



**AI-DRIVEN
MULTI-MODAL
RETRIEVAL-AUGMENTED GENERATION
FOR ADVANCED PATENT ANALYTICS**

<https://synapsepatents.com>



SIVAKUMAR
THIYAGARAJAN

Data Scientist
Boeing



ADITYA
MENGANI

Data Engineer
Amazon



VARUN
NAIDU

Software Engineer
Boeing



DR. MEIR
MARMOR

Orthopedic Surgeon
UCSF

MISSION

ACCELERATE GLOBAL INNOVATION WITH
ADVANCED PATENT ANALYTICS


TO DELIVER ACTIONABLE PATENT INSIGHTS, EMPOWER SMARTER
DECISIONS AND DRIVE INTELLECTUAL PROPERTY ADVANCEMENT





IS MY INVENTION

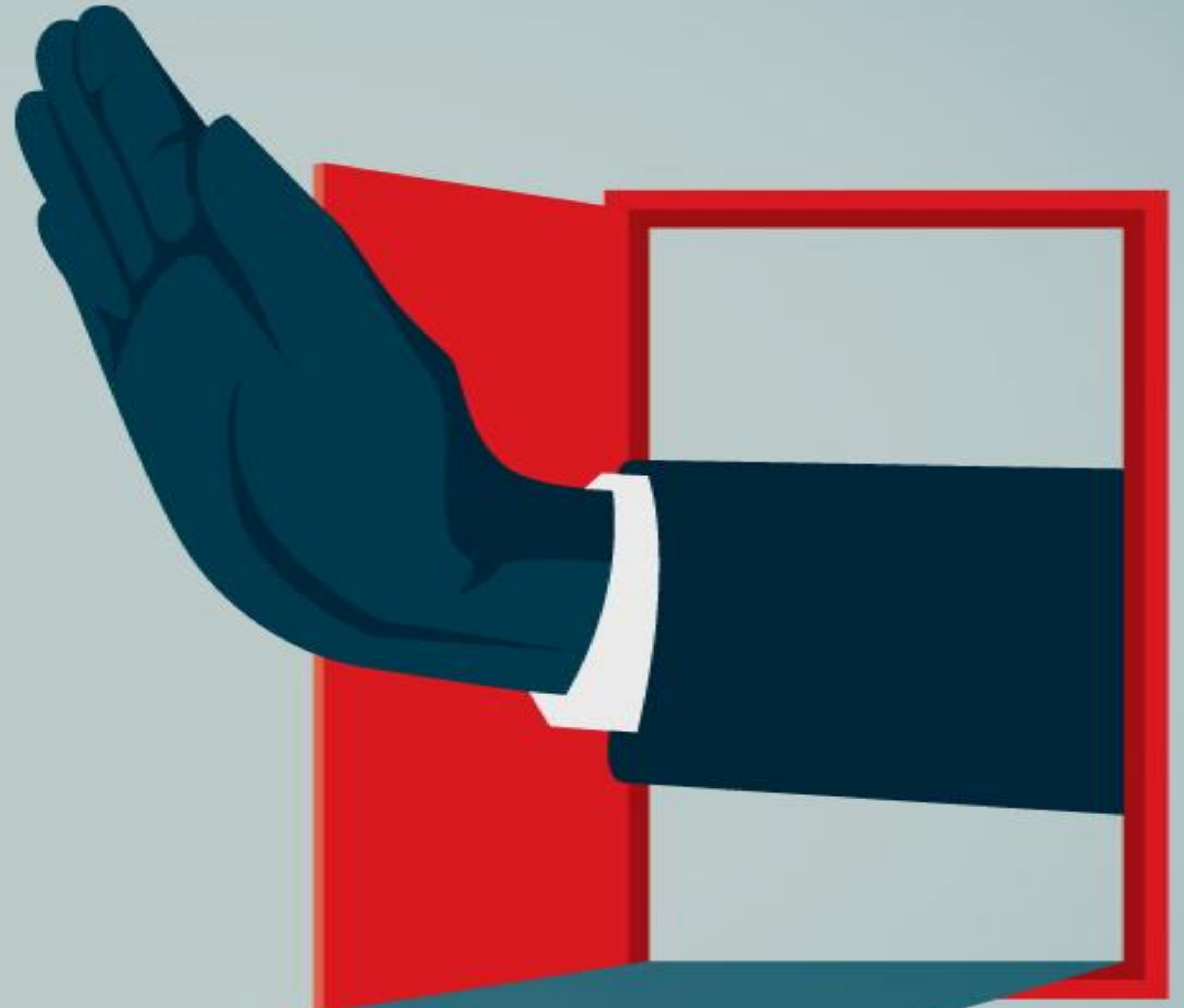
NOVEL?

