

Lambda News Digest

University of Washington Information School Capstone 2024

Read More

Who We Are

We are Team Lambda: Passionate about technology, dedicated to leveraging our expertise to extract value from data.



Sponsoring Organization

Glamazon, an affinity group at Amazon that helps connect, support, and empower communities of LGBTQ+ employees. Linbo Ruan is our sponsor, who is also the director of Glamazon.

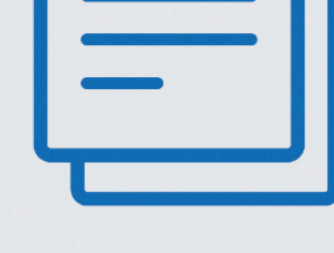


Problem Statement



Inefficient News Consumption

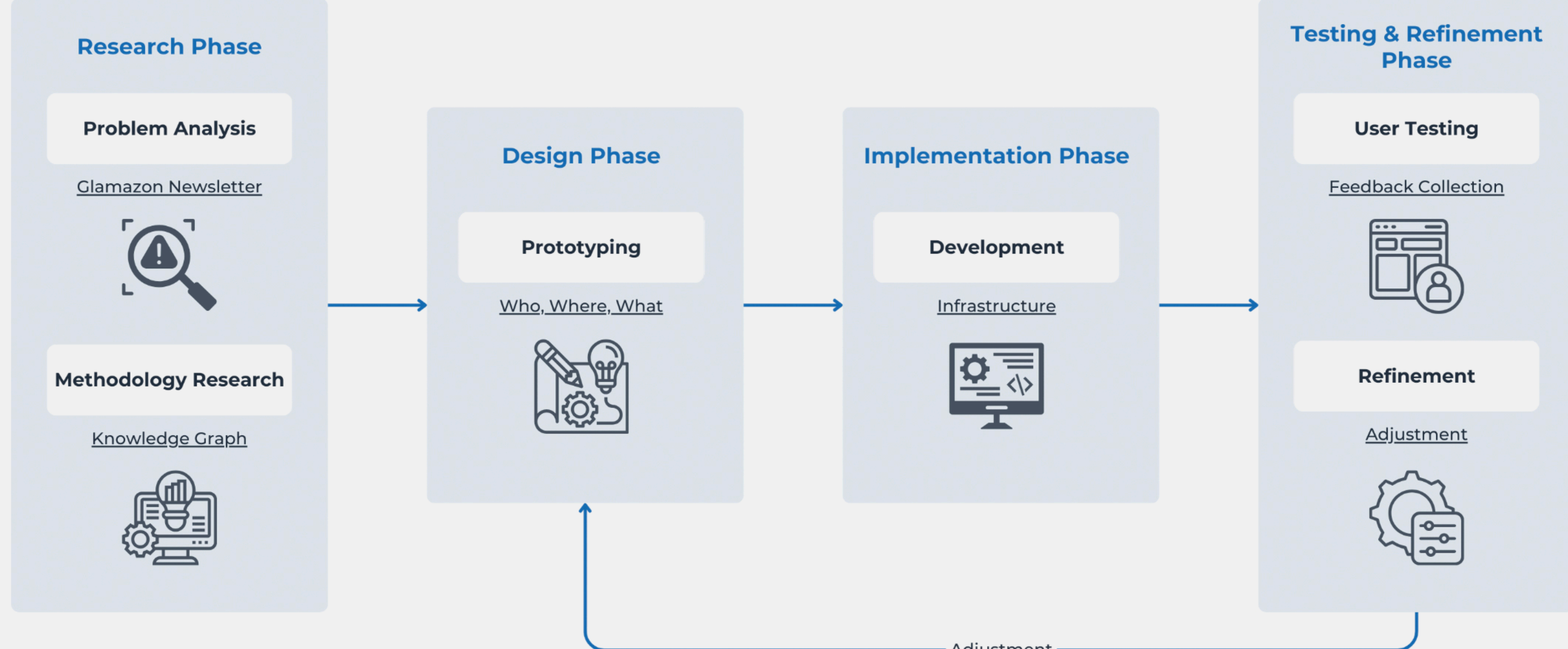
In today's fast-paced work environment, individuals struggle to stay informed due to the overwhelming volume of communications and the competitive attention economy. Moreover, LGBTQ+ related news often lacks a centralized, efficient distribution channel, leaving it scattered and hard to access. This fragmentation leads to a polarized perception of LGBTQ+ news, which tends to be categorized strictly into political or entertainment spheres.



