AIDA-Bot 2.0: Enhancing Conversational Agents with Knowledge Graphs for Analysing the Research Landscape

Antonello Meloni[1], Simone Angioni[1], Angelo Salatino[2], Francesco Osborne[2,3], Aliaksandr Birukou[4], Diego Reforgiato Recupero[1(✉)], and Enrico Motta[2]

[1] Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy
{antonello.meloni,simone.angioni,diego.reforgiato}@unica.it

[2] Knowledge Media Institute, The Open University, Milton Keynes, UK
{angelo.salatino,francesco.osborne,enrico.motta}@open.ac.uk

[3] Department of Business and Law, University of Milano Bicocca, Milan, Italy

[4] Springer-Verlag GmbH, Tiergartenstrasse 17, 69121 Heidelberg, Germany
aliaksandr.birukou@springer.com

Abstract. The crucial task of analysing the complex dynamics of the research landscape and uncovering the latest insights from the scientific literature is of paramount importance to researchers, governments, and commercial organizations. Springer Nature, one of the leading academic publishers worldwide, plays a significant role in this domain and regularly integrates and processes a variety of data sources to inform strategic decisions. Since exploring the resulting data is a challenging task, in 2021 we developed AIDA-Bot, a chatbot that addresses inquiries about the research landscape by utilising a large-scale knowledge graph of scholarly data. This paper presents the novel AIDA-Bot 2.0, which can both 1) support a set of predetermined question types by automatically translating them to formal queries on the knowledge graph, and 2) answer open questions by summarising information from relevant articles. We evaluated the performance of AIDA-Bot 2.0 through a comparative assessment against alternative architectures and an extensive user study. The results indicate that the novel features provide more accurate information and an excellent user experience.

Keywords: Conversational Agents · Knowledge Graphs · Scholarly Data · Science of Science · Scholarly Analytics

1 Introduction

The challenging task of analysing the dynamics of the research landscape and uncovering the latest insights from the scientific literature is of paramount importance to researchers, governments, and commercial organizations. Springer Nature (SN), one of the leading academic publishers worldwide, plays a significant role in this domain. Their Computer Science portfolio comprises over 170

T. R. Payne et al. (Eds.): ISWC 2023, LNCS 14266, pp. 400–418, 2023.
https://doi.org/10.1007/978-3-031-47243-5_22

journals and provides extensive coverage of the top conferences, resulting in approximately 900 proceedings volumes per year. Therefore, SN needs to continuously monitor the academic landscape to inform short- and long-term strategic decisions. For instance, editorial teams must regularly assess the quality of research venues, discover emerging communities and research areas, identify key researchers that could organise special issues or edit books on strategic topics, and evaluate the potential impact of new technologies on the industrial sectors. To this end, SN relies on a robust data pipeline that combines various large-scale academic datasets and provides analytical functionalities based on cutting-edge data mining and machine learning solutions. Semantic Web and Knowledge Graph technologies play a pivotal role in this infrastructure as they enable the integration and querying of diverse information from heterogeneous data sources [33]. Since 2014, SN and The Open University have collaboratively explored the application of semantic technologies in this space, resulting in numerous tools that have been integrated into the SN workflow [30,41–43]. In 2020, this collaboration led to the development of the AIDA Dashboard [4], a web application that allows users to assess and compare journals and conferences according to a comprehensive range of analytics. The AIDA Dashboard relies on the Academia/Industry DynAmics (AIDA) Knowledge Graph[1] [3], a knowledge base which integrates multiple data sources (e.g., OpenAlex, DBLP, DBpedia, CSO) and describes over 25 million scientific papers. A freely available version of the AIDA Dashboard was first launched at ISWC 2022 [2].

The AIDA Dashboard proved to be an excellent way to explore the multifaceted data in AIDA KG. However, it also suffers from two inherent limitations: 1) it mainly focuses on venues (journals and conferences), and 2) it only reports a fixed set of precomputed analytics. Therefore, it does not allow users to exploit the full range of information in the AIDA KG by formulating specific queries over all the described entities, including researchers, articles, organisations, countries, venues, and research topics. For example, an editor cannot retrieve the top researchers in a certain field, compare the academic impact of two organizations, or find all the articles in a journal that focus on a combination of topics. Implementing an interface based on a formal query language, such as SPARQL or SQL, was also not an option, as most users would not be comfortable with this solution. We thus decided to develop an alternative solution based on a conversational agent. The first prototype of this system was presented as a demo at ISWC 2021 [21].

In this paper, we introduce AIDA-Bot 2.0, a chatbot able to answer various questions about the research landscape and the scientific literature. This conversational agent has been designed to both 1) support a set of predetermined question types (e.g., "List all entities with a certain characteristic", "Compare two entities") by automatically translating them to formal queries on the knowledge graph, and 2) answer open questions (e.g., "What is a convolutional neural network?", "Define knowledge graph") by summarising information from relevant articles in the knowledge graph. This hybrid approach ensures that the

[1] Academia/Industry DynAmics Knowledge Graph - http://w3id.org/aida/.

responses provided are grounded in factual information that can be easily verified and, if necessary, corrected by updating the knowledge graph.

Recent advancements in natural language processing have led to the development of large language models, such as GPT 4.0 [29], which can generate coherent and eloquent responses to user queries. However, these models have raised concerns about the accuracy and reliability of the generated content, as they may produce text that is not based on factual knowledge, leading to what is known as hallucinations. For instance, asking the current version of ChatGPT for a list of prominent papers in Blockchain will result in a set of mostly fictional articles, typically generated by combining keywords and authors of real papers. Furthermore, recent studies suggest that recent GPT models exhibit limited accuracy in generating consistent responses to inquiries in the scientific domain [6]. In contrast, our aim in building AIDA-Bot 2.0 was to ensure that the system produces only accurate and verifiable information within a specific domain.

AIDA-Bot 2.0 boasts several significant improvements over its predecessor, AIDA-Bot 1.0 [22]. These include: i) a novel grammar-based approach for identifying question types, ii) the capacity to accommodate up to three filters in a query, and iii) the ability to respond to open queries by providing summaries of relevant articles.

