Fall 2024-2025

CS 491 – Senior Design Project I

Project Specification Report

T2422 - Article Lens

Eray Yapağcı - 22103242

Mehmet Dedeler - 22003168

Alper Göçmen - 22002948

Mustafa Berkan Yıkılmaz - 22003325

Serra Çapraz - 22102314

Table of Contents

1. Introduction

Researchers and professionals often face the challenge of identifying relevant studies, requiring them to go through vast amounts of information to identify the most relevant studies, which is both time-consuming and inefficient. Article-Lens aims to address these challenges with an AI-powered research assistant that streamlines academic workflows and helps users stay ahead in their fields.

1.1 Description

Article-Lens offers two core features which are a personalized newsletter and a research guide. The newsletter utilizes AI to filter, summarize, and rank new publications on platforms like arXiv, delivering customized daily updates based on user-defined categories and queries. This ensures users can stay informed without the burden of manual filtering. The research guide provides a structured learning path for specific topics or articles, ranking references by relevance, complexity, and citation count to create a learning journey from foundational to advanced knowledge.

With the integration of advanced large language models (LLMs), Article-Lens ensures summaries are accurate, concise, and personalized. By enabling researchers to focus on impactful studies and follow reading paths, Article-Lens significantly enhances the efficiency and depth of academic exploration.

1.2 High-Level System Architecture & Components of Proposed Solution

The high-level architecture of Article-Lens consists of three main components: Data Ingestion & Preprocessing, Core Processing Module, and User Interaction Layer.

The Data Ingestion & Preprocessing component is designed to gather and prepare data exclusively from arXiv and Google Scholar, ensuring it is ready for analysis. Its Data Collection Engine interfaces with the APIs of these platforms to fetch the latest publications based on user-selected categories. Once the data is retrieved, the Data Preprocessing stage cleans and structures the raw information by extracting metadata such as titles, abstracts, authors, PDFs, keywords, and publication dates, along with additional details about authors and their affiliations. Relevant papers are identified using a Relevance Filtering mechanism, which employs an initial binary classification model with LLM as a judge to evaluate the relevance of a paper based on its abstract. Irrelevant papers are discarded, streamlining subsequent processing steps. Processed data, including summarized versions or selected metadata, is stored in a local database, enabling efficient retrieval for personalized recommendations and further analysis.

The Core Processing Module powers Article-Lens's primary functionalities through the use of AI and advanced algorithms. It includes a Summarization Engine that utilizes large language models (LLMs) to create sectioned summaries with options for users to define custom sections like "Disadvantages." The Ranking Engine comprises two key submodules: Author Scoring, which evaluates authors based on their affiliations and h-index metrics, and Paper

Scoring, which assesses originality, impact, and relevance using weighted criteria adjustable by users. The Research Guide Module enhances this capability by extracting references from articles or topics and ranking them by factors such as publication date, citation count, and complexity, providing users with a structured reading path. A Personalization Module further refines scoring and ranking based on user preferences, ensuring tailored recommendations.

The User Interaction Layer ensures a seamless and engaging platform experience for users. It includes a responsive Web Interface compatible with desktops, tablets, and mobile devices, offering intuitive navigation for selecting categories, queries, and customization options. A Notification System keeps users informed through daily newsletters sent via email or platform notifications, providing timely updates. Additionally, the User Profile & Preferences feature allows users to manage their queries, categories, and scoring weights, securely storing search history and preferences to deliver a personalized and efficient research experience.

