

# LAMA (LLM-Powered Autonomous Marketing Agent)

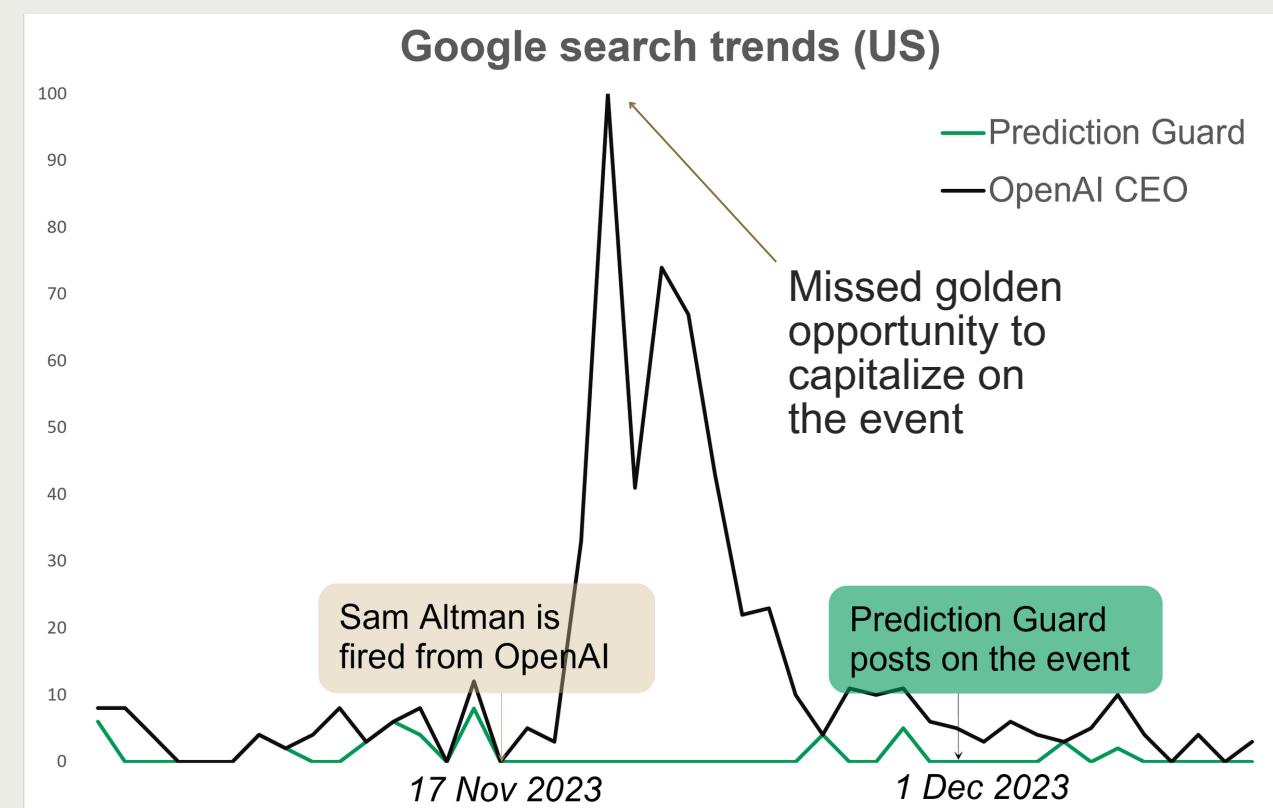


Abithaa Shree Venkatesh, Ann Mathew, Monika Madugula, Nagraj Shridhar Deshmukh, Pratik Mahesh Merchant, Matthew A. Lanham  
 venka155@purdue.edu; mathe121@purdue.edu; mmadugul@purdue.edu; deshmu37@purdue.edu; merchap@purdue.edu; lanhamm@purdue.edu  
 Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

## BUSINESS PROBLEM FRAMING

“70% of Chief Marketing Officers (CMOs) are already experimenting with Generative AI, and they say “inaction is not an option”.”  
 - BCG, June 2023

Prediction Guard faces the challenge of maximizing limited marketing resources amidst the rapidly changing landscape of digital trends, resulting in missed opportunities to capitalize on critical trends due to their short duration. To address this, Prediction Guard is seeking to develop an automated LLM-based solution that can enhance its lead generation and marketing efforts, while also serving as a template for a scalable product alongside its current offerings.