Growing Complexity and Technicality, and Lengthy Documents are leading to Patent Rejections

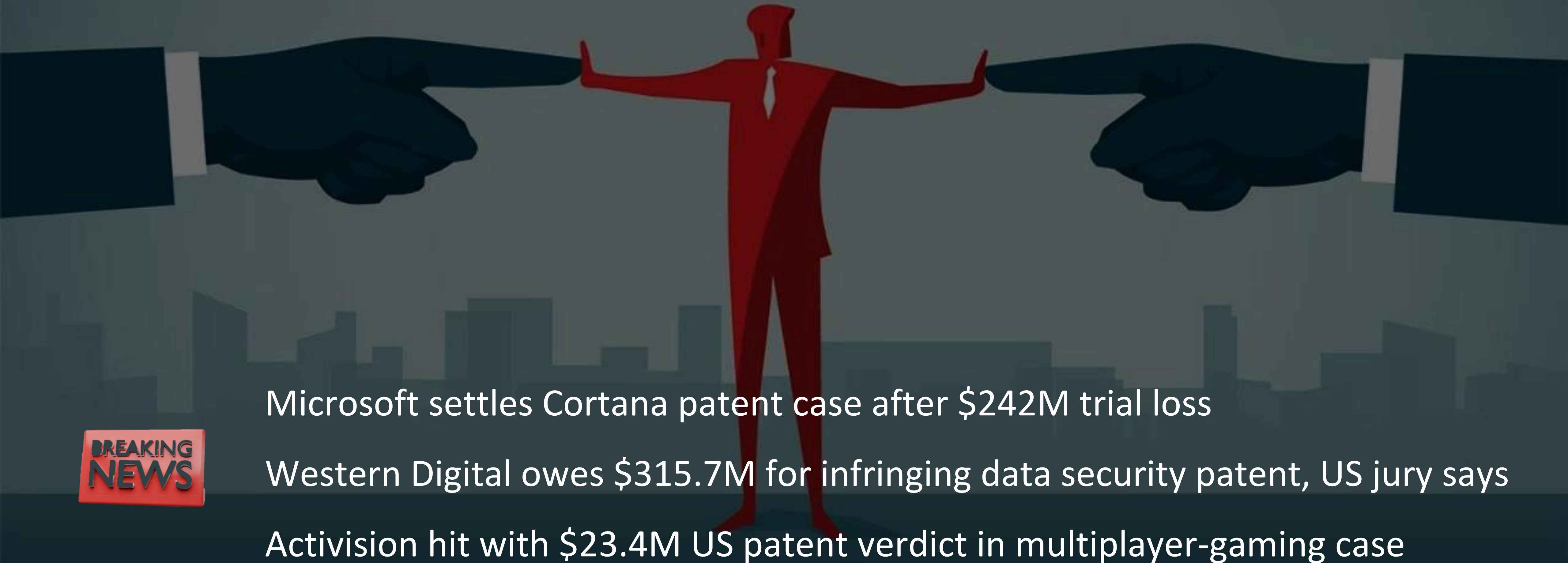
1  **1** — Not Patentable

1  **2** — Not Novel

1  **3** — Combination
of Prior Art



PATENT INFRINGEMENT CASES



Microsoft settles Cortana patent case after \$242M trial loss

Western Digital owes \$315.7M for infringing data security patent, US jury says

Activision hit with \$23.4M US patent verdict in multiplayer-gaming case

**BREAKING
NEWS**



RELATED SOLUTIONS

ONLINE TOOLS

LENS.ORG

 Patentfield

uspto

PATENT SEARCH FIRMS



FREELANCE PROFESSIONALS



ATTORNEYS & LAW FIRMS



KEY PROBLEMS IDENTIFIED BY TARGET USERS

Search through
a plethora of
patents

Identify
information in
images and
equations

Identify 101,
102 and 103
rejection risks



OUR SOLUTION

The SYNAPSE PATENTS ChatBOT (MVP)