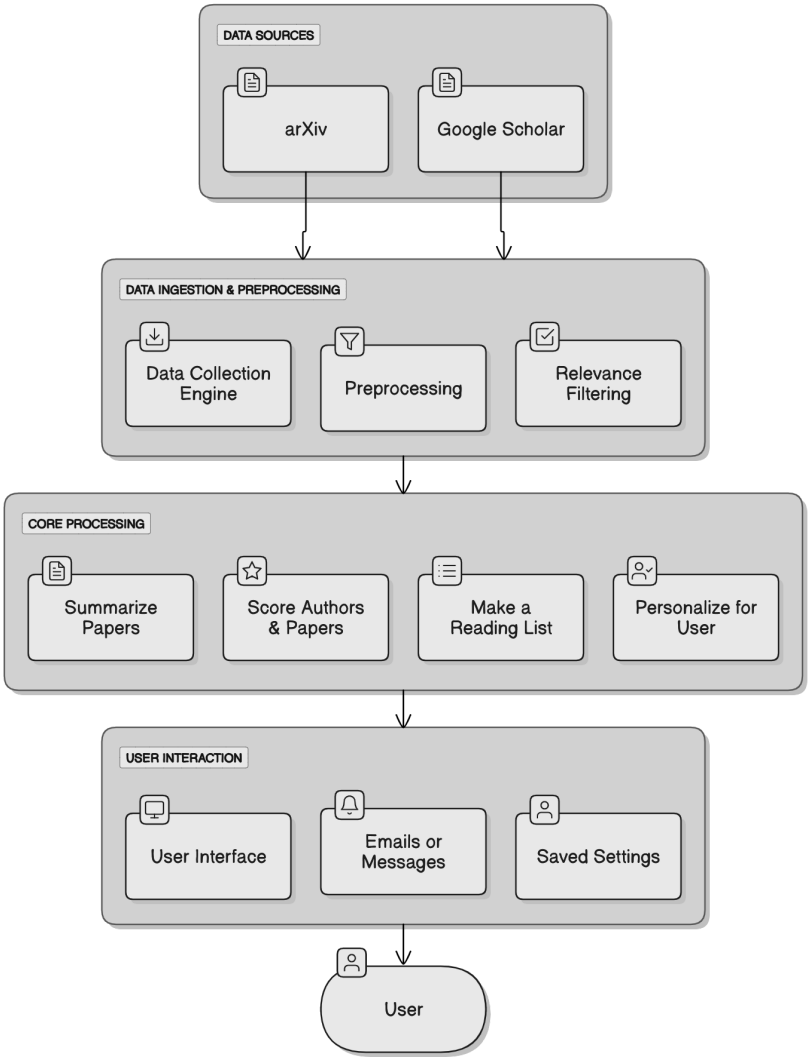

Figure 1: High-Level Diagram

1.3 Constraints

This project faces some constraints at various stages, both during the development and post-deployment stages. While some constraints depend on the user base that will be available after deployment, others are introduced due to the nature of the project, such as the ones imposed by using LLMs. These constraints can be categorized into three main areas: implementation, economic, and ethical constraints, as detailed below:

1.3.1. Implementation Constraints

- This project relies heavily on AI-based tools like LLMs for summarization and ranking, which may require high-performance computing resources. Platforms such as arXiv and Google Scholar may impose limitations on API usage and rate limits to integrate these tools with the necessary data.
- The effectiveness of relevance filtering and summarization depends on access to complete and accurate data, which might not be available for all articles and/or researchers. The project should be able to work correctly in such cases.
- Providing flexible user-defined scoring and ranking criteria increases system complexity, requiring a robust yet user-friendly interface design.

1.3.2. Economic Constraints

- While arXiv provides a public API that can be used freely, retrieving data from Google Scholar requires web scraping. Using paid tools and services for scraping might introduce significant recurring costs, especially as the user base grows.
- Cloud resources that might be used to serve the LLM, as well as storing large volumes of research data, must be scalable. This might introduce increased operational expenses.

1.3.3. Ethical Constraints

- As user data and preferences may be collected to enhance personalization, strict adherence to privacy laws is mandatory.
- Ranking algorithms may inadvertently favor some authors, institutions, or articles. Provisions must be made to mitigate these biases.
- The summarization and use of academic papers, as well as using some author data, must respect copyright laws, limiting access to open-access articles or those permitted under fair-use policies.

