

SciRAD

Beyond the Hype: Application of AI Agents to Streamline the Retrieval, Aggregation, and Delivery of Scientific Insights

Year-End Deliverable

Presented by: Saleh Alkhalifa

Project Plan

Number	Title	Description	Commitment	Status	Notes
SR-1	Project Plan	Create an overall plan and proposal and submit to professors	Committed	Complete	Submitted V2 with prof edits
SR-2	Create a single agent	Create a single agent that can query the database using keywords	Committed	Complete	Agent functional!
SR-3	Summarization model	Give the agent the ability to summarize content effectively and calculate metrics such as ROUGE and BLEU	Committed	Complete	BLEU and ROUGE were not good metrics, need agent instead.
SR-4	Cosine to filter abstracts	Implement a Cosine Similarity filter to reduce the number of articles	Committed	Complete	Professor recommendation.
SR-5	Judge Agent	Implement a "Agent as a judge" framework to assess the quality of the results	Committed	Complete	Agent working well, expanded to multiple criteria.
SR-6	MLFlow for Optimization	Incorporate MLFlow to optimize the different methods and parameters in the overall pipeline	Committed	Complete	MLFlow added, experiment is now functional
SR-7	Optimize the pipeline	Experiment with various parameters by changing the model type, prompt type, summary length, top P, temperature, and several others.	Committed	Complete	Pipeline optimized for NCBI
SR-8	Explore a Multiagent Framework	Explore modifying the architecture to a multi-agent framework by changing the tools to independent agents with an orchestrator agent to report to.	Optional	Pivoted	Explored, but could not achieve better performance compared to single agent.
SR-9	Explore Judge Re-Prompt Evaluation	Explore re-prompting the judge to determine whether the score changes	Professor Recommended	Complete	Complete, and evals are very similar
SR-10	Python Package	Wrap the package up as a Python package to deploy for people to use	Stretch Goal	Complete	Package created and functional
SR-11	Flask Application	Wrap package into a Flask API so individuals can deploy it (Lambda, API gateway, EC2, ...)	Stretch Goal	Complete	Flask API is functional
SR-12	Document the package	Fully document the package and the deployments	Committed	Complete	Package is fully documented

SciRAD

Beyond the Hype: Application of AI Agents to Streamline the Retrieval, Aggregation, and Delivery of Scientific Insights

Year-End Deliverable

Presented by: Saleh Alkhalifa

Executive Summary

Status: **Green**
Risks: **Mitigated**
Delivery: **On Time**



The Problem

- Researchers face **information overload** as millions of papers are published annually.
- It is **time-consuming** to manually filter, read, and extract relevant information.
- Existing tools **lack precision, personalization, and efficiency** in summarization.



The Solution

- **SciRAD** is an agentic assistant that summarizes scientific papers into a **personalized weekly digest**.
- Uses **LLM-based agent** to ensure **quality, consistency, & relevance**.
- Keyword recommendation + cosine similarity ranking **filter out noise**.



Mid-Point

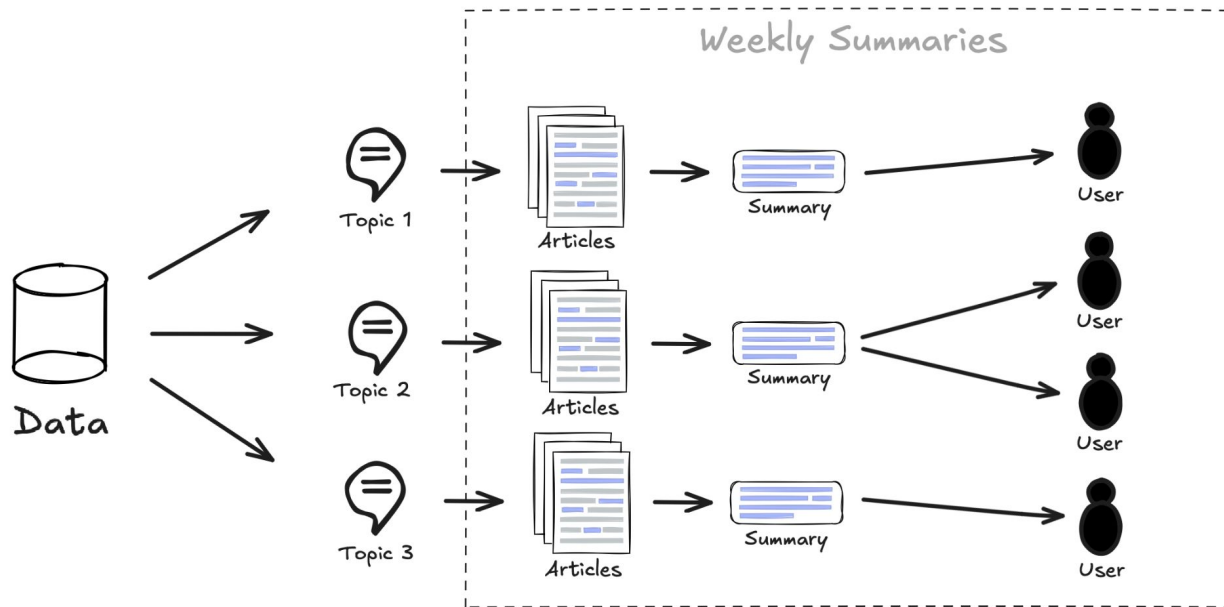
- **Mostly functional pipeline** allowing users to to run the end-to-end query.
- **High-quality summary** using agentic framework using LLM-as-a-Judge framework as an evaluation metric.
- **Single Agent** fully functional, but **Multi-Agent** did not outperform the others.



Year-End

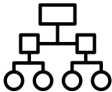



- **Fully functional E2E pipeline** allowing users to to run their queries, deployed via streamlit and PyPI for use.
- **High-quality summary** using agentic framework and LLM-as-a-Judge evaluated by experimental studies.
- **Single Agents** are fully functional, connected, and implemented.

How does SciRAD work?

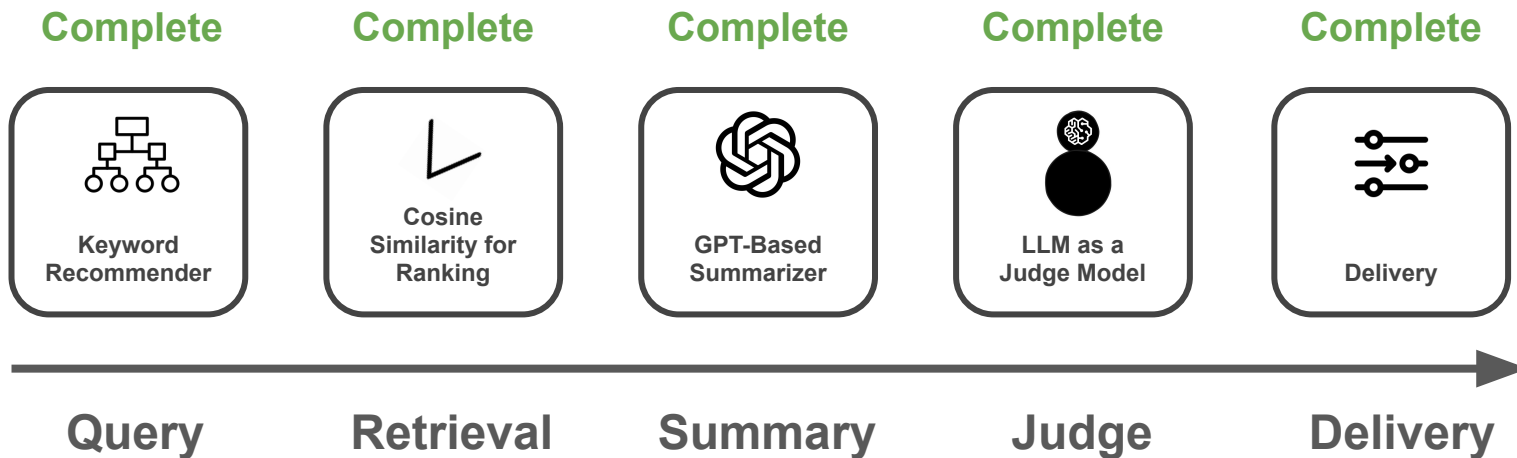


- **Quality:** Ensure that summaries reflect the original content precisely.
- **Relevance:** Filter out noise so users only receive content that directly impacts their work.
- **Efficiency:** Save time by quickly highlighting key insights, 80:20 rule.
- **Personalization:** Tailor content to each user's interests and needs.
- **Transparency:** Provide clear citations on which articles were used.

