

# Using Large Language Models to Analyze Intersectional Media Portrayals of Identity

Sean Palicki, Alice Beazer & Stefanie Walter  
Technical University of Munich, Germany

## Introduction

Recognition Crisis: Citizens' understanding of political world involves recognition and misrecognition of one's social group, and derogation of "others" (Wells, 2023; Davies, 2021).

Media portrayals of social groups impact personal & public opinions (Nisar & Bleich, 2020; Ahmed & Matthes, 2017).

Gap: Understanding complex layers of social privilege and exclusion, i.e. intersectionality theory (McCall, 2005; Crenshaw 1991).

### Goals

1. Develop and evaluate an LLM based query generation method for social identity.
2. Explore journalistic representations using big data across intersectional social identities and time.

## RQs and Data

### Research Questions

RQ 1. How effective are LLMs for generating and refining search queries about intersectional social identities compared to traditional approaches?

RQ 2. How does mainstream news coverage of intersectional social identities including sexuality, disability, religion, gender, and race vary in terms of salience and tone?

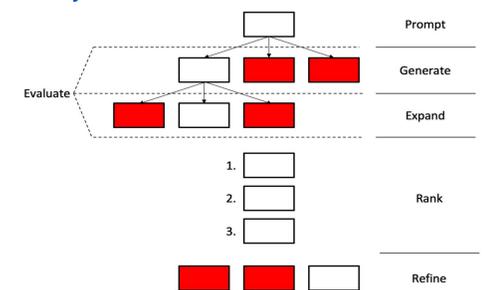
### Data

Full Sample: 4,842,484 UK news articles from The Guardian, The Telegraph, and The Sun (2012-2022).

Training Sample: 11,280 Labelled Guardian Online articles related to Race, Gender, Disability, Sexuality and Religion (2015-2022).

## Methodology

### Query Generation Process



Tree-of-Thoughts prompting using Human-in-the-Loop evaluation to quickly iterate through hundreds of queries.

### Prompt Template to Generate Queries

1. Explain Task: Retrieve documents about {Identity}.
2. Provide Seed Words: {Identity seed words}
3. Provide Seed Documents: n < 20 docs relevant to {Identity}.
4. Add Quality Instructions:
  - a. Include terms potentially relevant to {Identity}, including the Seed Words.
  - b. Be specific by including specific examples of {Identity} labels and communities.
  - c. Be inclusive, not biasing the query to any one community in particular.

### Query Evaluation Process

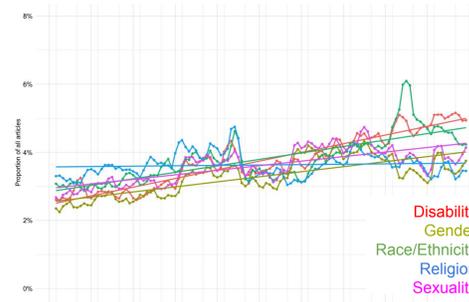
1. Iterative ChatGPT self-evaluation ("On a scale of 1-5, how well do the generated queries answer the task and follow the quality instructions?")
2. Iterative traditional retrieval evaluation using labelled seed documents.
3. Manual evaluation of top 3 prompts from Generation process.
4. Traditional retrieval evaluation of best prompt using Guardian training sample.
5. Comparisons between generated, simple keyword, and expanded queries.

## Results

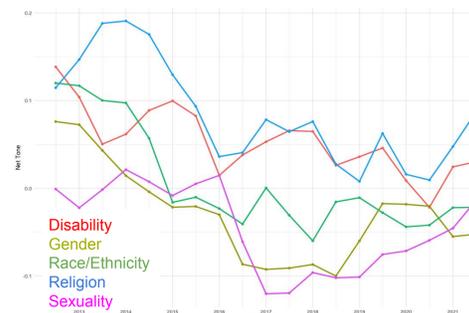
RQ 1: Generated queries have high precision and recall. F2 scores are higher than simple keyword and query expanded search queries. Evaluation metrics for document retrieval task using labelled Guardian Online articles.