1.4 Professional and Ethical Issues

The development and deployment of the "Article-Lens" project raise several professional and ethical considerations, primarily due to the handling of academic data and the use of AI in summarizing and ranking research content. These considerations include the following:

- **Data Privacy**: "Article-Lens" involves processing data about research articles and potentially user preferences. It is essential to ensure that any user data is handled in compliance with data privacy standards, such as GDPR [1] and KVKK [2], if user accounts or preferences are stored. Privacy-preserving measures, like anonymizing user data and applying stringent access controls, will be implemented to protect personal information and respect user confidentiality.
- **Intellectual Property**: Academic papers and their summaries are subject to copyright laws. "Article-Lens" must operate within these laws, only accessing or summarizing articles that are open access or permissible under fair use policies. The system will avoid storing or redistributing full-text copies of restricted access papers, focusing instead on summaries and metadata that respect intellectual property rights.
- **Bias and Fairness**: The scoring and ranking algorithms for research papers are based on criteria such as h-index, citation counts, and institutional affiliations. These metrics, while valuable, may introduce biases, particularly favoring more established authors or institutions. To mitigate this, "Article-Lens" provides users with the option to customize ranking criteria, allowing for a fairer and more tailored presentation of relevant research.
- **Transparency**: In line with ethical best practices, "Article-Lens" will maintain transparency regarding the criteria used for scoring and ranking research papers. Users will have access to information about the ranking methodology and the option to adjust scoring weights. This transparency is crucial for building trust and ensuring that users understand how recommendations are generated.
- **Accuracy and Reliability**: The quality of summaries and rankings directly impacts user trust. Inaccurate summaries could mislead researchers, so the system will employ validation techniques to ensure that summaries and rankings reflect the content accurately. Efforts will be made to periodically assess the quality of generated summaries and adapt the underlying algorithms as needed to maintain reliability.

1.5 Standards

To ensure the "Article-Lens" system is developed and maintained according to industry standards, the following standards and guidelines will be adhered to:

- **Data Security Standards**: The project will comply with **ISO/IEC 27001** for data security management, especially if user data or login credentials are stored [3]. This standard ensures that security risks are systematically assessed and mitigated to protect sensitive information.
- **Metadata Standards for Academic Content**: The **Dublin Core Metadata Standard** will be utilized for organizing and retrieving academic content effectively [4]. This standard provides a structured way to manage metadata (such as author, title, and publication year), enhancing the organization and discoverability of research articles.

- **Usability Standards**: "Article-Lens" will follow **ISO 9241** standards for usability, which are essential to ensure the platform is user-friendly and meets the needs of researchers [5]. This standard guides the design of interfaces to maximize ease of use and satisfaction, particularly for time-pressed researchers who need an intuitive tool.
- **Interoperability Standards**: To integrate with external databases and reference tools, "Article-Lens" will incorporate **OpenURL** and **DOI** standards [6],[7]. These standards facilitate direct linking to academic resources and consistent referencing, enhancing interoperability and access to research material.

By adhering to these standards, "Article-Lens" aims to provide a reliable, secure, and accessible tool that respects both user needs and the academic ecosystem.

2. Design Requirements

This section outlines the design requirements for the Article-Lens platform, detailing both the functional and non-functional aspects necessary to meet user needs and ensure system efficiency.

2.1. Functional Requirements

The functional requirements define the essential operations and features that the Article-Lens platform must provide.

User Initiation

- Users must be able to select one or more categories from arXiv, such as "Computation & Language," to tailor the scope of papers retrieved.
- They must also be able to input specific topics or queries of interest, like "Prompt Injection," to further refine the selection of relevant papers.

Daily Paper Retrieval

- The system must automatically collect all new papers daily from the selected arXiv categories for the newsletter.

Relevance Filtering

- The system must analyze the abstracts of retrieved papers to determine their relevance to the user's query using a binary classification (Yes or No).