? How can small businesses use limited marketing resources to keep up with latest news?

? Can Gen AI be used to optimize content creation, personalize marketing strategies and generate relevant brand stories?

Our solution, LAMA, can benefit small businesses, medium to large businesses in need of marketing support, and individuals looking to elevate their social media presence.



## ANALYTICS PROBLEM FRAMING

The goal is to develop a novel solution that leverages some of these capabilities of Gen AI to design an autonomous LLM-powered agent that identifies trends, generates aligned brand narratives, and creates posts for social media platforms in real-time.

It is further justified by an LLM's ability to generate human-like and customized text efficiently, enhancing brand engagement and personalization at scale.



### Assumptions

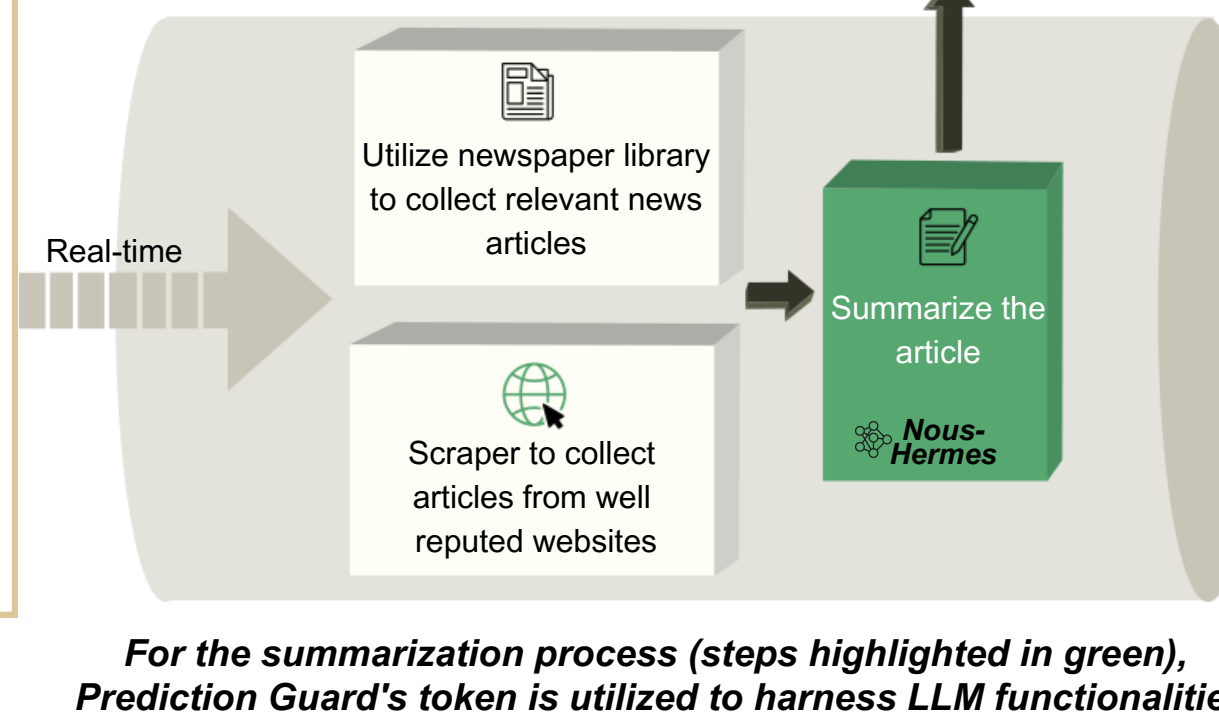
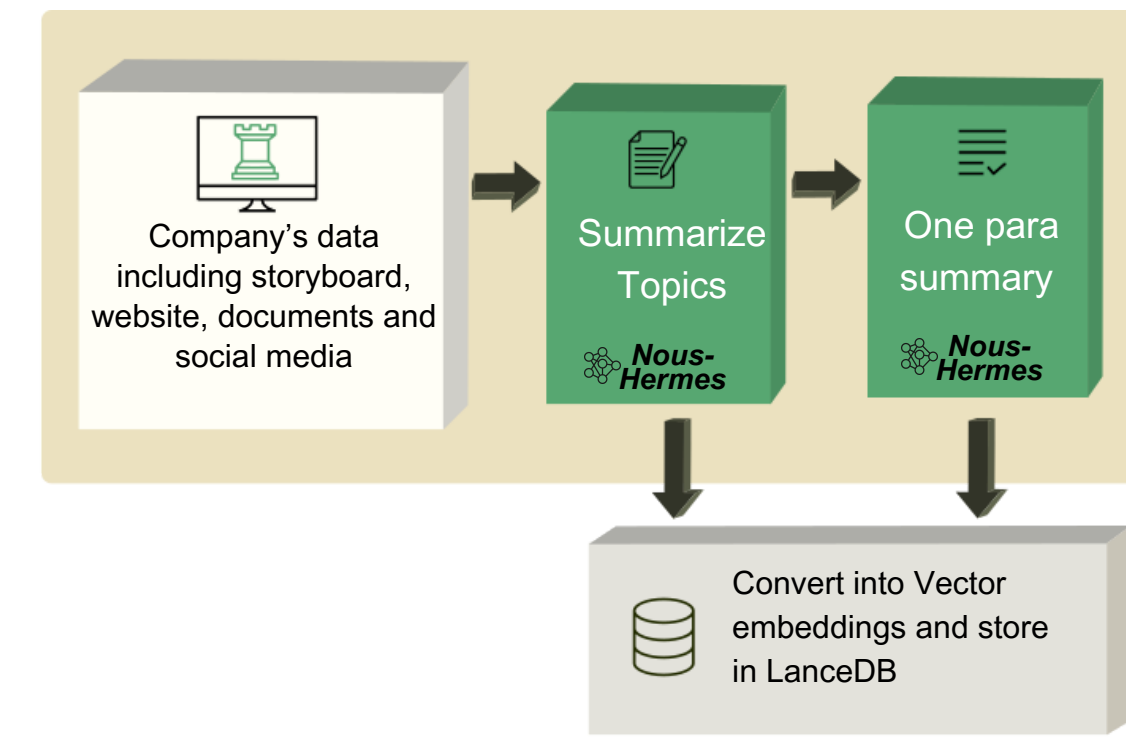
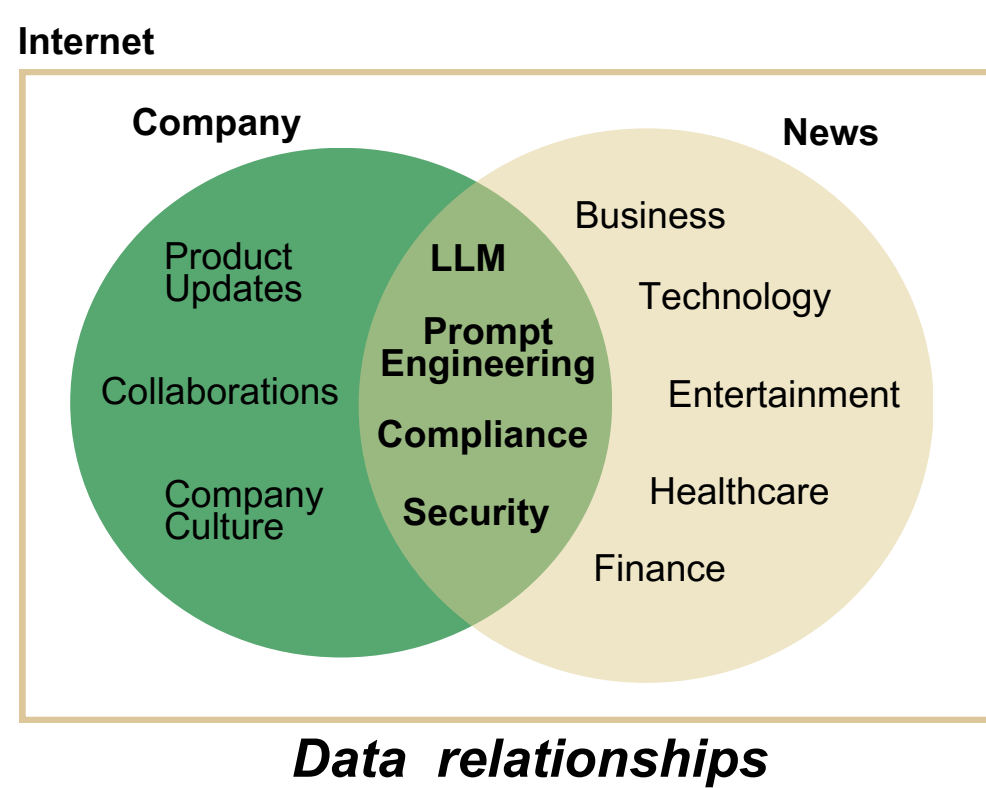
- 1 Every article holds equal significance
- 2 Sources provide complete AI news coverage
- 3 Data from these sources is reliable and factual

