source: The Trial of the Chicago 7 (2021 film)

NLP: Police Complaints

Federico Domínguez Molina, Jonathan Juarez, Matthew Jackson
Advanced Machine Learning for Public Policy
CAPP 30255
March 2024

Overview: Police Misconduct & Accountability in Chicago

- Chicago Police Department (CPD) -- allegations of misconduct
- 1974-2007: Office of Professional Standards (within CPD)
- 2007-17: **Independent Police Review Authority (IPRA)** (1st public reports)
- 2017-now: **Civilian Office of Police Accountability (COPA)**

Note: This student project does not in any way endorse any views expressed by COPA, CPD, The University of Chicago, or any employees thereof, or any other organization past or present.

Jurisdiction

IPRA investigates allegations of misconduct that concern:

USE OF EXCESSIVE FORCE

DOMESTIC VIOLENCE

VERBAL ABUSE, INCLUDING BIAS

COERCION

IPRA investigates certain incidents regardless of whether misconduct is alleged:

ALL CASES IN WHICH A CPD MEMBER DISCHARGES A FIREARM, STUN GUN OR TASER IN A MANNER THAT COULD POTENTIALLY STRIKE AN INDIVIDUAL

ALL DEATHS OR SERIOUS INJURY OF PERSONS IN POLICE CUSTODY

COPA INVESTIGATES ALLEGATIONS OF:

- Bias-based verbal abuse
- Coercion
- Death or serious bodily injury in custody
- Domestic violence
- Excessive force
- Improper search and seizure
- Firearm discharge
- Sexual misconduct
- Taser discharge that results in death or serious bodily injury
- Pattern or practices of misconduct
- Unlawful denial or access to counsel

- https://www.chicago.gov/dam/city/depts/ipra/general/IPRA%20Brochure.pdf; https://www.chicagocopa.org/investigations/jurisdiction/
- CPD Bureau of Internal Affairs reviews “All other complaints of police misconduct, including but not limited to: Criminal misconduct, Operational violations, Theft of money or property, Planting of drugs, Substance abuse, Residency violations, Medical roll abuse”

Example report pages

CIVILIAN OFFICE OF POLICE ACCOUNTABILITY **LOG#1089802**

SUMMARY REPORT OF INVESTIGATION

I. EXECUTIVE SUMMARY

Date of Incident:	June 7, 2018
Time of Incident:	18:30
Location of Incident:	[redacted]
Date of COPA Notification:	June 8, 2018
Time of COPA Notification:	15:12

Mr. [redacted] alleged that the accused, Officer [redacted] entered onto his property and shot his dog without lawful justification.

II. INVOLVED PARTIES

Involved Officer #1:	[redacted] Star # [redacted] Employee # [redacted] DOA [redacted], Unit 015, DOB [redacted], 1981, M, White
Involved Individual #1:	[redacted] DOB [redacted], 1975, M, Blk

III. ALLEGATIONS

Officer	**Allegation**	**Finding / Recommendation**
Officer [redacted]	1. Unlawfully entered [redacted] back yard in violation of Amendment IV of the United States Constitution.	Exonerated
	2. Shot the Involved Individual's dog, without justification, in violation of Rule 1, 38.	Exonerated

IV. APPLICABLE RULES AND LAWS

Rules
1. Rule 38: Unlawful or unnecessary use or display of a weapon.
General Orders
1. G03-02 – Use of Force- Effective October 16, 2017
Federal Laws

1

CIVILIAN OFFICE OF POLICE ACCOUNTABILITY **LOG#1089802**

1. United States Constitution, Amendment IV

V. INVESTIGATION

a. Interviews

In his statement to COPA[1] on June 8, 2018, [redacted] stated that on June 7, 2018 at approximately 6:30pm, he heard his dog growl and then heard one gun shot. Mr. [redacted] said after he heard the gun shot, he went onto his second-floor back porch and saw his dog on the ground. He saw Officer [redacted] walking along the side of the building and asked him what happened to his dog. Mr. [redacted] explained Officer [redacted] told him someone ran into the back yard and he was chasing this person. Officer [redacted] told him that when he entered the back yard, the dog bit him. Mr. [redacted] said Officer [redacted] showed him the area of his leg where he was bitten but he did not see any bite marks on the officer's leg. Mr. [redacted] explained he did not see anyone run through his back yard or the interaction between Officer [redacted] and his dog. Mr. [redacted] stated he was on the second-floor porch and could not see down into the yard, where the incident occurred. Mr. [redacted] said his dog was secured to a pole, with a two or three-foot leash, and the dog weighed 110 pounds.

In his statement to COPA[2] on July 9, 2018, Accused Officer [redacted] stated he was assisting officers with a foot pursuit of a subject. Officer [redacted] explained he ran parallel to the subject when he entered the front yard at 101 S. Parkside. He ran down the gangway of the building and entered the back yard of the residence. Officer [redacted] said he saw the dog when he reached the end of the gangway. He said the dog was approximately 10 feet away in the yard and he stopped running when he saw the dog. He explained that he looked down to avoid eye contact with the dog and turned his body sideways. Officer [redacted] said that action was to deescalate the dog's aggression. He stated, the dog then growled at him. He turned to retreat through the gangway. As he ran away, the dog charged him and bit him in the back of the knee. He explained he tried to pull his leg from the dog's mouth, but the dog was shaking its head violently. He feared great bodily harm by the dog, so he pulled out his service weapon and shot the dog. Officer [redacted] explained he was unaware the dog was restrained by a leash. Officer [redacted] reported that the back of his knee was red and bruised in the area where the dog bit him. He was transported to the hospital where he was treated for the dog bite.

b. Digital Evidence

The **Crime Scene Evidence Photographs**[3] include photographs of Officer [redacted] jeans with two holes on the right leg of the jeans, and red marks on the back of his right knee and leg.

