Current Workflows

- o

Arxiv Paper Retrieval: Given a list of keywords, retrieves the most relevant papers from arXiv.

Current Workflows

o

PDF to Text: Utilizes Nougat from Meta for OCR conversion of papers. Effective but requires significant computational resources.

Current Workflows

o

OCR Enhancer: Uses MuPDF to correct citation formats after initial OCR by Nougat. Compares and merges the best aspects of both text files.

Current Workflows

- o

Proof Remover: Attempts to remove proofs to clean texts before summarization. This workflow needs further improvement.

Current Workflows

- o

Keyword Extraction and Topic Summarization: Extracts keywords and summarizes the article page by page. Potential for improvement by performing multiple passes.

Current Workflows

o

Context-Based Translation: Translates the article into different languages using context from the summarization, ensuring appropriate terminology for the specific academic community.

Upcoming Features

o

Survey Creation: The LLM will identify the most relevant citations in a paper, retrieve related papers from arXiv or other sources, process each paper using the current workflows, and compile a comprehensive survey. This feature aims to streamline the creation of complex surveys.

Upcoming Features

o

Proof Explainer: This ambitious feature intends to analyze mathematical texts, generate context, and explain proofs. A secondary goal is to categorize proofs with similar arguments, enhancing understanding and learning.