Outdated Content Strategy

The traditional format of the Glamazon Newsletter has been criticized for its verbosity and lack of engagement, impairing its readability and overall effectiveness. Feedback underscores two primary concerns: the newsletter's excessive wordiness, which diminishes its readability, and a deficiency in news content, which compromises its informativeness. This calls for a strategic overhaul to ensure the content is both concise and richly informative, enhancing its utility and appeal to readers.

Our Solution



Research Phase



Problem Analysis

After analyzing the current Glamazon newsletter, we identified several opportunities for improvement. We observed issues related to verbosity, lack of visual elements, and the absence of updated news.



Methodology Research

To integrate news content into the newsletter, we started methodical research on various machine learning and AI technologies such as name entity recognition, topic modeling, knowledge graphs, and Large Language Models (LLMs).

Name entity recognition aids in news aggregation by extracting relevant entities. Topic modeling facilitates the classification and management of news topics for organized content delivery. Knowledge graph plays a key role in data integration and information retrieval by connecting information pieces. LLMs enable data summarization and semantic search, enhancing content comprehension. These technologies are the cornerstone of our implementation phase.

Design Phase



Positioning

Following identification of opportunities and exploration of existing technologies, we positioned our product as a Live News Digest Platform with Interactive Visualization. The overarching objective is to facilitate news retrieval and content understanding.

