Towards Intelligent Research Collaboration: A Hybrid AI Framework for Recommending Participants in Research Projects

Journal Title
XX(X):1–10
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Piermichele Rosati[1,2], Emanuele Laurenzi[1], and Michela Quadrini[2]

Abstract

The success of research project proposals largely depends on the quality of the consortium, which must possess strong expertise and experience aligned with the themes of the relevant funding calls, such as those under the EU's Horizon Europe programme. However, forming such a consortium remains one of the most difficult tasks, as it involves identifying suitable research collaborators. Traditional approaches typically rely on social networks or citation metrics, but these have shown limited effectiveness. This paper introduces an Agentic Graph-based Retrieval-Augmented Generation (RAG) approach that delivers contextualized and explainable collaborator recommendations, tailored to researchers' expertise and the relevance of proposed projects, offering improved performance over conventional methods. The approach integrates the strengths of Knowledge Graphs (KGs) and Large Language Models (LLMs), and has been developed using the Design Science research methodology. Its effectiveness was assessed using two of the top-performing LLMs currently available: Claude Sonnet 3.5 and GPT-4o.

Keywords

Hybrid AI, Knowledge Graphs, Large Language Models, Retrieval Augmented Generation, Research Collaboration.

Introduction

Research collaborations play a crucial role in tackling complex, multidisciplinary challenges, driving scientific advancement, and fostering innovation (Meißuner et al. (2021)). Traditionally, such collaborations have emerged through professional networks, conferences, workshops, and institutional ties (Katz and Martin (1997)). While effective, these methods are often limited by geographic and logistical constraints.

Digital platforms like ResearchGate* and LinkedIn† have increased opportunities for global networking, enabling researchers to connect, share outputs, and identify potential collaborators. However, these tools primarily support social networking rather than intelligent matchmaking based on complementary expertise or shared objectives.

Recommender systems (Lü et al. (2012)), widely adopted in commercial domains such as e-commerce and streaming, show strong potential for matching users with relevant content or entities (Hussien et al. (2021)). Traditional recommendation techniques, including Content-Based Filtering (CBF) and Collaborative Filtering (CF), have been enhanced by Deep Learning (DL) models, especially with the recent progress in Large Language Models (LLMs) (Zhao et al. (2024)). While LLMs can better capture user preferences and interpret diverse textual data, they still struggle with explainability, an essential requirement for high-stakes tasks like research collaborator recommendations.

Hybrid Artifical Intelligence (AI), a field that integrates machine learning and knowledge engineering (Rordorf et al. (2023); Prater and Laurenzi (2022); d'Avila Garcez and Lamb (2023)), offers a promising path forward. Two key technologies in this domain are LLMs and Knowledge Graphs (KGs), which have recently been combined to power intelligent and interpretable recommendation systems (Zhao et al. (2024)). KGs structure relationships between entities to support reasoning, while LLMs excel at understanding and generating complex textual content.

When integrated using Retrieval-Augmented Generation (RAG), LLMs and KGs can complement each other, providing contextual, explainable, and knowledge-grounded recommendations that reduce hallucinations and improve accuracy (Deldjoo et al. (2024)).

Motivated by this, our study investigates the use of LLMs and KGs for recommending research collaborators. The central research question is:

How can a KG and LLM-based approach enhance the process of suggesting collaborators for research projects?

To address this question, we adopted the Design Science Research (DSR) methodology (Hevner and Chatterjee

[1]University of Applied Sciences and Arts Northwestern Switzerland
[2]School of Sciences and Technology, University of Camerino, Italy

Corresponding author:
Piermichele Rosati, Emanuele Laurenzi, Michela Quadrini
Email: piermichele.rosati@students.fhnw.ch,
piermichele.rosati@studenti.unicam.it,
emanuele.laurenzi.fhnw.ch,
michela.quadrini@unicam.it

*https://www.researchgate.net
†https://www.linkedin.com

(2010)). In the problem awareness phase, we reviewed relevant literature and analyzed real-world practices to define system requirements. Based on these insights, we designed an artifact that integrates Agentic and Graph-RAG methods, leveraging both KGs and LLMs to generate explainable recommendations for potential collaborators in the context of European research projects.

The paper is structured as follows. First, we discuss related works and list relevant requirements. Then, we describe the selected dataset, and from this the analysis of five EU research projects. Next, we present the proposed system architecture and its implementation. Afterward, we detail the evaluation of our approach, which is carried out through experiments on the developed prototype. Finally, we conclude the paper by summarizing the main findings and outlining directions for future work.

Literature Review

This section reviews the state-of-the-art in both KGs and recommender systems as they relate to identifying project collaborators. It concludes with a set of design requirements that guided the development of the proposed artifact.

Knowledge Graphs in the Research Domain

KGs have become a foundational technology in data management and AI, enabling advanced data integration, retrieval, and analysis. Structured as directed graphs, KGs represent real-world entities (nodes) and their relationships (edges), making them well-suited for modeling complex domains and their semantics.

Numerous KGs and ontologies have significantly influenced the research field by providing structured representations of domain knowledge that facilitate advanced data analysis. A notable example was the Microsoft Academic Graph (MAG), which mapped academic publications, authors, institutions, and topics to uncover research trends and foster collaboration (Wang et al. (2020)). Following its discontinuation in 2021, OpenAlex emerged as its open-source successor, providing rich metadata connections among scholarly entities (Priem et al. (2022)).

The Allen Institute's Semantic Scholar Academic Graph (S2AG) is among the largest research-focused KGs, encompassing over 205 million publications and 121 million authors (Wade (2022)). It aggregates metadata from sources like Crossref and PubMed, supporting tasks such as citation analysis, research discovery, and Natural Language Processing (NLP).

Wikidata, maintained by the Wikimedia Foundation since 2012, is a collaboratively edited KG that spans diverse domains (Vrandečić and Krötzsch (2014)). It supports Linked Open Data (LOD) through globally recognized identifiers, enhancing interoperability and powering applications in NLP, recommender systems, and AI research.

