Analysis of public comments on the use of facial recognition and AI technologies by DHS

Kurt Williams, Sara Bano, Mir Rabby, Chetna Kuanr

Rationale

Can NLP make government processing of text data more efficient?

- To what extent does the public support use of facial recognition technology by DHS?
- What are the themes of concerns raised in these comments?
- How do the NLP results compare to what we expect of human coders?

Workflow:

- Scrape comments from regulations.gov
- Topic modeling
- Sentiment analysis
- Qualitative content analysis

DHS case: "*Public Perceptions of Emerging Technology*"

Information Collection Requests are one mechanism for public input in federal policymaking

In Nov. 2021, DHS submitted a request for comments on its use of facial recognition and AI technologies.

> understanding how the public perceives [facial recognition and AI], and then **designing and deploying them in a manner responsive to the public's concerns**, is critical in gaining public support for DHS's use of these technologies

Building the Corpus

Web Scraping Process

- **url queries** for **regulations.gov API**
- **`httr::GET()`** for sending requests
- **`tidyjson`** for parsing the output

Corpus Description

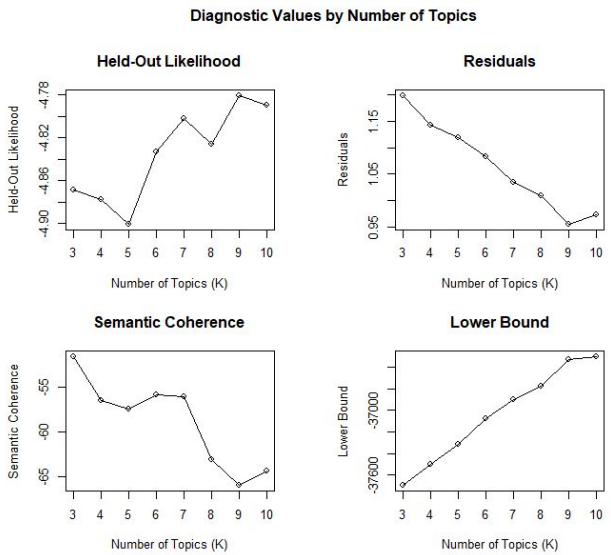
- 220 comments
- Named; Anonymous; Organizations
- Tokens IQR: 33–161 (39,000 total)
- Sentences IQR: 2–6 (1,475 total)

Most Common Words

	freq		freq		freq
facial	358	ai	206	dhs	125
recognition	357	government	157	information	124
technology	221	public	157	used	116
data	212	privacy	146	people	111
use	211	can	133	security	106

Topic Modeling

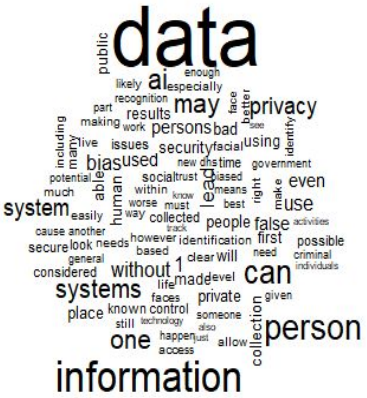
- Did the topics reflect qualitative topics?
- How many topics?



	Topics 1	Topics 2	Topics 3	Topics 4	Topics 5	Topics 6	Topics 7
1	americans	technology	great	technologies	dont	person	rights
2	american	surveillance	like	homeland	software	one	violation
3	tools	enforcement	never	dhs	tracking	data	want
4	will	recognition	think	department	every	may	right
5	used	law	idea	security	stop	information	constitutional
6	abused	facial	good	border	people	lead	privacy
7	years	intelligence	please	collection	way	systems	amendment
8	citizens	due	ai	individual	freedom	system	citizen
9	invasion	artificial	problem	comment	just	persons	us
10	liberties	must	long	public	personal	1	act
11	much	use	understand	research	overreach	false	violate
12	absolutely	search	another	including	anyone	can	allowed
13	concerned	police	future	manner	know	bias	say
14	civil	likely	especially	issues	government	bad	crime
15	agree	identify	issue	regarding	even	without	feel
16	already	find	potential	address	protect	able	believe
17	crime	risk	risk	protection	lives	results	citizens
18	look	social	society	individuals	federal	considered	everyone
19	new	concerns	far	information	united	place	abuse
20	government	state	country	identification	states	private	using

#wordcloud
cloud(model, topic = 1, min.freq=3)