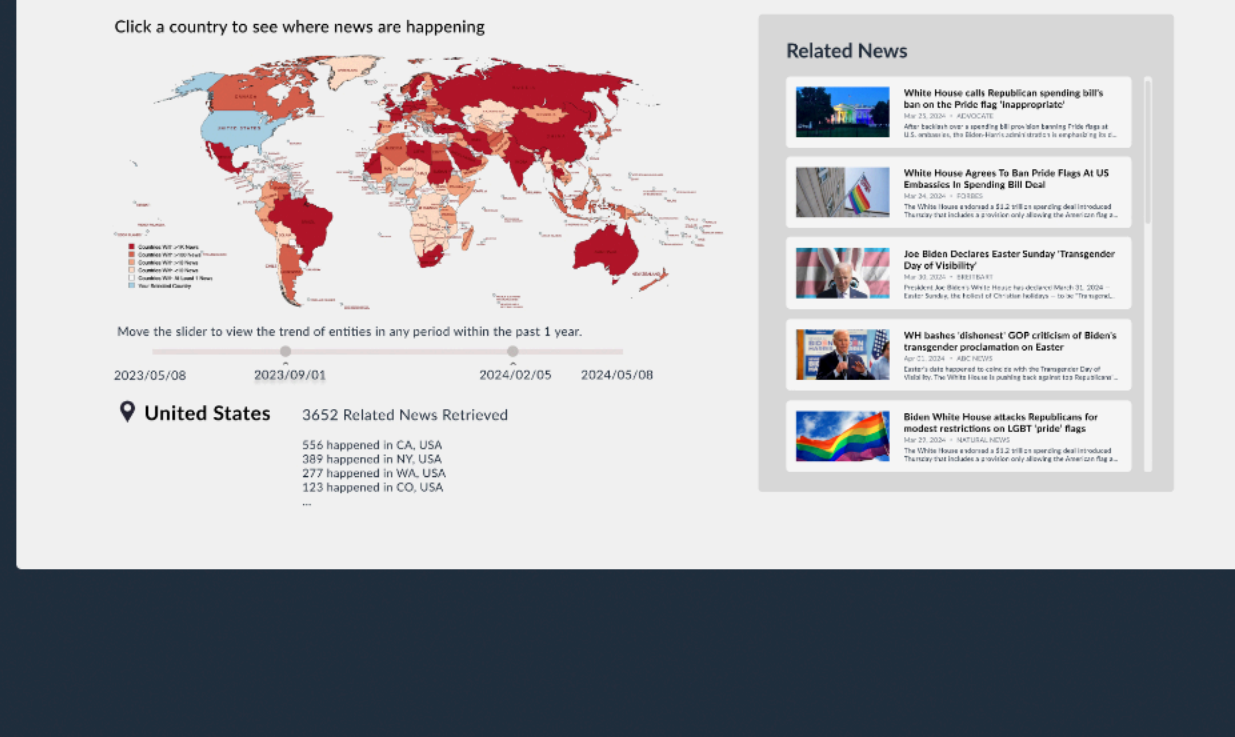


Prototype Design

Based on our positioning, we started product prototyping. Our design is concentrated on four main components, each centered around location, people, trends, and search functionality.

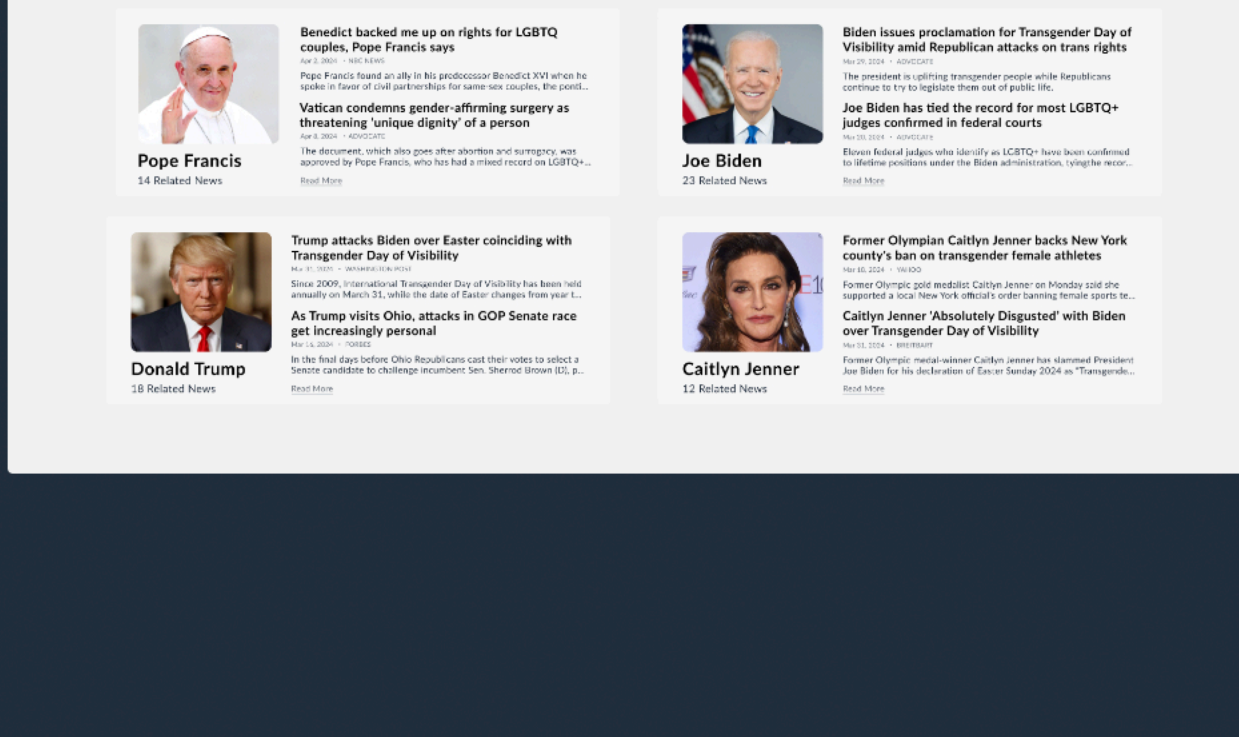
Location

The location page allows users to access news specific to their chosen area by clicking on regions or countries.