We evaluated AIDA-Bot 2.0 in terms of both accuracy and usability. We first conducted a comparative evaluation in which ten researchers posed 15 questions to AIDABot 2.0 and three alternative solutions and ranked the responses. We then performed a user study that involved five senior computer science researchers to obtain an in-depth evaluation of AIDABot 2.0 usability and usefulness.

In summary, the main contributions of this paper are the following:

- AIDA-Bot 2.0, a novel conversational agent that takes advantage of a large-scale knowledge graph to produce reliable answers in the research domain;
- a new hybrid architecture for addressing both pre-determined and open questions that greatly improves on AIDA-Bot 1.0;
- an evaluation comparing AIDA-Bot 2.0 against three alternative architectures;
- a user study further assessing AIDA-Bot 2.0's user experience;
- a discussion of the impact and uptake of this tool.

The remainder of this manuscript is structured as follows. Section 2 describes the pipeline for data integration and AIDA KG. Section 3 presents the AIDA-Bot 2.0 architecture. Section 4 reports the outcome of the evaluations. Section 5 describes the uptake and impact of AIDA-Bot and Sect. 6 discusses the development plans. Section 7 presents the related work on conversational agents. Finally, Sect. 8 concludes the paper and outlines future research directions.

2 The AIDA Knowledge Graph Pipeline

The SN Data Cloud Infrastructure, based on Google BigQuery[2] and Google Vertex AI Workbench[3], enables us to define complex data pipelines to integrate different data sources, update them regularly, and enrich them by applying machine learning models for classification and information extraction. It is also employed to facilitate data exchange between different systems within SN, ensuring that data and analytics are consistent.

In an effort to obtain a more comprehensive and detailed representation of research dynamics in Computer Science, we adopted this system to create and maintain the Academia/Industry DynAmics Knowledge Graph (AIDA KG) [3]. AIDA KG is a large-scale knowledge base that describes publications and patents in Computer Science according to their research topics, authors, conferences, journals, organisations, types of organisations (i.e., academia, industry, or collaborative), and industrial sectors (e.g., automotive, financial, semiconductors, manufacturing). It is generated by integrating data sources such as OpenAlex[4], DBLP, Research Organization Registry[5] (ROR), DBpedia, the Computer Science Ontology[6] (CSO), and the Industrial Sectors Ontology[7] (INDUSO).

The current version of AIDA KG describes 25M publications and 8M patents.

AIDA KG focuses on eight main classes: *paper*, *patent*, *author*, *affiliation*, *journal*, *conference*, *topic*, and *industrialSector*. All entities from these classes are interlinked via 22 unique relationships such as: i) *hasAffiliation*, to indicate the affiliations of the authors of a paper, ii) *hasTopic* to identify the topics of papers and patents, iii) *schema:creator* to indicate the author of a paper. The complete schema of AIDA KG is available at https://w3id.org/aida.

The pipeline used to generate AIDA KG comprises different stages. First, it downloads and prepares all the relevant data sources. Then, it integrates research papers from the OpenAlex and DBLP, unifying them using DOIs and title similarity. Next, it leverages the CSO Classifier [43] to annotate all research documents according to their relevant topics, drawn from the Computer Science Ontology [44]. It then uses the ROR IDs from OpenAlex, to determine whether documents are written by academic institutions, industrial organizations, or through a collaborative effort. To provide additional context to the AIDA KG, all documents created by industrial authors, including those resulting from collaboration with academia, are also annotated with information regarding the relevant industrial sectors from INDUSO. This is accomplished by utilising the description of the affiliation available on DBpedia. For example, to characterise the company 'Samsung', we retrieve the relevant entity in DBpedia[8], extract information about their products, and map them to relevant sectors

[2] Google BigQuery - https://cloud.google.com/bigquery.
[3] Google Vertex AI Workbench - https://cloud.google.com/vertex-ai-workbench.
[4] OpenAlex - https://openalex.org/.
[5] ROR - https://ror.org/.
[6] CSO - https://w3id.org/cso.
[7] INDUSO - https://w3id.org/aida/#induso.
[8] https://dbpedia.org/page/Samsung.

in INDUSO, in this case: "semiconductor", "telecommunications", and "home appliances".

The AIDA Knowledge Graph is publicly available[9] and distributed under the CC-BY 4.0 license. In addition, it can be queried using SPARQL from the main triplestore https://w3id.org/aida/sparql/. While the current version focuses on Computer Science, we are now expanding it to cover additional disciplines.

3 The Architecture of AIDA-Bot 2.0

The architecture of AIDA-Bot 2.0 consists of two main modules: Question Understanding and Response Generator.

The Question Understanding module analyses the user input with the aim of recognising one of the four predefined query types (count, list, describe, and compare) and converting the question into a formal query on the knowledge graph. AIDA-Bot 2.0 supports complex queries using up to three filters (e.g., "List the top five papers about *computer vision* and *machine learning* written by researchers from the *University of Cambridge*"), in contrast with the previous version, which allowed only one condition.

The module extracts a set of key terms and searches them in AIDA KG to identify the relevant entities and their types.

It then uses the resulting entities to generate all pertinent questions the system can automatically translate to queries over the knowledge graph. Finally, it computes the similarity between the user's input question and the set of generated questions. This solution allows us to detect an extensive array of formulations associated with each supported question, encompassing different linguistic expressions.

If the similarity score between the user input and the most similar generated question exceeds a threshold, the Response Generator module uses a template to translate the latter to a query on AIDA KG and retrieves the relevant information. Otherwise, the system retrieves from AIDA KG the set of articles containing in the title or the abstract the key terms extracted from the user question. It then applies a question-answering model to produce a response based on the articles.

In the following, we describe the two modules in detail and provide more information on the adopted transformer models.

