

# Acceleron: A Tool to Accelerate Research Ideation

Harshit Nigam<sup>1</sup>, Manasi Patwardhan<sup>2</sup>, Lovekesh Vig<sup>3</sup>, Gautam Shroff<sup>4</sup>

TCS Research

<sup>1</sup>h.nigam@tcs.com, <sup>2</sup>manasi.patwardhan@tcs.com, <sup>3</sup>lovekesh.vig@tcs.com, <sup>4</sup>gautam.shroff@tcs.com

## Abstract

Several tools have recently been proposed for assisting researchers during various stages of the research life-cycle. However, these primarily concentrate on tasks such as retrieving and recommending relevant literature, reviewing and critiquing the draft, and writing of research manuscripts. Our investigation reveals a significant gap in availability of tools specifically designed to assist researchers during the challenging ideation phase of the research life-cycle. To aid with research ideation, we propose ‘Acceleron’, a research accelerator for different phases of the research life cycle, and which is specially designed to aid the ideation process. Acceleron guides researchers through the formulation of a comprehensive research proposal, encompassing a novel research problem. The proposals motivation is validated for novelty by identifying gaps in the existing literature and suggesting a plausible list of techniques to solve the proposed problem. We leverage the reasoning and domain-specific skills of Large Language Models (LLMs) to create an agent-based architecture incorporating colleague and mentor personas for LLMs. The LLM agents emulate the ideation process undertaken by researchers, engaging researchers in an interactive fashion to aid in the development of the research proposal. Notably, our tool addresses challenges inherent in LLMs, such as hallucinations, implements a two-stage aspect-based retrieval to manage precision-recall trade-offs, and tackles issues of unanswerability. To showcase the ideation capabilities of ‘Acceleron’, we illustrate the execution of our motivation validation and method synthesis workflows on proposals from the machine learning and natural language processing domain, given as an input by 3 distinct researchers. Our observations and evaluations provided by the researchers illustrate the efficacy of the tool in terms of assisting researchers with appropriate inputs at distinct stages and thus leading to improved time efficiency.

## Introduction

With fast-paced research happening in every field, we are witnessing an exponential growth in the number of scientific articles and research papers on the web. It is difficult for an individual researcher or a small research team to keep abreast of the relevant advances amidst this information explosion. This has a downstream impact on the ability to be

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

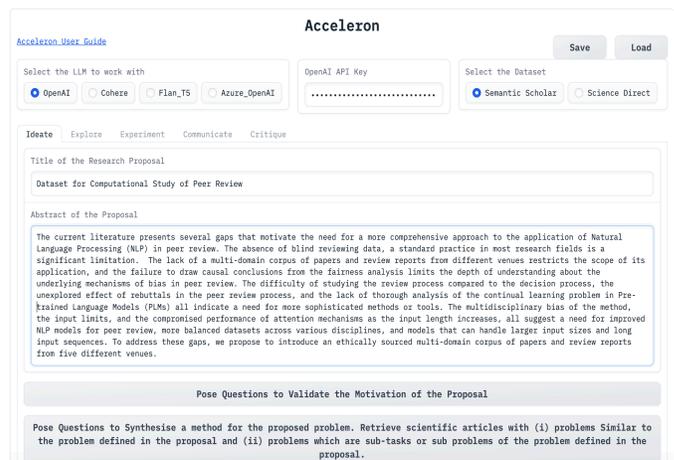


Figure 1: Acceleron Interface

consistently appraised and ensure novelty of a proposed solution at various stages of the research life cycle. Thus there is an urgent need for a tools that can aid researchers to 1) understand, evaluate and incorporate the latest developments in the literature and 2) Formulate/Modify the current proposed solution accordingly to ensure novelty and impact.

Existing tools facilitate searching of research papers relevant to one’s topic of interest based on a query (e.g. Elicit<sup>1</sup>, Raxter.io<sup>2</sup>, SciSpace<sup>3</sup>), keywords (e.g. Litmaps<sup>4</sup>, Iris.ai<sup>5</sup>), paper titles (e.g. inciteful<sup>6</sup>), through citation graphs (e.g. Litmaps<sup>4</sup>, inciteful<sup>6</sup>, Research Rabbit<sup>7</sup>, Connected Papers<sup>8</sup>) or facts and insights (e.g. Scite<sup>9</sup>) as inputs. They further help

<sup>1</sup><https://elicit.com>

<sup>2</sup><https://raxter.io/>

<sup>3</sup><https://typeset.io>

<sup>4</sup><https://www.litmaps.com>

<sup>5</sup><https://iris.ai>

<sup>6</sup><https://inciteful.xyz>

<sup>7</sup><https://researchrabbitapp.com>

<sup>8</sup><https://www.connectedpapers.com>

<sup>9</sup><https://scite.ai/>

in (i) extracting relevant information from the papers with predefined templates (e.g. Elicit<sup>1</sup>, Iris.ai<sup>5</sup>) or based on user queries (e.g. SciSpace<sup>3</sup>), (ii) sharing the papers in collaborative fashion (e.g. Litmaps<sup>4</sup>, Raxter.io<sup>2</sup>, Research Rabbit<sup>7</sup>), (iii) notifying researchers about articles relevant to their searches, tracking trends (e.g. Scite<sup>9</sup>, Research Rabbit<sup>7</sup>), etc. Some tools act as the facilitator for reading or writing (e.g. Paperpal<sup>10</sup>) manuscripts by allowing highlighting, extracting and documenting important aspects of the paper (e.g. Raxter.io<sup>2</sup>), communicating in more interactive fashion to analyze the paper in depth (e.g. SciSpace<sup>3</sup>, Iris.ai<sup>5</sup>), recommending relevant quality citations (e.g. Scite<sup>9</sup>, Raxter.io<sup>2</sup>) or summarizing papers (e.g. Elicit<sup>1</sup>, Iris.ai<sup>5</sup>). Thus, most of the existing tools focus on notifying and recommending researchers with relevant literature, facilitate exploration of existing literature and/or writing research manuscripts. Researchers have also proposed learning representations for retrieval of relevant scientific articles (Singh et al. 2022; Cohan et al. 2020; Ostendorff et al. 2022; Mysore, Cohan, and Hope 2021), literature Review Generation (Hu and Wan 2014; Kasanishi et al. 2023; Chen et al. 2021), Question Answering over scientific articles (Saikh et al. 2022; Dasigi et al. 2021; Lee et al. 2023), Scientific document summarization (Hayashi et al. 2020), citation recommendation (Ali et al. 2021, 2022; Medic and Snajder 2023) citation intent detection (Cohan et al. 2019; Berrebbi, Huynh, and Balalau 2022; Roman et al. 2021; Lauscher et al. 2021), critical review and rebuttal generation (Ruggeri, Mesgar, and Gurevych 2022; D’Arcy et al. 2023; Kennard et al. 2021; Dycke, Kuznetsov, and Gurevych 2022; Wu et al. 2022), etc. However, to the best of our knowledge, no tool or no approach in the literature facilitates a researcher during the most arduous ideation stage of the research life-cycle. Ideation involves: (i) Analyzing the existing literature to critically evaluate the motivation behind the research problem a researcher is trying to address to ensure that the mentioned research gap(s) still exist(s), (ii) Reformulating the proposed research problem and objectives based on the validation stage output and re-identification of research gaps, (iii) Identifying analogous research problems or sub-problems addressed in the literature and utilizing their solutions, available in the literature, to derive a set-of approaches or synthesizing a set-of plausible methods as a solution to the problem, (iv) Designing experimentation strategy for the given problem and selected methodology.