historical SEARCH

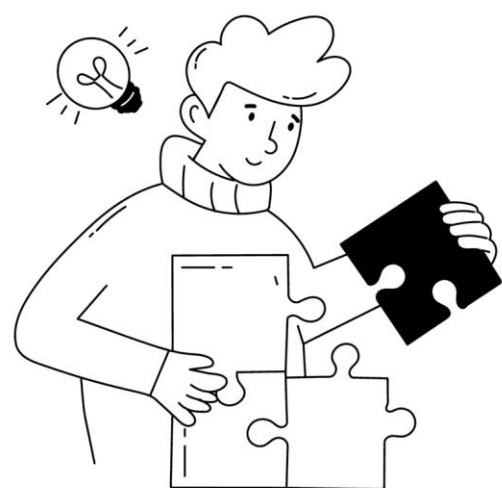
Tabulated list of matching patents

focused SEARCH

Up to 250 words summarization

patentability SEARCH

Rejection risk identification



EXPLORATORY DATA ANALYSIS

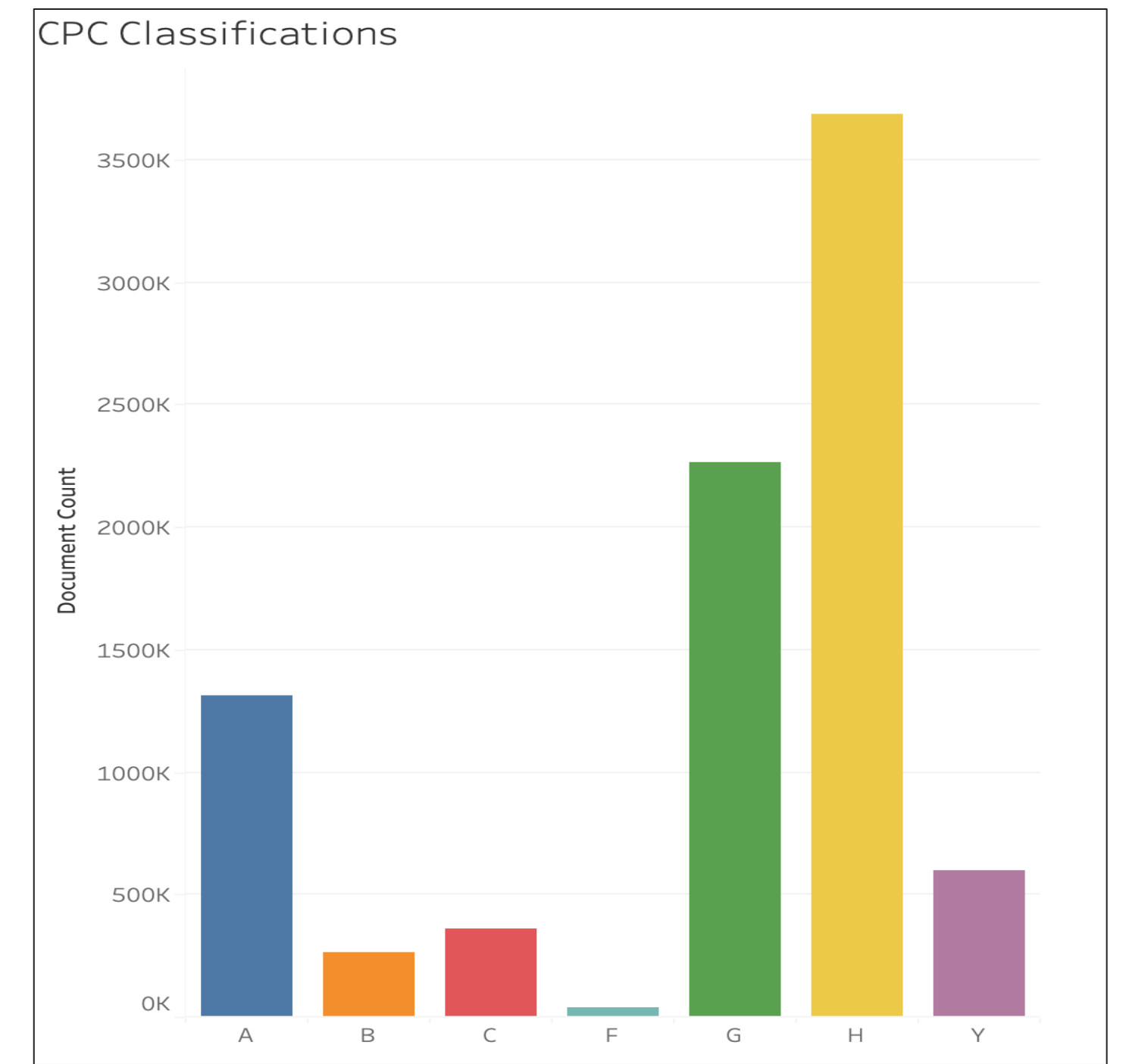


CPC - Cooperative Patent Classification

CPC Classification Codes

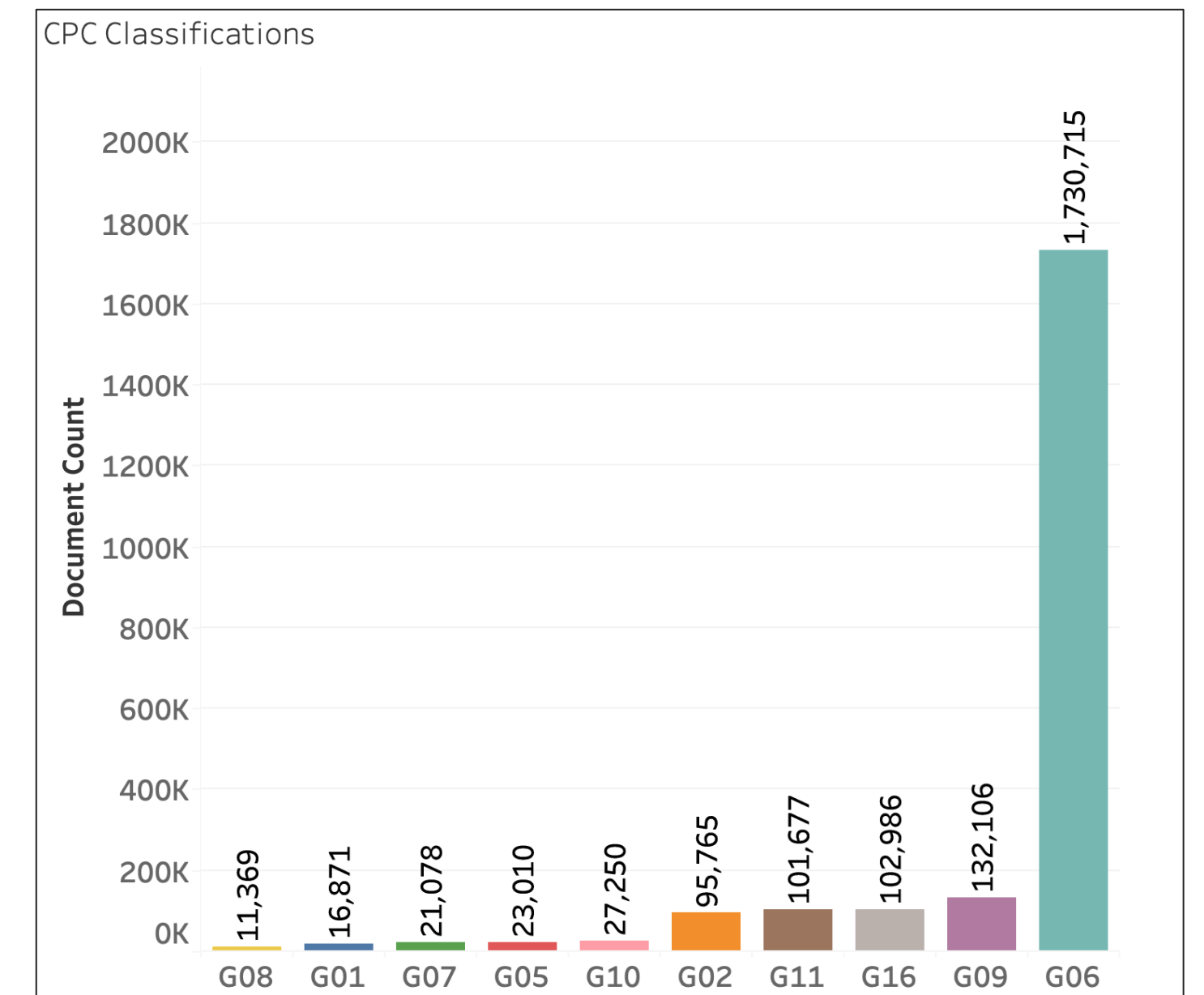
- A - Necessities
- B - Operations
- C - Chemistry
- D - Textiles
- E - Constructions
- F - Mechanical
- G - Physics
- H - Electricity