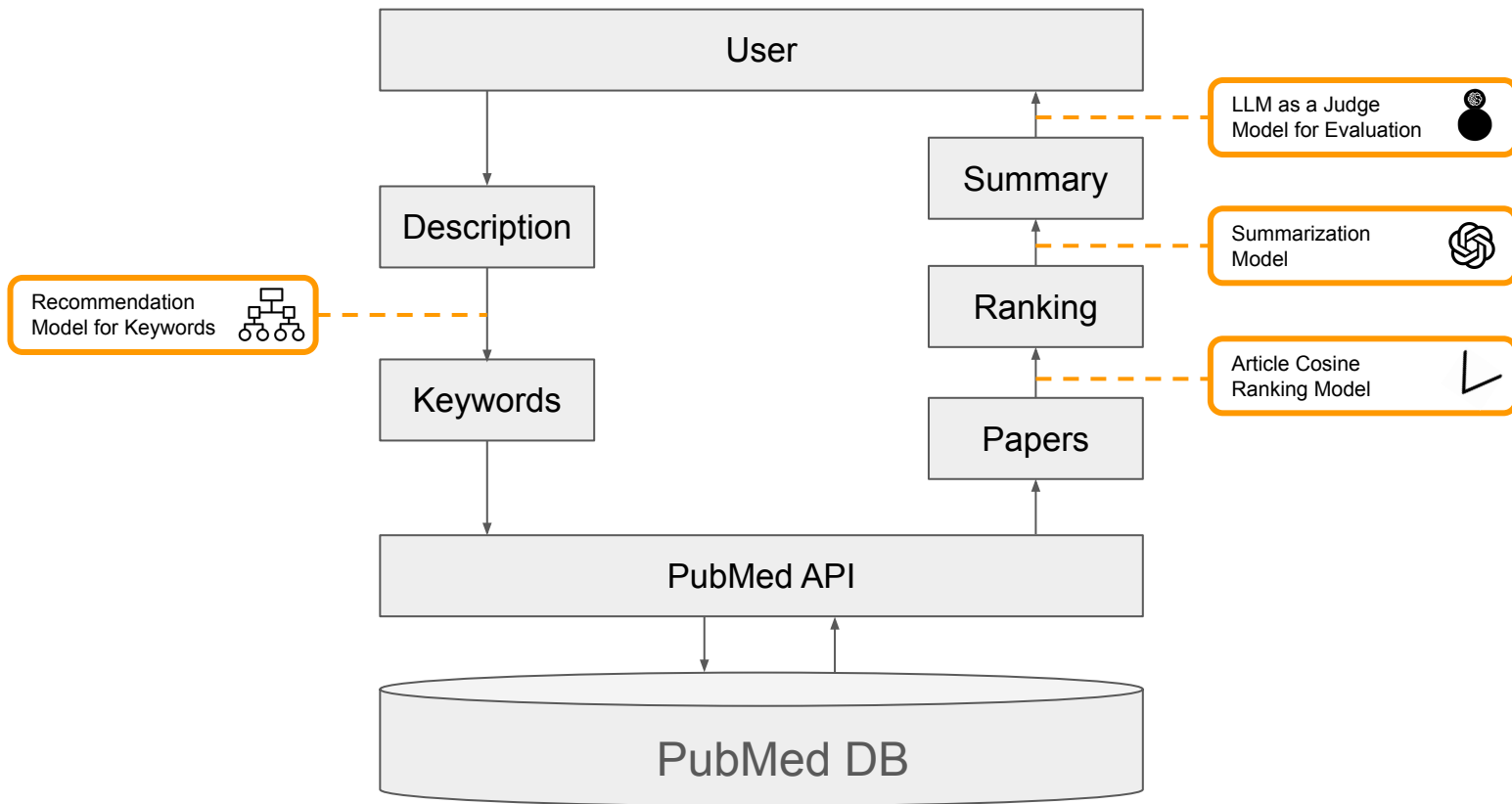
SciRAD Features

	Query	Retrieval	Summary	Delivery
Problem Statement	Insufficient keywords queries can result in misaligned search results	Limit large search results to a smaller subset of most relevant papers	A summary of all abstracts would make staying up to date much easier	Models are great, but only when they can be feasibly productionalized
Proposed Solution	 Keyword Recommender	 Cosine Similarity	 GPT-Based Summarizer	 Delivery
Resulting Feature	Recommends the top keywords for the user to pick from	Limits the scope of articles to the most relevant ones	Summarizes many abstracts to give the user a summary	Delivers to the user a summary of content for a topic

High-level Steps



High-level Steps



Experimental Results

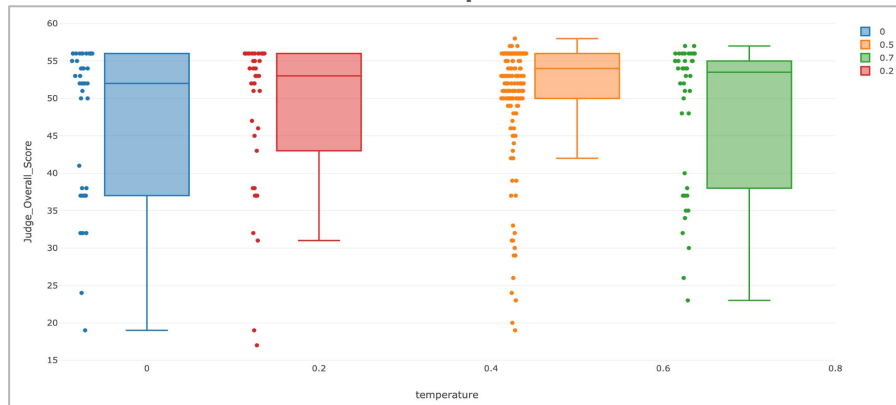
Optimal Value

Temperature	0.2	0.4	0.5	0.8	1.0
Top P	0.9	1.0			
Word Count	150	200	250	300	350
Prompting Mechanism	Role-Based	Instruction-Based	Zero-Shot	Tree of Thought	Chain of Thought
Summary Model Type	GPT-4	GPT-4o	GPT-3.5-Turbo		
Judge Model Type	GPT-4	GPT-4o	GPT-3.5-Turbo		
Framework	LC Single Agent	LG Single Agent	Multi-Agent		

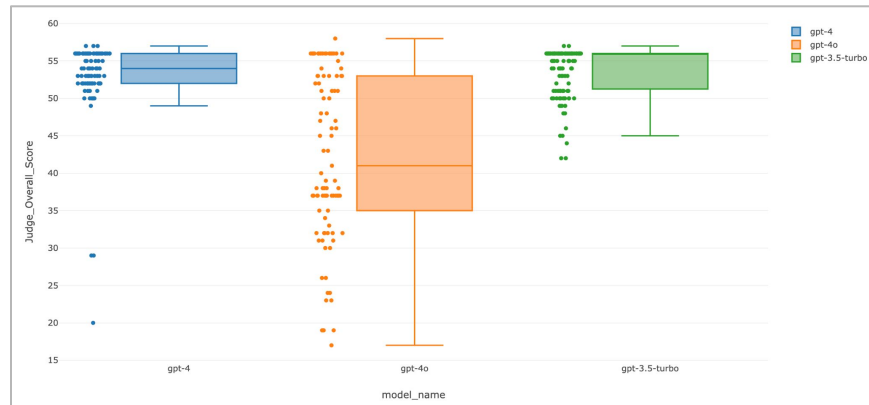
LC = LangChain Framework
LG = LangGraph Framework

Sample Experimental Results

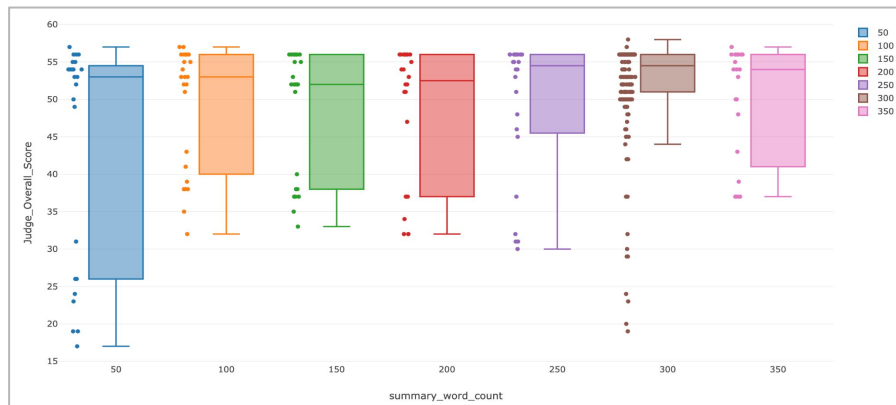
Score vs Temperature



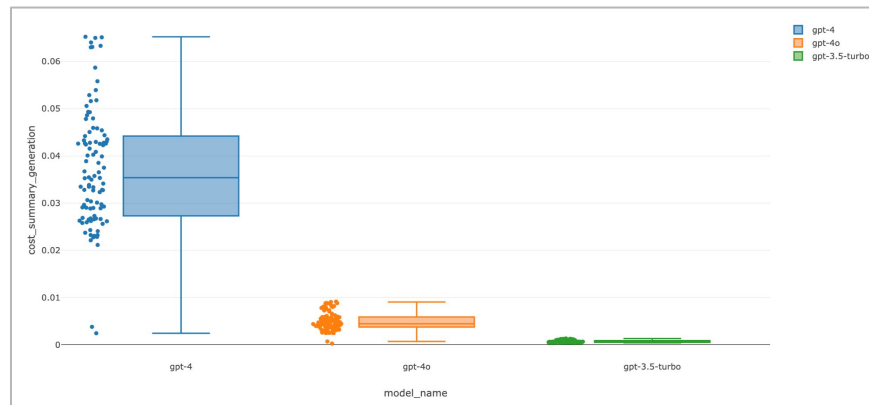
Score vs Model



Score vs Word Count



Model vs Cost



Streamlit Demo Environment



SciRAD

Welcome to SciRAD

This application is a Streamlit Agent designed to assist with research on content.

How to use:

- Admin Tab:** Enter a research description, generate recommended keywords, edit keywords, click Run Agentic Search, view results (input, judge evaluation, and total cost).
- User Tab:** Enter a research description, generate recommended keywords, edit them as needed, then click Approve and Get Summary to see the final summary and the search timeframe.

Useful Links:

- [GitHub Repository](#)
- [PyPI Library](#)
- [Slide Deck](#)

A DevNote © 2025 All Rights Reserved

Admin Panel - Development Mode

Enter Research Description

I am interested in GLP1 as it relates to weight loss medications being developed in the pharma industry

Generate Recommended Keywords

Edit Keywords (comma-separated):

GLP1, weight loss, medications

Run Agentic Search

Input & Summary

Research Description

I am interested in GLP1 as it relates to weight loss medications being developed in the pharma industry

Keywords

```
[{"id": 0, "keyword": "GLP1"}, {"id": 1, "keyword": "weight loss"}, {"id": 2, "keyword": "medications"}]
```

Summary

Recent research has shown that tirzepatide, a GLP1 agonist, is being used in real-world settings among individuals without type 2 diabetes in the United States. The study aimed to understand the treatment patterns and effectiveness of tirzepatide in this population. The findings revealed that tirzepatide was associated with significant weight loss and improvements in cardiometabolic risk factors, indicating its potential as a weight loss medication. The study utilized real-world data to assess the effectiveness of tirzepatide, providing valuable insights into its use outside of clinical trials. The implications of these findings suggest that GLP1 agonists like tirzepatide could be promising options for weight loss interventions in individuals without diabetes, highlighting the importance of further research in this area to address the growing obesity epidemic. Overall, the study contributes to the growing body of evidence supporting the use of GLP1 agonists for weight management and underscores the need for more personalized and effective treatments for obesity.

Articles:

- Title: Real-World Use and Effectiveness of Tirzepatide Among People Without Evidence of Type 2 Diabetes in the United States.