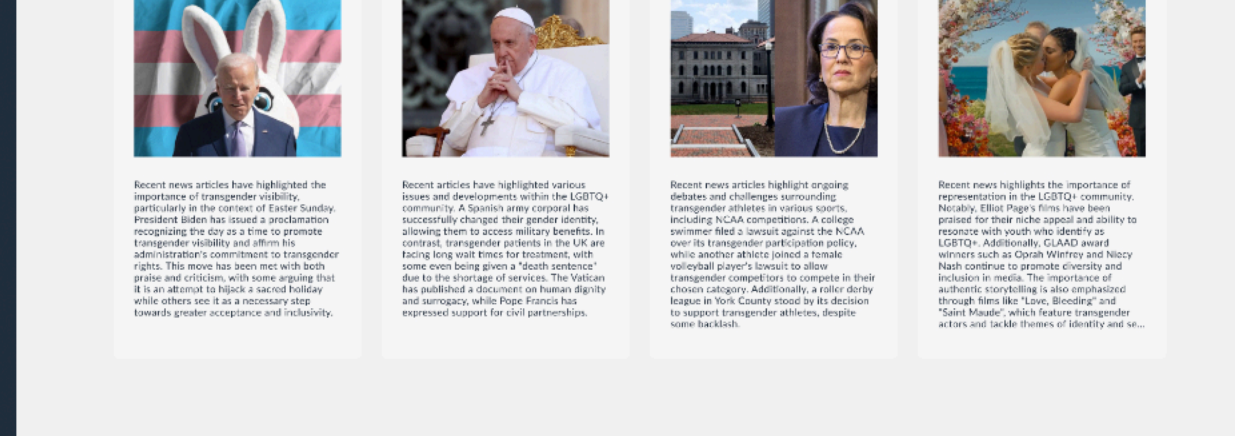
People

The people page showcases trending figures and provides news stories related to these individuals.



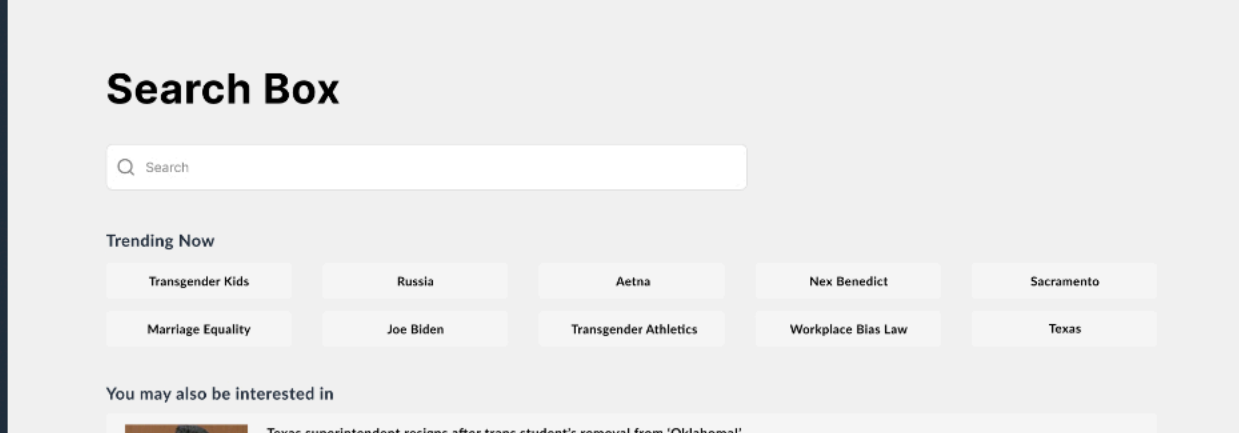
Trends

Users can view keywords and summaries on the trending page to discover what topics are currently popular in the news.



Search Box

The search page allows users to find news using trending search terms, enhanced by personalized recommendations based on past searches.