### Success Metrics

- 1 40% increase in post frequency
- 2 25% higher engagement on social media posts
- 3 Semantic alignment between agent's posts and company's established voice

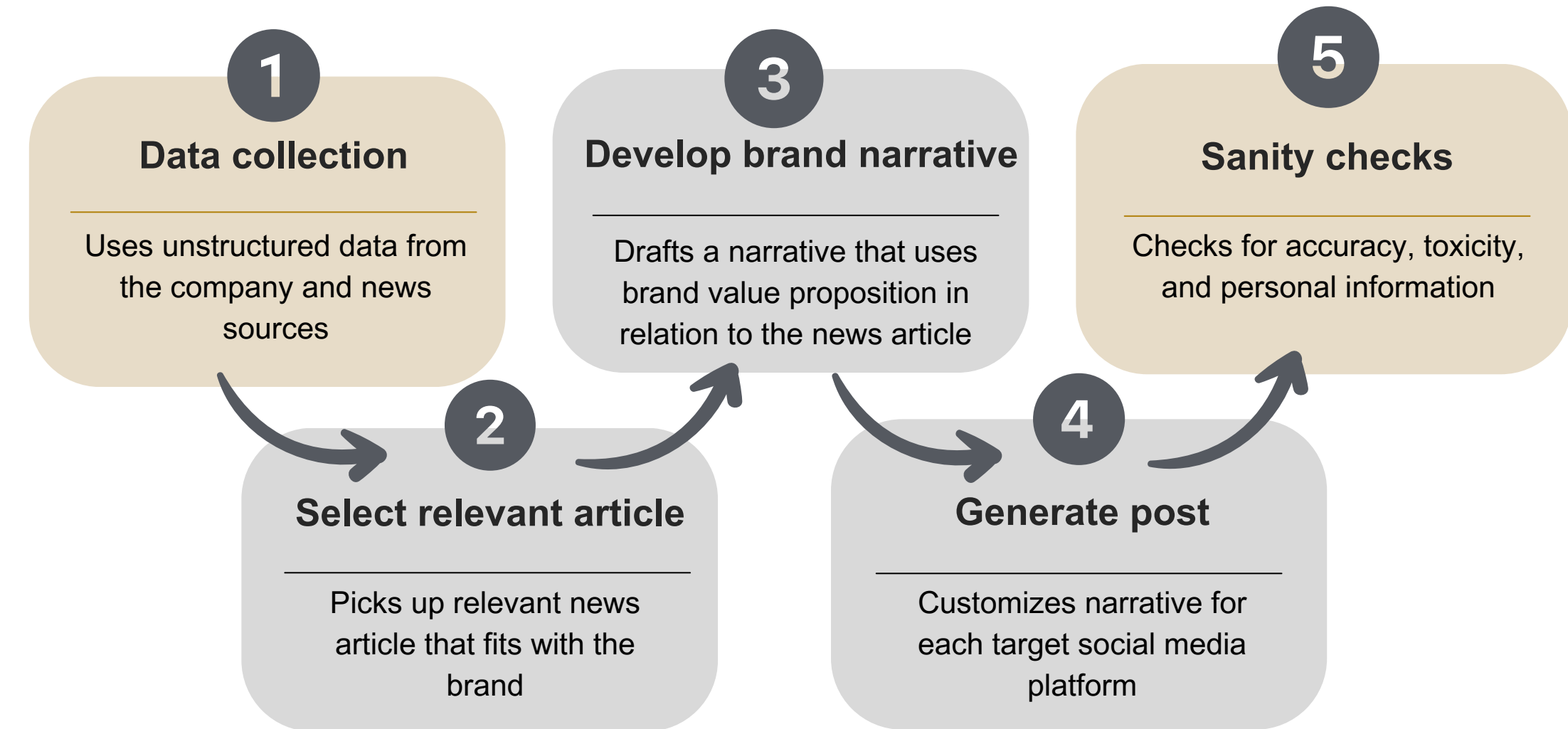
## DATA

Our project utilizes unstructured text data, including information about the company. To identify relevant trends, it is essential to