The Open Research Knowledge Graph (ORKG) represents scholarly contributions in a machine-readable format, streamlining systematic reviews and comparative analyses. By aligning methodologies and concepts, it enhances reproducibility, literature gap identification, and automated reasoning.

The VIVO ontology is the basis of the VIVO platform, modeling academic entities and relationships using semantic web technologies like Resource Description Framework (RDF) and Web Ontology Language (OWL) (Börner et al. (2012)). It integrates external vocabularies such as Friend of a Friend (FOAF), which supports the representation of researchers and social networks on the Semantic Web. Through adherence to LODs principles, VIVO facilitates research discovery, expert identification, and international academic collaboration.

Recommender Systems in the Research Domain

The greater number of recommender system applications in the research domain focus on suggesting academic papers to mitigate information overload and support scholarly discovery.

For instance, Tejeda-Lorente et al. (2015) introduced REFORE, a hybrid recommender system designed to deliver personalized paper suggestions. It combines bibliometric indicators (such as journal impact factor and author h-index) with a CBF approach that leverages paper metadata such as keywords, abstracts, and citations. Papers are represented as keyword-weighted vectors, and dynamic user profiles are built from previous publications and manual inputs. Additionally, CF is employed to incorporate feedback from similar users, enhancing recommendation quality.

Similarly, Kanwal and Amjad (2024) developed a paper recommendation system that integrates citation and collaboration networks to improve relevance. By constructing a multi-level citation graph, the system captures relationships up to six levels deep using bibliographic coupling and co-citation analysis. Centrality metrics, such as betweenness, degree, and eigenvector centrality, help rank papers by structural importance. The system also utilizes author collaboration networks to identify influential researchers and refine suggestions.

Murali et al. (2019) proposed a user-based CF recommender to address the challenge of information overload. It models user-paper interactions (e.g., ratings), computes cosine similarity to identify users with similar preferences, and predicts ratings for unseen papers based on peer preferences. A rating prediction mechanism prioritizes high-quality papers, while user-link formation enhances the CF process by capturing group behavior.

Beyond paper recommendations, some works focus directly on academic collaboration. Du and Li (2022) introduced ACR-ANE, a model for recommending research collaborators. It combines network topology with scholar attributes (e.g., research interests, h-index), encoding these features via a deep autoencoder. The model incorporates non-local neighbors through biased random walks and frequency filtering, capturing both local and global academic ties in a multi-type relational network.

In a similar direction, Zhu and Yaseen (2022) explored the use of Graph Neural Networks (GNNs) to recommend collaborators in academic settings. Their approach leverages both static and temporal dynamics of research networks, using MEDLINE data and two GNN-based models, GraphSAGE and Temporal Graph Networks (TGN), to capture evolving scholarly relationships.

Valluru et al. (2024) propose ULTRA, an AI-assisted framework for research team formation that leverages open data from funding calls and researcher profiles. The system extracts and normalizes technical skills using taxonomies and NLP techniques, and assembles teams by optimizing multiple objectives, including skill coverage, redundancy, and robustness. Extensive evaluation across institutions in the United States and India, using both quantitative experiments and large-scale user studies, demonstrates that ULTRA consistently produces higher-quality and more effective team recommendations compared to baseline methods.

Requirements from Relevant Literature. The following Literature Requirements (LRs) were selected from the relevant works in the literature. Several KGs exist in the research and scholarly fields, but no work uses them in the European research domain. The definition of an ontology and/or use of a KG in this domain is necessary in order to allow semantic representation of European projects (LR1). Among current research recommender systems, some exploit various filtering approaches including CBF. Generating content-based recommendation could be quite efficient (LR2), for example if based on a project description or abstract, recommend potential collaborators, who have collaborated on a project with that description similar to the one specified. Unfortunately, some DL-based recommender systems need data training to perform the assigned task. In some cases it may also be necessary to retrain the model. In these cases, generating recommendations without performing either training or other techniques such as fine-tuning helps to save computational power (LR3). Regarding the hop reasoning perspective, retrieving relevant information dynamically, using models such as GNNs, can require high computations on graphs, especially on large graphs (LR4).

Scenarios Analysis

This section describes the dataset used, and then introduces the relevant scenarios that were analysed to derive the application requirements.

Dataset and Scenarios

The Community Research and Development Information Service (CORDIS) dataset (Publications Office of the European Union (2015a,b, 2018)) was chosen as primary data source. CORDIS[‡] serves as the European Commission's primary public repository for distributing information about EU-funded research projects. This dataset is a valid resource for analyzing research trends, understanding research project collaborations, and identifying potential research partners. The chosen dataset contains information about projects funded under two major European Union research initiatives:

(i) The Seventh Framework Programme (FP7) for Research and Technological Development[§], covering projects funded from 2007 to 2013.

(ii) The Horizon 2020 (H2020) Programme for Research and Innovation[¶], covering projects funded from 2014 to 2020.

The dataset consists of five distinct subsets. The projects subset includes details on participating organizations, legal basis information, topic classifications, project URLs, and categorization using the European Science Vocabulary (EuroSciVoc). The project deliverables subset contains metadata and links to project deliverables. The project publications subset provides metadata and links to publications. Similarly, the report summaries, which include periodic or final publishable summaries. Reference data, including information on programs, topics, topic keywords, funding schemes (types of action), organization types, and countries are also stored in the dataset. Although the CORDIS dataset is a valuable resource for analysing research trends, project collaborations, and the distribution of funding, certain limitations must be taken into account. Data integration challenges arise because the dataset is distributed across multiple files in different formats (JSON, XML, CSV, XLSX), requiring careful preprocessing and linkage before meaningful insights can be extracted. Establishing relationships between different files (e.g., linking projects to participants and deliverables) is not straightforward and necessitates additional data processing steps. Inconsistent data representations further complicate the process, as some fields are formatted differently across various subsets of the dataset. For example, project participants may be listed as semicolon-separated values in some files, while in others they appear as separate records, making data consistency and automatic processing more difficult. Additionally, missing or incomplete data is a concern: some projects lack details on funding schemes, research topics, or coordinator organizations. Not all participant organizations have standardized names or unique identifiers, complicating entity resolution when merging data. Certain project deliverables or publications might also be unavailable due to proprietary restrictions or missing metadata.