- Most common G
- G06 - Computing
 - G09 - Display
 - G16 - Information
 - G02 - Optics

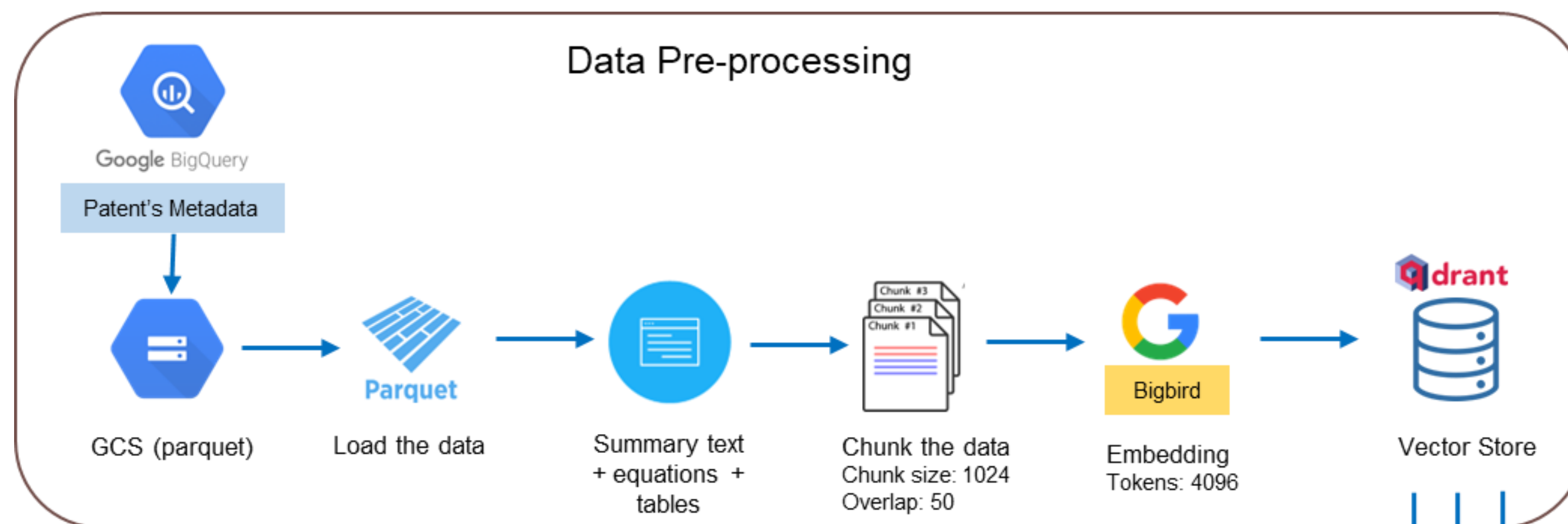


Key Takeaways:

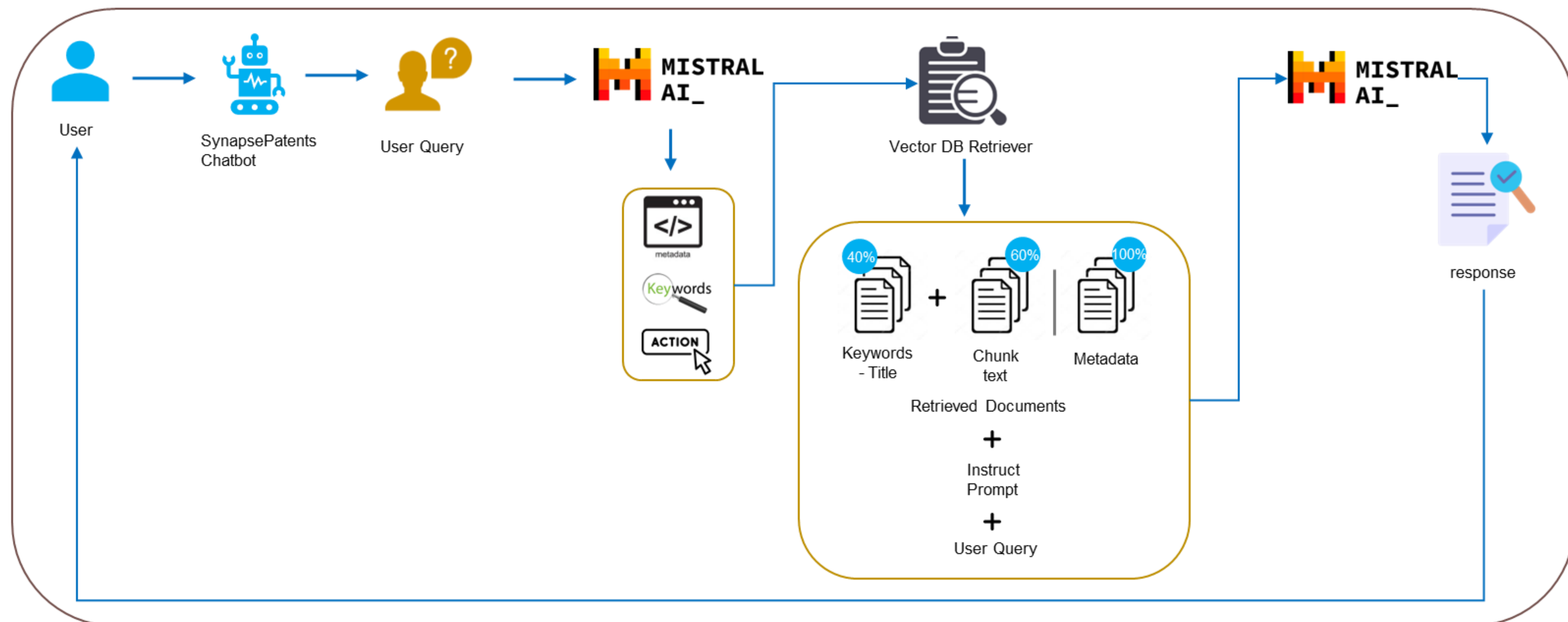
- Data source: Google Big Query Patents
- Last 10 years patents data too large for available resources (~1.2 TB)
- Focus on CPC G06 classification: Jan 1 2022 - Dec 31 2023 (~170K patents published in the USA)



Retrieval Augmented Generation



Cosine Similarity Ranking



EVALUATION DATASET

150 patents were chosen from the dataset at random

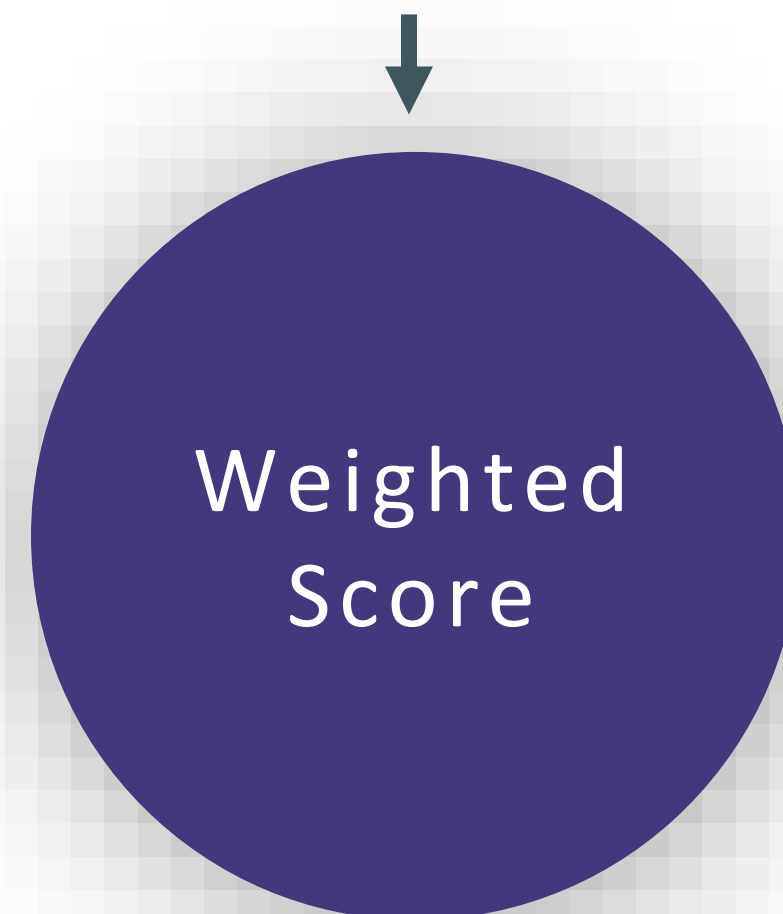
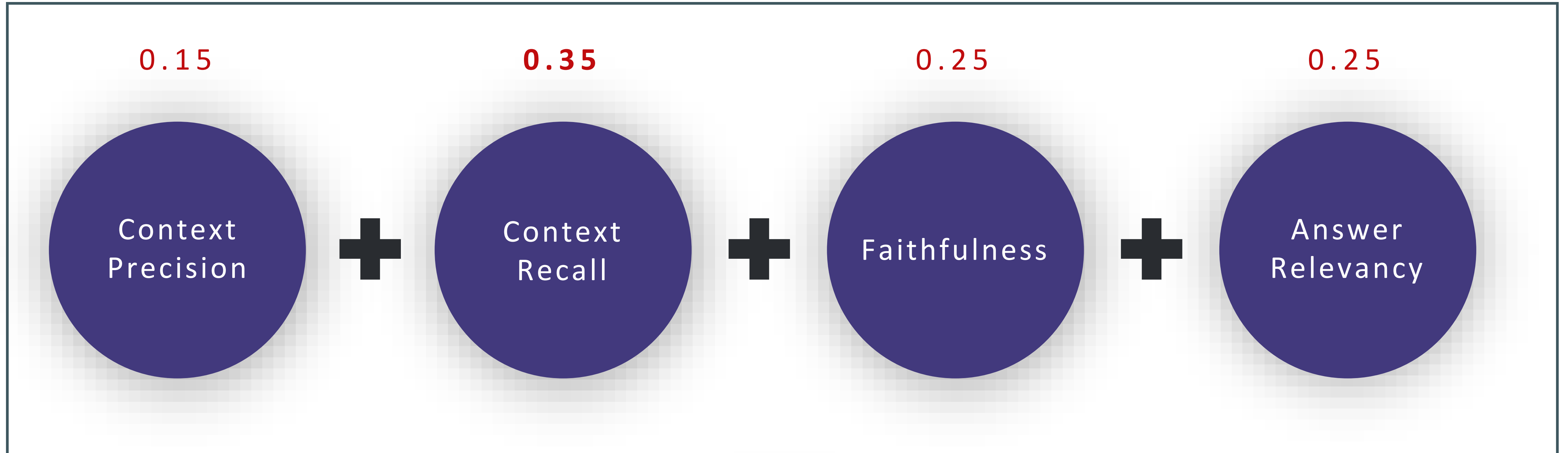
Keywords were generated from the Patent Abstract using ChatGPT

