Gender Representation and Opinion Detection in the Media

MIMS 2022 Final Capstone Project

May 12, 2022

Image Credit: ReThink Media Blog

1

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 ⅓ of news sources despite being **47%** of the labor market (Armstrong 2013).

- **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.
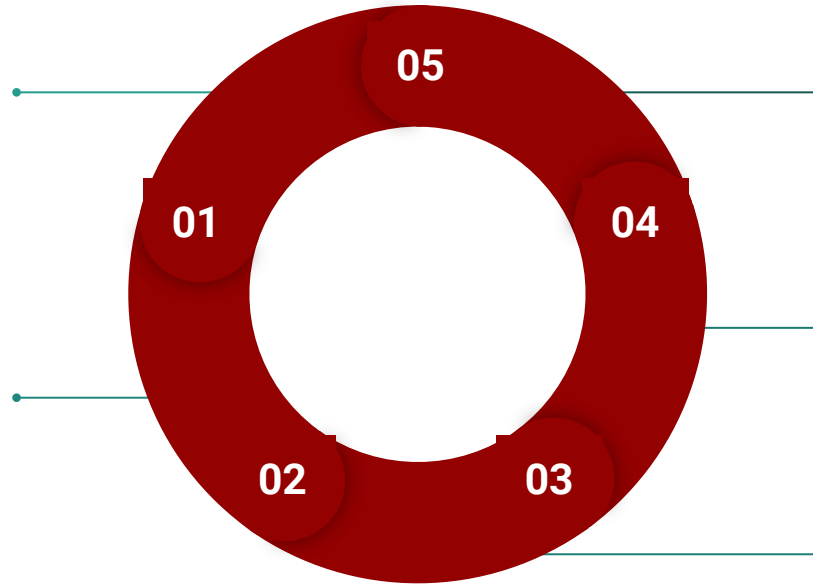
UX Design Process

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

05

01

04

02

03

Design Iterations

After the first round of usability tests, the team made edits and additions to the dashboard based on target user feedback
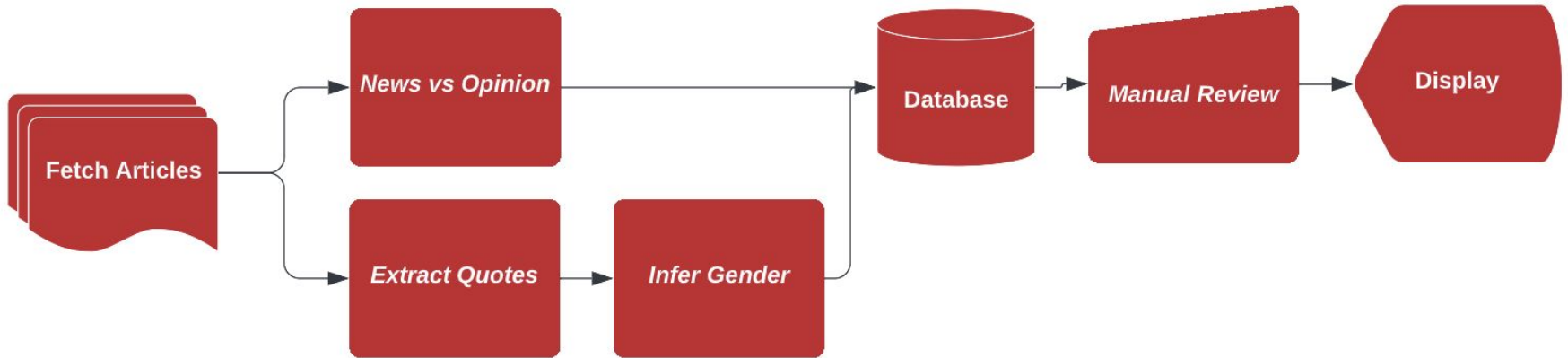
Usability Tests

Conducted two rounds of usability tests with technical and non-technical ReThink employees to test user friendliness

Design Initial Functionality

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

4

Dashboard Overview

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%

Iran rejects talks on missiles, but says no plan to increase range

By Reuters Staff   4 MIN READ   f   t

LONDON (Reuters) - Iran on Tuesday dismissed pressure from France and other Western powers for talks over its ballistic missile program, but said it had no plans to increase the range of the weapons.