continuously monitor articles from news outlets which have been prioritized based on their factual accuracy, popularity, and technical feasibility. Once data is extracted, text is summarized, transformed into embeddings and stored in a vector database in the subsequent preprocessing stage.

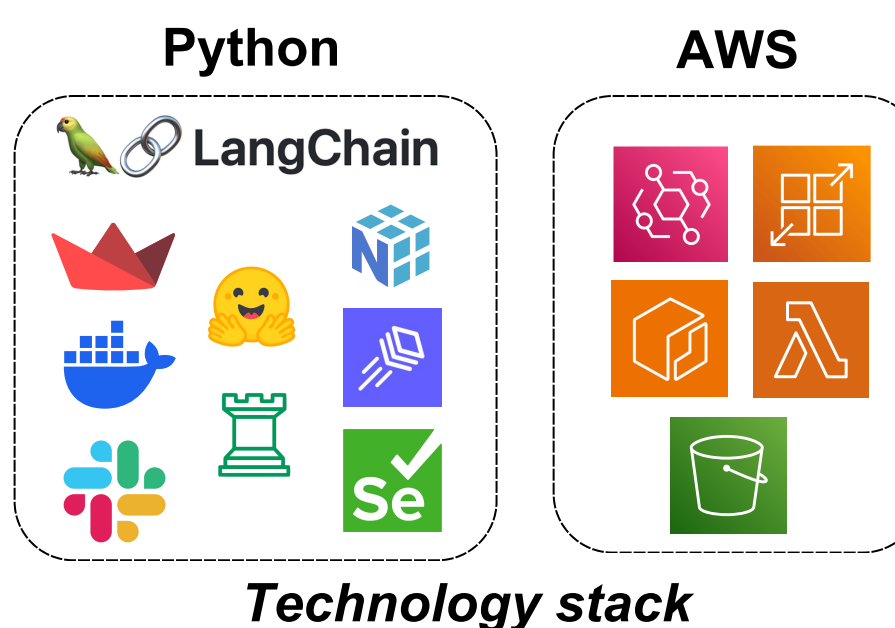
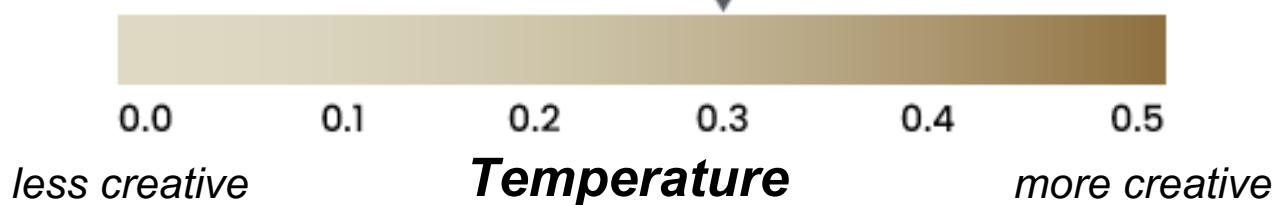