The **Body Worn Camera** footage for this incident does not capture Officer [redacted] interaction with the dog because he was not wearing a body worn camera. The footage captured is that of officers who responded to Mr. [redacted] 911call after the incident occurred.

[1] Att. 15
[2] Att. 22
[3] Att. 28

2

CIVILIAN OFFICE OF POLICE ACCOUNTABILITY **LOG#1089802**

c. Documentary Evidence

City of Chicago Fire Department records,[4] dated June 6, 2018, document Officer [redacted] cause as animal bite and stated that teeth marks are visible on the right calf and back of knee. The report states Officer [redacted] was transported to Northwestern Hospital for treatment.

The **Chicago Police Department Original Case Incident Report**,[5] RD# [redacted] states Officer [redacted] was assisting in a foot pursuit at the location of [redacted]. It reports that he was paralleling another officer involved in the chase when he entered the gangway of the residence, through an unsecured gate, and proceeded to the back yard. He entered the back yard where he encountered a female boxer dog. The dog charged toward him, causing Officer [redacted] to retreat through the gangway. During this retreat, the dog bit him and fearing being bitten again, Officer [redacted] discharged one round, from his service weapon, destroying the dog.

The **Chicago Police Department Original Case Incident Report**,[6] RD# [redacted] states that officers were patrolling an area plagued by gang violence and narcotic sales, when they observed an unknown male trying to solicit the sale of cannabis. The report explains the officers approached the unknown male and he began to flee, running north bound in the east alley of [redacted] The report states that during the foot chase, Officer [redacted] was bit by a dog and discharged his service weapon.

The medical records from Northwestern Memorial Emergency Department[7] lists the chief complaint as a patient with a dog bite and notes state that there are bite marks in the right pant leg and superficial abrasions behind right knee.

VI. ANALYSIS

COPA recommends a finding of **Exonerated** for allegations #1 against Officer [redacted]

The Fourth Amendment typically requires a warrant to conduct a search, especially of private property, but that requirement is excused when an officer faces exigent circumstances. *People v. Foskey*, 136 Ill. 2d 66, 74 citing *Payton v. New York*, 445 U.S. 573, 63 L. Ed. 2d 639 (1980). Key to the inquiry is whether it was reasonable for an officer on the scene to believe, considering the circumstances he faced, there was a compelling need to act and no time to obtain a warrant. *United States v. Williams*, 79 F. Supp. 3d 888, 894 citing *Sutterfield v. City of Milwaukee*, 751 F.3d 542, 557 (7th Cir. 2014). Relevant factors for determining whether exigent circumstances existed include whether: (1) the crime under investigation was recently committed; (2) there was any deliberate or unjustified delay by the police during which time a warrant could have been obtained; (3) a grave offense was involved, particularly a crime of violence; (4) there was reasonable belief that the suspect was armed; (5) the police officers were acting on a clear showing of probable cause; (6) it was likely that the suspect would escape if he was not swiftly

[4] Att. 26
[5] Att. 18
[6] Att. 13
[7] Att. 29

3

- Formulaic: section headings, Rules, signature, etc.
- Four possible findings per allegation: Sustained, Not Sustained, Unfounded, Exonerated
- Reports often contain disturbing or graphic content -- viewer discretion advised

Collecting the Corpus: Web Scraping, PDF Processing

- Can't just download a clean dataset from Kaggle!
- Public summary reports of IPRA/COPA investigations, date range 2008-2023
- Used `requests` & `BeautifulSoup` to scrape COPA website table (Feb. 8)
- Total n = 2,148 PDF documents. (~1.23 GB)
- Used `PyPDF2` library to recognize text and write each document in corpus to .txt files (~62.1 MB)
  - Issues: line breaks, footnotes, tables, redactions...
- (Later: Make `TextParser` class to build custom versions of corpus in Python: test stopword removal, stemming/lemmatization, etc. without mutating original files)

Finding focus: Supervised or Unsupervised?

- Four possible findings per allegation
  - Sustained, Not Sustained, Unfounded, Exonerated
- Most familiar to us: supervised learning to predict finding (label) based on report text. (e.g., which kinds of allegation get Exonerated most). But:
  - Exact format of document changes over time
  - Some reports have multiple allegations w/ different findings, discussed in interwoven order -- very hard to separate out
- Another path: *Unsupervised learning* to learn about corpus as a whole.
  - What are the bulk of reports about?
  - What do they tell us about police misconduct and accountability in Chicago? Generally?

Pre-exploratory: Big animating questions

- These are *unsupervised* tasks -- how to assess quality of results w/o labels?
- What tasks would shed light on each other as well as the corpus?