Opinion
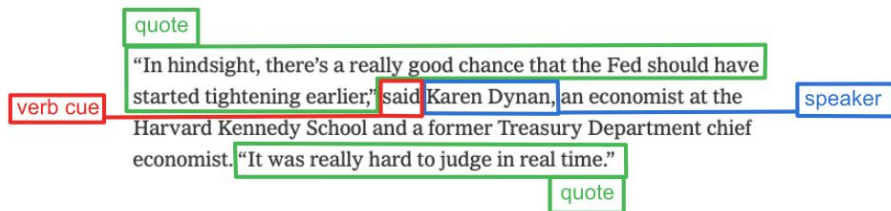
Israel's Image Issue

By Roger Cohen

Jan. 28, 2016

🎁 Give this article   ↗   🔖   💬 707

This is an interesting moment in relations between the United States and Israel. Call it a poisonous lull. The vitriol around the Iran nuclear deal has subsided. But something is rotten in the special bond.

6

Quote extraction

quote

"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."

verb cue   speaker   quote

quote	verb cue	speaker	Is floating quote
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

Coreference resolution and gender inference

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.
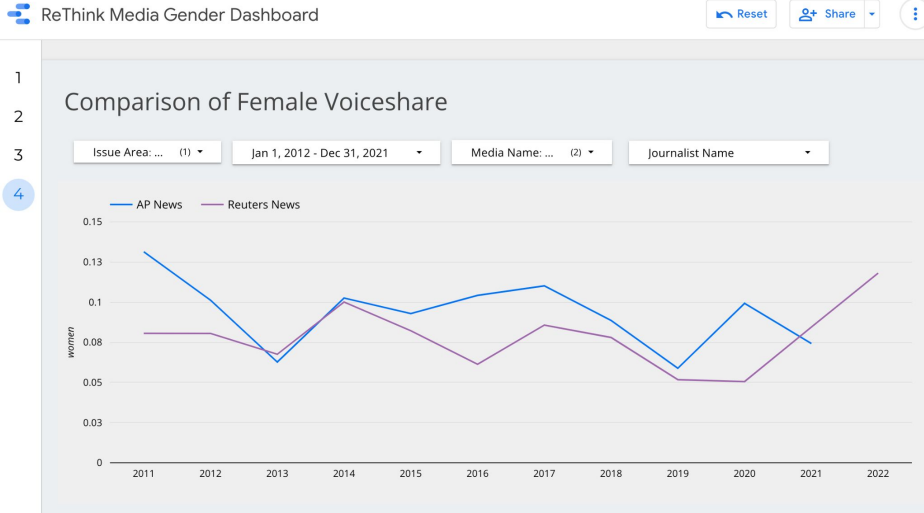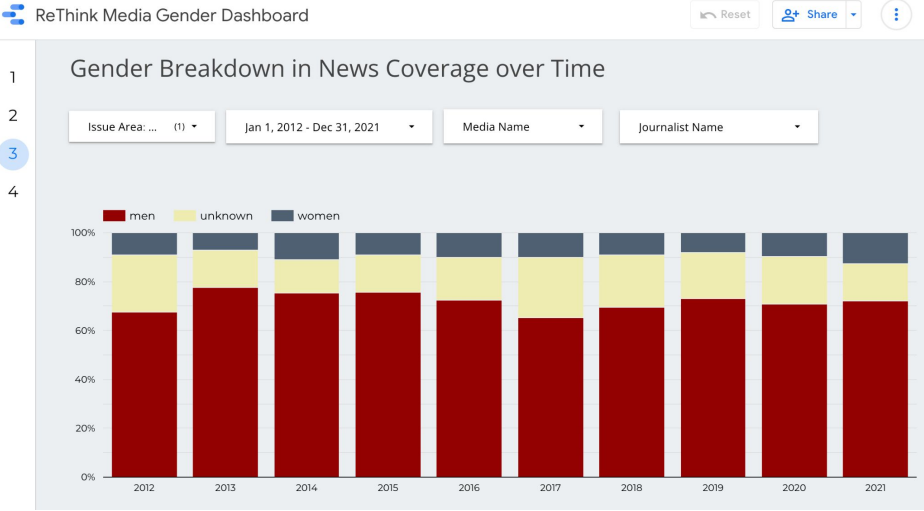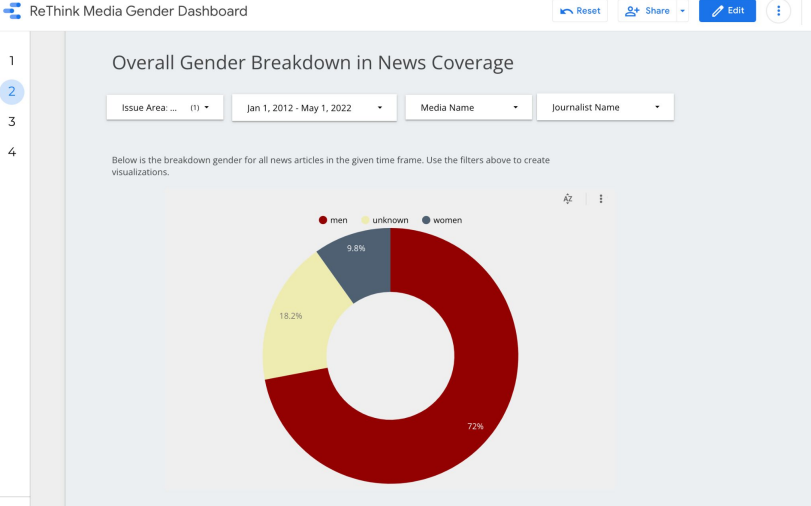
speaker	gender
The Biden administration	_
Dr. Philip Krause	man