## METHODOLOGY

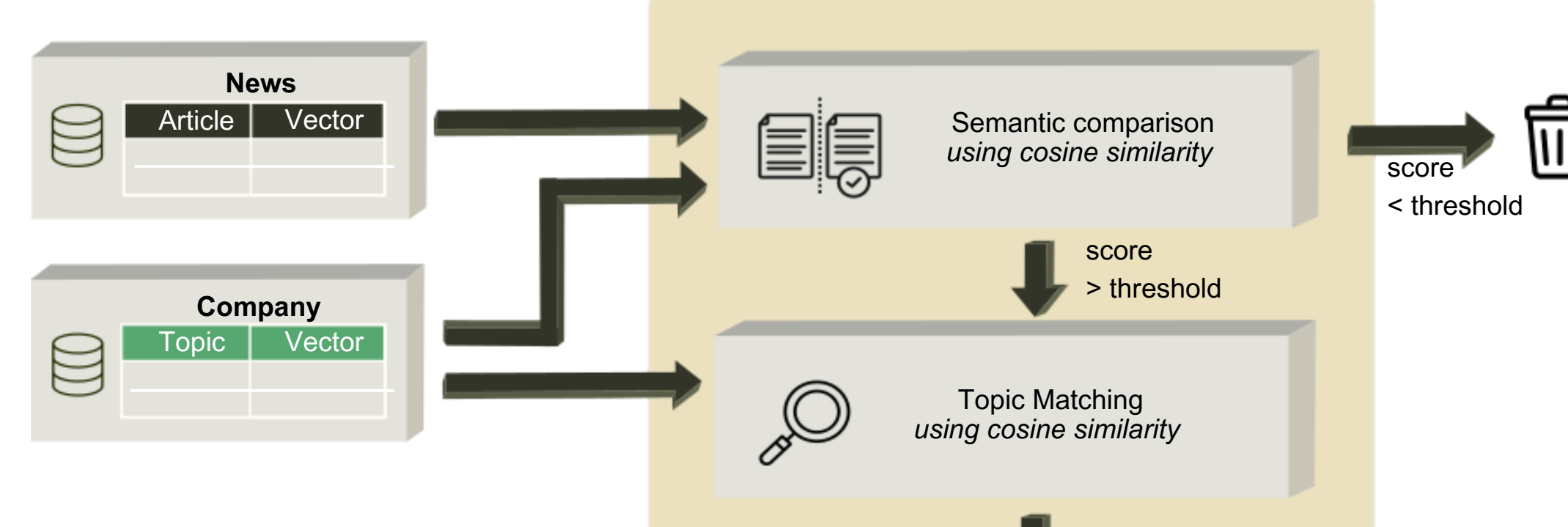
Our strategy adopts the Retrieval-Augmented Generation (RAG) framework as opposed to developing an LLM model from scratch or fine-tuning an existing one. This decision is primarily motivated by the desire to mitigate substantial computational resources required for model training and processing.



A test application was developed for experimenting with prompts, through which related parameters like temperature, maximum token count, and LLM models were finalized to determine the optimal conditions for accuracy and reliability.



## MODEL BUILDING



Vector embeddings stored in LanceDB allow us to compute cosine similarity and semantically compare news summaries with the company summary. If similarity score is above the threshold, we do a second level of in-depth semantic comparison to identify the top 5 topics related to the company that appear in the news article.

For content generation we employ role-prompting with the LLM assuming the role of a marketing expert, and provide the company summary and topic summaries as context. This generated content forms the baseline of our post, then checks are applied to sanitize this narrative.