Using data from CORDIS, the structures of five European research projects that are part of the H2020 programme were analysed. In order to perform an analysis of the projects' data and metadata, we restricted our focus selecting projects focused on **Building Information Modeling (BIM)** as the main topic. The first and last one are under the "Research and Innovation Action" funding scheme, while the other three are under "Innovation Action". All of these are actually closed projects, which means that they have already been completed. Moreover, these projects fall within the temporal range from 2015 to 2023. The selected projects are: "BIM-based holistic tools for Energy-driven Renovation of existing Residences", "Integrated and Replicable Solutions for Co-Creation in Sustainable Cities", "New integrated methodology and Tools for Retrofit design towards a next generation of ENergy efficient and sustainable buildings and Districts", "Proactive synergy of inteGrated Efficient Technologies on buildings' Envelopes", and "Adaptive

[‡] https://cordis.europa.eu
[§] https://cordis.europa.eu/programme/id/FP7
[¶] https://cordis.europa.eu/programme/id/H2020

Multimodal Interfaces to Assist Disabled People in Daily Activities".

All five projects analyzed share the same structure in data organization, components, and reporting style. To avoid redundancy, only the analysis of the first project is presented as a representative example.

BIM-based holistic tools for Energy-driven Renovation of existing Residences The BIMERR project, funded by Horizon 2020, developed BIM-based tools to drive the digital transformation of energy-efficient building renovations, focusing on enhanced BIM interoperability and workflow optimization. Its structure includes a Fact Sheet, Reporting, and Results sections.

Fact Sheet. BIMERR, coordinated by Fraunhofer Gesellschaft (Germany), ran from January 2019 to September 2022, fully funded by the EU with a budget of €6.93M. The project aimed to create an ICT-enabled Renovation 4.0 framework, improving digital building models and renovation workflows. The consortium included 20 partners from Germany, Greece, Austria, Poland, the UK, Spain, Cyprus, Belgium, Slovakia, and Italy.

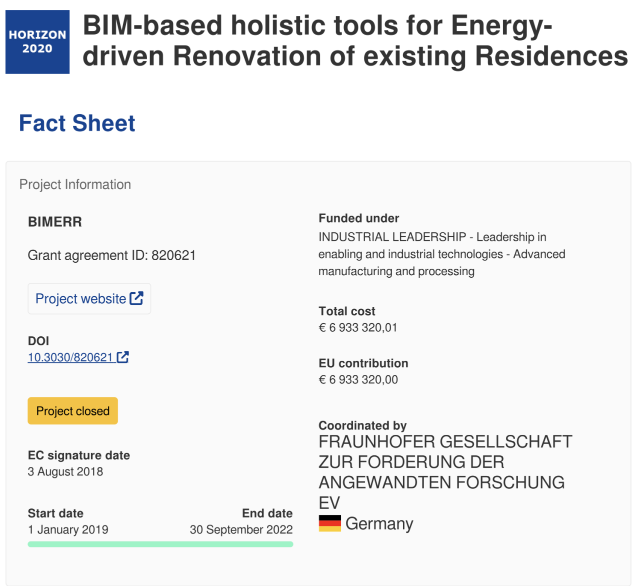


Figure 1. BIMERR Project Fact Sheet

Reporting. The project targeted five key objectives, including real-world demonstrations, BIM semantic interoperability, automated scan-to-BIM tools, decision-support systems, and broad adoption of BIMERR technologies. Deliverables included data models, middleware, AR tools, renovation decision systems, and workflow management toolkits. Commercialization strategies, including a business plan, were also developed.

Results. The project produced 22 reports and multiple demonstrators, such as the Interoperability Framework, AI-enabled Scan-to-BIM tools, adaptive workflow tools, and energy modeling modules. Dissemination activities resulted in 19 peer-reviewed articles, conference papers, training materials, and outreach efforts.

As an output of this analysis a list of Application Requirements (APRs) has been derived. In the following we report a short list of the main APRs: the project description (APR1), objectives (APR2), field of science, programmes, topics, proposal call, funding scheme, and keywords (APR3). In addition, there are the most important details concerning the coordinating organisation (APR4) of the project, and all the organisations participating (APR5) in that project. Each organisation presents secondary information such as postal address, location (APR6), and main type of activity.

The available information only concerns organisations. However, for the recommendation of collaborators it is important to also know which people from the relevant organisations are involved in the projects, therefore we considered it as an additional requirement (APR7).

Results

This section illustrates the architecture of the proposed system, describing the selected ontology and KG, and details the responsibilities of each component of the RAG approach. Finally, the technologies used to implement the system are shown.

Proposed Architecture

The proposed architecture integrates KGs and LLMs in a RAG pipeline, following and combining the GraphRAG and AgenticRAG paradigms together. This facilitates the recommendation of potential research collaborators without re-training the model (satisfied LR3).