Pipeline

Task	Requirement	Tool
● Fetch articles ● Run models ● Save results	● Automate process ● Keep costs low	
Manually review results	Use familiar tooling	
Display results in dashboard	● Make updates easy ● Allow customization	

Final Dashboard

[Link to Dashboard](#)

Findings from Usability Testing

Dashboard is user friendly and robustly creates customizable visualizations

ReThink employees enjoyed the customization of security and data permissioning

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

Desired an opportunity to compare Media Outlets to one another over time
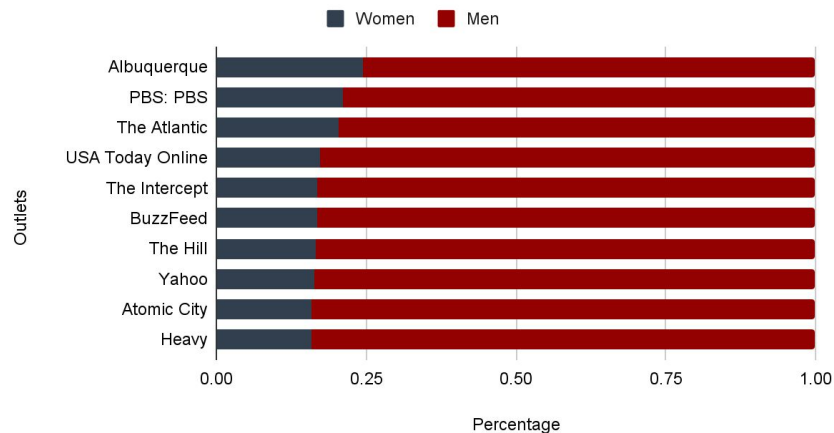
Major Changes Made:

1. Created a *Comparison of Female Voiceshare* page to allow comparison of female sourcing by Media Outlet
2. Included percentages in the visualizations to make interpretation of trends easier
3. Reconfigured chart type on *Over Time* page to better reflect results populated

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.**
- Gender word usage:
  - **Women:** her, women, husband, begging, security, council, elevated, actions
  - **Men:** he, his, deal, nuclear, missiles, uranium, bomb, capability, Islamic

Outlets with Highest Rates of Quoting Women

Chart legend: Women (dark blue), Men (red). Y-axis "Outlets" lists: Albuquerque, PBS: PBS, The Atlantic, USA Today Online, The Intercept, BuzzFeed, The Hill, Yahoo, Atomic City, Heavy. X-axis "Percentage" ranges 0.00 to 1.00.

rethink
Media for Security, Rights, and Democracy

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?