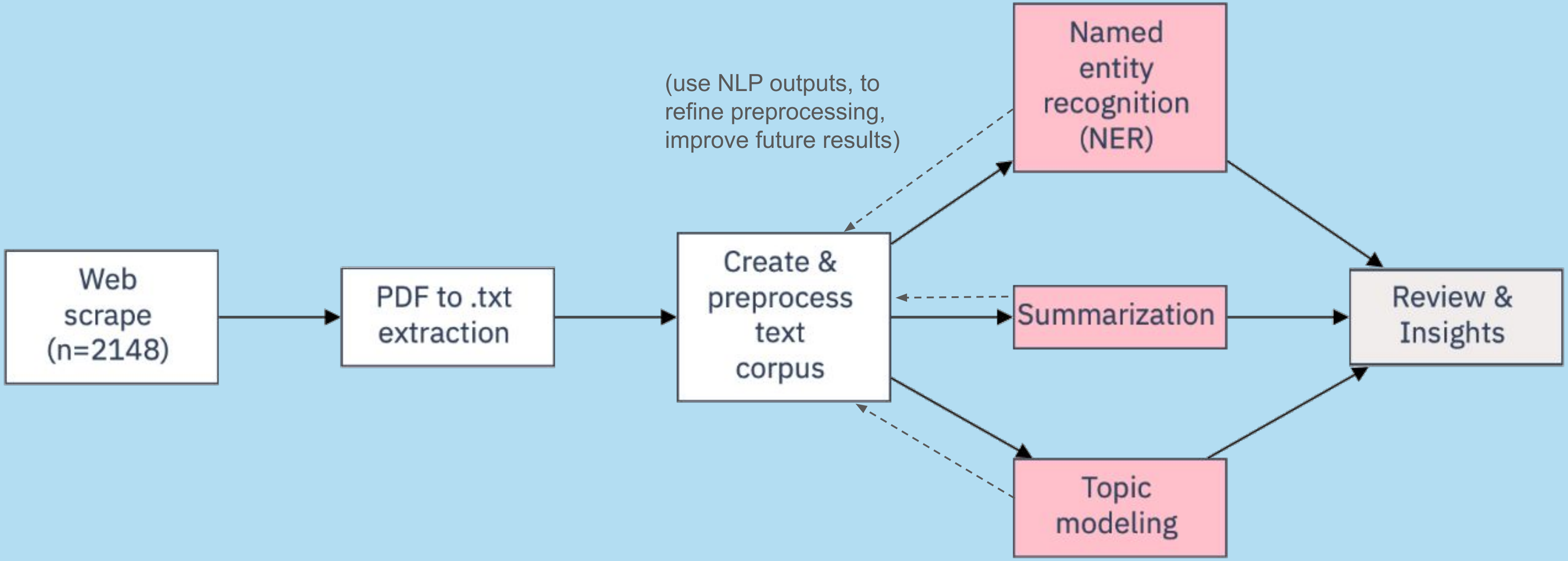
Literature review:

- Topic modeling of complaints against Carabineros police force, Chile
- CDO Council report: try NLP to process U.S. federal regulatory comments
- Using legal-BERT for NLP tasks of legal documents, Chalkidis

Settled on 3 parallel tasks: Named Entity Recognition (NER), summarization, topic modeling

- (*not sentiment analysis -- report tone is pretty neutral and in the same institutional voice across docs)*

Roadmap / Flowchart



Data ingestion and preprocessing

NLP tasks (unsupervised learning)

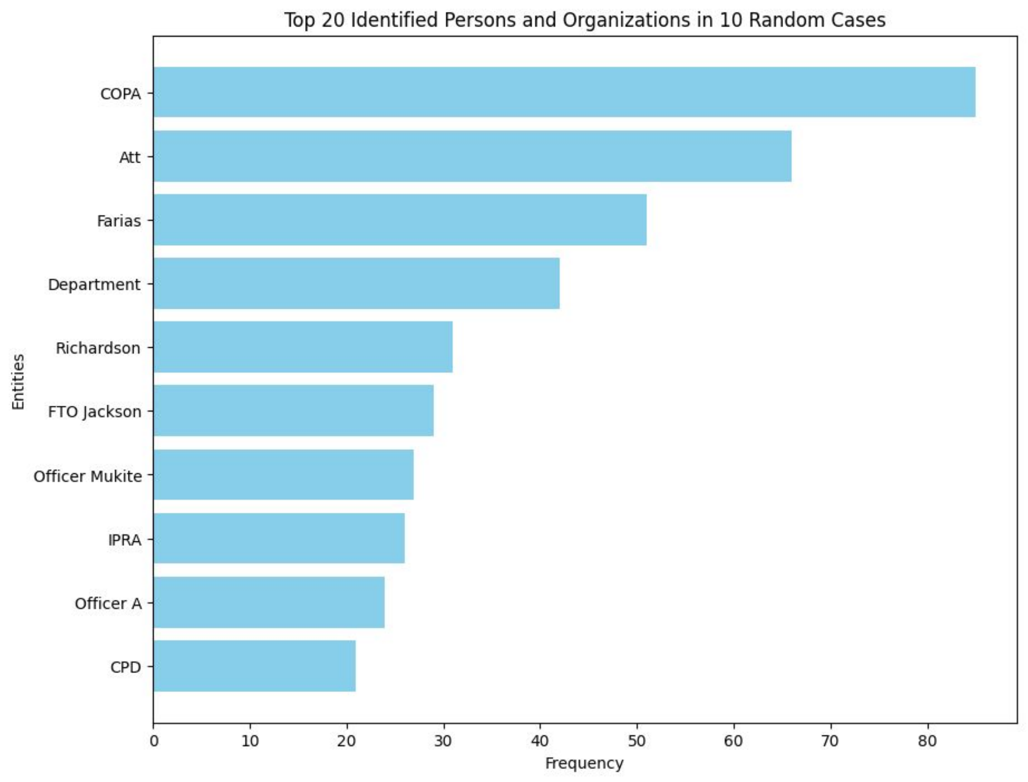
Human review

Named Entity Recognition (NER) (Jonathan)

Used three libraries for NER application

- spaCy
- NLTK
- HuggingFace Transformers (BERT)

Quick visual inspections of output entities showed good enough results for spaCy and BERT



spaCy CNN vs Hugging Face BERT

spaCy CNN Model:

- Optimized for speed and efficiency.
- Built-in support for NER with pre-trained models.
- Offers good performance with less computational resource usage.
- Extensible through custom training on new entities or fine-tuning.