Metrics

	Metric	Value
0	ROUGE-1	0.1569
1	ROUGE-2	0.1287
2	ROUGE-L	0.1569
3	BLEU	0.0741

Judge Evaluation

	Criterion	Score	Comment
0	Accuracy	10	The summary accurately reflects the key points and findings of the article.
1	Completeness	8	Most major findings are included, but some details could be more comprehensive.
2	Clarity	9	The summary is clear, well-structured, and easy to understand.
3	Conciseness	7	While concise, some additional details could be omitted for brevity.
4	Relevance	10	All included information is pertinent to the main topic without introducing unrelated details.
5	Objectivity	10	The summary is unbiased and presents information factually without personal opinion.
6	Overall Score	54	

Timeframe Searched: Last 7 days

Total Cost: \$0.000544



Keyword Recommender



Cosine Similarity for Ranking



GPT-Based Summarizer



LLM as a Judge Model

Promises Made, Promises Kept



The Agent

The agent can autonomously use the OpenAI API as needed to complete its task.



OpenAI

The agent can autonomously build queries and search the NCBI database



Database

Summary Agent



Metric Calc

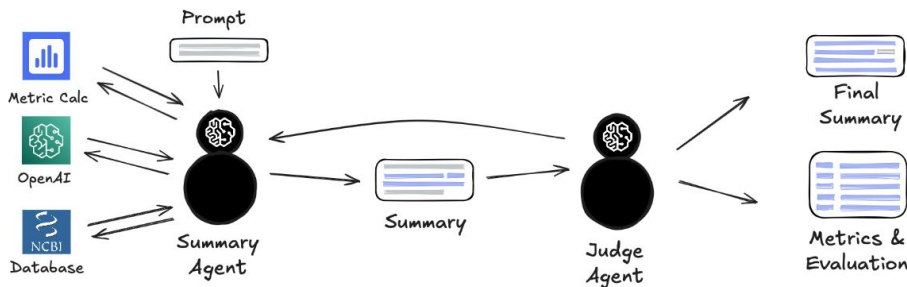
The agent can autonomously calculate metrics for a summary using ROUGE and BLEU



Judge Evaluator

The agent can autonomously ask an independent Agent-As-A-Judge to evaluate the summary

The Task



The Deployment



GitHub Repo



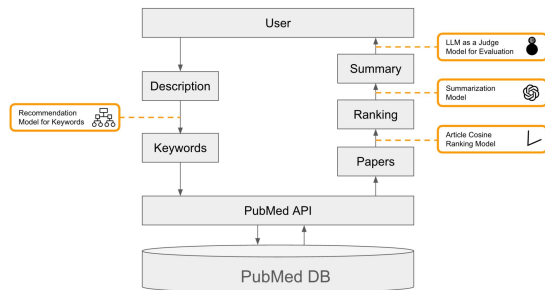
Flask API



Python Package

Lessons Learned

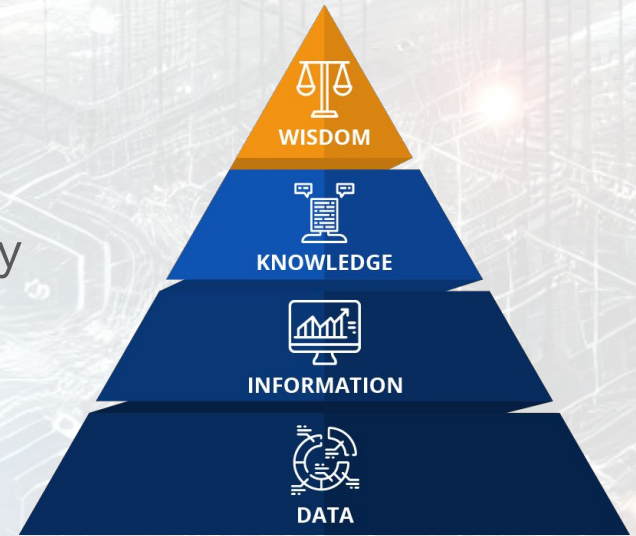
- It was important to have **clear objectives and maintain a line of sight of the final goal**. Being able to evaluate each feature in terms of its integration and impact within the broader system helped maintain focus.
- Actively **soliciting and incorporate feedback** throughout development was useful; the professor's guidance was helpful in refining the project scope to a practical, achievable scale given the duration of the term.
- Although the foundational agentic AI frameworks are well-established at this point, the **rapid evolution** of multi-agent architectures and **shifting best practices** make it challenging to implement a fully operational multi-agent system.



SciRAD

Streamlined Retrieval, Aggregation, and Delivery of Scientific Insights

- Enhanced Knowledge Accessibility
- Optimized Researcher Productivity
- Empowered Decision Making



References

1. Zhang, J., Zhao, Y., Saleh, M., & Liu, P. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. arXiv. **2019**.
2. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv. **2020**.
3. Hernandez-Leal, P., Kartal, B., & Taylor, M. E. A Survey of Multi-Agent Reinforcement Learning. arXiv. **2019**.
4. Bawden, D., & Robinson, L. The Dark Side of Information: Overload, Anxiety and Other Paradoxes and Pathologies. Journal of Information Science. **2009**.
5. Jackson, T., Dawson, R., Wilson, D., & Riddle, D. Overcoming Information Overload in the Digital Age: A Framework for Decision-Making. International Journal of Information Management. **2017**.
6. Chen, H., Hsieh, Y., & Wang, S. Reducing Information Overload in Healthcare: A Systematic Review. Journal of Medical Internet Research. **2005**.
7. OpenAI. Learning to Summarize with Human Feedback. OpenAI Blog. **2021**.
8. OpenAI. ChatGPT: Optimizing Language Models for Dialogue. OpenAI Blog. **2022**.
9. Fu, Y., et al. ReAct: Synergizing Reasoning and Acting in Language Models. arXiv. **2022**.
10. Schick, T., et al. Toolformer: Language Models Can Teach Themselves to Use Tools. arXiv. **2023**.
11. LangChain Team. LangChain: A Framework for LLM-Powered Applications. LangChain Blog. **2023**.
12. LangChain Team. Introducing LangSmith: Observability for LLM Applications. LangChain Blog. **2023**.
13. Park, J., et al. Generative Agents: Interactive Simulacra of Human Behavior. Conference Publication. **2023**.
14. Kim, S., et al. Big Data and Information Overload: A Survey. arXiv. **2020**.



Supporting



VC Pitch



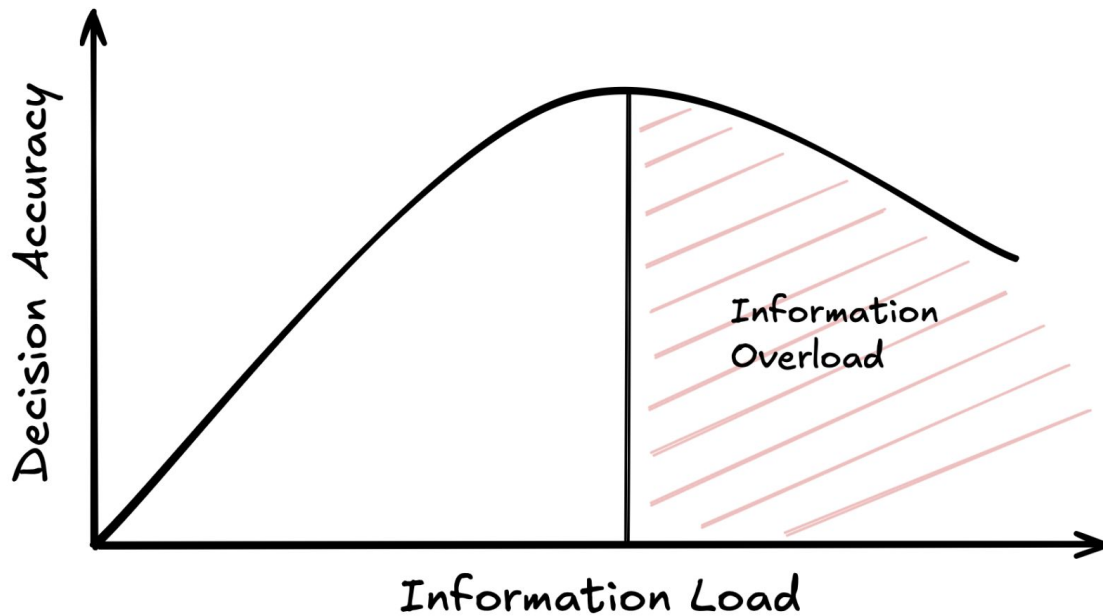
SciRAD

Beyond the Hype: Application of AI Agents to Streamline the Retrieval, Aggregation, and Delivery of Scientific Insights

VC Pitch

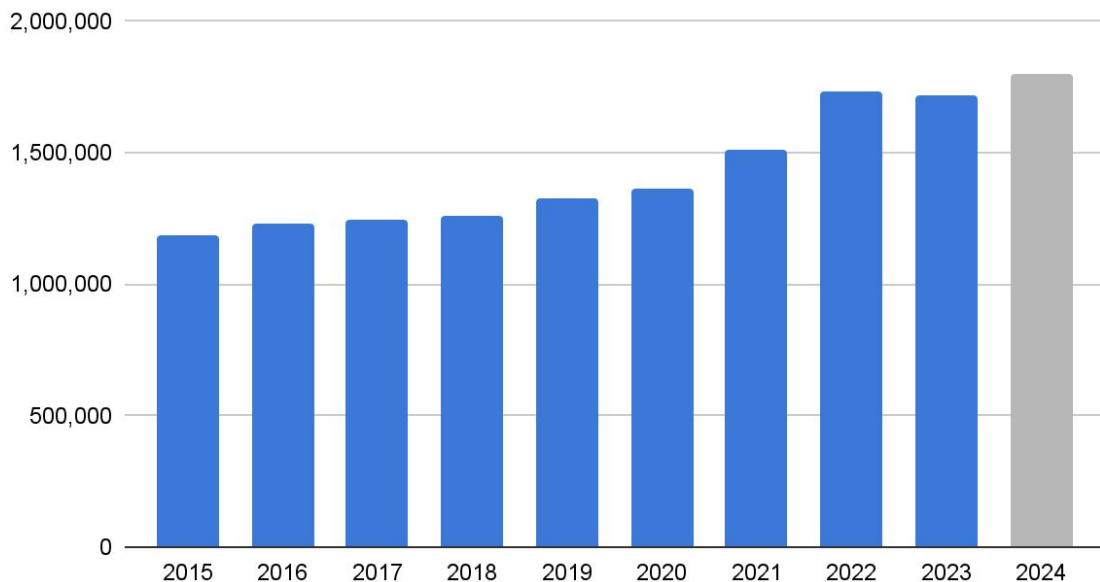
Presented by: Saleh Alkhalifa

What is Information Overload?