Summary Generation

- For each relevant paper, the system must generate summaries divided into sections like Title, Findings, and Framework.
- Users must have the ability to add or modify summary sections based on their needs, such as adding a "Disadvantages" section.
- The system must utilize a Language Learning Model (LLM) to generate tailored summaries according to the specified sections.

Paper Ranking

- Authors are assessed based on their affiliated institutions and their h-index, with individual author scores summed to contribute to the paper's overall ranking.
- The paper's originality and impact within its field are evaluated, assigning scores out of 10 for each criterion using a predefined framework prompted by the LLM.
- Scores are combined using a weighted sum to rank papers effectively.
- Users are allowed to add their classifications and adjust weightings within the scoring system to personalize rankings.

Personalized Newsletter Delivery

- The top-ranked papers are aggregated into a personalized newsletter format.
- Users are provided options to receive the newsletter via email or access it through the platform.

Research Guide Feature

- Users can input a specific article or topic of interest.
- The system extracts references or related works from external databases based on the provided input.
- References are ranked based on publication date, citation count, and complexity level (introductory or advanced).
- An ordered list of papers is presented, starting with foundational works and progressing to advanced research.
- Users can follow the suggested path or customize it according to their preferences.

User Account Management

- The platform provides functionalities for users to manage their profiles and preferences.
- Users can save queries and access their search history for convenience.

Notification System

- Users are notified of new relevant papers, updates in their fields of interest, or changes in the research guide.

2.2. Non-Functional Requirements

The non-functional requirements address the quality attributes of the system, ensuring it is user-friendly, reliable, efficient, and scalable.

2.2.1. Usability

- The platform must feature a clean, intuitive user interface that facilitates easy navigation and operation.
- Users must be able to customize summary sections and scoring criteria without technical difficulties.
- Compatibility across various devices and screen sizes, including desktops, tablets, and mobile devices, must be ensured.
- Accessibility standards must be adhered to, accommodating users with disabilities.

2.2.2. Reliability

- The system must be operational 99.9% of the time to provide consistent access.
- Data correctness and integrity must be ensured in the information retrieved and presented.

- Mechanisms must be implemented to handle and recover from system errors or failures without data loss.
- Frequent data backups should be scheduled to prevent loss in case of unexpected issues.

2.2.3. Performance

- The system must be able to send the newsletter daily as soon as the papers are out in arXiv and process them at least in 5 minutes for a user.
- Optimal performance levels must be maintained even as user numbers and data volume increase.
- Computational resources must be efficiently managed, especially when interacting with LLMs, to prevent bottlenecks. The LLMs should be not only selected in performance but also speed as well, looking for fast API providers.

2.2.4. Supportability

- A modular and well-documented codebase must be utilized to facilitate maintenance and future development.
- The system architecture must be designed to allow easy integration of new features or updates.
- Users must have access to help resources, FAQs, and customer support channels for assistance.
- Comprehensive logging and monitoring must be implemented to detect and address issues promptly.

2.2.5. Scalability

- The system must support scaling out by adding more machines or instances to handle increased load.
- Cloud services should be leveraged to dynamically allocate resources based on demand.
- Load balancing strategies must be implemented to distribute workloads evenly across servers, enhancing performance and reliability.
- Databases must be able to handle increased read/write operations without performance degradation.

3. Feasibility Discussions

3.1. Market & Competitive Analysis

The research and academic publishing industry has experienced unprecedented growth, driven by rapid technological advancements. With over 2 million scientific articles published annually across diverse domains, staying current with the latest developments has become increasingly challenging. Researchers face the critical need to efficiently identify, process, and comprehend vast amounts of new publications to implement novel projects and advance their research endeavors.

Article-Lens addresses this challenge by serving as a comprehensive research guide for four key market segments:

Academic Researchers: Professors, PhD students, and postdoctoral researchers who need to stay updated on the latest developments in their fields.

Industry Researchers: Companies engaged in AI, machine learning, and cutting-edge technology rely on the latest research for innovation and product development.