3.1 Question Understanding

The Question Understanding module analyses the input query and uses named-entity recognition (NER) to identify the key terms, which include nouns, noun phrases, named entities, and compound expressions in quotes. This information extraction step employs spaCy[10], an open-source Python library for Natural Language Processing[11].

[9] AIDA Knowledge Graph Download - https://w3id.org/aida.
[10] Spacy - https://spacy.io/.
[11] Specifically, we adopted the "en_core_web_sm" model.

We allow users to employ compound expressions in quotes to specify an exact match, similar to search engines. To avoid redundancy, nouns and noun phrases that appear in a named entity or an expression enclosed in quotation marks, are removed from the key terms. We also discard from key terms words that suggest questions (e.g., "who", "what") and terms that indicate an entity type (e.g., "papers", "articles", "citations"). For instance, the request "*Count papers about mathematics and matrix algebra written by authors from 'French Institute for research in computer science and automation'* ", contains the nouns *papers, mathematics, matrix, algebra, authors, French, Institute, research, computer, science, automation*, which are also included in the noun phrases *'mathematics', 'matrix algebra', 'authors', 'French Institute', 'research', 'computer science', 'automation'* and, therefore, are discarded. The words *papers* and *authors* are removed as they are types in AIDA KG. All the terms which appear in the quoted expression are also discarded. Therefore, the resulting key terms would be: *'mathematics', 'matrix algebra'* and *"French Institute for research in computer science and automation"*.

The key terms are searched in AIDA KG to retrieve the relevant entities and their types. In the previous example, all key terms would be found in AIDA KG: the "French Institute for research in computer science and automation" as *organization* while 'mathematics' and 'matrix algebra' as *topic.*

The Question Understanding module uses the resulting entities to generate a grammar for producing all compatible requests that can be translated to queries on the knowledge graph. A grammar is a set of production rules that describe how to generate valid sentences. These rules specify the allowable combinations of symbols or tokens and the order in which they appear.

In our system, the grammar is dynamically generated by using templates that include placeholders that are populated with the entities and their types. In the following, we report an example of a simple template for each query type.

1. count < sub_c > {}
2. list the <super> {num} <sub_l> {}
3. describe {}
4. compare {} vs {}

where:

- <sub_c> = papers | authors | conferences | organizations | citations | journals
- <super> = top | most important | main | most cited
- <sub_l> = papers | authors | conferences | organizations | topics | journals

Curly parentheses can only be filled with instances from the AIDA KG. Variables in angular parenthesis (e.g., <sub_c>) can only be filled with the previously defined items (e.g., papers, authors, conferences, and so on). Additionally, synonyms for these items, as pre-defined in a list, may also be employed. For example, <sub_c> would match both the words "papers" and "articles".

During the generation of the grammar, the system will produce all questions compatible with the set of detected entities. When considering the four templates defined above, if the system detects entities of type [“topics”, “conferences”, “organizations”, “authors”, “journals”], it will produce a range of questions of types 1 and 2. The module produces types 1 and 2 queries with up to three identified instances, allowing users to specify queries with three filters. Whenever at least one element from [“authors”, “conferences”, “organizations”] is found, the system will produce queries of type 3. Whenever it detects two items of the same class, it will generate queries of type 4.

In practice, each question type is supported by multiple templates since the same type of question can appear in several forms. For example, *how many* *<sub_c>* {} is another template for the query type *count* and would support questions such as “How many papers are there about the semantic web and machine learning?”. Therefore, from a modest number of initial templates covering the four query types (15 in the current implementation) and a set of identified entities, AIDA-Bot can generate a large number of candidate questions. Current templates were derived from use cases specific to SN and further improved through iterative refinement based on user feedback. Since developing new templates requires limited effort, the system can be easily adapted to other domains.

Next, the system computes the similarity between the original user request and the questions generated by the grammar. This step enables us to recognise a wide variety of formulations pertaining to the same question. In practice, we encode both the user’s input and the generated questions as sentence embeddings and then compute their cosine similarity. If the similarity score between the user input and the most similar generated question exceeds an empirically established threshold, the module designates the latter as the representative of the user query. As this question was derived from a template, the system knows how to translate it into a query on the knowledge graph.

Finally, the Question Understanding module sends all pertinent information for the next phase to the Response Generator, including key terms, entities, entity types, and query types.

3.2 Response Generator

The Response Generator distinguishes two main cases. If the user request matched one of the generated queries, it produces the equivalent query, runs it over AIDA KG, and retrieves the relevant data. To produce a natural language response, the module employs a response template tailored to each specific query type. These templates are populated with relevant data and further refined through the adjustment of singular and plural terms, ensuring grammatical correctness and coherence in the answer.

When the user question fails to match one of the generated queries, the module handles the user’s request as an ’open question’. In such cases, the module endeavours to generate a response by employing a question-answering model that operates on both the user request and the abstracts of relevant articles. To this end, it retrieves from AIDA KG the set of papers containing relevant key terms. If the query returns no paper, typically because the user request was out of scope, AIDA-Bot asks the user to reformulate or modify the request. Otherwise, the module selects the papers whose abstracts exhibit the highest similarity to the user query. This selection process utilises a transformer model designed for assessing sentence similarity. Subsequently, a summarisation model is applied to condense the abstracts into a more concise text. Finally, the module employs a question-answering model to generate a response to the user question based on the resulting information. The answer is further enhanced by providing a brief bibliography that lists the relevant articles. When feasible, the bibliography includes the Digital Object Identifiers (DOIs) and links to the open-access versions of the articles.

3.3 Transformer Models

AIDA-Bot 2.0 relies on transformer models for three main tasks: i) assessing the similarity between two texts, ii) summarising a text, and iii) question-answering.

The transformer model employed for measuring sentence similarity is the 'all-MiniLM-L6-v2', sourced from the Sentence-Transformers library[12]. This model was chosen due to its efficiency and compact size. Widely recognised as state-of-the-art technology, it is highly regarded for its effectiveness in addressing tasks pertaining to Semantic Textual Similarity [37]. To utilise this model, we leveraged the SentenceTransformers framework[13], which provides a convenient package for accessing BERT-based models and their variants, such as RoBERTa, MPNet, and ALBERT.