Annual Count of Articles Published in PubMed

PubMed Articles Published

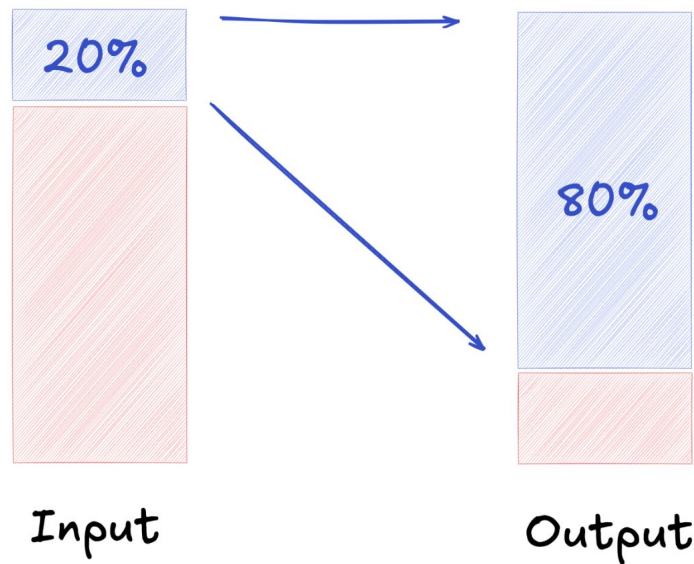


$$1,800,000 \times 0.001 \\ = \mathbf{1,800 \text{ per year}}$$

$$1,800 / 52 \\ = \mathbf{34 \text{ per week}}$$

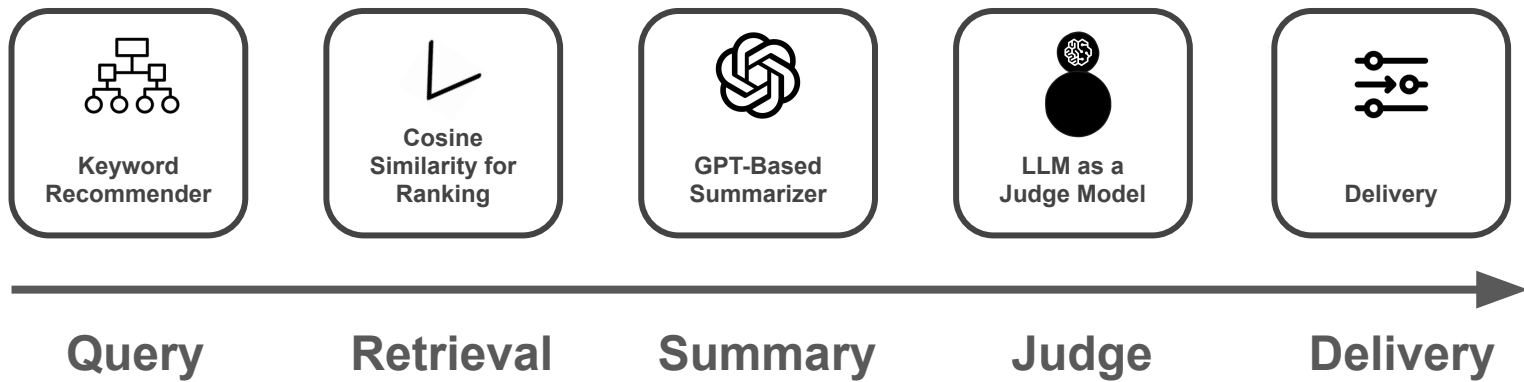
$$34 / 7 \\ = \mathbf{4.8 \text{ per day}}$$

The 80:20 Rule



80% of output results
from 20% of the input

How SciRAD Works



SciRAD in Action

SciRAD

Welcome to SciRAD

This application showcases an Intelligent Agent designed to help you stay up to date on content.

How to use:

- Admin Tab:** Enter a research description, generate recommended keywords, edit them as needed, then click **Run Agentic Search** to see full results (input, summary, metrics, judge evaluation, search timeframe, and total cost).
- User Tab:** Enter your description, generate recommended keywords, edit them as needed, then click **Approve and Get Summary** to see the final summary and the search timeframe.

Useful Links:

- [GitHub Repository](#)
- [PyPI Library](#)
- [Slide Deck](#)

[A DevNavigator Product](#)
© 2025 DevNavigator
All Rights Reserved

Admin Panel - Development Mode

Deploy

Enter Research Description

I am interested in GLP1 as it relates to weight loss medications being developed in the pharma industry

Generate Recommended Keywords

Edit Keywords (comma-separated):

GLP1, weight loss, medications

Run Agentic Search

Input & Summary

Research Description

I am interested in GLP1 as it relates to weight loss medications being developed in the pharma industry

Keywords

```
[{"id": 0, "keyword": "GLP1"}, {"id": 1, "keyword": "weight loss"}, {"id": 2, "keyword": "medications"}]
```

Summary

Recent research has shown that tirzepatide, a GLP1 agonist, is being used in real-world settings among individuals without type 2 diabetes in the United States. The study aimed to understand the treatment patterns and effectiveness of tirzepatide in this population. The findings revealed that tirzepatide was associated with significant weight loss and improvements in cardiometabolic risk factors, indicating its potential as a weight loss medication. The study utilized real-world data to assess the effectiveness of tirzepatide, providing valuable insights into its use outside of clinical trials. The implications of these findings suggest that GLP1 agonists like tirzepatide could be promising options for weight loss interventions in individuals without diabetes, highlighting the importance of further research in this area to address the growing obesity epidemic. Overall, the study contributes to the growing body of evidence supporting the use of GLP1 agonists for weight management and underscores the need for more personalized and effective treatments for obesity.

Articles:

- Title: Real-World Use and Effectiveness of Tirzepatide Among People Without Evidence of Type 2 Diabetes in the United States.

Metrics

	Metric	Value
0	ROUGE-1	0.1569
1	ROUGE-2	0.1287
2	ROUGE-L	0.1569
3	BLEU	0.0741

Judge Evaluation

	Criterion	Score	Comment
0	Accuracy	10	The summary accurately reflects the key points and findings of the article.
1	Completeness	8	Most major findings are included, but some details could be more comprehensive.
2	Clarity	9	The summary is clear, well-structured, and easy to understand.
3	Conciseness	7	While concise, some additional details could be omitted for brevity.
4	Relevance	10	All included information is pertinent to the main topic without introducing unrelated
5	Objectivity	10	The summary is unbiased and presents information factually without personal opinio
6	Overall Score	54	

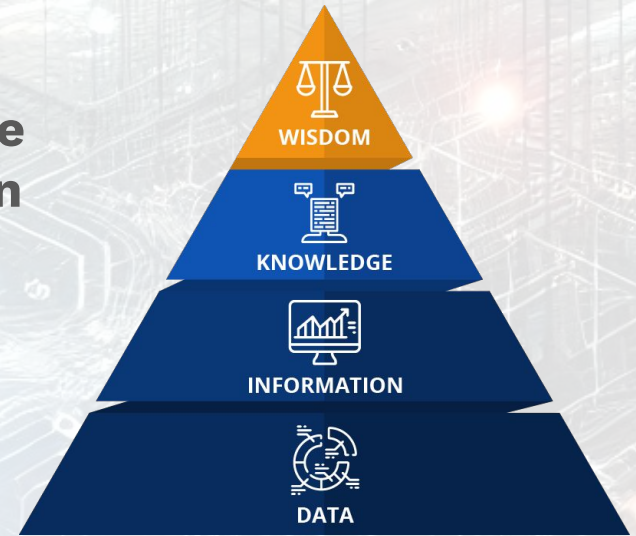
Timeframe Searched: Last 7 days

Total Cost: \$0.000544

SciRAD

Streamlined Retrieval, Aggregation, and Delivery of Scientific Insights

Help us fund our product, and scale our efforts to make a difference in the scientific domain today.



Elevator Pitch

Imagine a tool that cuts through the overwhelming noise of research updates and delivers only the most relevant breakthroughs to you right when you need them. Meet SciRAD. In today's fast-paced innovation landscape, staying current isn't just an advantage; it's a necessity. SciRAD solves the problem of information overload by curating bite-sized, personalized updates on the research areas you care about the most.

With SciRAD, you don't have to sift through endless academic journals, conference reports, or scattered news feeds. The platform smartly aggregates high-quality, targeted insights and delivers them in a succinct, easy-to-digest format, ensuring that you're always one step ahead. Whether you're a busy executive, an active researcher, or simply a lifelong learner, SciRAD provides clarity amid chaos, allowing you to make informed decisions quickly.

SciRAD works by allowing you to define exactly what research areas matter most to you. Each week, the platform taps into trusted databases like NCBI to fetch the latest articles and data related to your interests. These results are then aggregated using sophisticated Large Language Models to generate concise, targeted summaries. The result is an automated service that distills complex research into insights, delivered straight to you for an effortless and informed start to your day.

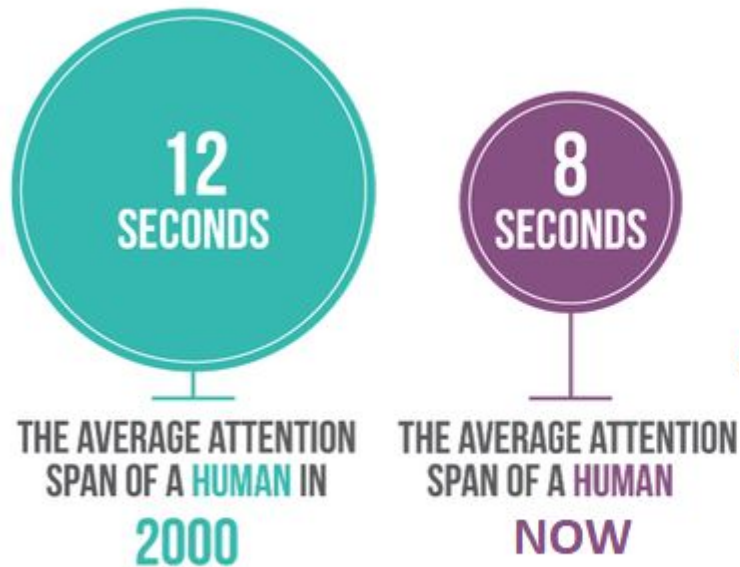
Imagine starting your day with a brief, focused update that highlights the most significant developments in your field. That's the power of SciRAD. It filters out the irrelevant and distills the essence of cutting-edge research, so you can invest your time where it truly matters.