Topic 1



#wordcloud

```
> cloud(model, topic = 2, min.freq=4)

> cloud(model, topic = 4, min.freq=4)

> #wordcloud> cloud(model, topic = 1,
min.freq=4)

> cloud(model, topic = 6, min.freq=4)

> #wordcloud

> cloud(model, topic = 2, min.freq=4)

> cloud(model, topic = 6, min.freq=4)

>
```

#plot two topics and compare

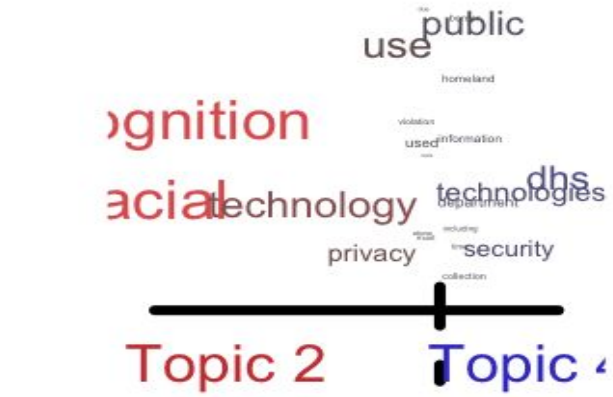
```
plot(model,
    type="perspectives",
    topics=c(2,4),
    plabels = c("Topic 2","Topic 4"))

#plot two topics and compare
plot(model,
    type="perspectives",
    topics=c(1,6),
    plabels = c("Topic 1","Topic 6"))

#plot two topics and compare
plot(model,
    type="perspectives",
    topics=c(2,6),
    plabels = c("Topic 2","Topic 6"))
```

Topics 2-4 : Wordcloud & Plot Comparison

Plot: Topics 2-4



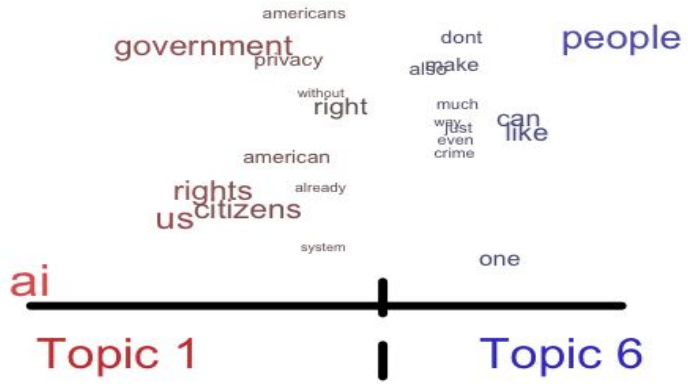
Topic Modeling

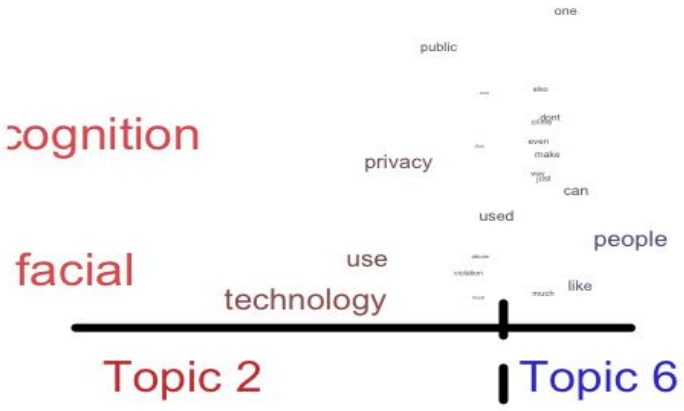
Topic: 1-6 (People)

Topic:2-6 (Data)