Ontology Selection. The architecture is built upon the EUropean Research Information Ontology (EURIO) KG, which provides a structured and semantically rich representation of (EU-funded) research projects (satisfied APR1, APR2), organizations (satisfied APR4, APR6), and participants (satisfied APR5, APR7). The EURIO ontology, developed by the Publications Office of the European Union[‖], was found as a data model that conceptualizes, formally encodes, and makes available in an open, structured, and machine-readable format data about research projects funded by the EU's framework programmes for research and innovation (satisfied LR1). CORDIS is responsible for publishing the results of these projects, while EURIO provides a semantic model that enhances transparency, reusability, and accessibility. The EURIO ontology is built on top of well-known ontologies and vocabularies to ensure interoperability and semantic richness. These include:

- Dublin Core (DC): used for metadata elements such as titles, descriptions, and identifiers;
- Data Catalog Vocabulary (DCAT): an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web;
- Data Integration for Grant Ontology (DINGO): an ontology expressly designed to provide an extensible interoperable framework for formally conceptualizing and expressing the relevant parts of the research/cultural landscape in relation to funding, such that they can easily be shared between different actors and platforms;

[‖] https://op.europa.eu/en/

- FRBR-aligned Bibliographic Ontology (FaBiO): facilitates the description of bibliographic entities and their relationships;
- Funding, Research Administration and Projects Ontology (FRAPO): an ontology for describing the administrative information of research projects, e.g., grant applications, funding bodies, project partners, etc.;
- FOAF: defines relationships between people and organizations;
- Simple Knowledge Organization System (SKOS): facilitates controlled vocabularies and classification schemes.

The EURIO Ontology also incorporates reference data, such as countries, funding schemes, types of action, the EuroSciVoc taxonomy, and the NUTS classification, to enhance the semantic representation of research information. It leverages the OWL 2 to formally define the semantics of domain-specific terms used to describe CORDIS entities (e.g., projects, organizations, etc.), their attributes (e.g., title, acronym, legal name, etc.), and their interrelations (e.g., the connection between a project and its participating organizations, etc.). The EURIO ontology defines multiple classes representing different concepts such as projects, organizations, funding schemes, grants, publications, and roles, along with associated data properties and object properties that define their relationships. Each class is characterized by a set of data properties, which describe its attributes, and a set of object properties, which define its relationships with other classes. In addition, each class, data property, and object property is associated with several annotations, the most relevant of which are: rdfs:label, providing a human-readable name for the entity; rdfs:comment, offering a human-readable description; and rdfs:isDefinedBy, specifying the ontology in which the entity is defined. For instance, the Project class includes data properties such as the title, description, start date, end date, and funding amount. The ontology also specifies relationships between classes, such as is funded by, which denotes the association between a project and its corresponding grant.

Recommendation Strategy. The proposed system architecture is designed to generate recommendations regarding potential research collaborators for research projects. The recommendation strategy in this study was designed to leverage the structured data and relationships within the EURIO KG to generate meaningful and context-aware recommendations. Given the rich metadata and interconnected entities in EURIO, we exploited its data properties and relationships to extract the most relevant information about researchers, organizations, and projects. To facilitate the efficient retrieval of information, we developed a mechanism to query and extract the most meaningful and useful relationships for our purposes, enabling the identification of essential project details, such as the participants in a project, a person's employing organisation, and the organisation's role in a project. Other useful information to be retrieved is the abstract, duration, status, and URL of a project given its title.

Types of recommendations. Based on this structured data retrieval, we designed the two following primary types of recommendations. By integrating these structured recommendations, this strategy enhances the discoverability of research partnerships and collaboration opportunities, leveraging the semantic richness of the EURIO KG to provide personalized and explainable suggestions. The approach ensures that both individual researchers and research organizations receive tailored recommendations based on contextual relevance and domain-specific alignment, making the system an effective tool for fostering research collaboration.

Research Collaborator Recommendations. The first approach focuses on recommending potential research collaborators based on a given project description and its objectives. By analysing the key attributes of a project, including the title, description and research objectives, the system identifies researchers with expertise in similar fields, ensuring that recommended collaborators are in line with the project's needs, and that they can contribute through their expertise in one or more specific areas, useful for achieving the project's objectives. In this type of recommendation, it is also useful to refer to projects similar to the one given as input, in which the suggested individual has been involved.

Research Consortium Recommendations. The second approach aims to suggest potential organizations suitable for forming a research consortium, given a project description and objectives. This recommendation process considers organizational expertise, prior involvement in similar research initiatives, and institutional capabilities, ensuring that the suggested consortium members complement the project's goals. This type of recommendation is similar to the first one, but here unlike suggesting a list of researchers who are suitable to collaborate on the project, one or more consortia are suggested. A suggested consortium consists of a group of organisations, which, on the basis of their past participation in other projects, have experience and expertise in achieving one or more of the project objectives.

Architecture Design. Based on the literature requirements and the application requirements derived in the problem awareness phase, the defined architecture and the approach we use allow us to fulfill these requirements. The fulfillment explanation of each requirement is given later in the description of the proposed architecture. As shown in Figure 2, the architecture consists of three main components: the *Retrieval Component* (red box), the *Augmentation Component* (purple box), and the *Generation Component* (blue box). Before describing each component, we provide a brief overview of the system's user interface, the graph database, the embedding model and vector database, used in the architecture.

The user interface of the system is designed as a chatbot to provide a user-friendly experience for researchers. The chatbot provides information retrieval and the two types of recommendations illustrated in previous subsection. The user can request information on people, projects and organisations details etc. in the context of European research projects. It allows users to input a project description and objectives, and receive recommendations for potential research collaborators and consortia. The interface is designed to be intuitive and easy to use, with clear instructions and guidance on how to input the required information. The recommendations are displayed in a clear and structured format, with detailed information about each recommended collaborator or consortium, including their expertise, experience, and relevance to the project.

The EURIO KG is stored as RDF format in a graph database, which provides a flexible and efficient way to represent and query the complex relationships between entities in the system. The graph database allows for the storage of structured data about research projects, organizations, and participants, and enables the retrieval of relevant information based on the relationships between these entities.

The embedding model and the vector database play a crucial role in the retrieval process of the proposed system. EURIO's KG, stored in a graph database, is transformed through an embedding model, which converts structured information into numerical vector representations. These embeddings capture the semantic relationships between entities in the KG, enabling efficient similarity searches. The generated vector embeddings are stored in a vector database, facilitating rapid retrieval of relevant knowledge. When a user submits a query, it is also converted into an embedding and compared against the stored vectors in a similarity search to identify relevant projects, participants, or research entities (satisfied LR2).