Most of the tasks involved in research require domain expertise and complex reasoning skills. The recent advancement in Large Language Models (LLMs) and Generative Artificial Intelligence (GenAI) has made it possible to partially automate some of these tasks (Liu and Shah 2023; Liang et al. 2023; Zhang et al. 2023b; Lahiri, Sanyal, and Mukherjee 2023; Kunnath, Pride, and Knoth 2023). In this work, we propose ‘Acceleron’ (Figure 1), a tool to accelerate the research life cycle. By exploiting the reasoning and domain specific skills of LLM based agents, the goal of the tool is to assist with research activities, alleviating the burden of researchers. The aim is not to replace a researcher, but to assist

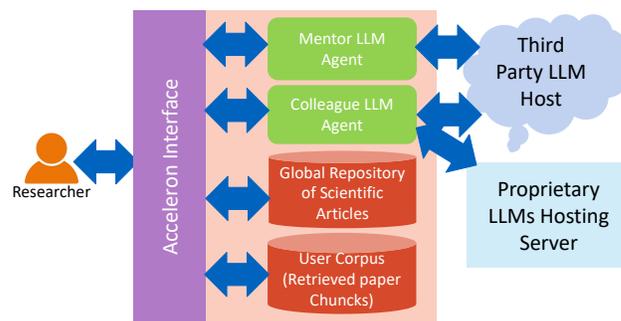


Figure 2: System Architecture

the researcher by providing relevant inputs in an interactive fashion, at various stages of research life cycle, viz. ideate, explore, codify, experiment, communicate, critique, inform; to expedite the process of meeting the research objectives. In this paper we are focusing on the ‘ideate’ module of the tool, which facilitates the ideation of a research problem, specified by the researcher. We expect the researcher to provide a short paragraph along with the title of the research problem the researcher plan to address, along with the description of what motivates the researcher to solve the problem. With LLM powered *mentor* and *colleague* agents, the tool, in an interactive fashion, helps the researcher to develop the research proposal consisting of a validated motivation, a well-defined research problem focusing of research gaps in the literature, a proposed approach selected from a set-of plausible synthesized methods and possible set-of experiments to be conducted to evaluate the approach for the research problem. To the best of our knowledge, we are the first ones to mimic the ideation process, followed by researchers, using the LLM agents. We demonstrate the ideation power of ‘Acceleron’ by illustrating examples from natural language processing.

## System Architecture

Acceleron provides a web-based interface for researchers to interact. The system architecture is illustrated in Figure 2. We define an LLM Agent based architecture (Wang et al. 2023b), with agents of two distinct types of profiles or personas<sup>11</sup>. A *Colleague* persona<sup>12</sup> performs less complex tasks including extraction of relevant information from user inputs, generation of relevant questions from extracted information or retrieval of relevant data from scientific documents. Whereas, *mentor* persona<sup>13</sup> performs more complex tasks requiring reasoning such as understanding the limitations or gaps of the existing work, identifying problems similar to the problem discussed in the proposal, identifying sub-tasks of the problem being solved in the proposal, solving similar problems and/or sub-tasks to synthesize a solution to the proposed problem and re-write the proposal

<sup>11</sup>We use Langchain framework <https://www.langchain.com> for implementation

<sup>12</sup>OpenAPI’s GPT-turbo-3.5 model

<sup>13</sup>OpenAPI’s GPT4 model

<sup>10</sup><https://paperpal.com>

given a plausible set-of approaches or possible limitations of related work. The architecture is flexible such that the LLM agents can interact with (i) LLMs using API calls<sup>14</sup> or (ii) open-source LLMs which reside on an internal hosting server.

We expect to have a global repository which is a vector store<sup>15</sup> of domain specific scientific articles<sup>16</sup> which are indexed by the Specter embeddings (Cohan et al. 2020) produced using the paper’s title and abstract. We also have a User Specific corpus which has chunks of all the retrieved papers relevant to the current proposal the researcher is working on. The paper chunks are created with our in-house parser<sup>17</sup> treating paragraphs as semantic segments. If a paragraph does not fit into the the maximum token length of LLM agents, while chunking it is further split to fit into the maximum token length. The chunks are further converted to vector embeddings and indexed<sup>18</sup> for efficient retrieval based on semantic similarity with a query. This user corpus acts as a shared ‘memory’ for the LLM agents.

### Approach

In this section we describe the ideation process adopted by Acceleron which simulates the ideation process followed by researchers using LLM agents. The process involves interaction between a researcher and the LLM agents, where the LLM agents perform actions based on the feedback received by the researcher or another agent. The process takes a proposal as an input from a researcher with a research problem description specified at a high level along with the motivation behind the problem. The output of the ideation process is the updated proposal with a (i) Validated motivation or updated research problem by identifying gaps mentioned in the literature addressing the motivation (ii) Plausible methods to address the research problem. The overall ideation task is split into two workflows: (i) Motivation Validation and (ii) Method Synthesis. The detailed prompts for the steps in each of the workflow are illustrated in the Appendix Section .

### Motivation Validation Workflow

This is the first phase of the ideation process. The workflow is elaborated in Figure 3. The steps of the workflow are explained in detail here: 1. The researcher provides the title and abstract of the proposal elaborating the motivation behind the proposal and a high level description of the problem statement the research wishes to solve. 2. The retriever functionality of Acceleron uses this title and abstract of the proposal as a query and gets a vector representation<sup>19</sup> of the