Keywords were passed to LENS.ORG with filters -

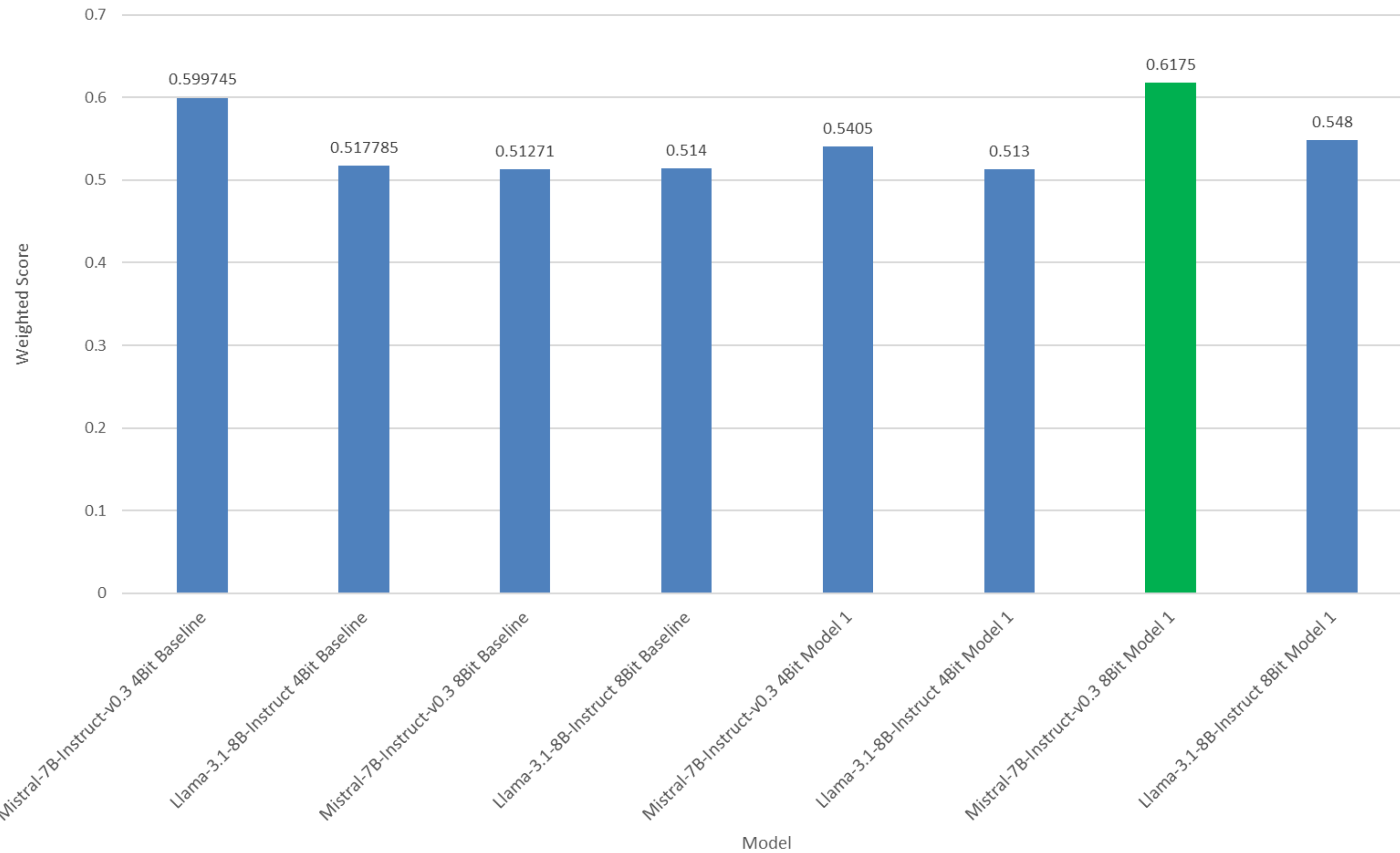
- CPC: *G06*
- Grant Date: 1st Jan 2022 to 31st Dec 2023
- US Patent Office
- Publication numbers of the results were saved