Hugging Face BERT Model:

- Provides state-of-the-art results
- Pre-trained models available through their library (used legal-BERT for fine tuning)
- Requires more computational resources but delivers higher accuracy.
- Supports fine-tuning on specific datasets to enhance model performance.

Fine-Tuning and Evaluations

Fined-Tuned spaCy 'en_core_web_lg' with Augmented Data

Label	Precision	Recall	F1-score
LOC	0.93	0.94	0.94
MISC	na	na	na
ORG	0.92	0.93	0.92
PER	0.98	0.99	0.99
weighted avg	0.98	0.99	0.98

Fine-Tuned base-bert with Legal-BERT

Label	Precision	Recall	F1-score
LOC	0.91	0.91	0.91
MISC	0.71	0.75	0.73
ORG	0.82	0.85	0.83
PER	0.95	0.94	0.94
weighted avg	0.87	0.88	0.87

- Training ‘en_core_web_lg’ with manually created context data led to overfitted model.
- Legal-BERT showed better results, but proved difficult to tag generated labels to respective labels.

	entity	label	report_number
0	Cedric Baileydric	PER	2020-0001354
1	Bailey	PER	2020-0001354
2	CO	ORG	2020-0001354
3	Airport Operations North - Unit	ORG	2020-0001354
4	the Chicago Police Academy	ORG	2020-0001354
5	IP	ORG	2014-1068753
6	A Sub 1	PER	2014-1068753
7	XX N . Hermitage	addr	2014-1068753
8	Officer A	PER	2014-1068753
9	Department	ORG	2014-1068753
10	CP	ORG	2017-1084536
11	Department	ORG	2017-1084536
12	John Graham Police	PER	2017-1084536
13	Gang Investigation Division	ORG	2017-1084536
14	Jason Acevedo Edwards	PER	2017-1084536
15	Office	ORG	2017-1084536
16	Police Authority	ORG	2017-1084536
17	Federal Bureau	ORG	2017-1084536
18	FBI	ORG	2017-1084536
19	Ben	PER	2017-1084536
20	7843 S . Her 3	addr	2016-1078999
21	the Civilian Office of Police Accountability	ORG	2016-1078999

Extracting Named Entities Using ‘ctrlbuzz/bert-addresses’

This model that extracts three types of entities (persons, organizations and addresses) proved most useable.

Using fuzzy matching with true labels:

Person Recognition Accuracy: 92.8%*

Address Recognition Accuracy: 86.5%*

*assuming match with fuzzy score of 85 or above

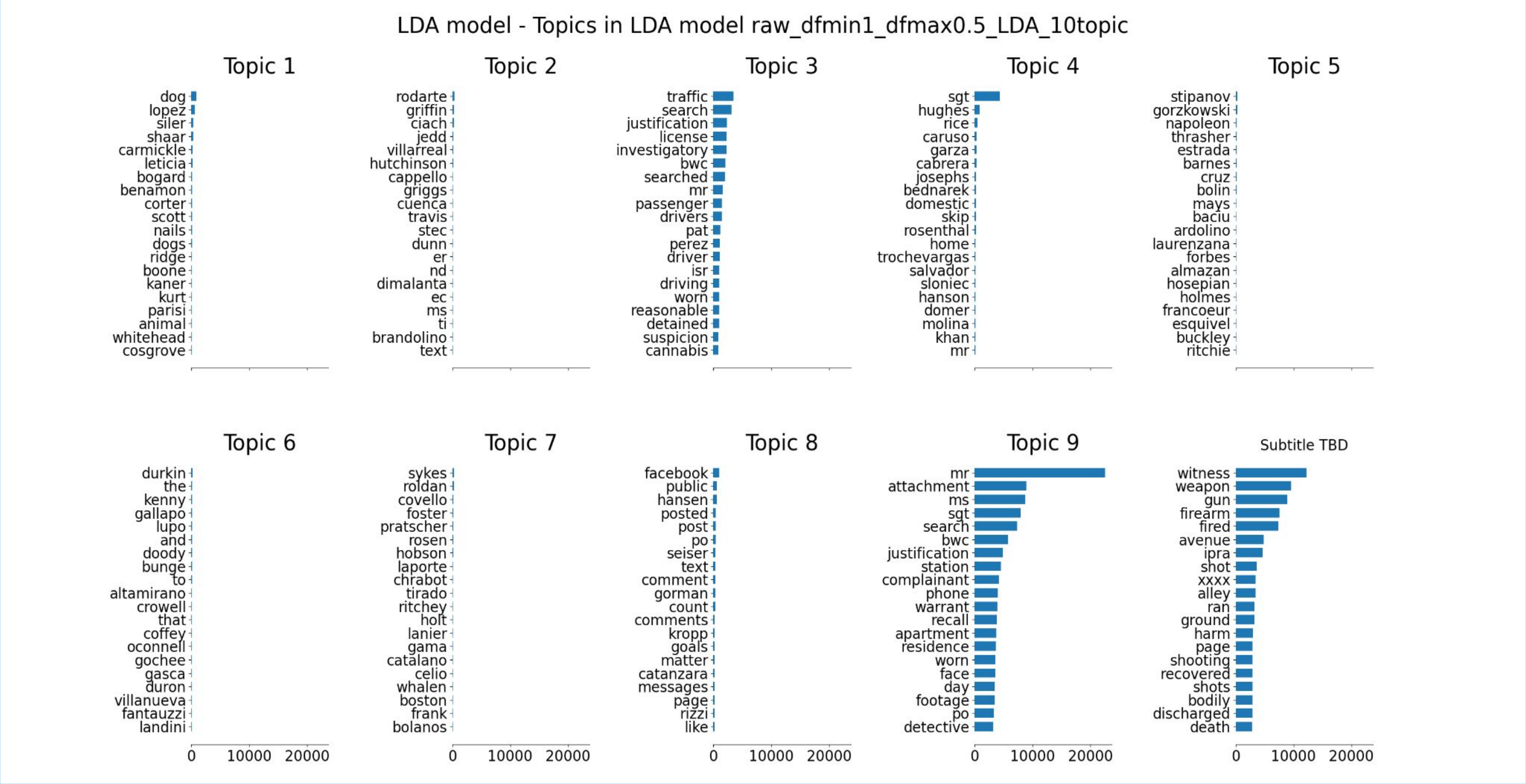
Topic modeling (Matt): three “families” of model