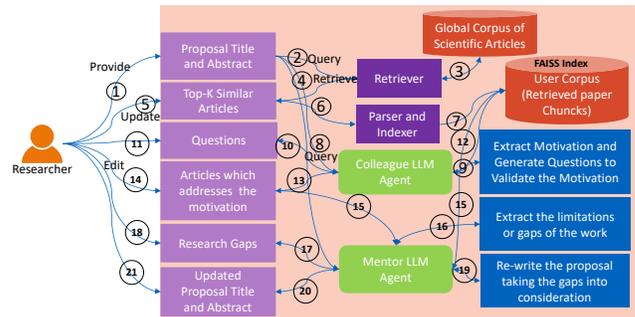


Figure 3: Motivation Validation Workflow

same 3. This representation is used to retrieve top-K articles similar to the contents of the proposal from the global corpus of scientific articles. 4. The top-K articles along with a description of the relevance of each article to the proposal is shown to the researcher. 5. These articles are editable by the researcher who can delete articles they find irrelevant or by manually adding relevant articles. 6. The parser and indexer functionality chunks the finalized set-of papers in paragraphs (semantic segments) 7. The chunks are stored in the user corpus with appropriate indexing mechanisms 8. The *colleague* agent fetches the proposal title and abstract provided by the researcher 9. It extracts the motivation out of the proposal (Prompt 1) and generates a list of questions to be posed on the shortlisted scientific articles to validate the motivation of the proposal (Prompt 2). The questions are binary and formulated such that if, for a scientific article, the answer to the question is ‘yes’, then the article is already addressing the motivation of the proposal mentioned by that question. For example, if the researcher proposes to develop a technique to solve a novel aspect of a problem, a question would be generated of the form ‘Does the research paper address that specific aspect of the problem?’. If this question is answered as ‘yes’ by a scientific article then it implies that the article addresses that aspect of the problem and hence the motivation behind the study is weak or invalidated. 10. The generated set-of questions are shown to the researcher 11. The researcher can edit the questions by updating the format of the questions, deleting questions that the researcher deems irrelevant or adding missing relevant questions 12. For each question and a retrieved paper stored in the user corpus, the *colleague* agent retrieves the chunks of the paper relevant to that question and tries to answer that question using Retrieval Augmented Generation (RAG) (Prompt 3). The answer can be ‘yes’, ‘no’ or unanswerable along with an explanation. 13. If all the papers answer either ‘no’ or ‘unanswerable’ for all the questions generated to validate the motivation; it indicates that the existing literature is not addressing the motivation behind the proposal and hence this phase is ended with a comment shown to the researcher that the motivation of the proposal is validated. Otherwise, the question-research paper pairs with only ‘yes’ as an answer, along with explanation are shown to the researcher. 14. The researcher can edit this output in terms of removing papers which he doesn’t agree to address the question based on the

<sup>14</sup>LLMs designed by OpenAI <https://platform.openai.com/docs/api-reference> or Cohere <https://cohere.com>

<sup>15</sup>We use pinecone <https://www.pinecone.io/>

<sup>16</sup>We use more than 2 million scientific articles in semantic scholar fetched using S2ORC dataset (Lo et al. 2020) as the global repository

<sup>17</sup>We have built the parser using pdfminer <https://pypi.org/project/pdfminer/>

<sup>18</sup>We use FAISS indexing <https://github.com/facebookresearch/faiss>

<sup>19</sup>Specter (Cohan et al. 2020)

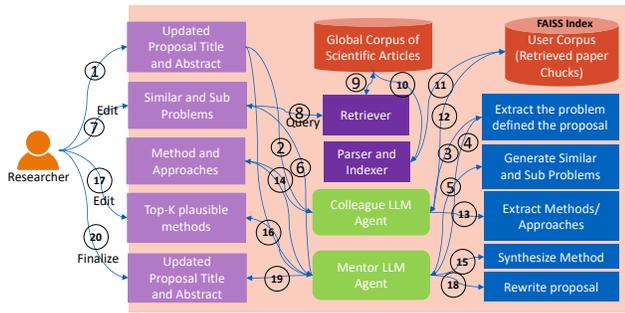


Figure 4: Method Synthesis Workflow

explanation provided. If the answers are hallucinated, this step enables the researcher to remove such papers. 15. The *mentor* agent uses the chunks of each of the shortlisted paper, the original proposal and the description of the prior question addressing the motivation 16. It extracts the limitations or gaps of each of the papers (Prompt 4), which has been identifying to be addressing the motivation of the proposal, such that the gaps can help redefining the problem in the proposal. 17. The research gaps are shown to the researcher 18. The researcher can ignore the gaps found to be irrelevant and selects some of the gaps that address part of the research problem. If the researcher does not agree with any of the specified research gaps, they can add their own set of research gaps 19. The *mentor* agent uses these gaps along with the original proposal to re-formulate the motivation (Prompt 5) and the problem statement of the proposal to address the new research gaps. 20. The modified proposal is presented to the researcher (Prompt 6). 21. The modified proposal can be edited by the researcher to finalize the same or he can reject the edits and go back to the prior proposal. This workflow can be applied to the proposal in an iterative manner. Meaning, the motivation behind the resultant updated proposal can be again validated by initiating the same workflow. This can be executed in an iterative fashion, until there are no scientific articles retrieved to be addressing the motivation behind the proposal, validating the novelty of the proposal.

## Method Synthesis Workflow

The steps of the method synthesis workflow are illustrated in Figure 4 and are elaborated here: 1. The method synthesis phase begins with the proposal whose motivation is validated based on reviewing the literature and accepted by the researcher 2. The *colleague* agent takes the proposal as an input 3. It extracts the problem defined in the proposal (Prompt 7) 4. The *mentor* agent takes this problem as an input and 5. uses its parametric knowledge to generate a plausible set-of similar research problems (Prompt 8). For example, if the problem defined in the proposal is ‘To design a reference-free evaluation metric for question answering task’. A similar problem can be ‘To come up with an evaluation metric for text summarization which can have multiple possible reference summaries’. The *mentor* agent also uses its parametric knowledge to decompose the prob-

lem defined in the proposal into sub-tasks (if any) (Prompt 11). For example, the problem of ‘question answering over scientific papers’ can be decomposed into ‘extraction of text from PDF document of the scientific paper’, ‘segmenting the paper and storing it for efficient retrieval’, ‘retrieval of paragraphs from the paper related to the question’, ‘answering the question using the retrieved paragraphs as context’, ‘evaluating the retrieved paragraphs’ and ‘evaluating the answers’. 6. The generated similar and sub problems are shown to the researcher 7. The researcher can edit them by removing the ones found to be irrelevant, updating them or adding missing ones. This helps in removing the problems hallucinated by the agent 8. Each of the edited similar problem or sub-problem is used as a query to retrieve scientific articles from the global corpus which address those respective problems. 10. The retrieved scientific articles are parsed and chunked and 11. are stored in the user corpus. 12. For each retrieved paper, the *colleague* agent extracts the paragraphs which talks about the method or approach taken by the paper to solve the respective problem and 13. generates a consolidate list of similar or sub problem and solution pairs. (Prompt 9, 10, 12) 14. This list is showcased to the researcher who makes edits to the same 15. The extracted problem of the proposal along with the consolidated list of similar or sub problems along with their solutions in the literature is provided to the *mentor* agent. (Prompt 14, 15) 15. The *mentor* agent uses this information along with its parametric knowledge to synthesize a list of plausible methods to solve the problem defined in the proposal. (Prompt 16) 16. The list of plausible methods is shown to the researcher. (Prompt 13) 17. Researcher can choose a subset of these methods which they deem most plausible and further edit these if need be. 18. The updated list is provided to the *mentor* agent along with the original proposal. The *mentor* agent re-writes the proposal including these methods. (Prompt 17) 19. This updated proposal is shown to the researcher 20. The researcher can further edit the proposal and finalize the same.

## Novel Components

**LLM agents for research ideation:** To the best of our knowledge, ours is the first LLM agent based tool which assists in the complex task of ideation for research. We have devised with two novel portfolios for LLMS, viz., *colleague* and *mentor*, allocating less complex tasks to the *colleague* agent and more complex reasoning based tasks to the *mentor* agent. The user corpus acts as the shared memory for the agents, whereas the agents perform fixed set of actions at various stages of the workflow based on the provided inputs as discussed in the prior sections. Rather than using a costly LLM like GPT4 for all the tasks involved in the workflows; dividing the tasks as per the difficulty level and leveraging less costly LLM such as GPT-turbo-3.5 for colleague agent, performing less complex tasks, provides a cost-effective solution for workflows.