In essence, SciRAD transforms the way you stay informed in a world inundated by data, turning overwhelming streams of information into actionable insights. It's not just an update service—it's your personal research curator, helping you leverage the fastest pace of innovation without the stress of information overload. Step into the future of smart research communication with SciRAD.



Supporting (Intro)

Today's Attention Span



Today's Attention Span



THE AVERAGE ATTENTION
SPAN OF A HUMAN IN
2000



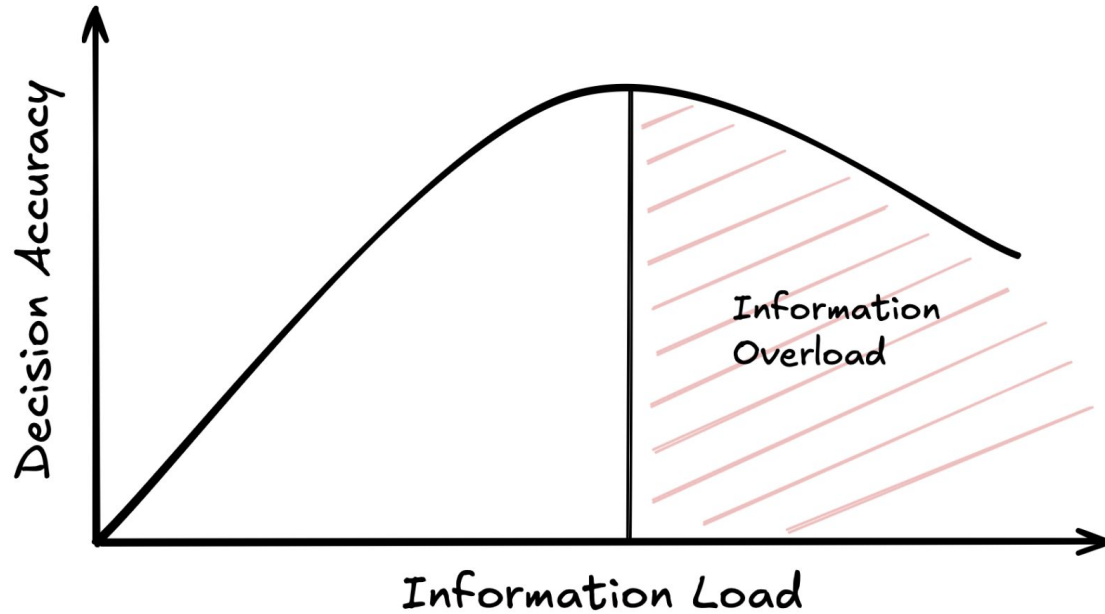
THE AVERAGE ATTENTION
SPAN OF A HUMAN
NOW

AND...

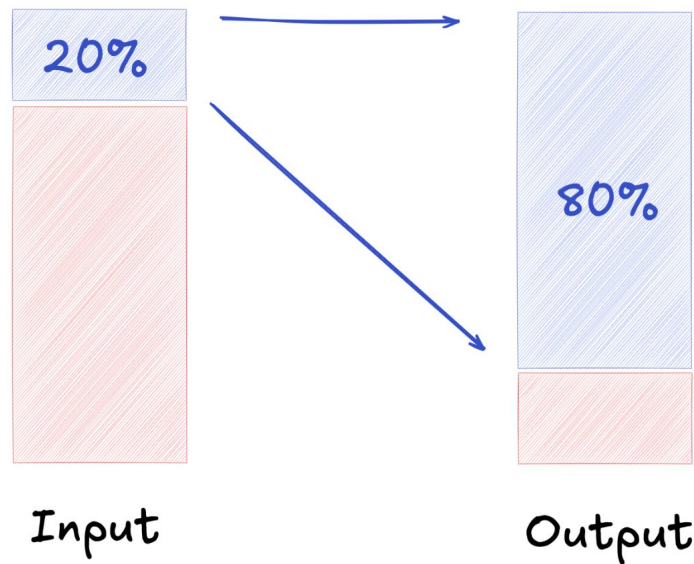


THE AVERAGE ATTENTION
SPAN OF A
GOLDFISH

What is Information Overload?



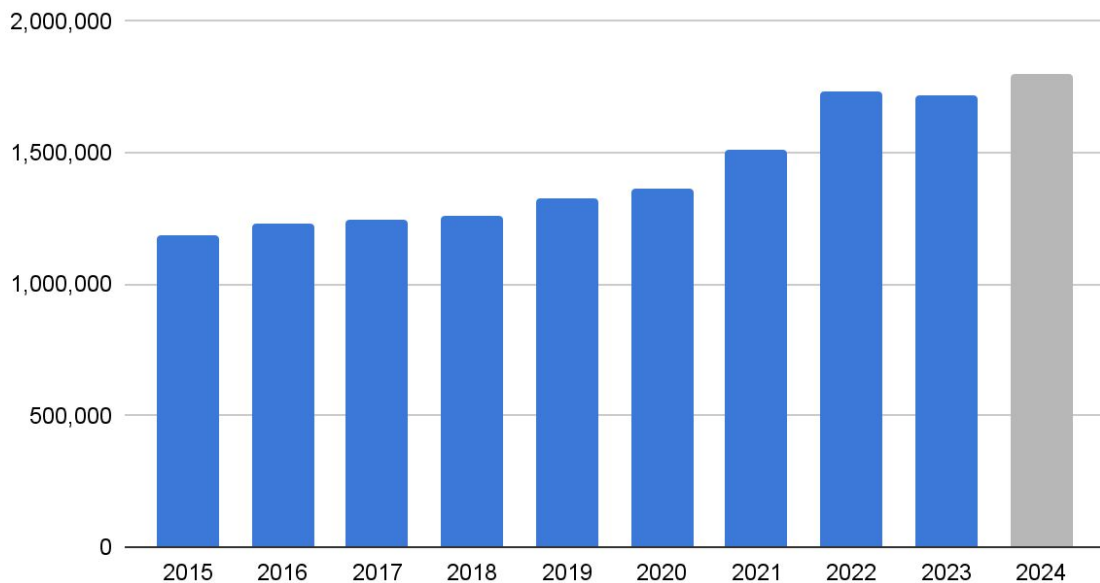
The 80:20 Rule



80% of output results
from 20% of the input

Annual Count of Articles Published in PubMed

PubMed Articles Published



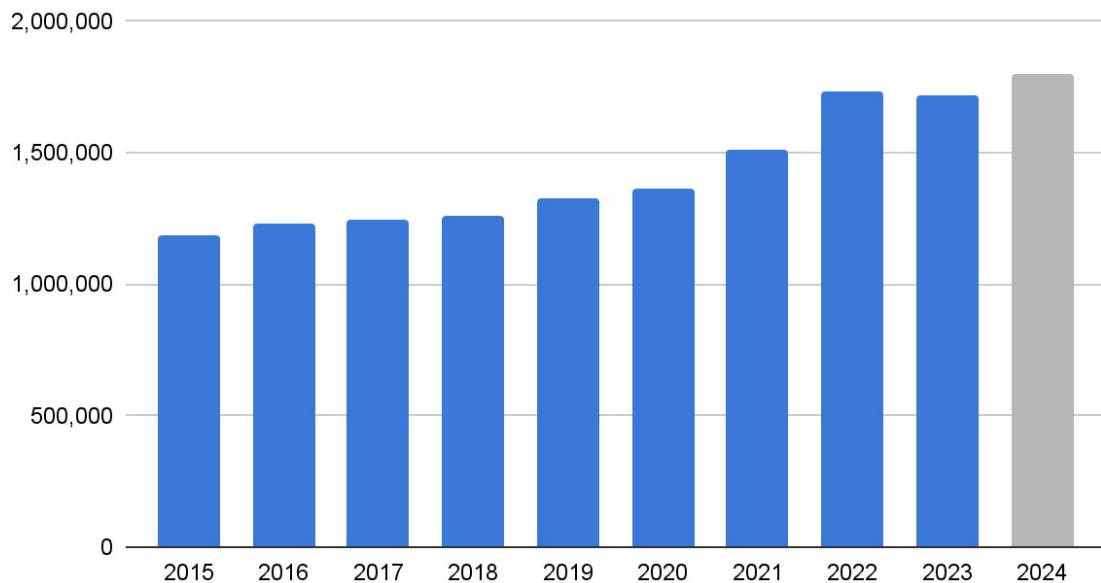
$$1,800,000 \times 0.001 \\ = \mathbf{1,800 \text{ per year}}$$

$$1,800 / 52 \\ = \mathbf{34 \text{ per week}}$$

$$34 / 7 \\ = \mathbf{4.8 \text{ per day}}$$

Annual Count of Articles Published in PubMed

PubMed Articles Published





Supporting (Results)