Search

The search page allows users to find news using trending search terms, enhanced by personalized recommendations based on past searches.

Implementation Phase



Data Engineering

We sourced news data from online platforms via the Google News API. To automate this task, we implemented Apache Airflow to schedule daily scraping activities.



Machine Learning

Location/People Page: We leveraged name entity recognition to extract entities from news articles and aggregated news based on these entities. In our testing product, we constructed an entity graph within our product, with entities comprising individuals and locations.

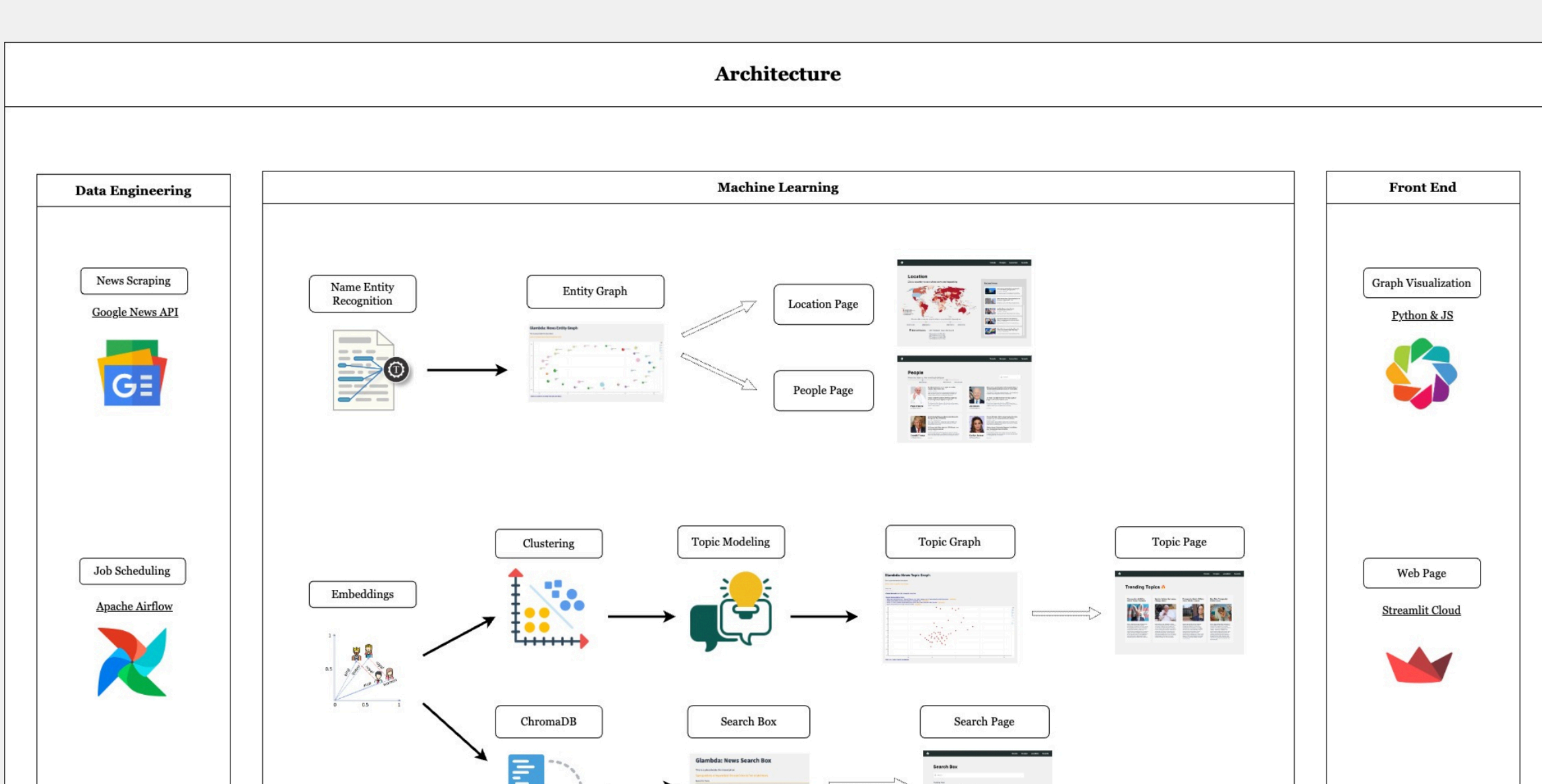
Topic Page: Utilizing Large Language Model embeddings, we developed a clustering model to categorize news articles into different groups. For each cluster, we employed topic modeling and LLM prompt engineering techniques to extract keywords and summaries.