Model	**Type of assignment**	**Based on**
Latent Dirichlet Allocation (LDA)	probabilistic	Raw term frequency (BoW)
Non-Negative Matrix Factorization (NMF)	probabilistic	tf-idf
K-means clustering	hard	tf-idf
Truncated SVD / latent semantic analysis (LSA) k-means	hard	tf-idf
BERTopic	Hard (one topic per doc)	embeddings (fine-tuning)
Top2Vec	Hard	embeddings (fine-tuning)

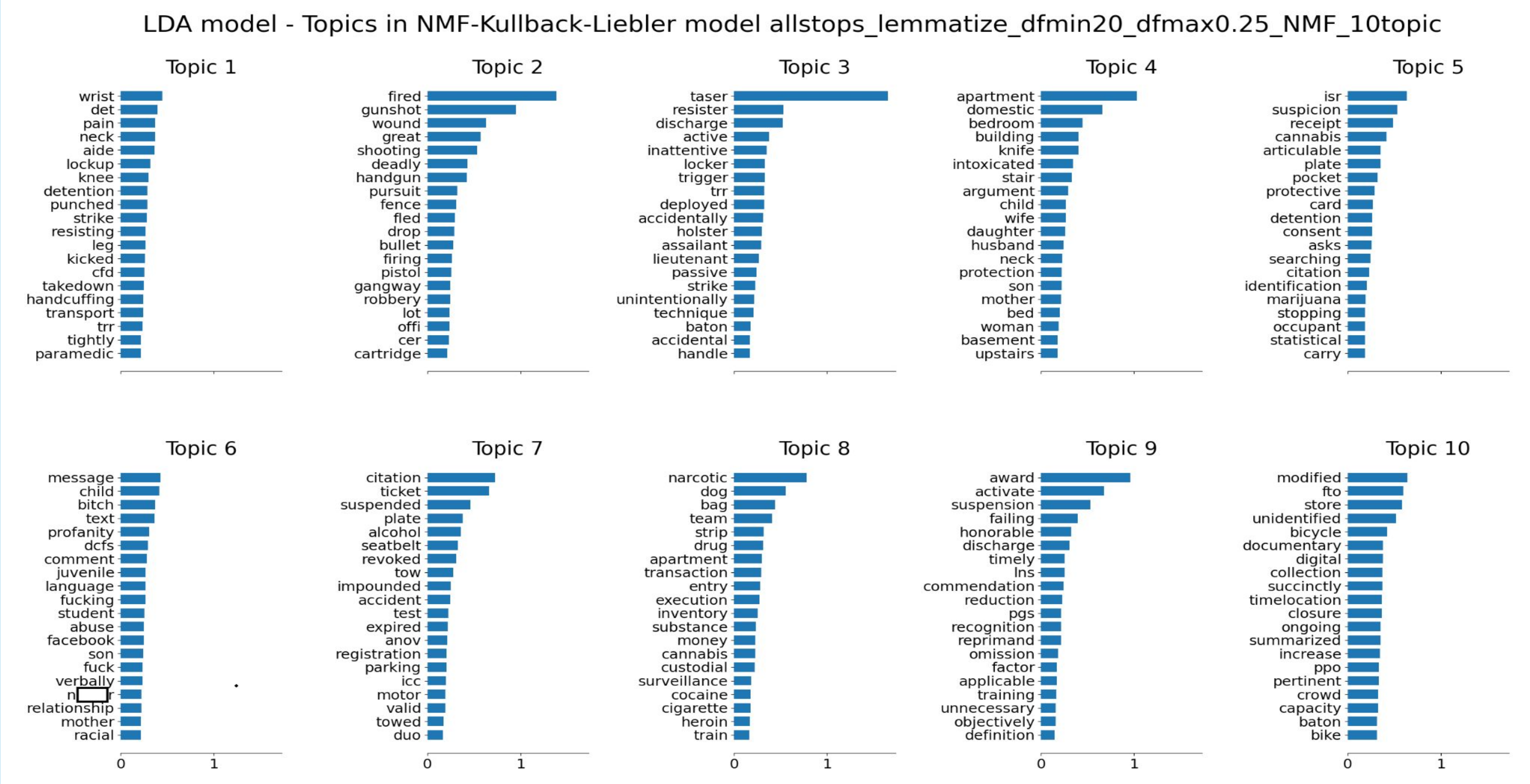
Topic modeling: results

- Big grid search → eyeball results → manually review best-performing options → further, visualization-assisted hyperparameter tuning
- LDA and NMF: **best *<u>if</u>*** *stopword removal <u>very aggressive</u>*
  - names, common words, titles, 2-letter words, “xxx/bbb”...
  - min_df > 1: eliminate words from only one document
  - max_df <= 0.25: limit to words in 25% or fewer of the documents
  - “Shortcut”: Invisible Institute names list and Chicago Open Data portal street names list
  - Probabilistic modeling best fit for multi-allegation and topic-ambiguous reports
- K-means: too much variation based on initial random seed
- Truncated SVD/LSA: more stable, but still prone to high topic overlap even slightly off peak “silhouette score”
- BERTopic and Top2Vec: initial attempts <u>very</u> disappointing
  - Found too few topics, often arbitrary and full of nonsense
  - Embedding based models were “least bad” with minimal stopword removal