Research Institutions and Libraries: Organizations that need efficient tools for managing research discovery, topic exploration, and citation management.

Research Enthusiasts: Individuals who are deeply interested in specific scientific fields and need a tailored solution to stay informed.

In total, the growing number of researchers globally is estimated at nearly 9 million active researchers [8]. The increasing volume of academic content on platforms like **Arxiv** has seen exponential growth due to the popularity of open-access preprints.

Scholarcy [9] utilizes AI to condense research papers into concise summaries, transforming complex texts into digestible flashcards. Its primary focus is on summarization, lacks advanced recommendation features and personalized reading paths that Article-Lens intends to offer.

Google Scholar [10] is a freely accessible web search engine that indexes scholarly articles across various disciplines and formats. It provides users with a broad search capability, citation metrics, and links to full-text articles when available. However, Google Scholar primarily functions as a search engine and lacks features such as personalized summaries, daily newsletters, and structured reading guides, which are central to Article-Lens's offerings.

Connected Papers [11] is a visual tool that generates graphs of research papers related to a user-provided seed paper, facilitating the exploration of academic literature through visual networks. While it excels in visualizing connections, it doesn't offer personalized summaries or daily updates, which are key features of Article-Lens.

ResearchRabbit [12] is an AI-powered research tool designed to assist users in discovering and organizing academic literature. It offers features such as personalized

recommendations, interactive visualizations of research networks, and the ability to create and share collections of papers. However, it does not provide structured reading guides that suggest an ordered path through the literature. Article-Lens aims to offer a research Guide feature that not only recommends relevant articles but also suggests a sequential reading path, progressing from foundational to advanced papers. This approach facilitates efficient knowledge acquisition and is particularly beneficial for users new to a specific topic or those seeking a structured learning experience.

3.2. Academic Analysis

In developing Article-Lens, we conducted extensive research into Large Language Models (LLMs) to inform our approach to summarization, evaluation, and ranking of academic literature. We aimed to leverage state-of-the-art AI technologies to enhance researchers' ability to navigate and comprehend the vast array of scholarly publications.

LLM-Based Summarization

Article-Lens utilizes LLMs to generate concise summaries of research papers, allowing users to customize sections based on their interests. This approach aligns with findings by [13], who introduced an element-aware summarization method using LLMs. Their technique involves prompting LLMs to generate summaries step by step, integrating fine-grained details that mirror human writing processes. This method demonstrated superior performance over state-of-the-art models, suggesting that Article-Lens's strategy is well-founded [13].

Li et al. [14] introduced SliSum, a summarization technique that enhances the faithfulness and coherence of Large Language Models (LLMs) through sliding generation and self-consistency. This method segments source articles into overlapping windows, generating local summaries for each segment, which are then aggregated using clustering and majority voting. By improving the accuracy of diverse LLMs such as LLaMA-2 and GPT-3.5, SliSum achieves higher faithfulness without additional fine-tuning. Article-Lens could adopt similar methodologies to enhance its summarization capabilities, ensuring that generated summaries maintain both reliability and comprehensiveness. This may be one of the most suitable methods because the method is designed to summarize long articles. This would align with the project's goal of delivering high-quality, user-tailored summaries. [14]

LLM Judgement

The paper "Replacing Judges with Juries: Evaluating LLM Generations with a Panel of Diverse Models" by Verga et al. (2024) proposes a novel evaluation method using a Panel of LLM Evaluators (PoLL) instead of relying on a single large model like GPT-4 for scoring outputs [15]. The study demonstrates that PoLL, composed of multiple smaller models from diverse families, reduces intra-model bias and aligns better with human judgments across tasks. PoLL is shown to outperform single models in reliability while being significantly more cost-efficient. This approach aligns with Article-Lens's goal of enhancing research evaluation by leveraging multi-model inputs for ranking and judgments. Incorporating a panel-based evaluation system could improve the fairness and robustness of paper rankings, ensuring

diverse perspectives and reducing systemic biases inherent in single-model evaluations. This method would further bolster Article-Lens's credibility in delivering nuanced and accurate research insights.