Search Page: We stored embeddings of news articles in a vector database. User search queries are encoded using the same embeddings, facilitating matching against the database to retrieve the most relevant news items.

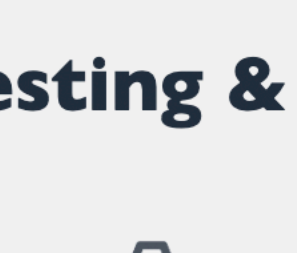


Front-end Development

Incorporating JavaScript and Python, we introduced numerous interactive functionalities, such as node-specific click interactions and cluster filtering, to enhance user engagement and experience.



Testing & Refinement Phase



User Testing & Adjustments

We invited classmates and friends to engage with our test product and prototype, gathering valuable feedback to refine our design. Based on this input, we iteratively adjusted our prototype to enhance its usability and effectiveness.

Benefits of Our Solution

Lambda News Digest provides readers with tailored, digestible insights into the world LGBTQ+ news. Backed by a well-established ML & LLM infrastructure, Lambda News Digest enables a highly efficient, interactive, and intriguing news reading experience, which is hard to find at most traditional news websites.

Lambda News Digest features the following signature attributes and benefits :

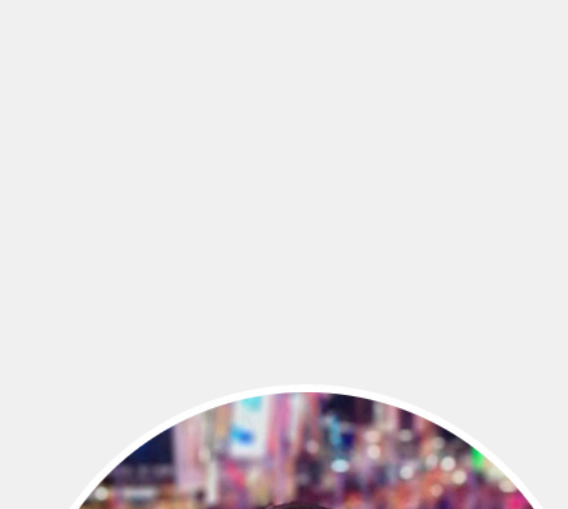
Diverse Selection of News Sources – Break the 'Echo Chamber' for Users Lambda News Digest sources news from over 15000 credible sources. The wide range of news sources breaks the 'echo chamber' caused by limited information, and endows the news reader with an unbiased view of the current LGBTQ+ trends.	Live News Combined With Retrospective Views – Better Exposure of Timeliness Timeliness is one critical consideration for many news readers, and Lambda News Digest addresses this concern. Supported by the flexible ML pipeline that can produce news digest for any specified timeframe. Readers can not only see the most timely up-to-date news digests, while learning the growing trend of LGBTQ+ events worldwide in the Lambda News Digest front-end.	Extracted News Metadata & Summary – Maximize News Reading Efficiency Readers get a straightforward view of the critical people, places, and topics relevant to LGBTQ+ in a few clicks within minutes, instead of reading thousands of news across thousands of different websites to understand the current highlights in LGBTQ+ world. If users are interested in a specific domain, they can always find the original news articles through Lambda News Digest.	Tailored Reading Experience – Deliver What Users Care Most About Lambda News Digest is powered by recommendation engine technologies including knowledge graph, topic modeling, etc. By recognizing users' reading preferences, Lambda News Digest can prioritize to present the news digests and news that users care more about.	Generalizable Infrastructure – Benefit Wider Audience Other Than LGBTQ Lambda News Digest is not just about LGBTQ+, or just news. It is fueled by a mature ETL & ML pipeline that can handle any category of text information. Lambda News Digest's infrastructure can be generalized to more fields, such as finance, health care, or anything else, and therefore benefits a much wider range of audiences who care about extracting valuable information.
---	---	---	--	--

The table below shows Lambda News Digest's Benefit over traditional news sources.