The question-answering and the summarisation models are 'distilbert-base-cased-distilled-squad'[14] and 'sshleifer/distilbart-cnn-12-6'[15] from Huggingface.

Their performances are comparable to those of BERT, but they use less computing power. The question-answering transformer, in particular, runs 60% quicker while retaining 95% of BERT's performance. It was developed by distilling the BERT base with 40% fewer parameters than the standard *textitbert-base-uncased*.

The summarisation model is based on DistilBART models, which are models created by removing the decoder layers from a Seq2Seq transformer and then producing high-quality student models through fine-tuning. We evaluated various BART models and observed that bart-large-cnn and distilbart-cnn-12-6 consistently generated superior summaries for our use cases. We adopted distilbart-cnn-12-6 since it is significantly lighter.

4 Evaluation

In this section, we present a systematic evaluation of the AIDA-Bot 2.0 against three alternative architectures (Sect. 4.1), which involved the participation of

[12] https://www.sbert.net/docs/pretrained_models.html.
[13] https://www.sbert.net/.
[14] https://huggingface.co/distilbert-base-cased-distilled-squad.
[15] https://huggingface.co/sshleifer/distilbart-cnn-12-6.

10 users. Furthermore, we present the findings of a user study conducted with five researchers in Computer Science (Sect. 4.2). The data produced during the evaluation and the user study are available online[16].

4.1 Comparative Evaluation

As previously discussed, AIDA-Bot 2.0 incorporates three main key enhancements over AIDA-Bot: i) the grammar-based method for detecting query types, ii) the ability to support queries with up to three filters, and iii) the capability to answer open queries by summarising relevant articles. To validate the efficacy of these improvements, we conducted a formal evaluation that compared AIDA-Bot 2.0, as described in Sect. 3, with three baselines:

- **AB** (AIDA-Bot, version 1.0) as originally presented at ISWC 2021 [21] and described in the subsequent journal paper [22]. This version adopts a simple token-aware approach to match user input with query types;
- **AB-G** (AIDA-Bot 1.0 with the grammar-based approach), a more advanced version of AB that employs the novel grammar-based approach;
- **AB-GF** (AIDA-Bot 1.0 with the grammar-based approach and filters), a further extension that also supports complex queries incorporating up to three filters.

To compare these four approaches, we organised individual sessions with ten researchers in Computer Science with an average of 12 years of academic experience. We instructed each researcher to generate 15 questions covering the five query types supported by our system: count, list, describe, compare, and open. Regarding the open-ended question, the researchers were requested to formulate queries that could realistically be covered in research articles. For each query, we listed the answers produced by the four chatbots in random order. The researchers were asked to rate their satisfaction with the responses on a Likert Scale ranging from 1 (very dissatisfied) to 5 (very satisfied).

Table 1. Average score of the four chatbots per query type. In bold, the best results.

Chatbot	Count	List	Describe	Compare	Open	Average
AB	2.33	1.37	1.62	2.60	1.00	1.79
AB-G	3.48	3.49	4.05	3.13	1.20	3.07
AB-GF	4.11	3.91	4.05	3.13	1.20	3.28
AIDA-Bot 2.0	**4.15**	**4.04**	**4.81**	**3.20**	**4.30**	**4.10**

Table 1 reports the average scores obtained by the four chatbots across the query types. The findings indicate that the implementation of the three new

[16] AIDA-Bot 2.0 evaluation data - https://w3id.org/aida/downloads#evaluation.

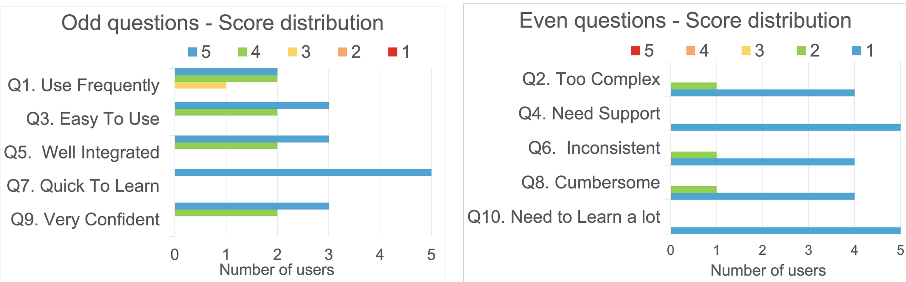
features has a favourable impact on the overall mean rating, which increases from 1.79 to 4.10. Furthermore, the average rating for each query type shows a steady increase when the new methods are integrated. Notably, users exhibited the highest level of satisfaction with 'describe' type queries, commonly utilized for zooming on researchers or topics, and 'open' type queries, indicating AIDA-Bot 2.0's substantial proficiency in handling non-predefined queries.

4.2 User Study

We performed a user study involving five computer scientists in order to assess the usability of the system and collect additional feedback. The users were selected among researchers in Computer Science at the University of Cagliari (IT), Gesis - Leibniz Institute for Social Science (DE), and The Open University (UK). Their areas of expertise include Artificial Intelligence, Natural Language Processing, Semantic Web, Complex Networks, Data Science, and Big Data.

We began each session with a 15-minute presentation of AIDA-Bot 2.0 and its capabilities. Then, we instructed the users to engage in an interactive session of about 45 min.

We asked them to complete a two-part survey describing their overall experience. The first section uses the standard *System Usability Scale* (SUS) questionnaire to assess the usability of AIDA-Bot 2.0. The second section includes five open questions regarding the strengths, weaknesses, and general feedback about AIDA-Bot 2.0. In what follows, we describe the outcome of these surveys.



(a) Odd questions. The higher the score, the better the system.

(b) Even questions. The lower the score, the better the system.

Fig. 1. SUS Questionnaire results.

SUS Questionnaire. The SUS questionnaire[17] provided excellent results, scoring 93.5/100, which is equivalent to an A grade, placing the AIDA-Bot 2.0 in the 95 percentile rank[18].