Our approach combines the GraphRAG and AgenticRAG paradigms to create a hybrid AI architecture that leverages the structured data in the EURIO KG and the vector database to generate contextually accurate and consistent recommendations. According to Singh et al. (2025), modern agents, such as LLM-powered and mobile agents, are intelligent entities that can perceive their environment, reason about it, and autonomously perform tasks. In our architecture, 3 agents were defined: the *Project-Participants-Information Agent*, the *Potential Collaborators Agent*, and the *Potential Consortium Organisations Agent*. These agents follow two agentic workflow patterns (Singh et al. (2025)): the tool use pattern and the multi-agent collaboration pattern (satisfied LR4). As explained in the following paragraphs, the proposed architecture consists of three main components: Retrieval, Augmentation, and Generation, and agents work together within these components, using external tools to expand their capabilities to achieve specific goals.

Retrieval Component. This component is responsible for retrieving relevant information from the EURIO KG and the vector database. The agent workflow is always initiated by the master agent: the *Project-Participants-Information Agent*. This agent is responsible for returning information about projects, such as the (e.g. project abstract), and also for returning information about the participants involved in that project (e.g. person full name, organization details). To perform the retrieval of this information, the agent was specified to follow a prompt template, in which it is instructed to generate SPARQL Protocol and RDF Query Language (SPARQL) queries to query the EURIO KG from which to extract data. Depending on the task to be performed, the master agent will delegate that task to the agent responsible for that task, if necessary.

Augmentation Component. In addition to combine the user query, prompt templates, and information retrieved from the agent master, with all relevant information obtained from the similarity search, this component adds additional data from the activity performed by the agent workflow. Cosine similarity was used as a metric for semantic search to determine the similarity of embeddings. In this component, the agents *Potential Collaborators Agent* and *Potential Consortium Organisations Agent* are responsible for creating the recommendations. The *Potential Collaborators Agent* has the task of recommending research collaborators, given a project description as input. The *Potential Consortium Organisations Agent* is tasked with creating several consortia formed by various organisations, which contribute to the consortium in a complementary way, i.e. there will be no consortia with organisations specialising in the same research area. These two agents use the *Recommend collaborators* tool, and the *Recommend consortium organisations* tool, respectively, to accomplish the tasks previously described, and they too are instructed with specific prompt templates to follow to achieve their goal. In addition, both use the *Search web* tool, to enrich the information obtained, such as a researcher's areas of interest (satisfied APR3), which are not present within the EURIO KG. In this way, agents act autonomously and are able to make dynamic decisions, resulting in better results than a graph.

Generation Component. Finally, the generation component combines all previously retrieved and augmented information with the pre-trained knowledge of the LLM to generate contextually accurate and consistent responses. In addition to provide relevant recommendations, this component ensures explainability by offering detailed justifications for suggested research collaborators, outlining the reasons behind their selection based on expertise, involvement in previous projects and research alignment. It also describes how the proposed research consortia are composed, specifying the complementary roles of the participating organisations and their collective ability to achieve the project objectives.

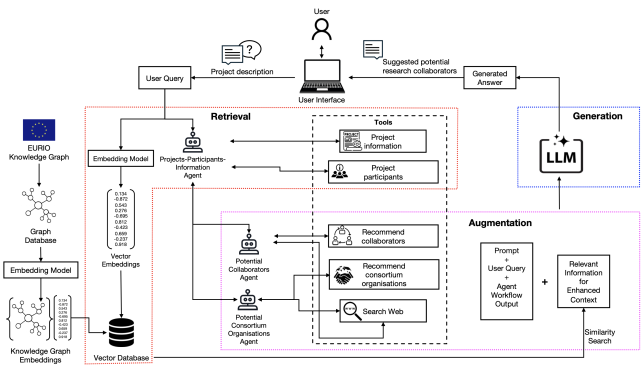


Figure 2. The proposed Knowledge-Driven and Hybrid AI Architecture for Research Collaboration

Technological Implementation

The proposed system is built entirely in Python, using Streamlit for both the back-end and front-end to simplify web app development and sharing. GraphDB by Ontotext was chosen as the graph database for our system because of its robust support for the RDF standard, enabling semantic data representation, efficient SPARQL querying, and easy integration with KGs for enhanced reasoning and retrieval.

Additionally, Chroma was selected as the vector database for storing KG embeddings due to its efficient similarity search, scalable storage, and suitable integration with machine learning pipelines, ensuring fast and accurate retrieval of relevant entities. To compute embeddings, we selected the all-MiniLM-L6-v2 model due to its efficiency and strong performance in semantic similarity tasks. This model is particularly well-suited for encoding short to medium-length text, making it ideal for our use case, where we store and retrieve project titles and abstracts from the EURIO dataset. By leveraging a lightweight transformer architecture, all-MiniLM-L6-v2 balances accuracy and computational efficiency, enabling fast and effective similarity search within our system. To implement our GraphRAG and AgenticRAG approach, we utilized two main frameworks: LlamaIndex was employed to build the agent workflow, coordinate agents, and integrate tools, while LangChain primarily facilitated the generation of SPARQL queries from user query inputs. For response generation, we utilized two state-of-the-art LLMs: GPT-4o, developed by OpenAI, and Claude 3.5 Sonnet, developed by Anthropic. The context window and maximum output tokens for both models are detailed in Table 1, which illustrates their suitability for handling long input queries and generating high-quality responses. The token limit represents the maximum number of tokens a model can handle in a single input, where a token typically corresponds to approximately $^3/_4$ of a word or four characters.

Table 1. Context window and token limit of models used in our work

Model	Context Window	Max Output Tokens
GPT-4o 2024-05-13	128K tokens	4,096 tokens
Claude 3.5 Sonnet 2024-10-22	200K tokens	8,192 tokens

Finally, we containerized the entire system using Docker to ensure portability, reproducibility, and scalability across different environments.

