consortium (temporary body) that is open to any interested party. The participation of other experts would be helpful and is desired. It is recommended that

  – Organizations that use or want to use language models in conjunction with knowledge graphs

  – Organizations that actively develop language models or knowledge graphs and possibly corresponding hybrid systems

  – etc.

  take part in the development of this DIN SPEC.

- **Organizations**[Fehler! Textmarke nicht definiert.] **that have registered for participation/Organisations that have adopted this business plan (consortium members):**

| Person | Organization |
|---|---|
| Daniel Baldassare | Doctima |
| Moritz Busch | Fraunhofer FIT |
| Thomas Usländer | Fraunhofer IOSB |
| Michael Wetzel | Coreon |

| Person | Organization |
|---|---|
| Sabine Mahr | WordBSign |
| Thorsten Trippel | Universität Tübingen |
| Harald Sack, Heike Fliegl, Wolfram Horstmann, Sven Hertling | FIZ Karlsruhe |
| Rene Pietzsch | Eccenca |
| Alan Akbik, Max Ploner | Humboldt-Universität zu Berlin |
| Desiree Heim, Markus Schröder, Fabio Barth, Julian Moreno Schneider, Leonhard Hennig, Georg Rehm | DFKI |
| Daniel Burkhardt | Ferdinand-Steinbis-Institut |
| Jan Rösler | DIN e. V. |
| Claudia Reinel | DIN e. V. |

# 3 Objectives of the project

## 3.1 General

While language models represent the state of the art in science and technology for a variety of language technology tasks, there are numerous knowledge bases, knowledge graphs and ontologies that contain symbolic knowledge or semantic knowledge in symbolic representation. There are currently no standards for how such knowledge bases and ontologies can be integrated into language models and made usable safely (evaluation of ”criticality”) according to the respective requirements. This aspect also concerns the merging and integration of different knowledge bases and knowledge packages.

The standardization of approaches such as knowledge graphs and ontologies in large language models, which can be integrated and made usable, serves the use of existing symbolic knowledge bases in the context of state-of-the-art research and technology of language technologies, which are typically based on large language models. The merging, integration and management of ontologies and ontology modules or ontology packages from different sources should also be considered. Approaches can also be considered as to how (more ontology-based) world knowledge can be integrated into (more document-based) knowledge graphs. These aspects are important and relevant for the use of symbolic knowledge bases (i.e. ontologies) in the context of knowledge graph-based applications.

## 3.2 Planned scope

This DIN SPEC is intended to define requirements for the use of large language models (LLMs) for the construction or validation of knowledge graphs (KGs) and - vice versa - for the use of knowledge graphs for generative large language models, e.g. to avoid hallucinations and generate factually more correct answers. Since this is a very dynamic and very active field of research, the planned DIN SPEC clarifies in particular the terminology, the possible applications as well as the basic application possibilities (LLMs for KGs and KGs for LLMs) and their respective advantages, disadvantages and challenges. The document is applicable to research, development and deployment of ML techniques, including but not limited to natural language processing (NLP),

knowledge management and domain-specific applications using LLMs and KGs in the above sense.

Occupational health and safety measures are not the subject of this document. This document does not specify any requirements for the safety (in the sense of safety and security) and health protection of employees. Furthermore, no safety levels and safety functions are defined for products or applications that contribute to the prevention of accidents or health hazards in the work context.

## 3.3 Related activities

The subject of the planned DIN SPEC is not at present the subject of a standard. However, there are committees, standards and/or other technical rules that deal with related subjects and thus need to be taken into account - and involved or incorporated, where necessary - in this project:

- NA 043-01-42 GA “DIN/DKE Gemeinschaftsarbeitsausschuss Künstliche Intelligenz“
- NA 105-00-06-01 GAK “Gemeinschaftsarbeitskreis NAT/NIA, Natural Language Processing“
- German Standardization Roadmap on AI (2nd Edition)
- DIN DKE SPEC 99001:2022-05 “Definition einer Erfolgsmethode zum Labelling von Daten zum Training künstlicher Intelligenz - Anwendungsfokus: Question-Answering“
- F.746.5: Framework for a language learning system based on speech and natural language processing (NLP) technology
- DIN SPEC 92005:2024-03 “Künstliche Intelligenz - Quantifizierung von Unsicherheiten im Maschinellen Lernen“
- GB/T 42131 “Artificial intelligence - Technical framework of knowledge graph”
- ISO/IEC AWI TR 23281 “Artificial intelligence — Overview of AI tasks and functionalities related to natural language processing”
- ISO/IEC AWI 23282 “Artificial Intelligence — Evaluation methods for accurate natural language processing systems