**Mitigation of hallucination:** Hallucination is one of the major difficulties of using LLMs for knowledge based tasks (Zhang et al. 2023a; Wang et al. 2023a). We mitigate this problem using a two-fold solution: (i) There are retrieval augmented components of the workflows, viz. the motiva-

tion validation workflow poses questions generated to validate the motivation of the proposal on the retrieved articles stored in the user corpus or extract limitations of the articles which address the proposal motivation or the method synthesis workflow extracts approaches used to solve similar or sub problems from the retrieved articles. For these retrieval augmented tasks through proper prompt engineering, we ensure that the answers are provided by restricting the knowledge to the retrieved context only. We observe this helps to mitigate hallucinations. (ii) There are components of the workflows which rely on parametric knowledge of LLMs, for example the motivation validation involves re-writing the proposal and the method synthesis involves generating similar sub problems for the problem defined in the proposal and synthesizing methods. For these tasks the output can not be restricted to the provided input. In such cases, there is a higher chance of hallucinated outputs. For such scenarios, we ensure mitigation of hallucinated outputs, by keeping the system semi-automated and allowing user-interactions at every step to edit or delete hallucinated outputs. Moreover at every stage of the workflow, the LLM agents are asked to justify their outputs and the provided justification is exposed to the researcher through the interface. This forces the model to apply Chain-of-Thoughts (COT) (Wei et al. 2022) and allows the researcher to validate the output and check if it is in sync with the justification provided. This assists in alleviating the effect of hallucinations.

**Two-stage aspect based retrieval:** The global corpus contains a large number of scientific articles stored with the Specter embedding of the title and abstract of the papers. The title and abstract of the papers contains information about motivation and problem statement of the papers and a high level mention of the methodology and the results. For ideation we require more in-depth information from the papers across various aspects such as methodology, limitations, etc. To achieve this we perform retrieval in two stages. In motivation validation workflow, we first retrieve top-K papers from the global corpus with the proposal as the query and high value of K for good recall. This allows us to have a set-of papers with similar motivation and problem statement to that of the proposal. These papers are chunked and stored in the user corpus for further aspect based retrieval, such as papers with similar motivation to that of the proposal and paper paragraphs mentioning the research gaps of these papers. In method synthesis workflow, we first retrieve top-K papers from the global corpus with similar sub problem statements as the query and high value of K for good recall. This allows us to have a set-of papers with problems similar to the problem described in the proposal or similar to any of the sub-tasks of the problem described in the proposal. These papers are chunked and stored in the user corpus for further aspect based retrieval such as extracting the approaches of the papers. Note that keeping high-recall for the first stage of retrieval ensures coverage of papers, whereas for the second stage we favor more precise outcomes for aspect based retrieval.

**Introduction of Unanswerability:** The output of aspect based retrieval is always top-K paragraphs from the retrieved and chunked papers. We keep the value of K low to get more

precise retrieval for the given aspect based query. However, there is a possibility that the retrieved paragraphs do not have the answer to the query (the query is unanswerable). For example, in the motivation validation workflow the retrieved paragraphs from the papers do not answer the question of whether the paper addresses a specific motivation of the proposal and does not specify the limitations of the paper which would help to refine the problem defined in the proposal. Similarly, for the method synthesis workflow the retrieved paragraphs may not have an approach to solve a similar problem. In such cases, the LLM based agents check the relevancy of retrieved paragraphs for the given query and identifies the query as ‘unanswerable’ in case if all the retrieved paragraphs are irrelevant, avoiding irrelevant outputs. Allowing unanswerability also assists in reduction of hallucinations.

### Qualitative Analysis of the Workflows

In the absence of an appropriate dataset for the tasks relevant to the ideation process, we provide a qualitative analysis of the workflows with 3 proposals from distinct researchers, specifically in the domain of Artificial Intelligence (AI), Machine Learning (ML) and Natural Language Processing (NLP). The topics of these proposals are: (i) Datasets for Computational Study of Peer Reviews (ii) Topic-based citation retrieval for research proposal and (iii) Reference-free evaluation metric for retrieval augmented question answering. These researchers are from our lab, working on distinct research problems, for which they intend to write proposals. We use Semantic Scholar data fetched using S2ORC dataset (Lo et al. 2020) as our global repository, which has a variety of papers in the AI, ML and NLP domain. We utilize the logging functionality of ‘Acceleron’ to keep track of the interactions between the researcher and the LLM Agents and derive the following observations.

The abstract of the proposal with topic ‘*Dataset for Computational Study of Peer Reviews*’ is: ‘*Peer review constitutes a core component of scholarly publishing and demands substantial expertise and training, and is susceptible to errors and biases. Various applications of NLP for peer reviewing assistance aim to support reviewers in this complex process, but the lack of clearly licensed datasets and multi-domain corpora prevent the systematic study of NLP for peer review. To remedy this, we plan to introduce an ethically sourced multi-domain corpus of papers and review reports from five different venues.*’. With the proposal topic and abstract as an input, the motivation validation workflow generates the question: ‘Is the research paper addressing the lack of clearly licensed datasets for studying natural language processing for peer review?’. As a part of this workflow, to validate the motivation of the proposal, the Colleague LLM agent poses this question on the chunks of top-50 scientific articles, retrieved to be similar to the proposal topic and abstract and stored in the user corpus. Out of these 50 scientific articles, 5 articles are retrieved with ‘yes’ as a answer, out of which the researcher agrees to the justification provided for four and disagrees with the justification of one. Following are the 4 research articles with a valid justification: (i) NLPeer: A Unified Resource for the Computational Study

of Peer Review (ii) A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications (iii) Investigating Fairness Disparities in Peer Review: A Language Model Enhanced Approach (iv) MOPRD: A multidisciplinary open peer review dataset. Whereas, the article with title ‘What Can We Do to Improve Peer Review in NLP?’ agrees with the motivation of the proposal, however, does not address the same. As the next part of the motivation validation workflow, the colleague LLM agent extracts the following research gaps from the scientific articles: (i) NLPeer: A Unified Resource for the Computational Study of Peer Review: (a) The paper does not contain blind reviewing data which is a standard practice in most research fields, (b) The paper does not perform extensive hyper parameter search and tuning of the models, which could limit the effectiveness of the models, (c) The paper acknowledges the risk of “lazy reading” where reviewers might only read the paragraphs suggested by the model, but does not provide a solution to prevent this. (ii) A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications: (a) The models used in the research are relatively simple, which may limit their effectiveness in complex peer review scenarios, (b) The paper leave room for further research in areas such as demographic biases in accept/reject decisions, (c) The research does not provide a multidomain corpus of papers and review reports from different venues, limiting the scope of its application. (iii) Investigating Fairness Disparities in Peer Review, A Language Model Enhanced Approach: (a) The research paper does not draw any causal conclusions from the fairness analysis, limiting the depth of understanding about the underlying mechanisms of bias in peer review, (b) The research paper does not investigate the effect of rebuttals in the peer review process, which could be a significant factor in the final decision on a paper, (c) The research paper does not provide a thorough analysis of the continual learning problem in Pre-trained Language Models (PLMs), which could be crucial for improving automatic review generation. (iv) MOPRD: A multidisciplinary open peer review dataset: (a) The research paper acknowledges that the multidisciplinary bias of their method could be a concern, indicating a need for more balanced datasets across various disciplines, (b) The paper points out that the input limits still exist as the maximum length of the input text only reaches 16,384 tokens, indicating a need for models that can handle larger input sizes, (c) The paper mentions that the performance of attention mechanisms will be compromised as the input length increases, indicating a need for more efficient attention mechanisms or alternative models for handling long input sequences. This showcases the quality of outputs provided by the workflow in terms of identifying the research gaps in the existing articles, partially eliminating the need of detailed literature survey of these articles to find gaps and thus alleviating research efforts.