### Confusion matrix - model performance

|              | Selected | Not Selected |
|--------------|----------|--------------|
| Relevant     | 17       | 0            |
| Not relevant | 5        | 331          |

### Prototype - Semantic Comparison & Post-Gen

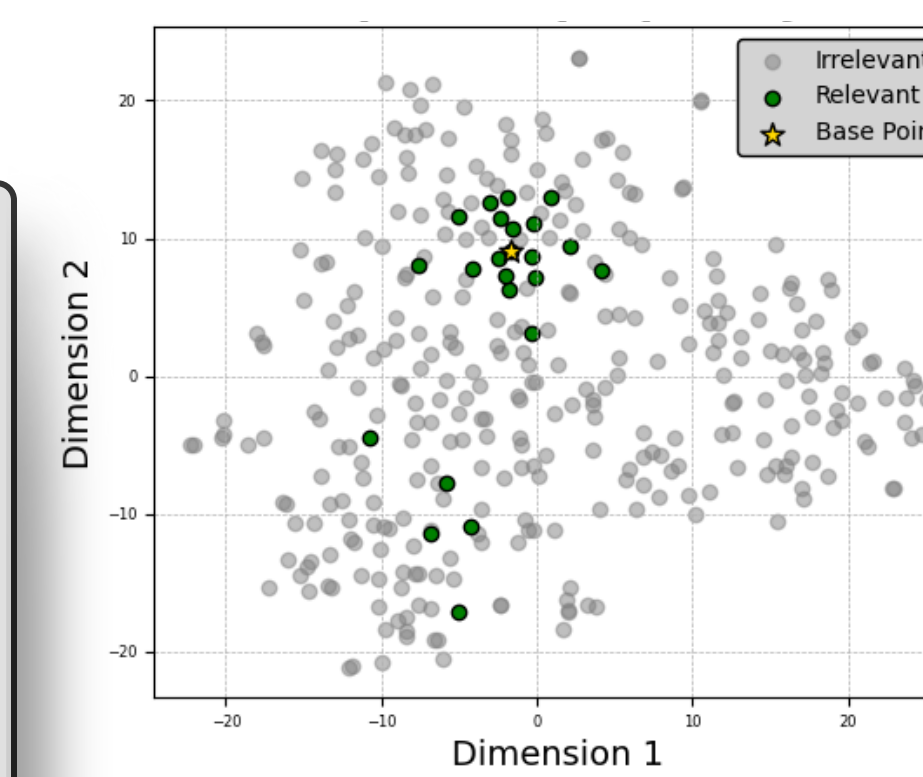
**You**  
Can you check which of the following articles is relevant to Prediction Guard:  
 1. [Generative AI's Biggest Security Flaw Is Not Easy to Fix](#)  
 2. [13 percent of AI chat bot users in the US just want to talk](#)  
 Then based on the relevant one, create a professional LinkedIn post for Prediction Guard

**LLM Agent**  
Article 1, with a similarity score of 0.6 is relevant to Prediction Guard's brand, and here is a LinkedIn post based on it:

With the recent news of indirect prompt injection attacks on AI systems, businesses are understandably concerned about the potential misuse of LLMs by hackers. At Prediction Guard, we offer features like de-risking inputs, validating outputs, and implementing private, compliant LLM systems, ensuring that your data remains secure. In light of the news, Prediction Guard is committed to providing innovative and secure approaches to enhance language models, including multi-step reasoning, chaining operations, toxicity checks, factuality checks, and more. Contact us to learn how we can help make your AI journey safe and secure. #LLM #PromptInjection

Message LLM Agent..