# 4 Work programme

The aim of the project is to develop a DIN SPEC according to the PAS procedure (see www.din.de/go/din-spec-en). The DIN SPEC shall be consistent with the body of German standards and shall not be in conflict with any DIN Standard.

The kick-off meeting has taken place virtually on **2024-09-18, 3pm to 5pm**. The project duration will be about 4 months.

At this kick-off meeting, the consortium for developing the DIN SPEC was constituted, further organizational issues will be decided on and clarified, and, where possible, work on the subject matter will be begun.

A draft for public commenting will not be published.

A total of 1 project meeting (besides the kick-off) and up to 3 web conferences will be held (including 1 meeting for the final approval by the consortium), during which the content of the DIN SPEC will be presented, discussed and approved. The content of

the DIN SPEC can be drawn up by individual consortium members or in working groups.

Dates of further meetings and/or web conferences are to be agreed on within the consortium in consultation with DIN.

The DIN SPEC will be drawn up in English (language of meetings, minutes, etc.). The DIN SPEC will be written in English.

NOTE The calculation covers only one language version. Please keep in mind the fact that other language versions involve additional expenses; for this reason, they shall be agreed on separately. If another language version is desired, Beuth/DIN can provide a translation. Requests for translations are to be submitted after the DIN SPEC manuscript has been approved for publication.

# 5 Resource planning

Each consortium member shall bear the expenses he/she incurs as a result of participation in the project.

If the DIN Executive Board approves the project, the initiator of the project will then conclude a contract with DIN.

Consortium membership and participation in the project meetings is free of charge, as the costs incurred by DIN throughout the performance of this project will be financed by funding from the research project ”Geschäftsstelle KI” funded by the Bundesministerium für Wirtschaft und Klimaschutz (BMWK) as per the funding announcement (funding reference no.: 46DIN21F5).

# 6 Rules of cooperation in the DIN SPEC consortium

This project is governed by the PAS procedural rules. All interested parties and consortium members are to inform themselves of these procedures by going to www.din.de/go/din-spec-en.

The consortium will be constituted during the course of the kick-off meeting. The kick-off meeting will not take place until the business plan has been published and approved by DIN's Management Board. The consortium shall comprise at least three members from different organizations[1]. It is not necessary that these members come from different areas and represent different stakeholders. By approving this business plan, the interested parties declare their willingness to participate in the consortium and will be formally named as consortium members, with the associated rights and duties. Participants at the kick-off meeting who do not approve the business plan are not given the status of a consortium member and are thus excluded from further decisions made during the kick-off meeting and from any other decisions regarding the project.

If an organization (e.g. an association) sends someone who is not an employee to the consortium, this person shall be authorized by the organization, who shall provide proof of this to DIN.

Each consortium member is entitled to vote and has one vote. If an organization sends several experts to the consortium, that organization has only one vote, regardless of

[1] Organizations are legal entities and natural persons, insofar as they participate in business transactions on a commercial or freelance basis. If several legal entities are part of a group or a corporate structure within the meaning of Section 15 of the German Stock Corporation Act (§ 15 Aktiengesetz) or Section 271 (2) of the German Commercial Code (§ 271 Absatz 2 Handelsgesetzbuch), they are deemed to be one organization.

how many consortium participants it sends. Transferring voting rights to other consortium members is not permitted. During voting procedures, decisions are passed by simple majority; abstentions never count.

As a rule, the consortium is closed once it is constituted. The current consortium members shall decide whether any additional members will be accepted or not.