We further observe that the researcher selects a subset of these research gaps, she finds to be relevant. The mentor LLM agent further uses these selected gaps to re-write the proposal. Following is the updated proposal provided as an outcome of the workflow taking into account the selected research gaps. *The current literature presents several*

*gaps that motivate the need for a more comprehensive approach to the application of Natural Language Processing (NLP) in peer review. The absence of blind reviewing data, a standard practice in most research fields is a significant limitation. The lack of a multidomain corpus of papers and review reports from different venues restricts the scope of its application. The difficulty of studying the review process compared to the decision process, the unexplored effect of rebuttals in the peer review process indicate a need for more sophisticated methods or tools. The multidisciplinary bias of the method, the input limits of Language Models, and the compromised performance of attention mechanisms as the input length increases, all suggest a need for improved NLP models for peer review, more balanced datasets across various disciplines, and models that can handle larger input sizes and long input sequences. To address these gaps, we propose to introduce an ethically sourced multidomain corpus of papers and review reports from five different venues. The mentor LLM agent takes into consideration the following selected research gaps and introduces them in the revised version of the proposal: (i) availability of multi-disciplinary, multi-venue, blind-review corpus, (ii) no studies of effect of rebuttals (iii) context limits of LMs to tackle long reviews of existing work in this domain. We further observe that the researcher makes a very few edits to the the revised proposal abstract and accepts the same. The total time required for validating the motivation and updating the proposal abstract accordingly is substantially (~5x for this proposal as mentioned by the researcher) reduced by this workflow.*

We receive an input from another researcher with a proposal titled ‘*Topic-based citation retrieval for research proposal*’ and the corresponding abstract ‘*Retrieval of research articles pertinent to a given query represents a thoroughly investigated research challenge. Typically, queries take the form of a title and abstract of a research article, or a specific sentence or paragraph from an existing research article requiring citation. However, existing approaches presuppose the availability of a well-constructed manuscript, an assumption that is inappropriate during the initial research proposal writing stage. At this initial phase, researchers seek pertinent literature for citing in their proposals, often focusing on specific topics or intents and further build the proposal. In this work, we aim to tackle the issue of topic-based citation retrieval for research proposals. We anticipate researchers providing the title and abstract of their research proposals, encompassing elements such as the research gap, problem statement, and a high-level overview of the proposed methodology and experiments. Additionally, researchers will furnish a list of topics for which relevant scientific articles need to be retrieved. Our proposed algorithm intends not only to fetch research articles pertinent to the given proposal from a corpus, but also to establish a crucial many-to-many mapping between these articles and the specified topics.*’ The colleague LLM agent generates the following questions for validation of the motivation: 1. Is the research paper specifically addressing the retrieval of research articles relevant to a topic of a research proposal? and 2. Is the research paper developing a technique to map research articles to specified topics in research proposals?. Out

of top-50 research articles used to validate the motivation of the proposal by posing the above mentioned questions, the following four got retrieved to be answering as ‘yes’ to at the least one of the above questions and thus invalidating the motivation behind the proposal: 1 Citation Recommendation: Approaches and Datasets 2. CitationIE: Leveraging the Citation Graph for Scientific Information Extraction 3. Content-Based Citation Recommendation and 4. unarXiv 2022: All arXiv Publications Pre-Processed for NLP, Including Structured Full-Text and Citation Network. However, the justifications provided for these papers highlight that paper no 1. and 3. introduce an approach for citation recommendations during the writing phase of the target manuscripts and not at the proposal writing stage. Also, scientific article 2. leverages contents of a target paper and citation graph to extract scientific information. The outcome of the scientific article 4. is a dataset which can be useful for the proposal, but does not address the task of ‘topic-based citation retrieval for research proposal’. Thus, we observe that after evaluating the retrieved scientific articles claimed to be invalidating the proposal, the researcher disagrees with the justifications provided for each of the retrieved articles for addressing the motivation behind the proposal, hence validating the novelty of the proposal. This exemplifies the need as well as the effectiveness of the user interaction facility provided by the tool for the workflow. This example demonstrates acceleration of motivation validation stage of the research-life cycle (~8x for this proposal as mentioned by the researcher), by eliminating the need for the researcher to manually go through multiple relevant research articles retrieved by generic or academic search engines to ensure that the literature does not have a solution for the specific problem the researcher is trying to address, leading to a time consuming process.

We receive input from another researcher with the proposal titled ‘*Reference-Free evaluation metric for Retrieval augmented question answering task*’ and the abstract ‘*We observe that questions with long answers on long documents do not have unique reference evidences (relevant paragraphs from the document) and answers. Rather, there is a distribution over reference answers, making expert based evaluation expensive and existing unique reference-based evaluation metrics inadequate. We also do not find any reference-free evaluation metric designed for evaluating retrieval augmented question answering task. Hence, this this work we propose to define this metric.*’. The colleague LLM agent generates the following question to validate the motivation of the proposal: Is the research paper proposes a reference-free evaluation metric designed for evaluating retrieval augmented question answering tasks?. We observe that out of top-50 retrieved scientific articles relevant to the proposal none of the articles provides answer as ‘yes’ to the question, leading to retrieval of no paper which invalidates the motivation of the proposal. Manual analysis of the top-50 retrieved articles performed by the researcher (as well as other relevant articles manually visited by the researcher) to evaluate this outcome of the workflow, substantiates the results.

For the next workflow of method synthesis for the above proposal, the mentor LLM agent generates following set of

research problems similar to the problem defined in the proposal: 1 Evaluating complex tasks where there is no unique correct answer or reference. 2. Designing evaluation metrics for tasks that involve retrieval and interpretation of large amounts of data. 3. Creating reference-free evaluation metrics for tasks where reference-based metrics are inadequate or impractical. 4. Assessing the quality of answers in tasks where the answers can be long and drawn from extensive documents. The mentor LLM agent also generates the following sub-tasks for the problem defined in the proposal: 1. Defining a new metric that can effectively evaluate retrieval augmented question answering tasks. and 2. Overcoming the inadequacy of existing unique reference-based evaluation metrics for questions with long answers on long documents. With these similar and sub-problems as queries, the colleague LLM agent first retrieves Top-10 similar scientific articles per statement (total 40 articles with some overlap as same article may be retrieved for multiple queries) and then poses the question that ‘if the article provides a methodology or an approach to solve the above defined problem or sub-task’. The researcher receives total 17 scientific articles which answers ‘yes’ to the question along with a description of the methodology implemented for each of the above problems or task. Out of these papers, the researcher accepts the following 11 scientific articles, finding them to be more relevant to the problem the researcher is trying to address: 1. AVA: an Automatic eValuation Approach to Question Answering Systems. 2. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 3. Re-visiting Automated Topic Model Evaluation with Large Language Models. 4. SacreROUGE: An Open-Source Library for Using and Developing Summarization Evaluation Metrics. 5. Quantified Reproducibility Assessment of NLP Results. 6. Revisiting the Gold Standard: Grounding Summarization Evaluation with Robust Human Evaluation. 7. A Critical Evaluation of Evaluations for Long-form Question Answering. 8. Think you have Solved Direct-Answer Question Answering? Try ARC-DA, the Direct-Answer AI2 Reasoning Challenge. 9. More Than Reading Comprehension: A Survey on Datasets and Metrics of Textual Question Answering. 10. Hurdles to Progress in Long-form Question Answering. 11. A Critical Evaluation of Evaluations for Long-form Question Answering.