### TSNE - visualizing embeddings

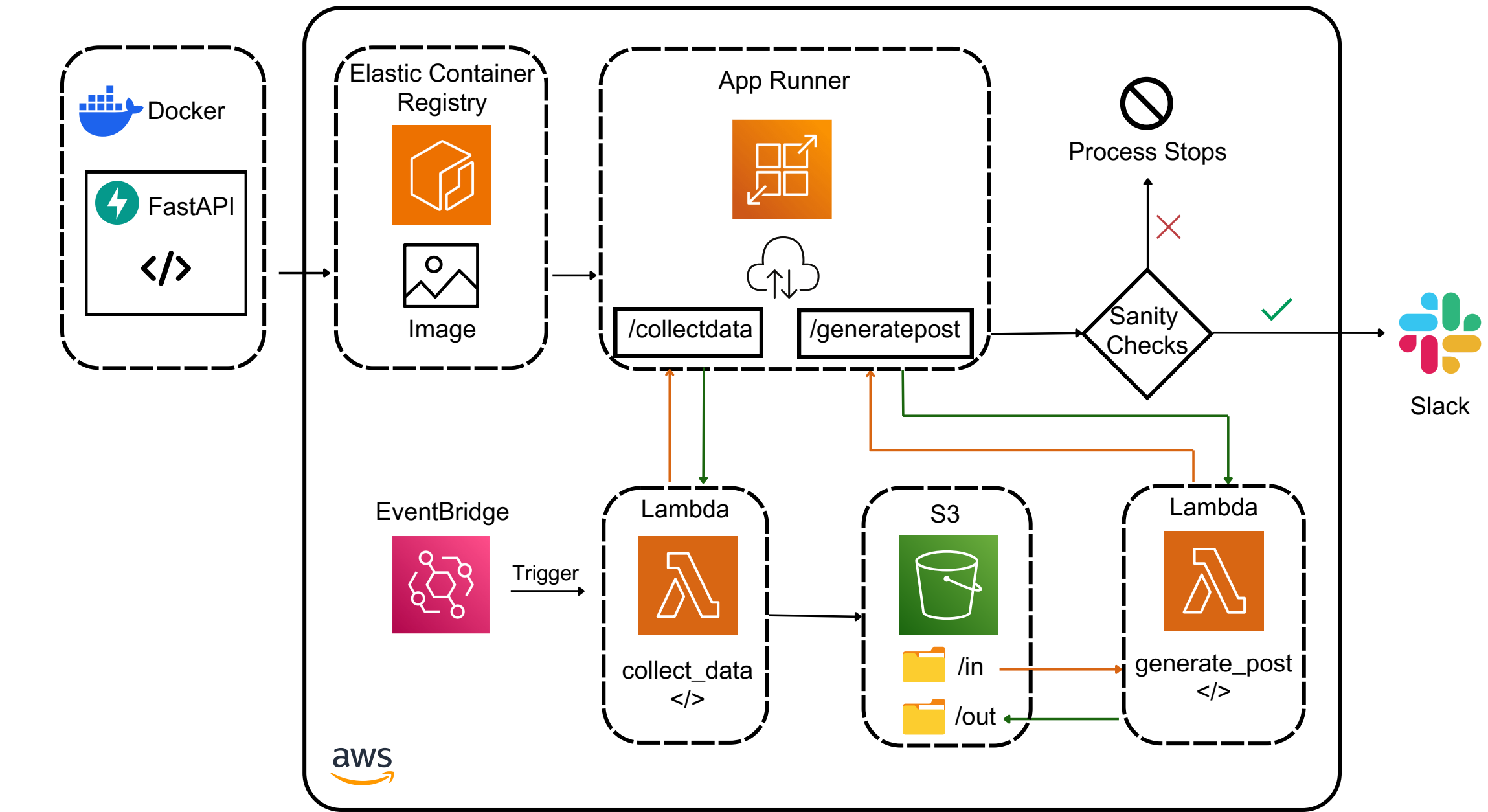


### Areas of Improvement

- 1 Chunking for lengthier articles
- 2 Few-shot learning for model updates
- 3 Assigning priorities to sources

## DEPLOYMENT AND LIFE CYCLE MANAGEMENT

To package our solution into a usable product and deploy it at scale, we propose an architecture that makes use of Docker containers and AWS services for automated data and content management.



### LAMA's impact on post frequency



### Client Testimonial

“Excellent work by the team! This innovative tool not only positions us at the forefront of latest AI developments but also opens the door to creating compelling marketing LLM applications for our customers.”

- Prediction Guard, April 2024

Moving forward, we can incorporate multimodal capabilities, such as image generation or utilizing articles with images to enhance user experience.

This solution can also be expanded to support multiple languages, which will greatly benefit non-English speakers and encourage further innovation in this field.

## ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham, Purdue University, INFORMS and our industry partner Prediction Guard for this opportunity, guidance, and support on this project.