Experiments and Discussion

The effectiveness of the approach has been evaluated by testing the performance of the prototype in terms of how well the collaborators are recommended. For this, an evaluation dataset, for both collaborators and consortia recommendation types, consisting of 10 user queries and 10 associated potential answers was built. Fig. 3 shows an example of a query presented in our evaluation dataset (Fig. 3a), and the corresponding response (Fig. 3b) generated using GPT-4o in our system. (Fig. 3c) shows the response for a similar prompt but asking to find potential collaborators. Since the recommendations generated consist of one or more lists of consortia, the figure presents only part of this response. For a full inspection of the code, evaluation dataset and the results obtained, they can be found at the repository link[**]. This section describes the metrics used for the evaluation. Finally, we present a comparison of the performance of the candidate LLMs for the evaluation of our approach.

Evaluation Metrics

Since RAG-based systems rely on structured data retrieval, evaluating their effectiveness requires assessing both the accuracy of the retrieval and the responses generated by the LLM. In this work, we focus on the consortia recommendations produced by our system using the metrics suite of the Retrieval-Augmented Generation Assessment (RAGAs) framework (Es et al. (2024)). Our evaluation aims to measure the relevance and consistency of the suggested contributors, ensuring that the retrieved knowledge is in line with user demands. The metrics used and the results of our evaluation, are discussed as follows.

Faithfulness evaluates the factual consistency of the generated response with respect to the provided context. It is derived from both the answer and the retrieved information, with scores normalized to a $(0, 1)$ range, where higher values indicate better consistency.

$$\text{Faithfulness} = \frac{\text{Number of claims in the response supported by the retrieved context}}{\text{Total number of claims in the response}}$$

Answer Relevancy (AR) measures how well the generated response aligns with the given prompt. Lower scores indicate incomplete or redundant answers, while higher scores reflect better relevance. It is computed as the mean cosine similarity between the original question and artificially generated questions derived from the answer. Though typically ranging from 0 to 1, the score is not strictly bounded due to cosine similarity's -1 to 1 range.

$$\text{AR} = \frac{1}{N}\sum_{i=1}^{N}\frac{E_{g_i}\cdot E_o}{\|E_{g_i}\|\|E_o\|}$$

where:

- E_{g_i} is the embedding of the generated question i.
- E_o is the embedding of the original question.
- N is the number of generated questions, which is 3 by default.

Context Precision (CP) measures how well ground-truth relevant items appear at the top ranks in the retrieved contexts. It is computed using the question, ground truth, and retrieved contexts, with values ranging from 0 to 1, where higher scores indicating better precision.

$$\text{CP@K} = \frac{\sum_{k=1}^{K}(\text{Precision@k}\times v_k)}{\text{Total number of relevant items in the top } K \text{ results}}$$

$$\text{Precision@k} = \frac{\text{true positives@k}}{\text{true positives@k} + \text{false positives@k}}$$

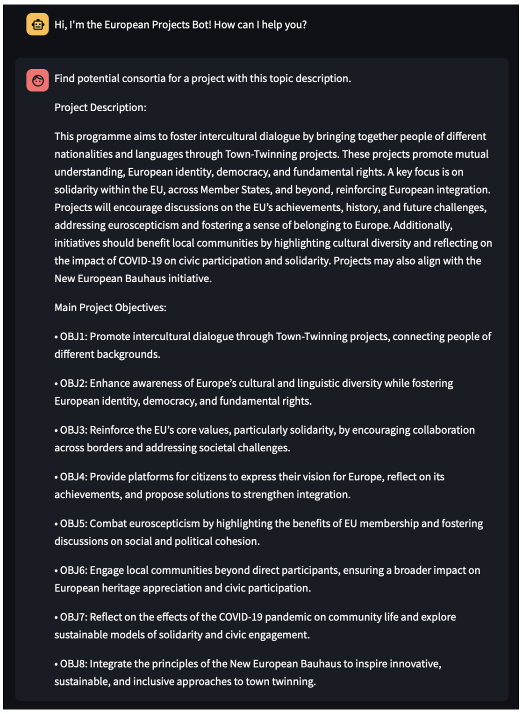
Where K is the total number of retrieved chunks, and $v_k \in \{0,1\}$ indicates relevance at rank k.

Context Recall (CR) evaluates how well the retrieved context aligns with the ground-truth answer. It is computed using the question, ground truth, and retrieved context, with values ranging from 0 to 1, where higher scores indicating better alignment. Each claim in the ground-truth answer is checked for attribution to the retrieved context, with ideal recall achieved when all claims are supported.

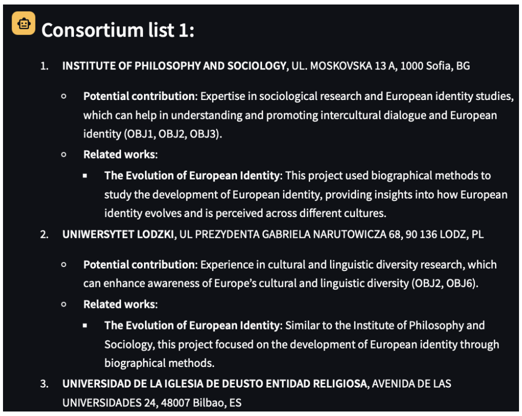
$$\text{CR} = \frac{\text{Number of claims in the reference supported by the retrieved context}}{\text{Total number of claims in the reference}}$$

Context Entity Recall (CER) measures how well retrieved contexts capture entities from the ground truth. It is defined as the fraction of entities in the ground truth that are also found in the retrieved contexts. This metric helps assess retrieval effectiveness in entity-focused tasks. To compute

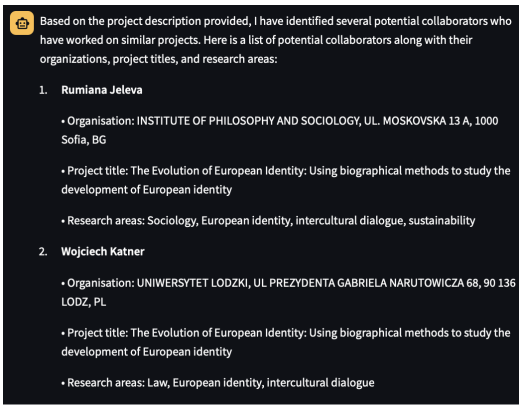
[**]Code and our experiments are available at `https://github.com/Piermuz7/MasterThesisProject.git`