Based on approaches discussed in these scientific articles to solve the above enlisted similar problems and/or sub-tasks, the mentor LLM agent suggests the following 10 plausible methodologies for the proposal. These suggestions with the exact verbatim provided by the agent are as follows: 1. AVA (Automatic Evaluation Approach): This approach uses Transformer-based language models to encode the question, answer, and reference text. By measuring the similarity between the reference and an automatic answer, biased towards the question semantics, AVA can estimate the system accuracy. This approach can be used to evaluate retrieval augmented question answering tasks and overcome the inadequacy of existing unique reference-based evaluation metrics. 2. Informedness and Markedness Measures: These new probabilistic and information theoretic measures can be used to evaluate complex tasks where

there is no unique correct answer or reference. They can also be used to design evaluation metrics for tasks that involve retrieval and interpretation of large amounts of data.

3. Large Language Models (LLMs): LLMs can be used as automated metrics for evaluating tasks that involve retrieval and interpretation of large amounts of data. They can also be used to assess the quality of answers in tasks where the answers can be long and drawn from extensive documents.
4. SacreROUGE: This open-source library dedicated to the development of summarization evaluation metrics can be used to develop reference-free evaluation metrics for tasks where reference-based metrics are inadequate or impractical.
5. Quantified Reproducibility Assessment (QRA): This method, based on concepts and definitions from metrology, can be used to develop reference-free evaluation metrics for tasks where reference-based metrics are inadequate or impractical.
6. RoSE Benchmark: This benchmark can be used to provide a more robust evaluation across different tasks. It can also be used to develop reference-free evaluation metrics for tasks where reference-based metrics are inadequate or impractical.
7. GENIE6: This human-in-the-loop leaderboard framework can be used to score answers in tasks where the answers can be long and drawn from extensive documents.
8. Sparse Attention and Contrastive Retriever Learning: This system conditions answer generation on identified documents and uses a dense retriever trained by scaling up a distantly supervised algorithm. It can be used to overcome the inadequacy of existing unique reference-based evaluation metrics for questions with long answers on long documents.
9. Unified Evaluation Benchmark for Long-form Answers: This approach involves conducting a thorough study of evaluation, including both human and automatic evaluation protocols. It can be used to overcome the inadequacy of existing unique reference-based evaluation metrics for questions with long answers on long documents.
10. Training an LFQA Evaluation Metric Directly on Human-Annotated Preference Judgments: This approach involves fine-tuning pre-trained Language Models based on human judgement scores for the task. This output showcases the quality of method recommendations provided by the tool for the given proposal. Though mentioned at high-level, the researcher agrees that most of these methods are well-suited as a plausible approach for the given proposal. Though there is a need for further work to finalize the most appropriate plausible method for proposal, the researcher finds this first cut of output provided by the tool to be relevant and the overall process to be  $\sim 10$  times more efficient than the regular process followed by the researcher for constructing a plausible set-of approaches for the given problem, by searching through the relevant literature from scratch.

These examples illustrating the outcomes of the motivation validation and method synthesis phases of the ideation workflow of the tool, demonstrates the efficacy of the tool, in terms of providing relevant outputs at each stage of the workflow. The observations made in terms of time saved by the researchers with the tool usage for the respective tasks demonstrates the power of the tool with regards to time efficiency gains.

## Conclusion

In this work, we have demonstrated a tool called ‘Acceleron’, developed to accelerate the ideation phase of the research life-cycle. To the best of our knowledge this is the first tool which addresses the tasks involved in the ideation stage. To emulate the ideation process, we use LLM agents with colleague and mentor personas to execute two workflows, viz. motivation validation and method synthesis, which engage researchers in an interactive fashion to develop the research proposal. Our workflow involves novel components to (i) alleviate the hallucinations of LLMs, (ii) ensure relevant outcomes by two-stage aspect based retrieval, where first stage introduces higher recall reducing False Negatives and False Positives are corrected by user interaction, second stage of more precise fine-grained aspect-based retrieval and introduction of unanswerability. The qualitative analysis performed with three proposals from distinct researchers, in the domain of Machine Learning and Natural Language Processing, demonstrates precise outcomes for various stages in the workflow with  $\sim 7.5x$  gains in the time efficiency for various stages of the ideation phase.

## Future Works

This is an ongoing work. We plan to augment the ideation functionality of the tool for other scientific domains such as life-sciences or material sciences, etc. We plan to emulate the domain specific aspects of the ideation process for these domains, which may differ from the current process specifically designed for Machine Learning and Artificial Intelligence based research projects in Computer Science. This would allow us to define a meta-process for ideation, which is domain independent and domain specific instances of this meta-process. Currently ‘Acceleron’ supports OpenAI models. We plan to augment the tool with Open-Source LMs such as Llama-2(Touvron et al. 2023), Zephyr(Tunstall et al. 2023), etc. The logging functionality of ‘Acceleron’ keeps track of every input provided to the researcher as well as LLM agents and every output from them along with the corresponding timestamps. We are saving these logs for each user interactions for all the sessions. We plan to use these logs with treating user validated inputs as ground truth annotations, to develop a datasets for the ideation process. The logs would be used for developing datasets for tasks such as: (i) retrieval of research papers with similar motivation (ii) proposal re-writing with addressing research-gaps (iii) retrieval of research papers with similar problems and/or (iv) method-synthesis from a set-of relevant papers. The datasets will be used to instruction-tune the above mentioned Open-Source LMs, which can replace the existing LLMs yielding more cost-effective solution. We plan to extend the implementation of current ideation phase to generate a list of experiments to be performed for the problem defined in the proposal and the methodology selected by the researcher. This would lead to generation of a (set-of) results table(s) in a semi-automated fashion, with baseline approaches, planned experiments (ablations) and appropriate metric(s) used for evaluation.