[17] SUS Questionnaire Questions: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

[18] Interpreting a SUS score - https://measuringu.com/interpret-sus-score/.

Figures 1a and 1b show the score distribution of the users. Specifically, Fig. 1a focuses on the odd questions (the positive ones, which should obtain a high score), while Fig. 1b reports on the even ones (the negative ones, which should obtain a low score). According to the users' feedback, the AIDA-Bot 2.0 was found to be easy to use (with an average rating of 4.6 ± 0.5[19]) and its features were well-integrated (4.6 ± 0.5). They stated that it was not complex to use (1.2 ± 0.4) and that they would not require any assistance to use it in the future (1.0 ± 0.0). Additionally, the SUS results indicated that most users felt highly confident while using our system (4.6 ± 0.5) and would be willing to use it frequently (4.2 ± 0.8).

Open Questions. In this section, we summarise the answers to the open questions.

Q1. What are the main strengths of AIDA-Bot 2.0? Three users stated that the main strength is the simplicity of the system in quickly providing all the information needed. One user considered the main strength the possibility to retrieve and explore scholarly information without having to search internal databases or the Web. The last user identified the primary strength of AIDA-Bot 2.0 as its capability to compare entities according to predefined metrics.

Q2. What are the main weaknesses of AIDA-Bot 2.0? The main weakness pointed out by most evaluators is the response time, as the current prototype can take several seconds to produce an answer. Two users expressed reservations about the quality of the open-ended responses, noting occasional instances where the formulation appeared peculiar. This issue can be attributed in part to the heightened expectations of users in light of the recent release of advanced GPT models, which have raised the bar for text generation quality.

Q3. Can you think of any additional features to be included in AIDA-Bot 2.0? The users suggested several interesting features, such as i) the possibility to generate a bibliography based on user inputs, ii) improving the approach for answering open questions by including GPT-like models, and iii) the ability to remember what the user said earlier in the conversation.

Q4. Can you think of any additional types of queries for AIDA-Bot 2.0? The users put forward several suggestions to enhance the system's functionality. First, they suggested the possibility to answer arbitrary questions about the content of a specific paper. They also recommended augmenting the system's ability to compare entities by allowing the definition of arbitrary metrics for comparison. Finally, they suggested incorporating a feature that enables the identification of articles referencing particular analysis techniques, algorithms, or datasets.

Q5. What would you add to increase the accuracy/comprehensiveness of the information returned by AIDA-Bot 2.0? Two users proposed

[19] With the notation $X \pm Y$, we specify that X is the average score and Y the standard deviation.

enhancing the entity detection methodology by incorporating the complete conversational context, encompassing previous messages as well as a user model. One user recommended utilizing the full text of research papers instead of just abstracts to improve the precision and comprehensiveness of the extracted information.

In summary, the user study demonstrated that AIDA-Bot 2.0 is highly usable and perceived as a valuable tool for providing accurate information about the research landscape. However, the emergence of modern GPT models has raised user expectations regarding the utilization of contextual information to comprehensively comprehend queries and generate highly coherent open-domain responses in real time. Although AIDA-Bot 2.0 is designed specifically for answering questions about the research landscape, it may be beneficial to integrate some of these new solutions. The primary challenge going forward is to do so without compromising the accuracy of the resulting analytics or deviating from the verifiable information in the knowledge graph.

5 Uptake and Impact

Since 2014, the collaboration between The Open University and SN has produced a wide range of tools based on AI and semantic technologies that have had a transformative effect on SN workflow, reducing the cost, improving the quality of the metadata, and supporting decision-making. These include intelligent services for automatically classifying articles [43] and proceeding books [41], recommending publications [48], evolving domain ontologies [30], and predicting the emergence of research topics [42].

In 2020, we generated AIDA KG by integrating multiple data sources in order to offer a very granular representation of research articles in terms of both research topics and the industrial sectors. The AIDA Dashboard, a web application to assess and compare journals and conferences, was the first tool to take advantage of this new resource [2].

In 2021, we released the first prototype of AIDA-Bot with the aim of allowing editors, analysts, and ultimately researchers to formulate complex natural language queries over AIDA KG. In 2023, we released AIDA-Bot 2.0, which marked a significant advancement over its predecessor, mainly due to its enhanced capacity to process complex queries and deliver responses to open-ended inquiries using information from the scientific literature. AIDA-Bot 2.0 was designed to cater to a range of specific use cases. Specifically, editors need to identify, assess, and compare the key researchers, organizations, venues, and trends within a certain field in order to evaluate the scope of the SN catalogue and initiate new editorial endeavours. The resulting information needs to be reliable but also verifiable, in order to be included in relevant reports. This process typically requires time-consuming analyses conducted by senior professionals in tandem with data analysts who must repeatedly procure accurate data from databases, resulting in delays and substantial expenses. A task such as identifying the key researchers at the intersection of deep learning and human-computer interaction would then take several days. AIDA-Bot 2.0 overcomes this bottleneck by allowing users to directly formulate natural language queries and receive real-time verifiable answers that refer to the internal KG or specific research articles.

The adoption of these solutions within SN is a testament to the advantages of incorporating knowledge graph technologies in this domain, yielding notable improvements in several key aspects and providing numerous benefits to the organization. Specifically, our technology has considerably reduced the amount of time required for performing complex analyses, leading to enhanced operational efficiency and reduced costs. Moreover, it has contributed to a reduction in the number of personnel involved in the analysis process. This, in turn, has freed up the time of analysts, enabling them to focus on other critical tasks. Finally, our system's ability to efficiently assess pertinent information has greatly enhanced the reliability and accuracy of the analyses, positively impacting the velocity and quality of the decision-making process.

6 Continuous Development Plan

We aim to keep developing and expanding AIDA-Bot 2.0 in the following years and we have several exciting developments in the works.