Risks and Challenges

Resolved

Database
access with
appropriate
metadata

Deployment
Plan via PyPi

LLM as a
Judge
Inconsistent

Owned

Lack of
Feedback
from Real
Users

Accepted

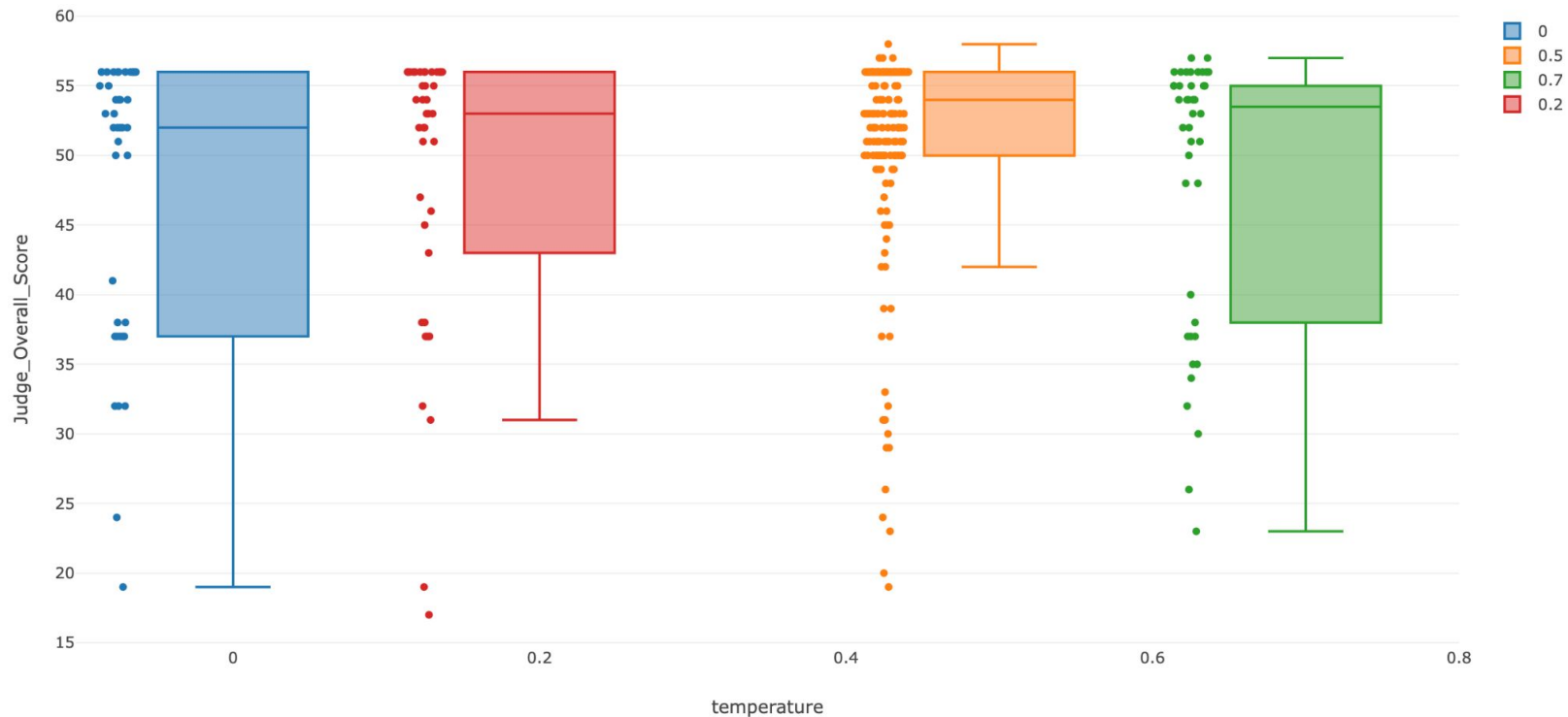
Mitigated

API
Limitations

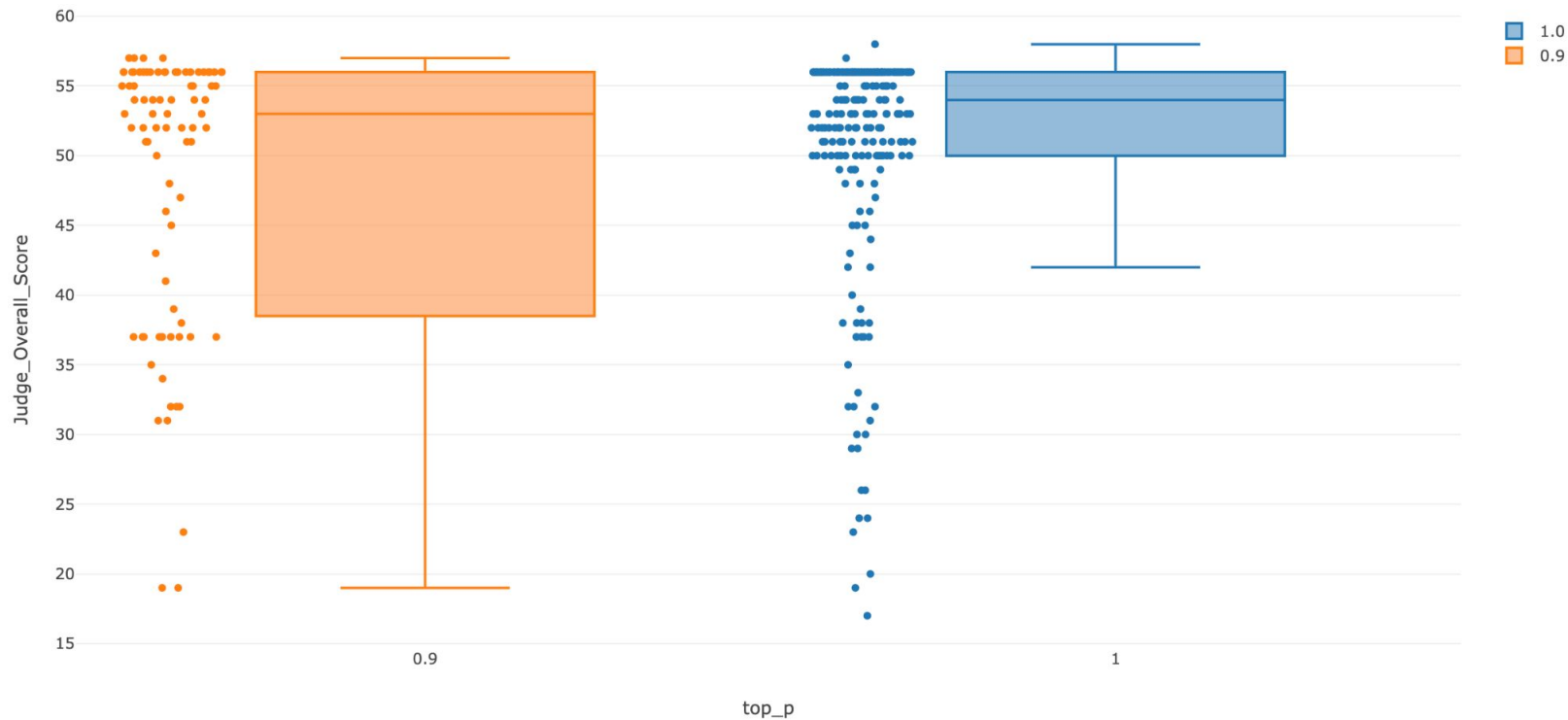
OpenAI
Dependency

Summarization
Accuracy

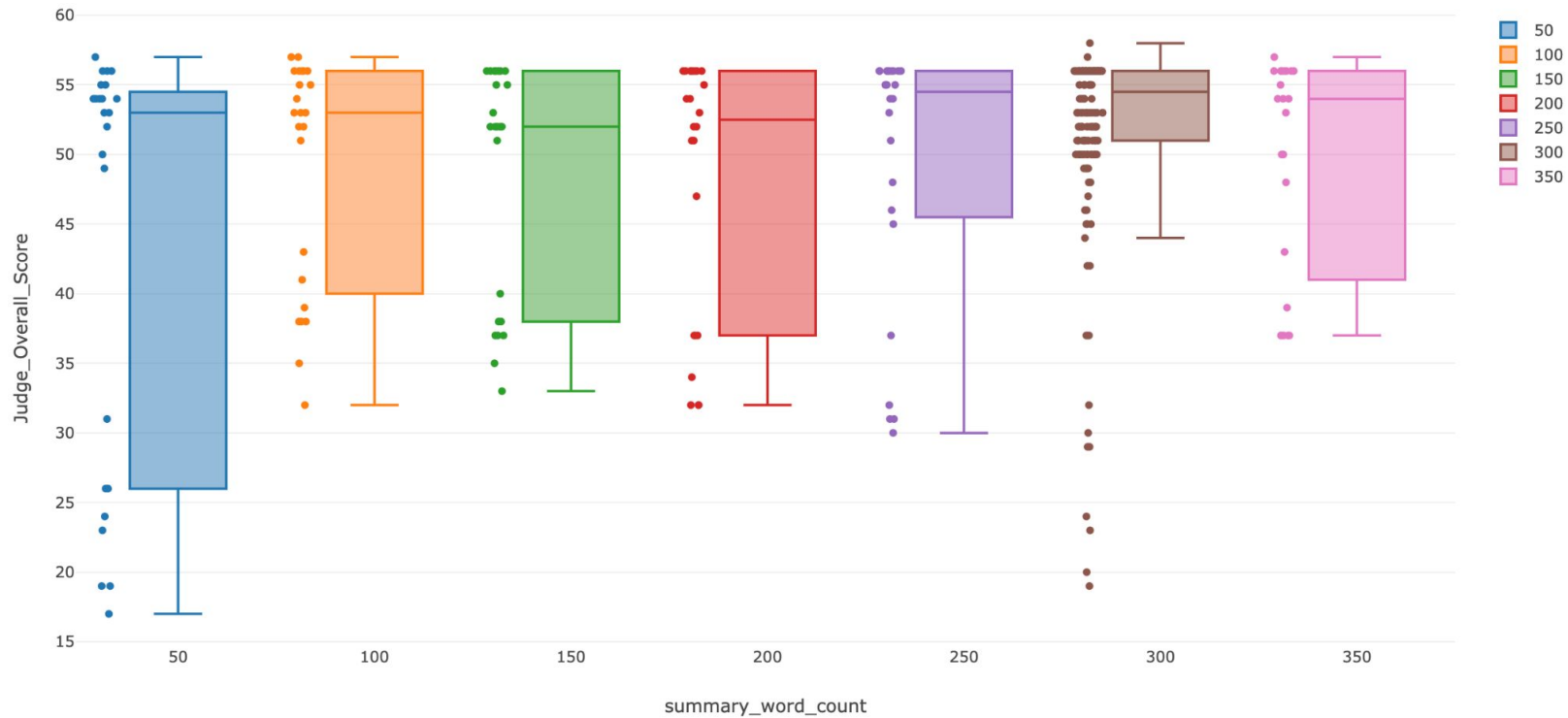
Score vs Temperature



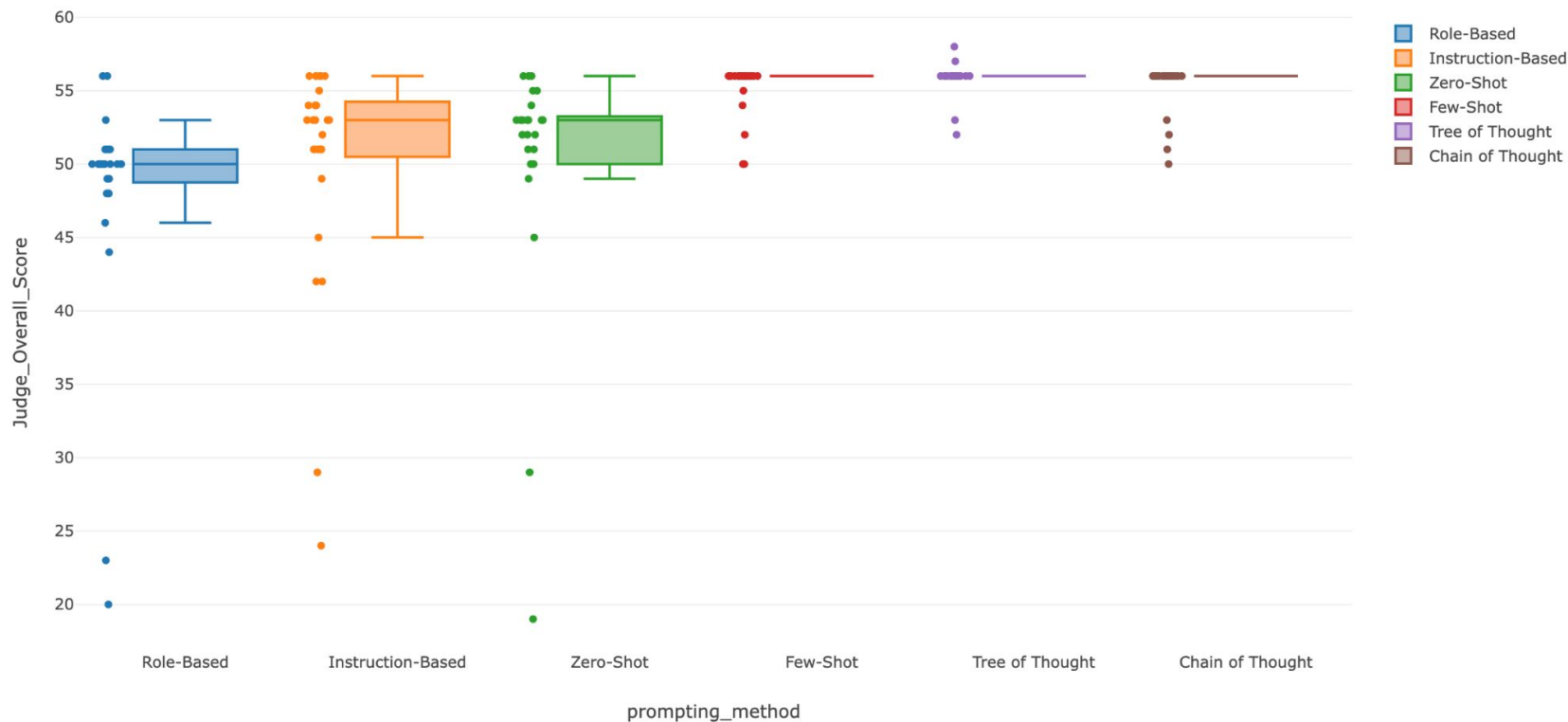
Score vs Top P



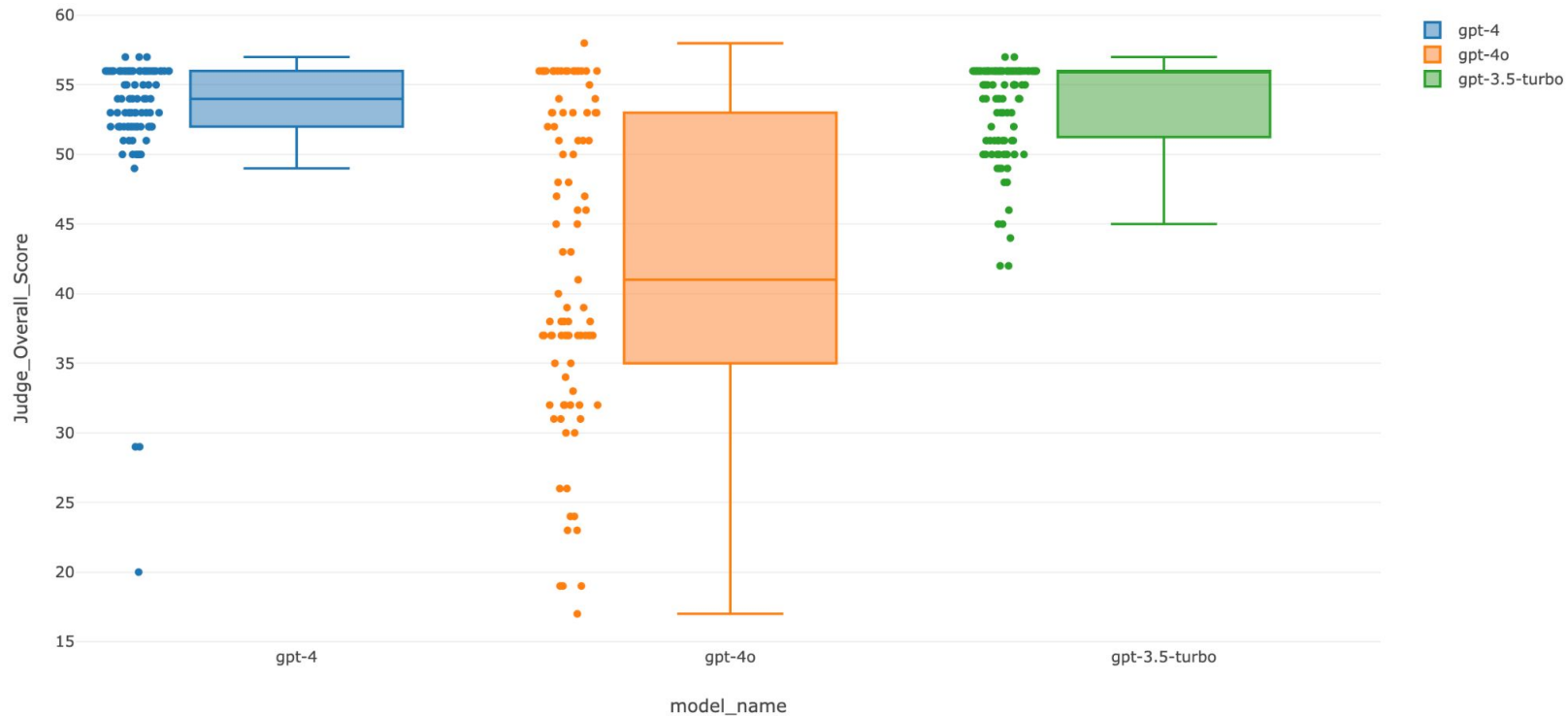
Score vs Word Count



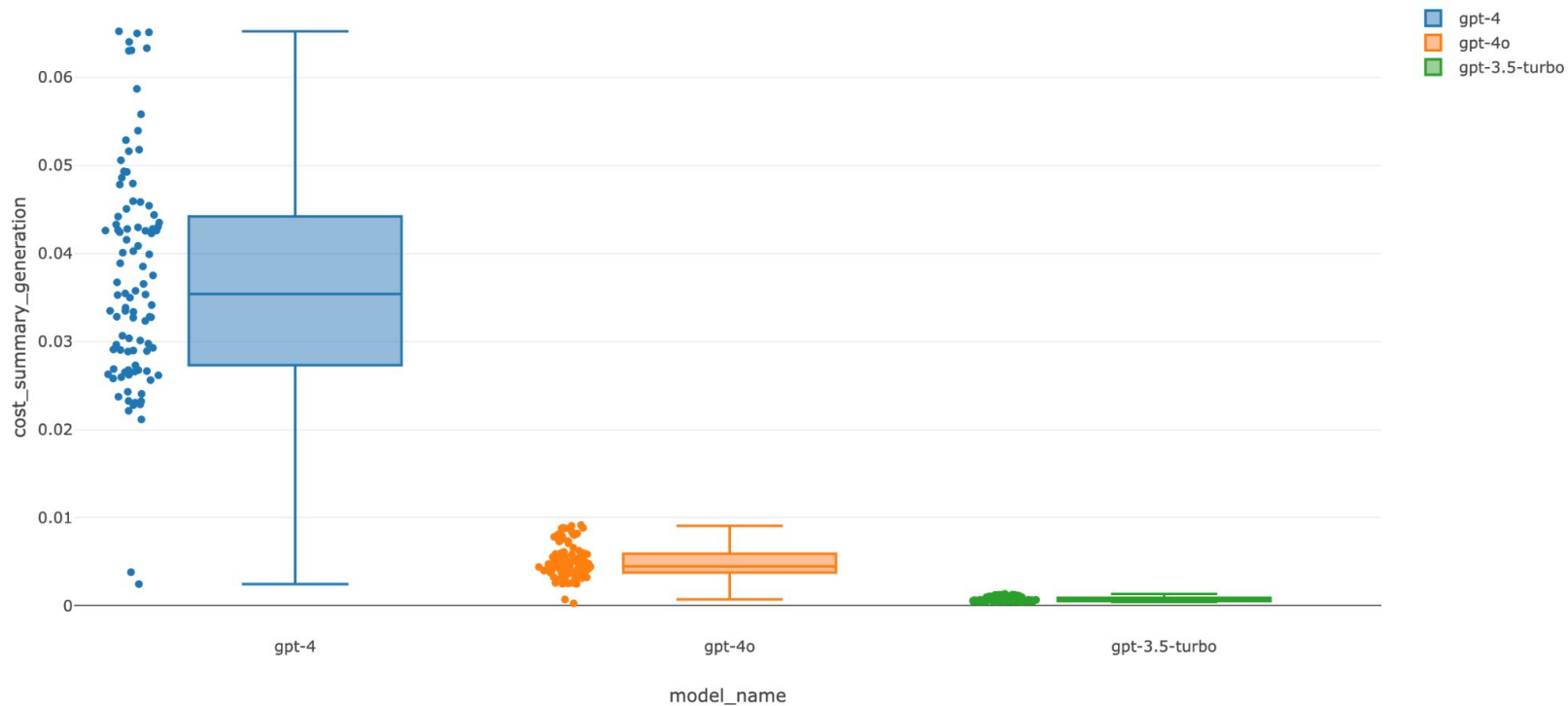
Score vs Prompt Method



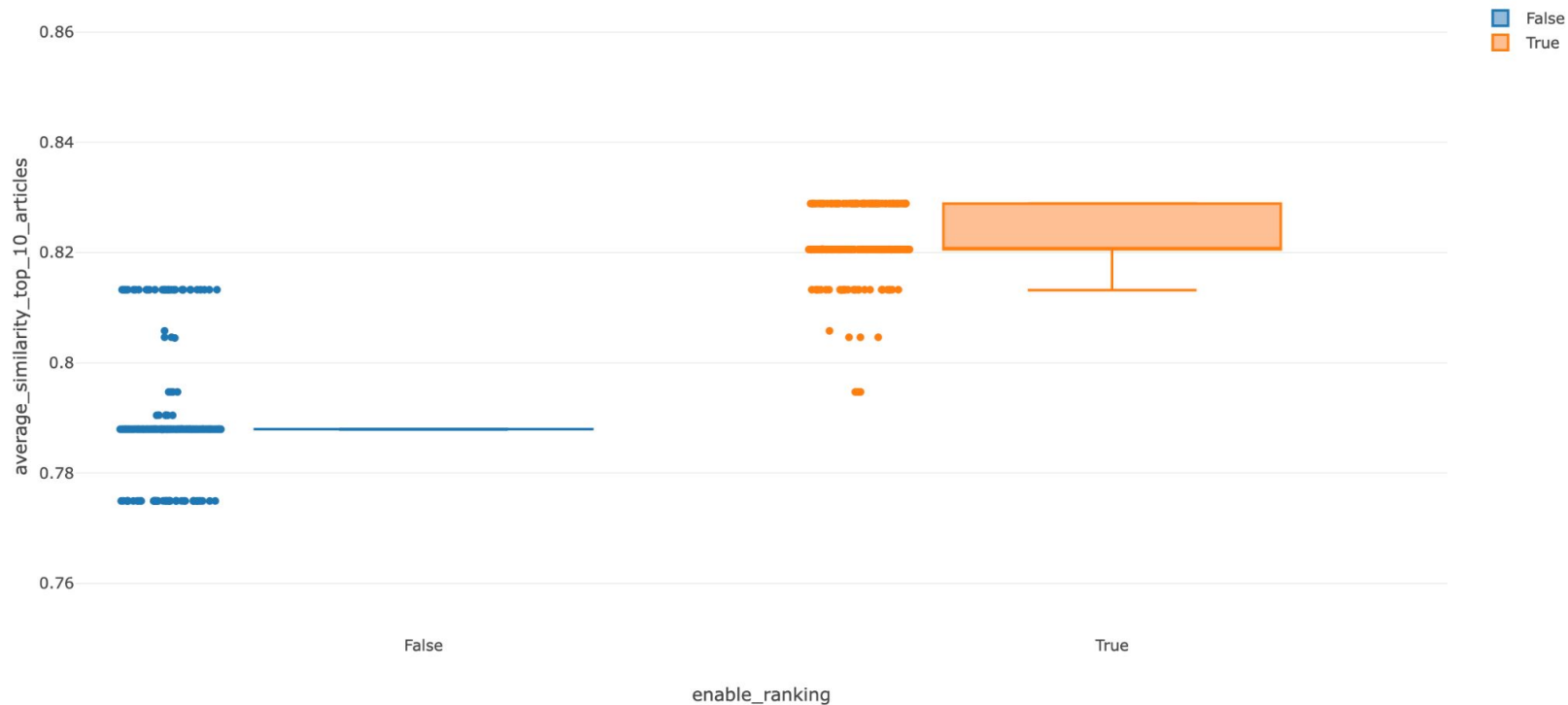
Score vs Model



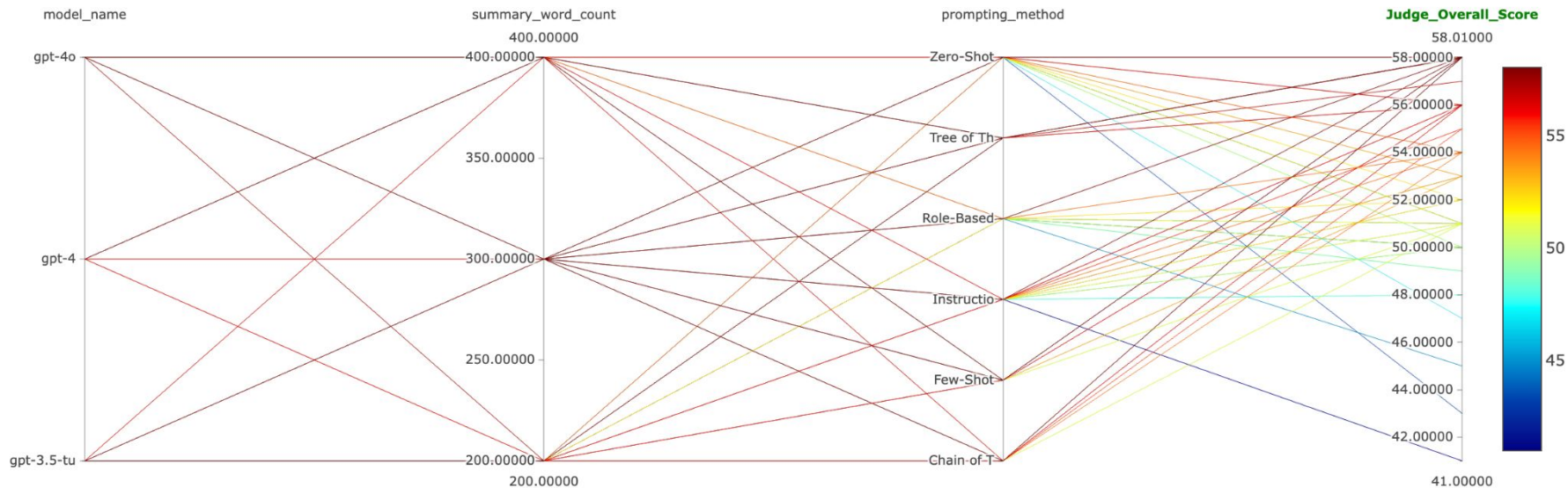