## References

- Ali, Z.; Qi, G.; Kefalas, P.; Khusro, S.; Khan, I.; and Muhammad, K. 2022. SPR-SMN: scientific paper recommendation employing SPECTER with memory network. *Scientometrics*, 127: 6763–6785.
- Ali, Z.; Qi, G.; Muhammad, K.; Kefalas, P.; and Khusro, S. 2021. Global citation recommendation employing generative adversarial network. *Expert Syst. Appl.*, 180: 114888.
- Berrebbi, D.; Huynh, N.; and Balalau, O. 2022. GraphCite: Citation Intent Classification in Scientific Publications via Graph Embeddings. *Companion Proceedings of the Web Conference 2022*.
- Chen, X.; Alamro, H.; Mingzhe, L.; Gao, S.; Zhang, X.; Zhao, D.; and Yan, R. 2021. Capturing Relations between Scientific Papers: An Abstractive Model for Related Work Section Generation. In *Annual Meeting of the Association for Computational Linguistics*.
- Cohan, A.; Ammar, W.; van Zuylen, M.; and Cady, F. 2019. Structural Scaffolds for Citation Intent Classification in Scientific Publications. *ArXiv*, abs/1904.01608.
- Cohan, A.; Feldman, S.; Beltagy, I.; Downey, D.; and Weld, D. S. 2020. SPECTER: Document-level Representation Learning using Citation-informed Transformers. *ArXiv*, abs/2004.07180.
- D’Arcy, M.; Ross, A.; Bransom, E.; Kuehl, B.; Bragg, J.; Hope, T.; and Downey, D. 2023. ARIES: A Corpus of Scientific Paper Edits Made in Response to Peer Reviews. *arXiv preprint arXiv:2306.12587*.
- Dasigi, P.; Lo, K.; Beltagy, I.; Cohan, A.; Smith, N. A.; and Gardner, M. 2021. A Dataset of Information-Seeking Questions and Answers Anchored in Research Papers. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4599–4610.
- Dycke, N.; Kuznetsov, I.; and Gurevych, I. 2022. NLPeer: A Unified Resource for the Computational Study of Peer Review. *ArXiv*, abs/2211.06651.
- Hayashi, H.; Kryscinski, W.; McCann, B.; Rajani, N.; and Xiong, C. 2020. What’s New? Summarizing Contributions in Scientific Literature. *ArXiv*, abs/2011.03161.
- Hu, Y.; and Wan, X. 2014. Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach. In *Conference on Empirical Methods in Natural Language Processing*.
- Kasanishi, T.; Isonuma, M.; Mori, J.; and Sakata, I. 2023. SciReviewGen: A Large-scale Dataset for Automatic Literature Review Generation. In *Annual Meeting of the Association for Computational Linguistics*.
- Kennard, N. N.; O’Gorman, T. J.; Sharma, A.; Bagchi, C.; Clinton, M.; Yelugam, P. K.; Das, R.; Zamani, H.; and McCallum, A. 2021. DISAPERE: A Dataset for Discourse Structure in Peer Review Discussions. *ArXiv*, abs/2110.08520.
- Kunnath, S. N.; Pride, D.; and Knoth, P. 2023. Prompting Strategies for Citation Classification. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*.
- Lahiri, A.; Sanyal, D. K.; and Mukherjee, I. 2023. CitePrompt: Using Prompts to Identify Citation Intent in Scientific Papers. *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 51–55.
- Lauscher, A.; Ko, B. R.; Kuehl, B.; Johnson, S.; Jurgens, D.; Cohan, A.; and Lo, K. 2021. MultiCite: Modeling realistic citations requires moving beyond the single-sentence single-label setting. *ArXiv*, abs/2107.00414.
- Lee, Y.; Lee, Y.; Lee, K.; Park, S.; Hwang, D.; Kim, J.; Lee, H. H.; and Lee, M. 2023. QASA: Advanced Question Answering on Scientific Articles. In *International Conference on Machine Learning*.
- Liang, W.; Zhang, Y.; Cao, H.; Wang, B.; Ding, D.; Yang, X.; Vodrahalli, K.; He, S.; Smith, D.; Yin, Y.; McFarland, D. A.; and Zou, J. 2023. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *ArXiv*, abs/2310.01783.
- Liu, R.; and Shah, N. B. 2023. ReviewerGPT? An Exploratory Study on Using Large Language Models for Paper Reviewing. *ArXiv*, abs/2306.00622.
- Lo, K.; Wang, L. L.; Neumann, M.; Kinney, R.; and Weld, D. 2020. S2ORC: The Semantic Scholar Open Research Corpus. In Jurafsky, D.; Chai, J.; Schluter, N.; and Tetreault, J., eds., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4969–4983. Online: Association for Computational Linguistics.
- Medic, Z.; and Snajder, J. 2023. Paragraph-level Citation Recommendation based on Topic Sentences as Queries. *ArXiv*, abs/2305.12190.
- Mysore, S.; Cohan, A.; and Hope, T. 2021. Multi-Vector Models with Textual Guidance for Fine-Grained Scientific Document Similarity. *ArXiv*, abs/2111.08366.
- Ostendorff, M.; Rethmeier, N.; Augenstein, I.; Gipp, B.; and Rehm, G. 2022. Neighborhood Contrastive Learning for Scientific Document Representations with Citation Embeddings. In *Conference on Empirical Methods in Natural Language Processing*.
- Roman, M.; Shahid, A.; Khan, S.; Koubâa, A.; and Yu, L. 2021. Citation Intent Classification Using Word Embedding. *IEEE Access*, 9: 9982–9995.
- Ruggeri, F.; Mesgar, M.; and Gurevych, I. 2022. ArgSci-Chat: A Dataset for Argumentative Dialogues on Scientific Papers. *ArXiv*, abs/2202.06690.
- Saikh, T.; Ghosal, T.; Mittal, A.; Ekbal, A.; and Bhattacharyya, P. 2022. ScienceQA: a novel resource for question answering on scholarly articles. *International Journal on Digital Libraries*, 23: 289 – 301.
- Singh, A.; D’Arcy, M.; Cohan, A.; Downey, D.; and Feldman, S. 2022. SciRepEval: A Multi-Format Benchmark for Scientific Document Representations. *ArXiv*, abs/2211.13308.
- Touvron, H.; Martin, L.; Stone, K. R.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; Bikel, D. M.; Blecher, L.; Ferrer, C. C.; Chen, M.; Cucurull, G.; Esiobu, D.; Fernandes, J.; Fu, J.; Fu, W.; Fuller, B.; Gao, C.; Goswami, V.; Goyal, N.; Hartshorn,

A. S.; Hosseini, S.; Hou, R.; Inan, H.; Kardas, M.; Kerkez, V.; Khabsa, M.; Kloumann, I. M.; Korenev, A. V.; Koura, P. S.; Lachaux, M.-A.; Lavril, T.; Lee, J.; Liskovich, D.; Lu, Y.; Mao, Y.; Martinet, X.; Mihaylov, T.; Mishra, P.; Molybog, I.; Nie, Y.; Poulton, A.; Reizenstein, J.; Rungta, R.; Saladi, K.; Schelten, A.; Silva, R.; Smith, E. M.; Subramanian, R.; Tan, X.; Tang, B.; Taylor, R.; Williams, A.; Kuan, J. X.; Xu, P.; Yan, Z.; Zarov, I.; Zhang, Y.; Fan, A.; Kam-badur, M.; Narang, S.; Rodriguez, A.; Stojnic, R.; Edunov, S.; and Scialom, T. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. *ArXiv*, abs/2307.09288.