Topic modeling: Let’s Go Look At Some Plots (bad run)



Topic modeling: Let's Go Look At Some Plots (better)



Topic modeling: Let’s Go Look At Some Plots

INTERTOPIC DISTANCE NOTEBOOK

https://github.com/FedericoDM/NLP-Police-Complaints/blob/main/src/exploratory/matt_good_topic_model_evaluation.ipynb

Topic modeling: Why did BERTopic (and Top2Vec) Fail So Hard?

- Layers of “Lego block” tower not set up right?
- Reports way longer than 512 tokens / models designed for sentence-length
- Couldn’t put bounds on number of topics generated
- Can’t do some forms of fine-tuning with unlabeled data

Example Top2vec output

```
print(model_l_pp.get_num_topics())
topic_sizes, topic_nums = model_l_pp.get_topic_sizes()
topic_words, word_scores, topic_nums = model_l_pp.get_topics()
print(topic_words)
```

```
2
[['kn' 'thrashers' 'speeding' 'malfunction' 'point' 'bruise' 'yanked'
  'thrasher' 'commotion' 'truth' 'n[redacted]rs' 'contusion' 'ay' 'points'
  'headlights' 'idiot' 'almost' 'excuse' 'denton' 'headlight' 'bump'
  'plead' 'disprove' 'whi' 'outcry' 'sudden' 'mad' 'said' 'motherfuckers'
  'poi' 'general' 'fail' 'arr' 'memory' 'eberhart' 'cited' 'answering'
  'falsified' 'th' 'mb' 'disobeying' 'bruising' 'deny' 'garner'
  'malicious' 'pl' 'refute' 'flash' 'effort' 'summit']
 ['uuw' 'giiib' 'ement' 'vaci' 'borjas' 'parisi' 'depa' 'accuseds'
  'encarnacions' 'caraballo' 'shafer' 'perezs' 'hreno' 'cers' 'altenbach'
  'obser' 'arekat' 'tohatan' 'rimsky' 'flailed' 'denie' 'kessem' 'bject'
  'incar' 'memorialized' 'haran' 'inventoried' 'esquivel' 'conlan'
  'zamorano' 'neylon' 'psik' 'uldrych' 'offic' 'durkin' 'cascone'
  'sloniec' 'brideson' 'loretto' 'accountabili' 'tion' 'nolle' 'stec'
  'ipras' 'radomski' 'grazziano' 'alsely' 'reentered' 'arguendo'
  'leavitt']]
```

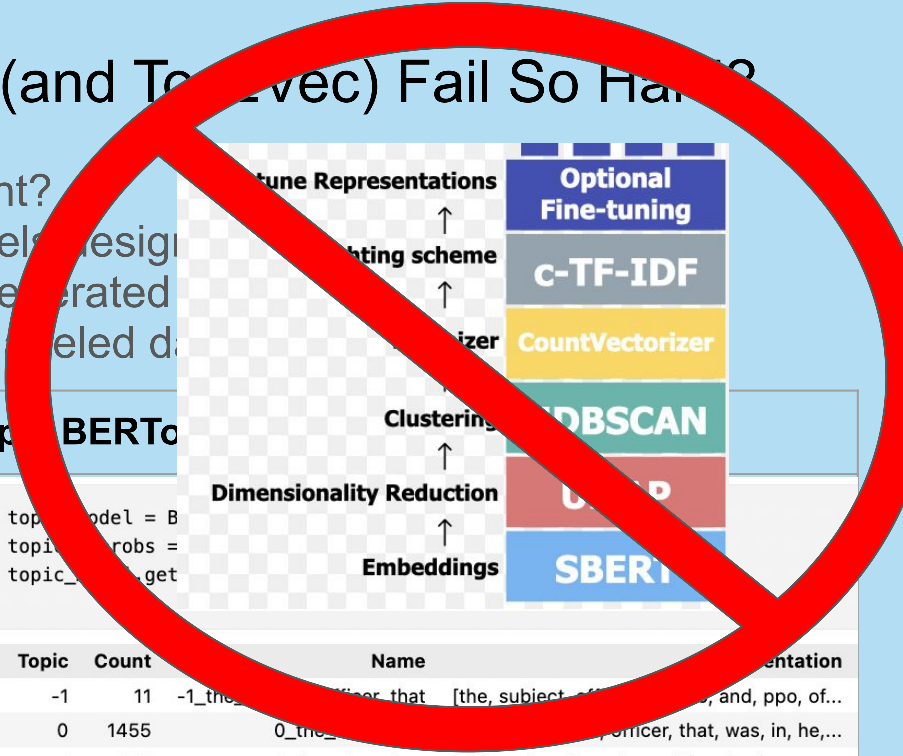
Example BERTopic output

```
topic_model = BERTopic()
topics, probs = topic_model.fit_transform(corpus)
topic_model.get_topic_info()
```