Cost vs Model



Similarity vs Ranking



Similarity vs Ranking





Supporting (Strategy)

Is it Doable?

S

Specific



M

Measurable



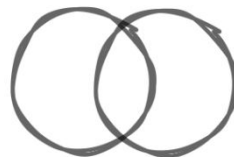
A

Achievable



R

Relevant



T

Time-Based



Tech Stack

Data:

PubTator³

Recommend
Papers.xyz

Cloud:



LLMs:



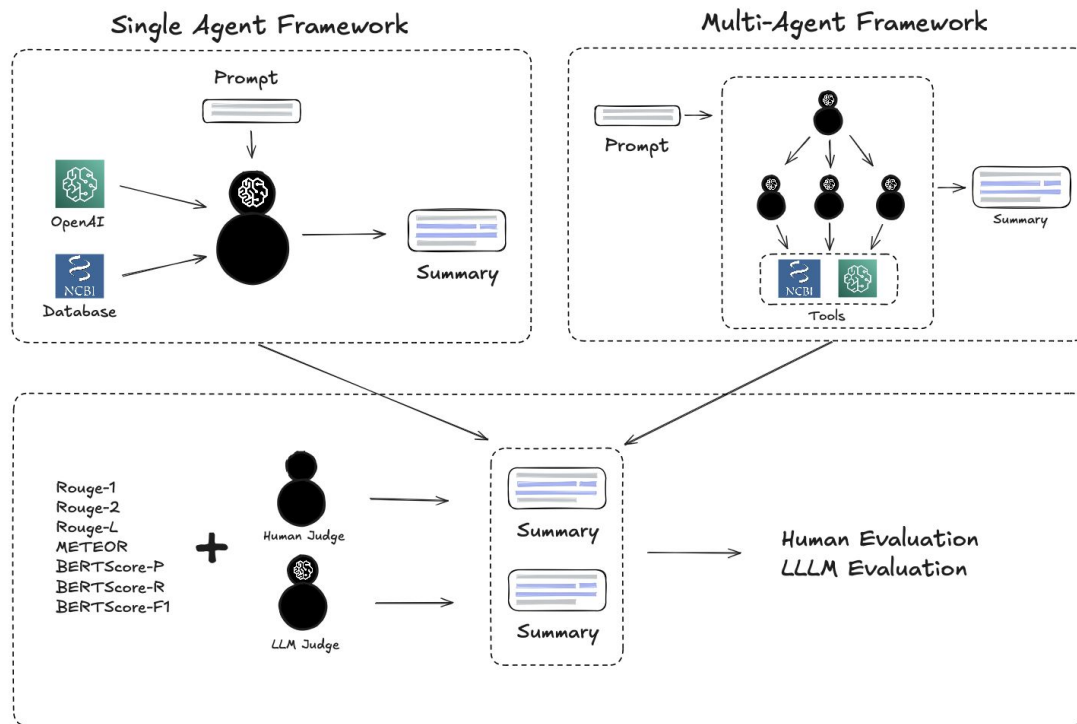
Hugging Face

Frameworks:



LangChain

Agent-Based Architecture



Risks and Challenges

Database access with appropriate metadata	OpenAI Dependency
API Limitations	Lack of Feedback from Users
Summarization Accuracy	Data Privacy

Resolved

Owned

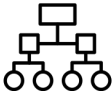



Accepted

Mitigated



Product

SciRAD Features

	Query	Retrieval	Summary	Delivery
Problem Statement	Insufficient keywords queries can result in misaligned search results	Limit large search results to a smaller subset of most relevant papers	A summary of all abstracts would make staying up to date much easier	Models are great, but only when they can be feasibly productionalized
Proposed Solution	 Keyword Recommender	 Cosine Similarity	 GPT-Based Summarizer	 Delivery
Resulting Feature	Recommends the top keywords for the user to pick from	Limits the scope of articles to the most relevant ones	Summarizes many abstracts to give the user a summary	Delivers to the user a summary of content for a topic

Sequence Diagram