	Precision	Recall	Accuracy	F1	F2	GPT
Sexual Identity	95.51	88.62	92.23	91.94	5.97	4
Disability	90.19	91.88	91.11	91.03	5.74	4
Religion	89.15	92.10	90.75	90.60	5.76	4
Gender Identity	90.31	83.32	86.11	86.67	5.21	3.5
Racial/Ethnic Identity	88.46	82.90	85.10	85.59	5.18	3.5

RQ 2a: Increased salience of identity in the news over time with periods of focus on particular groups. Variations in positive tone for news articles related to identities.



Salience of social identities in full sample of UK newspapers as percent of published articles per month.



Net tone of articles mentioning social identities in full sample of UK newspapers using latent semantic scaling.

## Results Continued

RQ 2b: Tone of Muslim related news is more negative than other religious identities across intersections. Comparison of tone for articles mentioning selected religious groups in UK news.

Tone	Muslims		Catholics		Jews		Atheists		Hindus	
	n	%	n	%	n	%	n	%	n	%
Positive Tone	5351	24	3591	39	3385	34	458	35	672	36
Negative Tone	7810	35	1969	21	2747	27	277	21	511	27
No Tone	9389	41	3686	40	3960	39	584	44	705	37
Net Tone	-0.11		+0.18		+0.06		+0.14		+0.09	
Total Headlines	22,548		9,246		10,092		1,319		1,888	

Ranking of tone for articles using joint-mentions of selected religious groups and non-religious identities in UK news.

Aggregate	Ethnicity/Race	Sexuality	Gender	Disability
Catholics	Catholics	Jews ~ Catholics	Catholics	Catholics ~ Hindus*
Atheists	Atheists	Hindus	Jews ~ Atheists ~ Hindus	Atheists* ~ Jews
Jews ~ Hindus	Hindus	Atheists	Muslims	Muslims
Muslims	Jews	Muslims		

## Conclusions

LLM-generated search queries may help researchers systematically develop accurate search queries to describe diverse intersectional identities.

Using accurate search queries to directly compare media representations of intersectional identities, we found an overall increased representation of identity in UK newspapers and large variations in tone for all groups over time.

Layering individual religious identities with non-religious identities, we found significantly more negativity in articles about Muslims than in articles about other religious groups.

## LLM Generated Identity Search Queries

<b>Sexual Identity</b>	(sexuality OR sexual identity OR sexual orientation OR sexual preference OR sexual behavior OR sex OR intimacy OR romantic relationship OR sexual activity OR sexual relationship OR partnership OR lgbt OR lgbtq OR heterosexual OR bisexual OR homosexual OR lesbian OR gay OR queer OR pansexual) AND (people OR individuals OR persons OR humans OR citizens OR inhabitants OR residents OR population OR man OR woman OR men OR women OR community)
<b>Disability</b>	(disability OR disabled OR differently abled OR deaf OR blind OR autism OR accessible OR ableism OR people with disabilities OR impairment OR learning disability OR dyslexia OR mental health OR physical disability OR intellectual disability OR developmental disability OR cognitive disability OR sensory disability)
<b>Religion</b>	(religion OR religious OR muslim OR christian OR jewish OR non-religious OR atheist OR hijab OR mosque OR rabbi OR catholic OR buddhist OR hindu OR sikh OR church OR synagogue OR temple OR faith OR belief OR worship OR spiritual OR prayer OR monotheistic OR polytheistic) AND (people OR individuals OR communities OR groups OR society OR population OR residents OR citizens OR inhabitants OR public OR man OR woman OR men OR women)
<b>Gender Identity</b>	(biological sex OR gender OR sex OR identity OR orientation OR LGBTQ OR LGBTQ OR LGBTQIA+ OR queer) AND (man OR woman OR transsexual OR nonbinary OR transgender OR genderqueer OR cisgender OR male OR female OR transgender OR genderfluid OR agender OR men OR women)
<b>Racial/Ethnic Identity</b>	(white OR english OR black OR chinese OR welsh OR scottish OR northern irish OR british OR irish OR gypsy OR irish traveller OR black caribbean OR black african OR asian OR asian british OR indian OR pakistani OR bangladeshi OR african OR caribbean OR black british OR arab OR bame OR ethnic minority OR racial minority OR biracial OR multiracial) AND (people OR individuals OR individual OR persons OR person OR population OR community OR society OR citizen OR citizenry OR residents OR inhabitants)