	Topic	Count	Name	Representation
0	-1	11	-1_the_subject_officer_that	[the, subject, officer, that, to, and, ppo, of...
1	0	1455	0_the_and_to_of	[the, and, to, of, officer, that, was, in, he,...
2	1	351	1_the_of_and_to	[the, of, and, to, that, officer, is, in, copa...
3	2	212	2_the_subject_and_officer	[the, subject, and, officer, to, witness, of, ...
4	3	36	3_the_subject_and_officer	[the, subject, and, officer, to, of, that, was...
5	4	30	4_nan___	[nan, , , , , , , , ,]
6	5	26	5_the_to_of_that	[the, to, of, that, officer, and, in, on, he, or]
7	6	15	6_taser_subject_the_to	[taser, subject, the, to, officer, and, that, ...
8	7	12	7_subject_the_officer_and	[subject, the, officer, and, to, of, his, in, ...

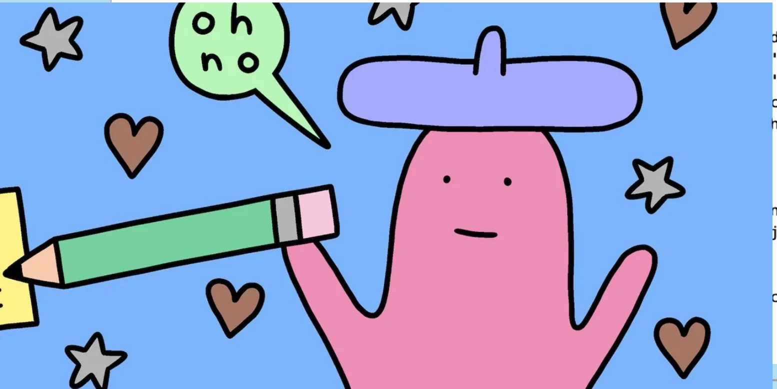
Topic modeling: Why did BERTopic (and To 2vec) Fail So Ha ?

- Layers of "Lego block" tower not set up right?
- Reports way longer than 512 tokens / models desig
- Couldn't put bounds on number of topics ge erated
- Can't do some forms of fine-tuning with unla eled da



Example Top2vec output	Examp BERTo

```
print(model_l_pp.get_num_topics())
topic_sizes, topic_nums = model_l_pp.get_topic_sizes()
topic_words, word_scores, topic_nums = model_l_pp.get_topics()
print(topic_words)
```



	Topic	Count	Name	entation
0	-1	11	-1_the	[the, subject	, and, ppo, of...
1	0	1455	0_the	officer, that, was, in, he,...
2	1	351	1_the_of_and_to	[the, of, and, to, that, officer, is, in, copa...
3	2	212	2_the_subject_and_officer	[the, subject, and, officer, to, witness, of, ...
4	3	36	3_the_subject_and_officer	[the, subject, and, officer, to, of, that, was...
5	4	30	4_nan___	[nan, , , , , , , , ,]
6	5	26	5_the_to_of_that	[the, to, of, that, officer, and, in, on, he, or]
7	6	15	6_taser_subject_the_to	[taser, subject, the, to, officer, and, that, ...
8	7	12	7_subject_the_officer_and	[subject, the, officer, and, to, of, his, in, ...

Summarization (Fede)

Goal: Create summary that captures essential information of a report.

Three models used:

i) T5 Small Transformer (finetuned for summarization).

ii) Facebook BART Large (finetuned with CNN Daily data)

iii) Google FLAN-T5 Large model

Random summaries to detect which model was best: BART was the winner

Summarization Models - Performance

- **BART** performed **the best** → Followed by T5-Small (finetuned for summarization)
- **FLAN-T5** did **poorly** (probably because it was not finetuned)
- Tweaked parameters for BART

Parameters are:

```
max_length=1200
min_length=40
length_penalty=2.0
no_repeat_ngram_size=2
num_beams=4,
early_stopping=True
```

Example: BART-Large

Model: facebook/bart-large-cnn

Summary: *Officer charles sykes and sergeant dennis graber engaged in a verbal altercation that turned physical in that she was pushed to the ground. copa obtained an affidavit override from the chicago police department in order to proceed with this investigation. It is alleged that the accused was intoxicated while off-duty in violation of rule 15.*

Example: T5-Small

`Model: Falconsai/text_summarization`

`Summary:` *`officer unit of assignment 116, dob: 1976 gender: male, race: black, black 1 ii. allegations officer allegation finding / recommendation charles sykes sergeant dennis graber 1. it is alleged that on or about july 11, 2019 at or near the location of chicago, il, the accused was intoxicated and allowed him to operate a motor vehicle in violation of rule 3 & 6. 4. he failed to provide adequate police service in his interaction with involved`*