One of the key objectives is to integrate AIDA-Bot 2.0 with other tools at SN to provide a seamless and more comprehensive data exploration experience. For instance, AIDA-Bot 2.0 will be fully integrated into the AIDA Dashboard [2], which currently allows editors to explore and compare research venues. The main objective is to foster more contextual and refined data explorations by allowing users to ask additional questions to expand or clarify the analytics. For example, when visualising the list of researchers active in a conference, the user will be able to ask the system to describe a researcher in detail or report their paper in a specific field.

Similarly, we are also planning to integrate the AIDA-Bot 2.0 with other internal dashboards, such SN Insights[20], a tool for exploring and aggregating their own published content. We aim to assist the editor by allowing them to formulate more specific questions on SN Insight's large set of analytics.

To further improve AIDA-Bot 2.0's functionalities, we are exploring the use of GPT 4.0 APIs to replace the local question-answering model. We are also investigating the possibility of developing an internal model based on LLaMA [49] or similar solutions. These deployments will allow for even more accurate and effective information retrieval while reducing response time. We also plan to switch the NER pipeline to a more robust transformer-based model, in order to solve occasional inaccuracies.

Finally, a substantial avenue for large-scale deployment is expanding AIDA-Bot 2.0 toward other disciplines. This is especially important given that research endeavours are becoming increasingly multidisciplinary, and the most compelling insights and potential for new editorial initiatives are often found at the intersections of two or more fields. In light of this, we are presently engaged in efforts

[20] SN Insights - https://sn-insights.dimensions.ai/.

to integrate Engineering, Material Science, and Biomedical Science, which are other fields with a high presence of conference proceedings.

7 Related Work

Chatbots are often classified into two main classes, based on their objectives: task-oriented [36] and non-task-oriented [11]. Task-oriented chatbots are made to perform certain tasks, such as booking a hotel room, airfare, or other travel-related accommodations, putting an order for goods, planning events, or helping users find information [19]. Despite their limited scope, they are designed to help users accomplish a certain task within a specific scope. On the other hand, non-task-oriented chatbots operate in the open domain and aim to emulate the characteristics of human-human unstructured dialogue [38].

Additionally, chatbots can be categorized according to their engine (AI-based vs. rule-based). Rule-based chatbots employ a structured flow, reminiscent of a tree structure, to address user queries and provide appropriate responses [46]. This flow-based approach involves organizing a set of predefined rules or decision nodes that guide the chatbot's conversational path. They typically lead the user with more inquiries in order to eventually get the proper response. Conversely, AI-based chatbots seek to deduce the user's intent directly from the input text by employing artificial intelligence (AI) and natural language processing techniques [23].

Chatbots are often designed to focus on specific domains, such as i) education [28], ii) business [7], and iii) healthcare [18]. In the education domain, chatbots help teachers in a variety of disciplines, such as English [45], Medicine [8], and Business [39]. Other chatbots are capable of responding to questions concerning institutions that are typically covered in FAQs [35]. The reader is directed to [28] for a survey of articles on the use of chatbots in education. In the business domain, several chatbots assist enterprises with routine tasks [7]. For example, chatbots were developed to support commercial customer service and e-commerce [12]. In the health domain, chatbots often focus on specific health-related questions in domains such as mental health [27] and child health [50]. For instance, Divya et al. [14] developed a medical chatbot that allows users to self-diagnose illnesses and get comprehensive descriptions of the problems. Two more healthcare chatbots are Mandy [26] and MedChatbot [8]. The former is used by healthcare practitioners to automate patient intake. The latter supports medical students financially. Some chatbots also collect information on users' eating habits or provide businesses with a way to access allergy data based on the users' allergies [17].

Recent years have seen the emergence of a number of conversational agents and question-answering systems based on knowledge graphs and semantic web technologies [1,5,16,31,47]. These systems are capable to run complex queries on heterogeneous data from a variety of sources [15], including big general knowledge bases like Wikidata [24] and DBpedia [5], yielding a competitive advantage. Furthermore, it is possible to continuously update and improve these knowledge graphs by applying a variety of techniques for link prediction [25,40] and information integration [13,32]. As a result, a number of well-known conversational agents now utilise extensive knowledge graphs such as the Google Knowledge Graph [9] and the Alexa Knowledge Graph [51].

Recently, the focus shifted to the creation of sophisticated conversational agents that took advantage of transformers. GPT-2 (Generative Pre-trained Transformer) [34], GPT-3 [10], and the recent GPT-4 [29] are three examples in this field. GPT-3 was released in 2020 and became one of the largest language models to date, with 175 billion parameters. It was trained on a large corpus of 45 terabytes of text data from the internet, including books, articles, and websites, and can perform a wide range of natural language tasks, such as language translation, summarisation, and question-answering. GPT-4, is the next iteration of the GPT that is used in ChatGPT, but the exact details of its architecture and training data have not been disclosed. For a more thorough summary of relevant literature, we refer to Mariani et al. [20].

In the last few months, GPT models have been used to power several prototypical chatbots targeted at the scholarly domain, such as Scite[21], Elicit[22], and CoreGPT[23]. These systems aim to assist users with a variety of tasks, such as identifying trends in the literature, choosing a venue for sharing their work, finding suitable collaborators, searching relevant articles, and more. However, it is not clear yet to what extent these new solutions can produce accurate answers about the academic landscape.

AIDA-Bot 2.0 adopts a hybrid architecture that combines easily extendable templates to translate user requests to queries over a knowledge graph and question-answering models for answering open questions with information from the literature. The main objective is to avoid hallucinations and provide verifiable and accurate information about the scholarly domain.

8 Conclusions

In this paper, we presented AIDA-Bot 2.0, a novel conversation agent designed to provide accurate and factual information about the research landscape, and discussed its role within Springer Nature workflow. AIDA-Bot 2.0 builds on top of the Academia/Industry Dynamics Knowledge Graph (AIDA KG), a large knowledge graph containing over 1.5B triples obtained by integrating data about 25M papers from OpenAlex, DBLP, DBpedia, ROR, CSO, and INDUSO. It offers two main capabilities: 1) the ability to identify a set of pre-determined question types and translate them to formal queries over the knowledge graph, and 2) the ability to answer open questions by summarising relevant information from articles. We conducted two evaluations that demonstrated the benefits of AIDA-Bot 2.0's new features and proved its excellent usability.