(a)



(b)



(c)

Figure 3. Excerpts of the experiments. (a): The user need is to find organizations forming a consortium on intercultural dialogue. (b): The response recommends 4 organisations, their potential contributions, and related works, for the user query of Fig. 3a. (c): The response, related to a user query similar to Fig. 3a, recommends 2 research collaborators, their employment organisations, project titles similar to the one asked in the user query, and their research areas.

it, we use two sets: GE (entities in ground truths) and CE (entities in retrieved contexts).

$$\text{Context Entity Recall} = \frac{|CE \cap GE|}{|GE|}$$

Answer Semantic Similarity (SS) measures how closely the generated answer aligns with the ground truth. Scores range from 0 to 1, with higher values indicating better alignment. This evaluation uses a cross-encoder model to compute the similarity, providing insights into response quality.

Answer Correctness (AC) measures how accurately the generated answer aligns with the ground truth, with scores from 0 to 1, where higher values indicating better correctness. It considers both semantic and factual similarity, combined using a weighted scheme. A threshold can be applied to convert the score to a binary value if needed.

Evaluation Datasets

A RAG pipeline is evaluated based on ground truths. RAGAs is particularly distinguished by the introduction of a "reference-free" evaluation approach (Es et al. (2024)). This means that, instead of relying solely on human-annotated ground truth labels, the framework uses LLMs to perform automatic evaluations, reducing the need for manual intervention. To evaluate the performance of a RAG pipeline, RAGAs requires an evaluation dataset with the following inputs:

- **question**: the user query that serves as the input to the RAG pipeline.
- **answer**: the response generated by the RAG pipeline.
- **contexts**: the retrieved contextual information from an external knowledge source that supports answering the question.
- **ground truths**: the human-annotated reference answer to the question.

We also integrated GPT-4o within the RAGAs framework to evaluate the performance and quality of our RAG pipeline. As mentioned earlier, our evaluation focuses on the recommendation of collaborators, and the recommendation of consortia. For this purpose, we created two evaluation datasets, one for each recommendation type, each consisting of 10 user queries paired with 10 corresponding ground truth answers. The datasets were created by manually annotating the ground truths for each question. Both were created based on the actual descriptions of European projects on the EU website calls. Fig. 3a is an example question within the evaluation datasets. Due to their length, the ground truth responses are not included in this thesis. However, as mentioned earlier for the codebase, for a complete inspection of the evaluation datasets, including the ground truths, a detailed repository is available on GitHub at `https://github.com/Piermuz7/MasterThesisProject.git`.

LLMs Performance Comparison

The evaluation results provide a comparative analysis of GPT-4o 2024-05-13 and Claude 3.5 Sonnet 2024-10-22 on the Agentic Graph RAG approach, based on key retrieval and answer quality described in the previous subsection. Table 2 shows the results, which highlight differences in the LLM ability to retrieve relevant context and generate semantically accurate consortia recommendations. Even though the token limits of context windows and the maximum output tokens are lower (see Table 1), GPT-4o outperforms Claude 3.5 Sonnet in AR, AR, SS and AC, being more effective in generating relevant and accurate responses. Claude, while slightly better in faithfulness, struggles significantly in CP and CER, leading to less effective retrieval. Overall, GPT-4o provides more aligned, complete and contextually correct responses and recommendations, while Claude is more grounded but less effective in retrieving, structuring relevant information and thus recommending consortia. A better performance of GPT-4o also comes from the use of agents during experiments execution, where the *potential consortium organisations agent* was used correctly for all 10 queries. In contrast, Claude 3.5 Sonnet incorrectly called *potential collaborators agent*. As can be seen from the responses generated, Clause Sonnet's recommendations are much more concise than GPT's, in particular Sonnet has no relevant works associated with the recommended organisation and in several responses the potential contributions the organisation could make to the consortium are not even suggested. As a result of the Agentic Graph RAG approach, neither LLMs hallucinates.

Table 2. LLMs performance comparison in the used Agentic Graph RAG approach

Model	Faithfulness	AR	CP	CR	CER	SS	AC
M1	0.314	0.93	0.076	0.266	0.298	0.716	0.469
M2	0.336	0.767	0.000	0.147	0.278	0.653	0.260

M1 = GPT-4o 2024-05-13, M2 = Claude 3.5 Sonnet 2024-10-22

Conclusions and Future Work

This study presents an Agentic Graph RAG approach that delivers contextualized and explainable recommendations for research collaborator selection. The method integrates the strengths of KGs and LLMs and was developed using the Design Science Research methodology. To carry out this approach, a user interface in the form of a chatbot web application was developed. In particular, the development to manage the agent flow was realised via the LlamaIndex library. To enable SPARQL queries to be generated from the user's text, the Langchain library was used instead. Regarding the evaluation of the artefact, experiments of both collaborators and consortia recommendations indicate that using this Agentic Graph RAG approach results in high quality retrieval, contextual reasoning and reduced hallucinations. However, aspects concerning consistency and context retrieval do not present relatively positive results, indicating that they should be improved. Experiments indicate that GPT-4o outperforms other LLMs in RAG-based recommendation metrics, demonstrating superior retrieval quality, contextual reasoning, and reduced hallucinations.

The recommendations are specifically aligned with researchers' areas of expertise and project relevance, making them more effective than traditional methods.

This work contributes to the field of hybrid AI, particularly by advancing AI-assisted research networks aimed at enhancing collaboration opportunities and enabling scalable, automated, and interdisciplinary research connections.

Future research directions include the automated updating of the KG, incorporating data from Horizon Europe projects to ensure the system remains current with emerging research trends. Additionally, integrating academic KGs could further enrich the EURIO knowledge base, providing greater contextual depth for both researcher and publication recommendations.