During the kick-off meeting, the consortium members shall elect a consortium leader, who is responsible for content management and any decision-making and voting procedures. The leader is supported by the responsible DIN Project Manager, whereby DIN will always remain neutral regarding the content of the DIN SPEC. Furthermore, the DIN Project Manager shall ensure that DIN's rules of procedure, rules of presentation, and the principles governing the publication of DIN SPEC have been observed. Should a consortium leader no longer be able to carry out his/her duties, the DIN Project Manager shall initiate the election of a new leader.

The DIN Project Manager is responsible for organizing and leading the kick-off meeting, in consultation with the initiator. Further project meetings and/or web conferences shall be organized by the DIN Project Manager in consultation with the consortium leader.

If consortium members cannot be present when the DIN SPEC or its draft is approved, an alternative means of including them in the voting procedure shall be used (e.g. in writing, electronically).

All consortium members who voted for the publication of the DIN SPEC or its draft will be named as authors in the Foreword, including the organizations which they represent. All consortium members who voted against the publication of the DIN SPEC or its draft, or who have abstained, will not be named in the Foreword.

Any expansion of the consortium at a later date is decided on by the members making up the consortium at that time. It is particularly important to consider these aspects:

a) expansion would be conducive to shortening the duration of the project or to avoiding or averting an impending delay in the planned duration of the project;

b) the expansion would not result in the project taking longer to complete;

c) the new consortium member would not address any new or complementary issues beyond the scope defined and approved in the business plan;

d) the new consortium member would bring complementary expertise into the consortium in order to incorporate the latest scientific findings and state-of-the-art knowledge;

e) the new consortium member would actively participate in the drafting of the manuscript by submitting concrete, not abstract, proposals and contributions;

f) the new consortium member would ensure wider application of the DIN SPEC.

To allow the legal reproduction and distribution of results for the purposes of project work, the consortium members grant DIN rights of use on the basis of the copyright that will accrue to them for the results of their work on the DIN SPEC. The transfer of these utilization rights does not prevent the consortium members from using and further developing the knowledge, experience and findings they bring to the project.

Consortium members are requested to inform DIN of all patent rights known to them to be relevant to this DIN SPEC project.

Subsequent changes to the scope (Section 3.2) or to the resource planning (Section 5) require, in addition to a two-thirds majority of all votes cast, the approval of DIN.

# 7 Contacts

- **Consortium leader:**
  Prof. Dr. Georg Rehm
  Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
  Alt-Moabit 91c
  10559 Berlin
  Tel.: +49 (0) 30 23895 1833
  E-Mail: georg.rehm@dfki.de

- **Project manager:**
  Jan Rösler
  DIN German Institute for Standardization
  Am DIN-Platz
  Burggrafenstraße 6
  10787 Berlin
  Tel.: + 49 (0) 30 2601-2815
  E-Mail: Jan.Roesler@din.de

- **Initiator:**
  Prof. Dr. Georg Rehm
  Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
  Alt-Moabit 91c
  10559 Berlin
  Tel.: +49 (0) 30 23895 1833
  E-Mail: georg.rehm@dfki.de

## Annex: Project schedule (preliminary)

| DIN SPEC Project | 2024 | | | | | | | | | | | | 2025 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jul | | Aug | | Sep | | Oct | | Nov | | Dec | | Jan | |
| **Initiation** | | | | | | | | | | | | | | |
| 1. Request and review | | | | | | | | | | | | | | |
| 2. Business plan drawn up | | | | | | | | | | | | | | |
| 3. Publication of business plan | | | | | | | | | | | | | | |
| **Development phase** | | | | | | | | | | | | | | |
| 4. Kick-off meeting/consortium constituted | | | | | | | | | | | | | | |
| 5. DIN SPEC drawn up | | | | | | | | | | | | | | |
| 6. DIN SPEC approved by consortium | | | | | | | | | | | | | | |
| **Publication** | | | | | | | | | | | | | | |
| 7. Review and release by DIN | | | | | | | | | | | | | | |
| 8. Publication of DIN SPEC | | | | | | | | | | | | | | |
| **Milestones** | | | | | | **K** | **W** | | **M/W** | **A** | | | | |

**K** - Kick-off **M** - Project meeting **W** - Web conference **A** - Adoption of DIN SPEC