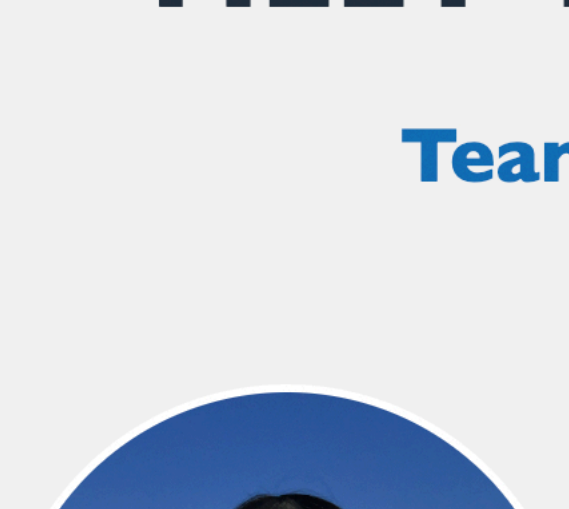
Lambda	Newsletter	News Website
Diverse selection of news sources	Limited sources	Limited compilation of news from different sources
Live news combined with retrospective views	Mostly live news	Mostly live news
News digest with original news	Only original news	Only original news
Tailored reading experience	No tailored reading experience	Only few websites provide tailored reading experience
Generalizable infrastructure	N/A	N/A

MEET THE TEAM

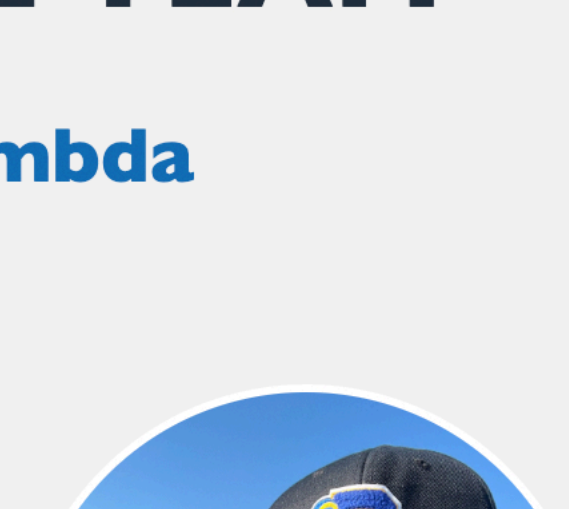
Team Lambda



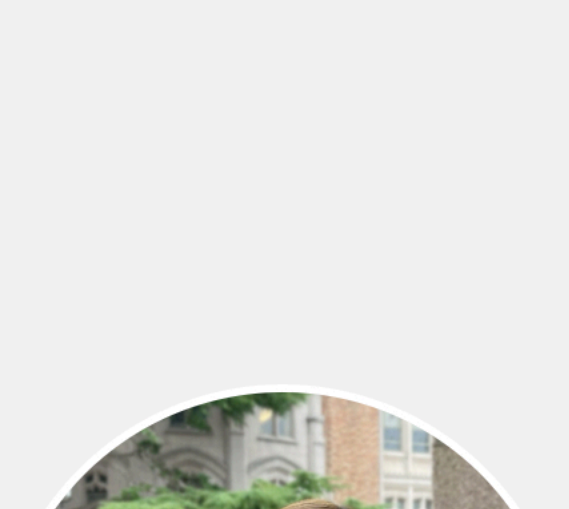
Jishu Chen
Machine Learning Engineer



Qianqian Liu
Software Developer



Juntong Wu
Data Engineer



Chesie Yu
Front-End Developer