Wordcloud

Comparison of Topics





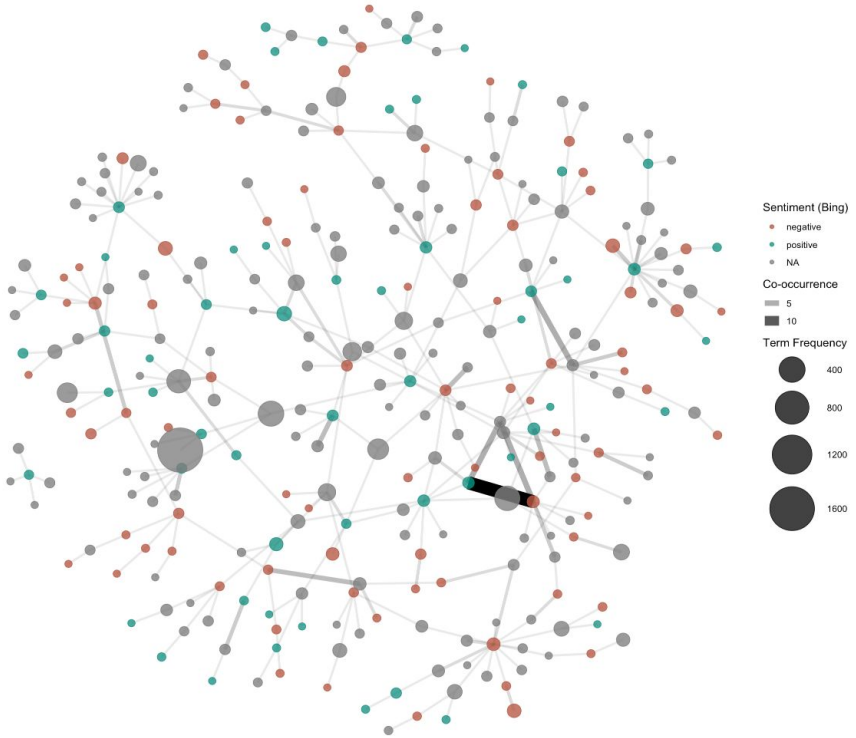
Plots

Sentiment Analysis

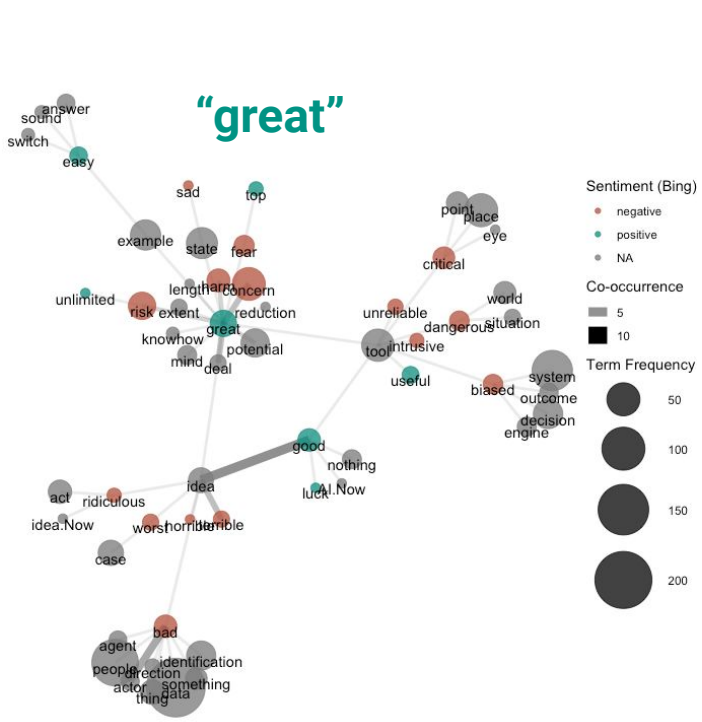
Number of **negative** tokens: **1349**
Number of **positive** tokens: **1110**
36,589 tokens not included

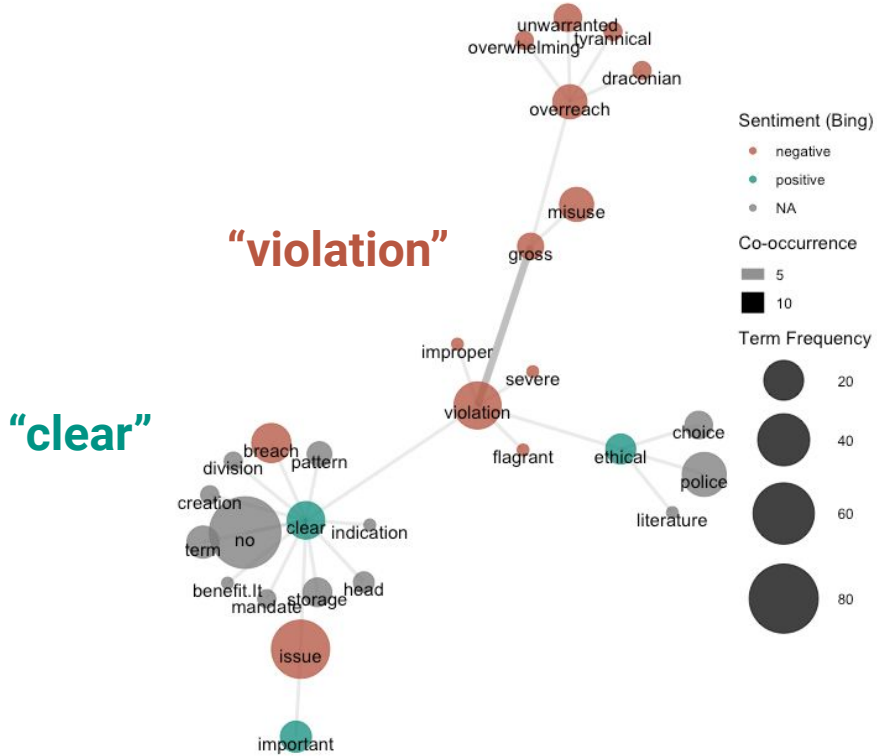
Highest degree words:

Negative	Positive	N.I.
criminal	great	rate
bad	clear	tool
false	free	idea
negative	better	technology
violation	accurate	information



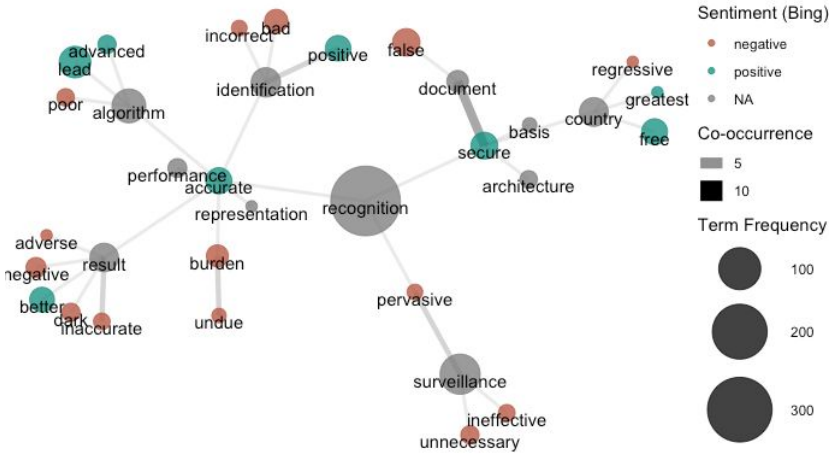
Sentiment Analysis



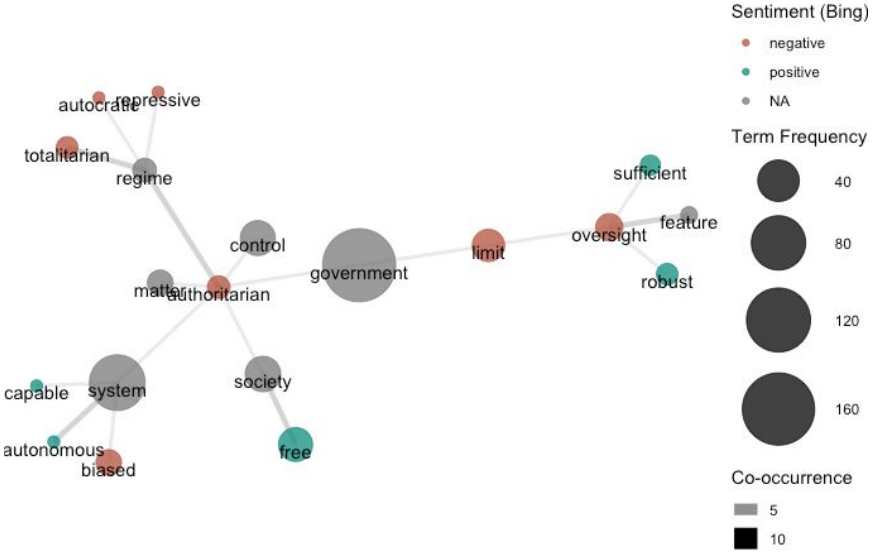


Sentiment Analysis

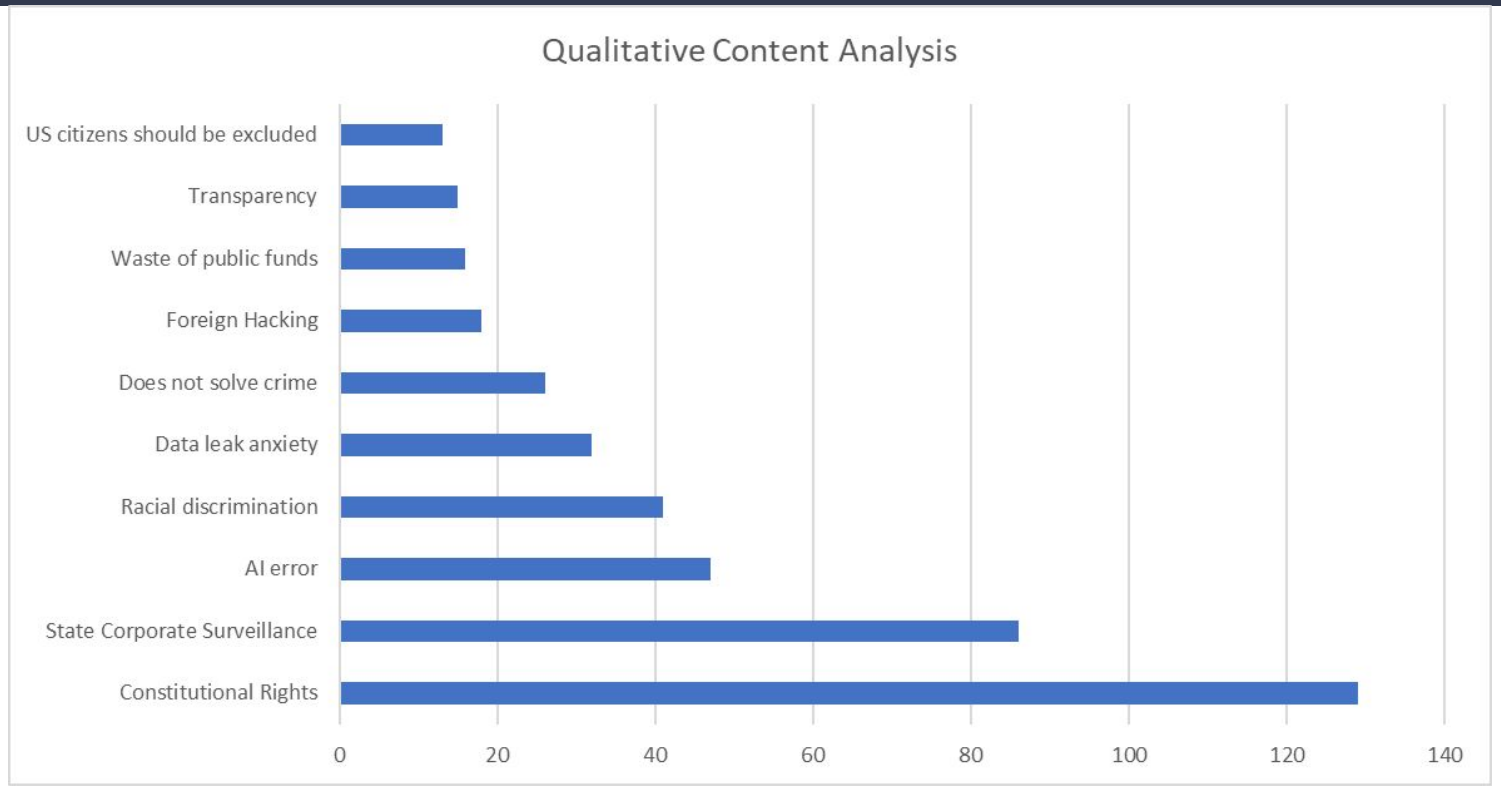
“recognition”



“government”



Content Analysis



Final Reflections

- Potential for using in higher education research.
- Great opportunity to learn about research methods used in variety of fields.
- Wonderful experience to meet such a diverse group of scholars from different fields.
- Overall, a great learning experience.

Questions

Thanks for this wonderful learning experience :)

Research Questions

1. To what extent does the US public (as represented by this collection of public comments) support the use of facial recognition technology by DHS?
2. What are the themes of the concerns raised in these comments about the use of facial recognition technology by DHS?
3. Are there meaningful subgroups of commenters who share similar concerns?
4. Do commenters from different stakeholder groups (e.g., organizations vs. individuals) or with different characteristics have different concerns ?
5. Are the results of the NLP approach reliable compared to standards of what we would expect of human coders? Realistically, what are the pros and cons of substituting the computational approach in place of traditional qualitative analysis?

```
topfeatures(comments_dtm, n=50)
```

facial	recognition	technology	data	use	ai	government	public
358	357	221	212	211	206	157	157
privacy	can	dhs	information	used	people	security	technologies
146	133	125	124	116	111	106	90
rveillance	us	citizens	law	even	enforcement	rights	person
88	77	76	74	72	71	65	65
without	like	right	collection	also	systems	bias	american
62	62	60	58	56	53	47	46
many	using	may	one	department	time	make	system
45	44	44	44	44	44	43	43
human	artificial	including	already	way	individuals	states	personal
42	42	42	41	40	40	40	40
americans	crime						
39	39						