References

Börner K, Conlon M, Corson-Rikert J and Ding Y (2012) Vivo a semantic approach to scholarly networking and discovery. *Synthesis Lectures on the Semantic Web: Theory and Technology* 2: 1–180.

d'Avila Garcez A and Lamb LC (2023) Neurosymbolic AI: the 3rd wave. *Artif. Intell. Rev.* 56(11): 12387–12406.

Deldjoo Y, He Z, Mcauley J, Korikov A, Sanner S, Ramisa A, Vidal R, Sathiamoorthy M, Kasirzadeh A and Milano S (2024) A review of modern recommender systems using generative models (gen-recsys). In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery. ISBN 9798400704901, pp. 6448–6458.

Du O and Li Y (2022) Academic collaborator recommendation based on attributed network embedding. *Journal of Data and Information Science* 7: 37–56.

Es S, James J, Espinosa Anke L and Schockaert S (2024) RAGAs: Automated evaluation of retrieval augmented generation. In: Aletras N and De Clercq O (eds.) *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. St. Julians, Malta: Association for Computational Linguistics, pp. 150–158.

Hevner A and Chatterjee S (2010) *Design Science Research in Information Systems*. Boston, MA: Springer US. ISBN 978-1-4419-5653-8, pp. 9–22.

Hussien FTA, Rahma AMS and Wahab HBA (2021) Recommendation systems for e-commerce systems an overview. In: *Journal of Physics: Conference Series*, volume 1897. IOP Publishing Ltd, p. 012024.

Kanwal T and Amjad T (2024) Research paper recommendation system based on multiple features from citation network. *Scientometrics* .

Katz J and Martin BR (1997) What is research collaboration? *Research Policy* 26(1): 1–18.

Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK and Zhou T (2012) Recommender systems.

Meißuner F, Weinmann C and Vowe G (2021) Understanding and addressing problems in research collaboration: A qualitative interview study from a self-governance perspective. *Frontiers in Research Metrics and Analytics* 6.

Murali MV, G VT and Victor N (2019) A collaborative filtering based recommender system for suggesting new trends in any domain of research. In: *Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE. ISBN 978-1-5386-9533-3, pp. 550–553.

Prater R and Laurenzi E (2022) A hybrid intelligent approach for the support of higher education students in literature discovery. In: Martin A, Hinkelmann K, Fill H, Gerber A, Lenat D, Stolle R and van Harmelen F (eds.) *Proceedings of the AAAI 2022 Spring Symposium on Machine Learning and Knowledge Engineering for Hybrid Intelligence (AAAI-MAKE 2022), Stanford University, Palo Alto, California, USA, March 21-23, 2022*, *CEUR Workshop Proceedings*, volume 3121. CEUR-WS.org.

Priem J, Piwowar H and Orr R (2022) Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts.

Publications Office of the European Union (2015a) Cordis - eu research projects under fp7 (2007-2013). Data set.

Publications Office of the European Union (2015b) Cordis - eu research projects under horizon 2020 (2014-2020). Data set.

Publications Office of the European Union (2018) Cordis reference data. Data set. Originally published in 2015.

Rordorf D, Käser J, Crego A and Laurenzi E (2023) A hybrid intelligent approach combining machine learning and a knowledge graph to support academic journal publishers addressing the reviewer assignment problem (RAP). In: Martin A, Fill H, Gerber A, Hinkelmann K, Lenat D, Stolle R and van Harmelen F (eds.) *Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023), Hyatt Regency, San Francisco Airport, California, USA, March 27-29, 2023*, *CEUR Workshop Proceedings*, volume 3433. CEUR-WS.org.

Singh A, Ehtesham A, Kumar S and Khoei TT (2025) Agentic retrieval-augmented generation: A survey on agentic rag. URL https://arxiv.org/abs/2501.09136.

Tejeda-Lorente A, Porcel C, Bernabé-Moreno J and Herrera-Viedma E (2015) Refore: A recommender system for researchers based on bibliometrics. *Applied Soft Computing Journal* 30: 778–791.

Valluru S, Widener M, Srivastava B, Natarajan S and Gangopadhyay S (2024) Ai-assisted research collaboration with open data for fair and effective response to call for proposals. *AI Magazine* 45: 457–471. DOI:10.1002/aaai.12203.

Vrandečić D and Krötzsch M (2014) Wikidata: A free collaborative knowledgebase. *Communications of the ACM* 57: 78–85.

Wade AD (2022) The semantic scholar academic graph (s2ag). In: *Companion Proceedings of the Web Conference 2022*, WWW '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391306, p. 739.

Wang K, Shen Z, Huang C, Wu CH, Dong Y and Kanakia A (2020) Microsoft academic graph: When experts are not enough. *Quantitative Science Studies* 1: 396–413.

Zhao Z, Fan W, Li J, Liu Y, Mei X, Wang Y, Wen Z, Wang F, Zhao X, Tang J and Li Q (2024) Recommender Systems in the Era of Large Language Models (LLMs) . *IEEE Transactions on Knowledge & Data Engineering* 36(11): 6889–6907.

Zhu J and Yaseen A (2022) A recommender for research collaborators using graph neural networks. *Frontiers in Artificial Intelligence* 5.

Copyright statement

Please be aware that the use of this LaTeX 2_ε class file is governed by the following conditions.

Copyright

Rules of use

This class file is made available for use by authors who wish to prepare an article for publication in a *SAGE Publications* journal. The user may not exploit any part of the class file commercially.

This class file is provided on an *as is* basis, without warranties of any kind, either express or implied, including but not limited to warranties of title, or implied warranties of merchantablility or fitness for a particular purpose. There will be no duty on the author[s] of the software or SAGE Publications Ltd to correct any errors or defects in the software. Any statutory rights you may have remain unaffected by your acceptance of these rules of use.

Acknowledgements

This class file was developed by Sunrise Setting Ltd, Brixham, Devon, UK.
Website: http://www.sunrise-setting.co.uk