Example:FLAN-T5-Large

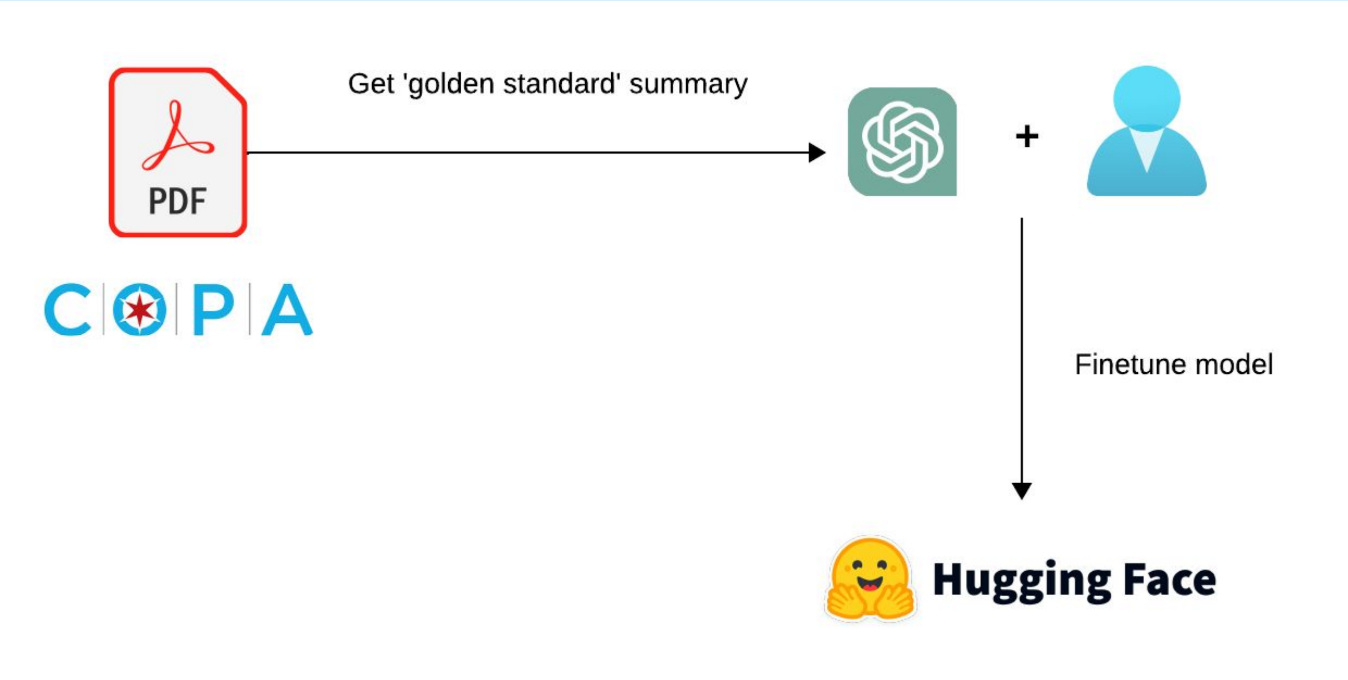
`Model: google/flan-t5-large`

`Summary:` *sykes is an alcoholic and blacks out when he drinks.10.'responding officers activated their body-worn cameras (bwc) 2 as they investigated the incident and interacted with the involved parties.*

Finetuning BART-Large

Create 25 ‘ideal’ summaries and finetune model

Outputs are similar, with slight differences



Putting it Together: do best summary & topic model match?

- Looked at outputs of probabilistic topic model (NMF-KL) on same 25 document mini-corpus as fine-tuned summaries
- Sadly, highest-probability topics often unexpected / didn't reflect (top 30 words in) topic. E.g.:

Our human-written summary	**ML summary**	**Topic model (15-topic NMF-KL)**	**First topic top words**	**Second topic top words**
On July 25, 2011, Officer A shot Subject 1, who was fleeing on foot and allegedly pointed a firearm at the officer. Despite Subject 1's claim of being unarmed, evidence supported the officer's account, including recovery of a loaded pistol from Subject 1. Subject 1 was treated for gunshot wounds and charged, later found guilty of Aggravated Unlawful Use of a Weapon but acquitted of Aggravated Assault against a Peace Officer. Witness accounts varied, with some contradicting the police narrative. The investigation concluded Officer A's use of deadly force was justified, aligning with department policy for preventing harm or resisting arrest under threat.	The incident occurred on 25 July 2011 at 600 n. sawyer avenue, chicago. The suspect, 21-year-old "subject 1," was taken to a local hospital with two gunshot wounds to his upper left back and left buttock. He was later charged with aggravated assault and unlawful use of a weapon. An investigation was launched by the Chicago Police Department.	Topic 4: domestic violence / violence against women (84%) Topic 8: department awards/merit (16%)	Apartment, domestic, bedroom, stair, entry, basement, upstairs, argument, living, bathroom, wife, daughter, child, hallway, son, knocked, husband, bed, downstairs, porch, neck, damage, neighbor, mother, threw, woman, kicked, entering, family, lieutenant	Award, activate, suspension, failing, isr, honorable, receipt, recognition, reprimand, reduction, commendation, omission, train, line, timely, detention, responsibility, deactivated, suspicion, meritorious, performance, emblem, attendance

Avenues for Future Work: Topic Modeling / NER

- Improve stemming/lemmatization (can this be trained?)
- Further stopword tailoring, including with NER results on corpus documents
- Consider **hierarchical** topic modeling (for e.g. different kinds of violence)
- Cross-reference names across documents to tally officers with multiple allegations, street addresses for geographic distribution of allegations, etc.
- Save timestamp of corpus documents, track topic distribution over time
  - IPRA vs. COPA - did relative weight of topics change when administrative body changed?

Avenues for Future Work: Summarization

- Overcome BART's input limitation
- Create a larger dataset for fine tuning (~10% of dataset)
- Finetune T5-Small → may be better than BART
- Customized preprocessing for reports of different years
- Separate report into sections, summarize each, then put together (so all major kinds of info end up in all summaries)

- Improve/automate unified pipeline: select mini-corpus → make summaries → run topic model with NER stopwords → compare → tune all models until outputs converge / match actual document

Thank you! Questions?

Photo credit: Spencer Bibbs, Hyde Park Herald (May 2020), https://tinyurl.com/mr2d6738