[21] Scite - https://scite.ai/.
[22] Elicit - https://elicit.org/.
[23] CoreGPT - https://tinyurl.com/mvrk2z4x.

In future work, we aim to incorporate the valuable user feedback received from the evaluations into the development of AIDA-Bot 2.0. Furthermore, we will explore the integration of other knowledge sources and research fields to improve the quality and coverage of the information provided. We also plan to produce a lightweight version that will be made freely available to the research community, similar to what we did for the AIDA Dashboard. Finally, we plan to further investigate how modern language models can be integrated with knowledge graphs in order to produce verifiable information in the scientific domain.

References

1. Expert System for Question Answering on Anomalous Events and Mitigation Strategies Using Bidirectional Transformers and Knowledge Graphs, Abu Dhabi International Petroleum Exhibition and Conference, vol. Day 3 Wed, 02 November 2022 (2022). https://doi.org/10.2118/211855-MS, d031S084R002
2. Angioni, S., Salatino, A., Osborne, F., Birukou, A., Recupero, D.R., Motta, E.: Leveraging knowledge graph technologies to assess journals and conferences at springer nature. In: Sattler, U., et al. (eds.) The Semantic Web - ISWC 2022. LNCS, pp. 735–752. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19433-7_42
3. Angioni, S., Salatino, A., Osborne, F., Recupero, D.R., Motta, E.: Aida: a knowledge graph about research dynamics in academia and industry. Quant. Sci. Stud. **2**(4), 1356–1398 (2021)
4. Angioni, S., Salatino, A.A., Osborne, F., Recupero, D.R., Motta, E.: Integrating knowledge graphs for analysing academia and industry dynamics. In: Bellatreche, L., et al. (eds.) TPDL/ADBIS -2020. CCIS, vol. 1260, pp. 219–225. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-55814-7_18
5. Athreya, R.G., Ngonga Ngomo, A.C., Usbeck, R.: Enhancing community interactions with data-driven chatbots-the dbpedia chatbot. In: Companion Proceedings of the The Web Conference 2018, Lyon, France, pp. 143–146 (2018)
6. Auer, S., et al.: The SciQA scientific question answering benchmark for scholarly knowledge. Sci. Rep. **13**(1), 7240 (2023). https://doi.org/10.1038/s41598-023-33607-z
7. Bavaresco, R., et al.: Conversational agents in business: a systematic literature review and future research directions. Comput. Sci. Rev. **36**, 100239 (2020)
8. Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., Gangwani, A.: Medbot: conversational artificial intelligence powered chatbot for delivering tele-health after covid-19. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES), Budva, Montenegro, pp. 870–875 (2020). https://doi.org/10.1109/ICCES48766.2020.9137944
9. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, New York, NY, USA, pp. 1247–1250. Association for Computing Machinery (2008). https://doi.org/10.1145/1376616.1376746

10. Brown, T., et al.: Language models are few-shot learners. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901. Curran Associates, Inc. (2020). https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
11. Chen, H., Liu, X., Yin, D., Tang, J.: A survey on dialogue systems: recent advances and new frontiers. SIGKDD Explor. Newsl. **19**(2), 25–35 (2017). https://doi.org/10.1145/3166054.3166058
12. Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., Zhou, M.: SuperAgent: a customer service chatbot for E-commerce websites. In: Proceedings of ACL 2017, System Demonstrations, Vancouver, Canada, pp. 97–102. Association for Computational Linguistics (2017). https://aclanthology.org/P17-4017
13. Dessí, D., Osborne, F., Recupero, D.R., Buscaldi, D., Motta, E.: Scicero: a deep learning and NLP approach for generating scientific knowledge graphs in the computer science domain. Knowl.-Based Syst. **258**, 109945 (2022)
14. Divya, S., Indumathi, V., Ishwarya, S., Priyasankari, M., Devi, S.K.: A self-diagnosis medical chatbot using artificial intelligence. J. Web Dev. Web Des. **3**(1), 1–7 (2018)
15. Fensel, D., et al.: Knowledge graphs methodology, tools and selected use cases (2020). https://lib.ugent.be/catalog/ebk01:4100000010122122
16. Höffner, K., Walter, S., Marx, E., Usbeck, R., Lehmann, J., Ngonga Ngomo, A.C.: Survey on challenges of question answering in the semantic web. Semantic Web **8**(6), 895–920 (2017)
17. Hsu, P., Zhao, J., Liao, K., Liu, T., Wang, C.: Allergybot: a chatbot technology intervention for young adults with food allergies dining out. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, pp. 74–79 (2017)
18. Laranjo, L., et al.: Conversational agents in healthcare: a systematic review. J. Am. Med. Inform. Assoc. **25**(9), 1248–1258 (2018)
19. Li, L., Lee, K.Y., Emokpae, E., Yang, S.B.: What makes you continuously use chatbot services? evidence from Chinese online travel agencies. Electron. Mark. (2021). https://doi.org/10.1007/s12525-020-00454-z
20. Mariani, M.M., Hashemi, N., Wirtz, J.: Artificial intelligence empowered conversational agents: a systematic literature review and research agenda. J. Bus. Res. **161**, 113838 (2023). https://doi.org/10.1016/j.jbusres.2023.113838
21. Meloni, A., Angioni, S., Salatino, A., Osborne, F., Reforgiato Recupero, D.: Aida-bot: a conversational agent to explore scholarly knowledge graphs. CEUR-WS (2021). https://ceur-ws.org/Vol-2980/paper310.pdf
22. Meloni, A., Angioni, S., Salatino, A.A., Osborne, F., Recupero, D.R., Motta, E.: Integrating conversational agents and knowledge graphs within the scholarly domain. IEEE Access **11**, 22468–22489 (2023). https://doi.org/10.1109/ACCESS.2023.3253388
23. Mohan, S., Chowdhary, C.: An AI-based chatbot using deep learning. In: Intelligent Systems: Advances in Biometric Systems, Soft Computing, Image Processing, and Data Analytics, chap. 12, London, UK, pp. 231–242. Apple Academic Press (2019). https://doi.org/10.1201/9780429265020-12
24. Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E.: A systematic literature review on wikidata. Data Technologies and Applications (2019)
25. Nayyeri, M., et al.: Trans4e: link prediction on scholarly knowledge graphs. Neurocomputing **461**, 530–542 (2021)