Tunstall, L.; Beeching, E.; Lambert, N.; Rajani, N.; Rasul, K.; Belkada, Y.; Huang, S.; von Werra, L.; Fourrier, C.; Habib, N.; Sarrazin, N.; Sanseviero, O.; Rush, A. M.; and Wolf, T. 2023. Zephyr: Direct Distillation of LM Alignment. *arXiv*:2310.16944.

Wang, C.; Liu, X.; Yue, Y.; Tang, X.; Zhang, T.; Jiayang, C.; Yao, Y.; Gao, W.; Hu, X.; Qi, Z.; Wang, Y.; Yang, L.; Wang, J.; Xie, X.; Zhang, Z.; and Zhang, Y. 2023a. Survey on Factuality in Large Language Models: Knowledge, Retrieval and Domain-Specificity. *ArXiv*, abs/2310.07521.

Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; ran Yang, H.; Zhang, J.; Chen, Z.-Y.; Tang, J.; Chen, X.; Lin, Y.; Zhao, W. X.; Wei, Z.; and rong Wen, J. 2023b. A Survey on Large Language Model based Autonomous Agents. *ArXiv*, abs/2308.11432.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; hsin Chi, E. H.; Xia, F.; Le, Q.; and Zhou, D. 2022. Chain of Thought Prompting Elicits Reasoning in Large Language Models. *ArXiv*, abs/2201.11903.

Wu, P.-C.; Yen, A.-Z.; Huang, H.-H.; and Chen, H.-H. 2022. Incorporating Peer Reviews and Rebuttal Counter-Arguments for Meta-Review Generation. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*.

Zhang, Y.; Li, Y.; Cui, L.; Cai, D.; Liu, L.; Fu, T.; Huang, X.; Zhao, E.; Zhang, Y.; Chen, Y.; Wang, L.; Luu, A. T.; Bi, W.; Shi, F.; and Shi, S. 2023a. Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *ArXiv*, abs/2309.01219.

Zhang, Y.; Wang, Y.; Wang, K.; Sheng, Q. Z.; Yao, L.; Mahmood, A.; Zhang, W. E.; and Zhao, R. 2023b. When Large Language Models Meet Citation: A Survey. *ArXiv*, abs/2309.09727.

## Appendix

### Prompts for different stages of the Workflows

#### 1. Motivation Extraction Prompt

System Message:

You are a researcher and trying to understand the following proposal written by another researcher: {proposal}

Human Message:

Describe in a bulleted list what is not addressed in the current literature which serves as the Motivation behind solving the above research problem proposed in the Proposal. Answer without a heading line and just the bullet points. Each bullet should mention one gap in the literature as a bullet point and not a sentence.

#### 2. Motivation Question Generation Prompt

System Message:

You are a researcher and trying to understand the following proposal written by another researcher: {proposal}

Human Message:

Describe in a bulleted list what is not addressed in the current literature which serves as the Motivation behind solving the above research problem proposed in the Proposal. Answer without a heading line and just the bullet points. Each bullet should mention one gap in the literature as a bullet point and not a sentence.

AI Message:

{motivation}

Human Message: Convert each of the above bullets in to a binary question. The question should begin with 'Is the research paper'.

#### 3. Ask Question for Motivation Validation Prompt

System Message:

You are a researcher. You have been given a context, which are paragraphs from a research paper. You have been given a question. Answer the given Question in 'Yes' OR 'No' OR 'Unanswerable'. Answer solely based on the provided context of the research paper. If the question can not be answered with the facts mentioned in the available context or there is any ambiguity in answering the question answer as 'Unanswerable'.

Answer as 'Yes' only when the question can be very clearly answered considering the facts in the research paper provided in the context. Do not repeat the question as the part of the answer.

Provide a concise explanation about how the answer to the question is 'Yes' mentioning the paragraphs used in the context to answer it as 'Yes'. If the answer is 'No' or 'Unanswerable' only output that with NO description or elaboration.

Human Message:

Question: {question}

Research Paper Context: {paper\_chunks}

#### 4. Extract Limitation Prompt

System Message:

You are a researcher. You have been given the following proposal: {proposal}

A different research paper provided in the context already addresses the gap mentioned as the motivation behind the proposal.

{descriptions}

Human Message:

Research Paper: {paper\_chunks}

Identify the limitations or gaps of this research paper which can serve as the new motivation for the proposal. Provide a bulleted list of limitations, where each bullet is concise. Answer WITHOUT a heading line and just the bullet points.

#### 5. Re-write Research Proposal Prompt

System Message:

You are a researcher and have written a proposal: {proposal}

Human Message:

Re-write the proposal by taking into consideration the mentioned gaps in the current literature as the new motivation behind of the problem defined in the proposal.

Answer in a Single detailed paragraph WITHOUT any bullet points or list.

Gaps in the current literature: {limitations}

#### 6. Research Problem Extraction Prompt

System Message:

You are a researcher and trying to understand the following proposal written by another researcher: {proposal}

Human Message:

What is the problem solved in the proposal?

#### 7. Similar Problem Generation Prompt

System Message:

You are a researcher and trying to understand the following proposal written by another researcher: {proposal}

Human Message:

What is the problem solved in the proposal?

AI Message:

{problem\_statement}

Human Message:

Give me a bulleted list of a more generalised or similar problems to the problem defined in the proposal. Don't give a heading just the answer in a bulleted list.

## 8. Sub Problem Generation Prompt

System Message:

You are a researcher and trying to understand the following proposal written by another researcher:  
{proposal}

Human Message:

What is the problem solved in the proposal?

AI Message:

{problem\_statement}

Human Message:

Provide a bulleted list of sub-problems or sub-tasks involved to solve the problem. Don't give a heading just the answer in a bulleted list.

## 9. Similar and Sub Problem Question Creation Prompt

Human Message:

{statement}

For the statement given above generate a question to be posed on a research paper to find out if the paper is proposing an approach or method to perform the task defined by the statement. Start the question with: 'Is the research paper proposing an approach or method to'.

## 10. Methodology Extraction Prompt

System Message:

You are a researcher and trying to answer the question posed on a research paper provided as the context.

Research Paper: {paper\_chunks}

Human Message:

Answer the given Question in 'Yes' OR 'No' OR 'Unanswerable'. Answer solely based on the provided context of the research paper. If the question can not be answered with the facts mentioned in the available context or there is any ambiguity in answering the question, answer as 'Unanswerable'. Answer as 'Yes' only when the question can be very clearly answered considering the facts in the research paper provided in the context. Do not repeat the question as the part of the answer. If the answer to the question is 'Yes', provide detailed approach or methodology to perform the task. If the answer is 'No' or 'Unanswerable' only output that with NO description.

Question: {question}

## 11. Method Synthesis Prompt

System Message:

You are a researcher and have been given a proposal and the research problem the proposal is trying to solve. You have been given the approaches in the literature trying to solve, similar problems and sub problems or sub tasks of the problem defined in the proposal. Your task is to synthesize and propose a possible set of methods or approaches to solve the problem defined in the proposal.

Proposal: {proposal}

Research Problem in the Proposal: {problem}

Human Message:

{method\_context}

Based on the above information suggest the top 3 possible methods or approaches to solve the problem defined in the proposal.