Zheng et al. [16] demonstrated that GPT-based evaluations of text quality exhibit a strong correlation with human annotations, particularly in tasks involving summarization and conversational agents. The study highlights that GPT-4 achieves over 80% agreement with human evaluators, indicating its potential as a scalable and reliable tool for assessing content. This finding supports Article-Lens's use of GPT models for ranking and evaluating research papers. This also makes the backbone of the system, which is heavily reliant on using LLMs as a judge, trustworthy [16].

5. Glossary

LLM: Large Language Model

GPT: Generative Pre-trained Transformer

GDPR: General Data Protection Regulation

KVKK: Kişisel Verileri Koruma Kanunu

ISO: International Organization for Standardization

IEC: International Electrotechnical Commission

DOI: Digital Object Identifier System

6. References

[1] Personal Data Protection Authority," KVKK, [Online]. Available: https://www.kvkk.gov.tr/. [Accessed: Nov. 20, 2024].

[2] General Data Protection Regulation (GDPR) – Official Legal Text," GDPR Info, [Online]. Available: https://gdpr-info.eu/. [Accessed: Nov. 20, 2024].

[3] ISO/IEC 27001: Information security, cybersecurity, and privacy protection — Information security management systems — Requirements," ISO, [Online]. Available: https://www.iso.org/standard/27001. [Accessed: Nov. 20, 2024].

[4] Dublin Core Metadata Element Set, Version 1.1," Dublin Core Metadata Initiative, [Online]. Available: https://www.dublincore.org/specifications/dublin-core/dces/. [Accessed: Nov. 20, 2024].

[5] ISO 9241:Ergonomics of human-system interaction," ISO, [Online]. Available: https://www.iso.org/standard/77520.html. [Accessed: Nov. 20, 2024].

[6] OpenURL Standards," OCLC Research, [Online]. Available: https://www.oclc.org/research/activities/openurl.html. [Accessed: Nov. 20, 2024].

[7] Digital Object Identifier System," DOI, [Online]. Available: https://www.doi.org/. [Accessed: Nov. 20, 2024].

[8] G. Naujokaitytė, “Number of scientists worldwide reaches 8.8M, as global research spending grows faster than the economy,” *Science|Business*, Jun. 14, 2021. https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy (accessed Nov. 16, 2024).

[9] Scholarcy, “Scholarcy”, *Scholarcy.com*. https://www.scholarcy.com/ (accessed Nov. 16, 2024).

[10] Google Scholar, “Google Scholar,” *Google.com*, 2024. https://scholar.google.com/

[11] Connected Papers, “Connected Papers | Find and explore academic papers,” *www.connectedpapers.com*, 2023. https://www.connectedpapers.com/

[12] Research Rabbit, “ResearchRabbit,” *ResearchRabbit*, 2023. https://www.researchrabbit.ai/

[13] Wang, Y., Zhang, Z., & Wang, R. (2023a). Element-aware summarization with large language models: Expert-aligned evaluation and chain-of-thought method. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 8640–8665. https://doi.org/10.18653/v1/2023.acl-long.482

[14] Li, T., Li, Z., & Zhang, Y. (2024, July 31). *Improving faithfulness of large language models in summarization via sliding generation and self-consistency*. arXiv.org. https://arxiv.org/abs/2407.21443

[15] P. Verga et al., “Replacing judges with juries: Evaluating LLM generations with a panel of diverse models,” arXiv.org, https://doi.org/10.48550/arXiv.2404.18796 (accessed Nov. 20, 2024).

[16] L. Zheng et al., “Judging LLM-as-a-judge with MT-bench and Chatbot Arena,” arXiv.org, https://arxiv.org/abs/2306.05685 (accessed Nov. 20, 2024).