26. Ni, L., Lu, C., Liu, N., Liu, J.: MANDY: towards a smart primary care chatbot application. In: Chen, J., Theeramunkong, T., Supnithi, T., Tang, X. (eds.) KSS 2017. CCIS, vol. 780, pp. 38–52. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-6989-5_4
27. Oh, K.J., Lee, D., Ko, B., Choi, H.J.: A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In: 2017 18th IEEE International Conference on Mobile Data Management (MDM), KAIST, Daejeon, pp. 371–375. IEEE (2017)
28. Okonkwo, C.W., Ade-Ibijola, A.: Chatbots applications in education: a systematic review. Comput. Educ. Artif. Intell. **2**, 100033 (2021). https://doi.org/10.1016/j.caeai.2021.100033
29. OpenAI: Gpt-4 technical report (2023)
30. Osborne, F., Motta, E.: Pragmatic ontology evolution: reconciling user requirements and application performance. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11136, pp. 495–512. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00671-6_29
31. Pang, R.Y., et al.: QuALITY: question answering with long input texts, yes! In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, United States, pp. 5336–5358. Association for Computational Linguistics (2022). https://doi.org/10.18653/v1/2022.naacl-main.391, https://aclanthology.org/2022.naacl-main.391
32. Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. Semantic web **8**(3), 489–508 (2017)
33. Peng, C., Xia, F., Naseriparsa, M., Osborne, F.: Knowledge graphs: opportunities and challenges. Artif. Intell. Rev. **56**, 1–32 (2023)
34. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners (2019)
35. Ranoliya, B.R., Raghuwanshi, N., Singh, S.: Chatbot for university related FAQs. In: 2017 International Conference on Advances in Computing. Communications and Informatics (ICACCI), Manipal, India, pp. 1525–1530. IEEE (2017)
36. Rastogi, A., Zang, X., Sunkara, S., Gupta, R., Khaitan, P.: Towards scalable multi-domain conversational agents: the schema-guided dialogue dataset. In: Proceedings of the AAAI Conference on Artificial Intelligence, New York, USA, vol. 34, pp. 8689–8696 (2020)
37. Reimers, N., Gurevych, I.: Sentence-Bert: sentence embeddings using SIAMESE Bert-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (2019). https://arxiv.org/abs/1908.10084
38. Roller, S., et al.: Recipes for building an open-domain chatbot. In: EACL. Online (2021)
39. Rooein, D., Bianchini, D., Leotta, F., Mecella, M., Paolini, P., Pernici, B.: achat-wf: generating conversational agents for teaching business process models. Softw. Syst. Model. (2021). https://doi.org/10.1007/s10270-021-00925-7
40. Rossi, A., Barbosa, D., Firmani, D., Matinata, A., Merialdo, P.: Knowledge graph embedding for link prediction: a comparative analysis. ACM Trans. Knowl. Disc. Data (TKDD) **15**(2), 1–49 (2021)
41. Salatino, A.A., Osborne, F., Birukou, A., Motta, E.: Improving editorial workflow and metadata quality at springer nature. In: Ghidini, C., et al. (eds.) ISWC 2019. LNCS, vol. 11779, pp. 507–525. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30796-7_31

42. Salatino, A.A., Osborne, F., Motta, E.: Augur: Forecasting the emergence of new research topics. In: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries. JCDL 2018, pp. 303–312, New York, NY, USA. ACM (2018). https://doi.org/10.1145/3197026.3197052
43. Salatino, A.A., Osborne, F., Thanapalasingam, T., Motta, E.: The CSO classifier: ontology-driven detection of research topics in scholarly articles. In: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (eds.) TPDL 2019. LNCS, vol. 11799, pp. 296–311. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30760-8_26
44. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: The computer science ontology: a large-scale taxonomy of research areas. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11137, pp. 187–205. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_12
45. Sarosa, M., Kusumawardani, M., Suyono, A., Wijaya, M.H.: Developing a social media-based chatbot for English learning. IOP Conf. Ser. Materials Sci. Eng. **732**(1), 012074 (2020). https://doi.org/10.1088/1757-899x/732/1/012074
46. Singh, J., Joesph, M.H., Jabbar, K.B.A.: Rule-based chabot for student enquiries. J. Phys. Conf. Ser. **1228**(1), 012060 (2019). https://doi.org/10.1088/1742-6596/1228/1/012060
47. Stasaski, K., Hearst, M.: Semantic diversity in dialogue with natural language inference. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, United States, pp. 85–98. Association for Computational Linguistics, Seattle (2022). https://doi.org/10.18653/v1/2022.naacl-main.6 , https://aclanthology.org/2022.naacl-main.6
48. Thanapalasingam, T., Osborne, F., Birukou, A., Motta, E.: Ontology-based recommendation of editorial products. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11137, pp. 341–358. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_21
49. Touvron, H., et al.: Llama: open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
50. Vaira, L., Bochicchio, M.A., Conte, M., Casaluci, F.M., Melpignano, A.: Mamabot: a system based on ml and NLP for supporting women and families during pregnancy. In: Proceedings of the 22nd International Database Engineering & Applications Symposium, Villa San Giovanni, Italy, pp. 273–277 (2018)
51. Zhu, Q., et al.: Collective knowledge graph multi-type entity alignment. In: The Web Conference 2020 (2020). https://www.amazon.science/publications/collective-knowledge-graph-multi-type-entity-alignment