Team

George McIntire
MIMS Focus: Applied Data Science & NLP
Data Science

Lia Chin-Purcell
MIMS Focus: Machine Learning & NLP
Data Science

Kailin Koch
MIMS Focus: Data Science & Product Management
Data Engineering

Siqi Wu
MIMS Focus: Software Engineering
Data Engineering

Jade Clarke
MIMS Focus: Product Management
UX Research and Design
Motivation

● Women make up 34% of news sources in the U.S. despite being 47% of the labor market (GMMP 2020, BLS).

● **Goal**: quantify and visualize disparity in gender representation in U.S. news media

● **Partner**: ReThink Media, a local Berkeley-based nonprofit that strengthens movements through communications and media training.

● **Use Cases**: benchmark representation, hold media outlets accountable, and encourage underrepresented people to serve as spokespeople.
UX Design Process

**Stakeholder Interviews**
Conducted Stakeholder Interviews with various levels of ReThink Senior Leadership to agree on common goals, expectations and limitations of the project.

**Decide Dashboard Tooling**
Chose Google Data Studio for familiarity and relatively simple sharing and security features.

**Design Initial Functionality**
Drafted initial pages based on stakeholder interview feedback. Integrated ReThink Media Style Guide into the aesthetic UI.

**Design Iterations**
After the first round of usability tests, the team made edits and additions to the dashboard based on target user feedback.

**Usability Tests**
Conducted two rounds of usability tests with technical and non-technical ReThink employees to test user friendliness.
Final Dashboard

Link to Dashboard
Findings from Usability Testing

Dashboard is user friendly and robustly creates customizable visualizations

Filters included in the dashboard are relevant and easily interpreted by all employees

ReThink employees enjoyed the customization of security and data permissioning

Desired an opportunity to compare Media Outlets to one another over time

Major Changes Made:

1. Created a *Comparison of Female Voiceshare* page
2. Included percentages in the visualizations
3. Reconfigured chart type on *Over Time* page
Dashboard Overview

Fetch Articles → News vs Opinion → Database → Manual Review → Display

Extract Quotes → Infer Gender

7
News Classification

Objective: Train a text classification model that classifies articles as either “opinion” or “news” for the purpose of filtering out opinion articles.

Product: An ensemble model in which the classification is the average probability of a MLP trained sentence-embeddings and a fine-tuned DistilBert.

Performance: Recall: 93%, Precision: 91%, F1: 92%
“In hindsight, there’s a really good chance that the Fed should have started tightening earlier,” said Karen Dynan, an economist at the Harvard Kennedy School and a former Treasury Department chief economist. “It was really hard to judge in real time.”

In hindsight, there’s a really good chance that the Fed should have started tightening earlier. said Karen Dynan false

It was really hard to judge in real time. - Karen Dynan true

More than a dozen studies are underway to find the next generation of vaccines, with the first results expected in May or June. If all goes well, that would allow enough time to produce new doses before the fall. One major hitch is that the Biden administration says it does not have the money it needs to reserve its place in line by paying vaccine manufacturers for doses in advance.

As it was in the fall, when boosters were first rolled out, the broader scientific community is divided over what to do. “I am not persuaded there is substantial waning of protection against severe disease after the third dose.” Dr. Philip Krause, a former senior regulator at the F.D.A., said in an interview.

<table>
<thead>
<tr>
<th>speaker</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Biden</td>
<td>administration</td>
</tr>
<tr>
<td>Dr. Philip Krause</td>
<td>man</td>
</tr>
</tbody>
</table>
### Pipeline

<table>
<thead>
<tr>
<th>Task</th>
<th>Requirement</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Fetch articles</td>
<td>● Automate process</td>
<td><img src="https://via.placeholder.com/150" alt="docker" /></td>
</tr>
<tr>
<td>● Run models</td>
<td>● Keep costs low</td>
<td><img src="https://via.placeholder.com/150" alt="amazon" /></td>
</tr>
<tr>
<td>● Save results</td>
<td></td>
<td><img src="https://via.placeholder.com/150" alt="GitHub Actions" /></td>
</tr>
<tr>
<td>Manually review results</td>
<td>Use familiar tooling</td>
<td><img src="https://via.placeholder.com/150" alt="mongoDB" /></td>
</tr>
<tr>
<td>Display results in dashboard</td>
<td>● Make updates easy</td>
<td><img src="https://via.placeholder.com/150" alt="Google Data Studio" /></td>
</tr>
<tr>
<td></td>
<td>● Allow customization</td>
<td></td>
</tr>
</tbody>
</table>
Analysis

● 3/4 of articles quote 0 women, 0.8 women per article.
● Men:Women ratio 8:1
● Ratio is consistent across outlet types and time
● 15% of journalists with at least 5 articles quoted 0 women.
● Detected no significant differences in the length and content of quotes from men and women.
Next Steps

➔ ReThink Media will operate this dashboard moving forward
➔ Provided documentation, run-through and how-to videos

➔ Potential future work includes:
  ◆ Use more sophisticated model such as AllenNLP for coreference resolution
  ◆ Infer journalist gender
  ◆ Use pronouns for non-binary gender
  ◆ Improve performance of pipeline on full dataset
  ◆ Further automate pipeline from MongoDB into Google Data Studio
Thank You!

Special Thanks to:

➔ Laura Nixon at ReThink Media
➔ Lana Elauria at ReThink Media
➔ Professor David Bamman at Berkeley School of Information
➔ ReThink Media stakeholders who participated in UX Research
Questions?