EVALUATION METRICS (RAGAs)



Weighted Metrics of Mistral & Llama 4Bit & 8Bit Models

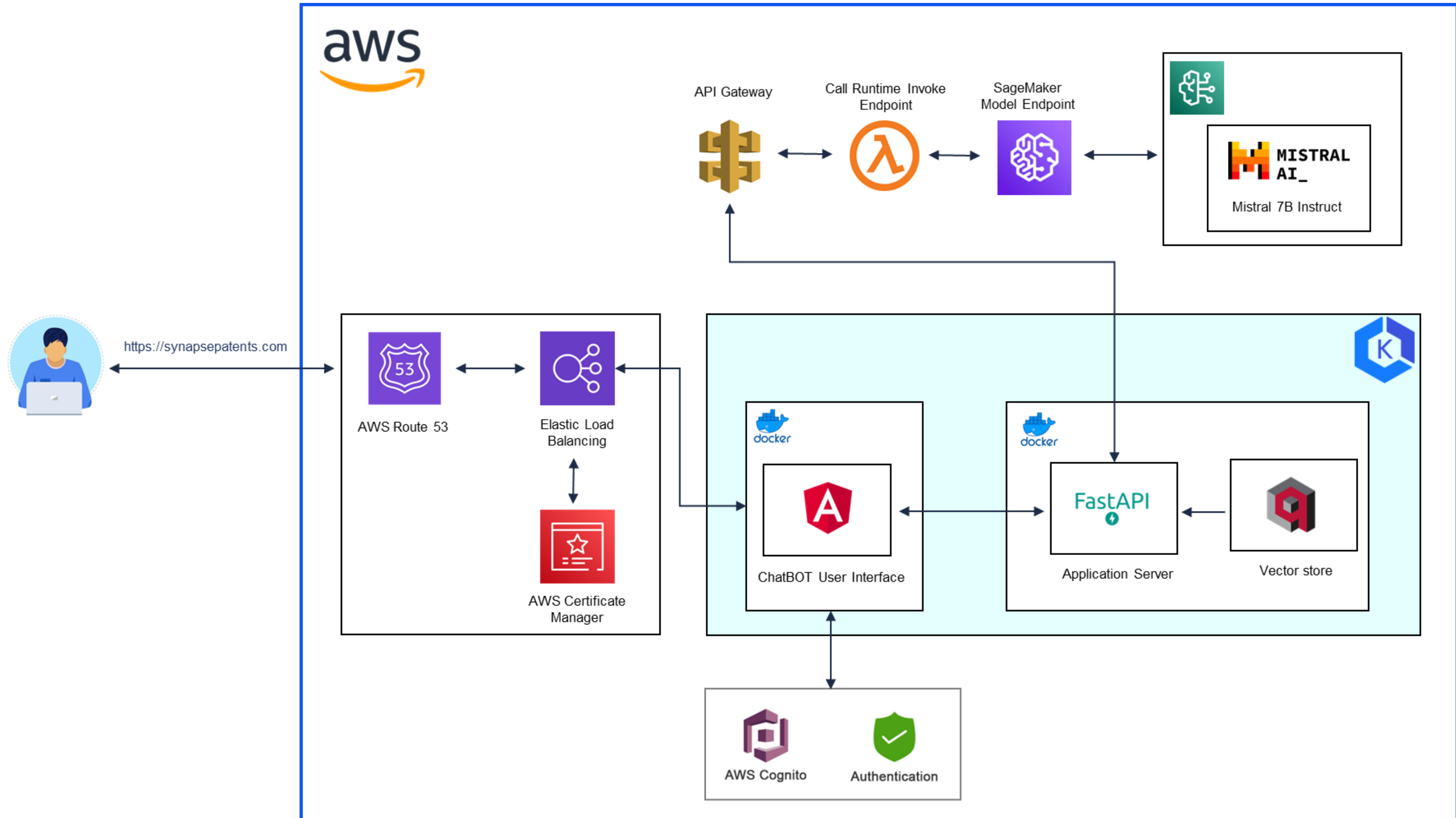


Model Card:

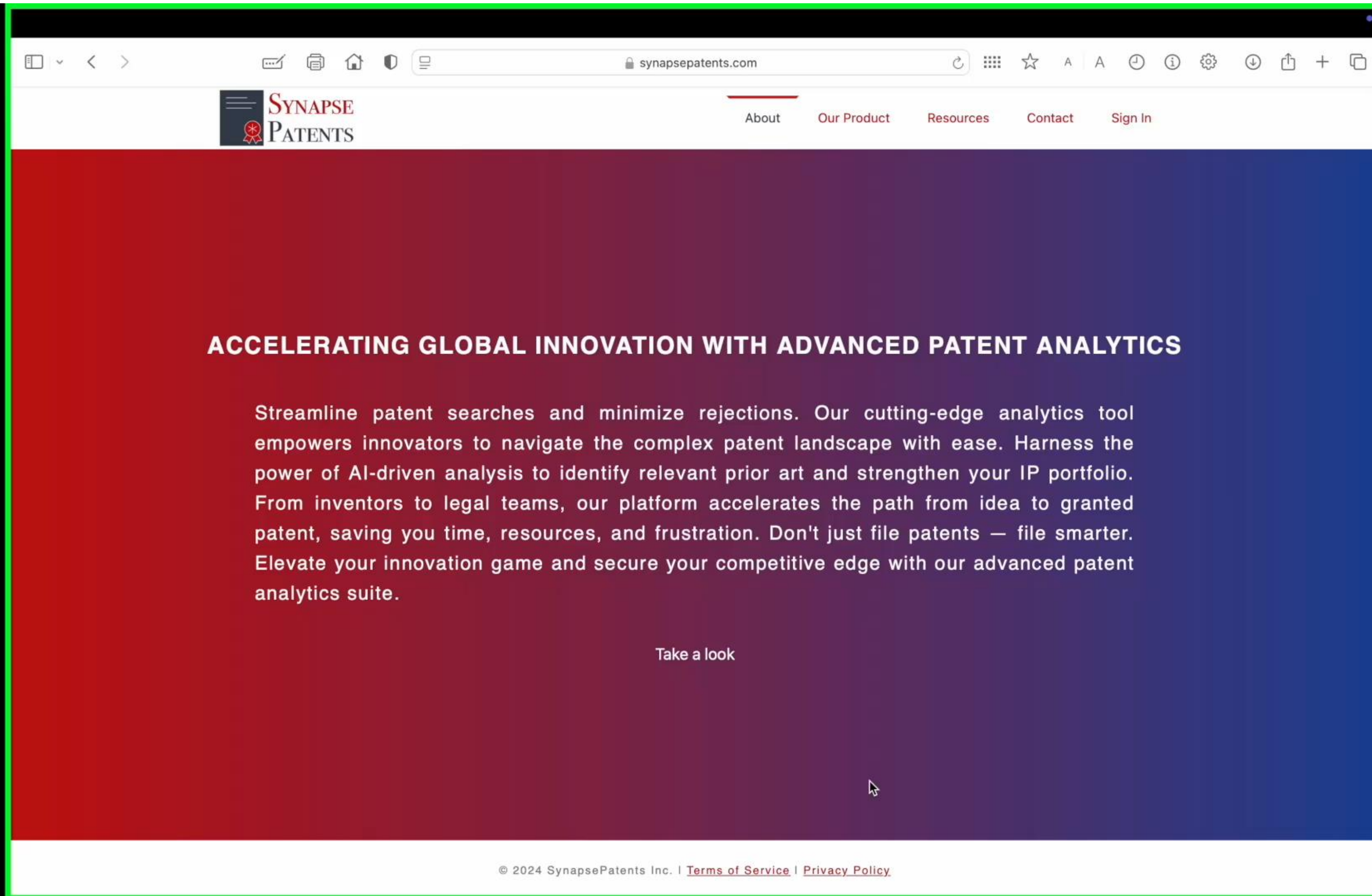
- LLM Model: Mistral-7B-Instruct-v0.3
- Quantization: 8 Bit
- Embedding Model: google/bigbird-roberta-base
- Chunk Size: 1024
- Chunk Overlap: 50
- Vector Store: QDrant
- Vector Store similarity search: cosine (Top 10)

- Mistral consistently follows prompt instructions, making it ideal for tasks requiring strict adherence to user input.
- In contrast, Llama often draws on its pre-trained memory, even when the prompt explicitly instructs it to avoid doing so.

Model Deployment Architecture







Link to Demo: <https://youtu.be/iPrQxjfehSw>

KEY TECHNICAL TAKEAWAYS

Challenges Faced

- Model Deployment in AWS
- Identifying right AWS Services / Model Inference Architecture
- AWS Services Costs / Compute Constraints

Learnings

- Patent IP search processes
- Generative AI techniques
- Hosting web application in AWS with domain

CONCLUSION

Despite challenges, this MVP of our AI-powered RAG ChatBOT has demonstrated strong potential in transforming patent searches, making them more accurate.

As the patent landscape continues to grow larger and more complex, we are confident that our product will empower organizations and individuals to stay ahead in the competitive world of innovation.

We aim to take this MVP to the next level through seed funding. Additionally, we plan to bring a subscription model for our business.

ACKNOWLEDGEMENTS

Capstone Instructors:

Kira Wetzel

Puya Vahabi

Subject Matter Experts:

Neal Cohen, Patent Attorney, UCSF, CA

Jay Chesavage, Freelance Patent Search
Professional, CA

REFERENCES

- Chen, W., Hu, H., Chen, X., Verga, P., & Cohen, W. W. (2022). MuRAG: Multimodal retrieval-augmented generator for open question answering over images and text. <https://aclanthology.org/2022.emnlp-main.375.pdf>
- Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, et. al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", Facebook AI Research; University College London; New York University. <https://arxiv.org/pdf/2005.11401>
- Damian Gil. "Advanced Retriever Techniques to Improve Your RAGs". <https://towardsdatascience.com/advanced-retriever-techniques-to-improve-your-rags-1fac2b86dd61>
- Llama Team, AI @ Meta. The Llama 3 Herd of Models. <https://arxiv.org/pdf/2407.21783>
- Jiang, H., Xu, C., Liang, P., & Xiao, C. (2023). Mistral 7B: A 7-billion-parameter language model for efficient and high-performance natural language processing. arXiv preprint arXiv:2310.06825. <https://arxiv.org/pdf/2310.06825>
- Adithya S K. (2023, October 9). Deploy Mistral/Llama 7b on AWS in 10 mins. Medium. <https://adithyask.medium.com/deploy-mistral-llama-7b-on-aws-in-10-mins-cc80e88d13f2>

THANK YOU

QUESTIONS?