Computational Approaches to Narratives

Maria Antoniak

University of Copenhagen, Pioneer Centre for AI
University of Colorado Boulder, Computer Science (Fall 2025)

Hello! I'm Maria Antoniak

I've been working in NLP for a decade both in academia and industry.

Right now, I'm a postdoc at the Pioneer Centre for AI at the University of Copenhagen. Next year I'll be starting as a professor of Computer Science at the University of Colorado Boulder.

There is a lot going on in AI right now. But for our fields of text analytics, text as data, cultural analytics, computational social science — many of the research questions have stayed the same.

Tell me about yourself!

Mentimeter report from the workshop:

https://maria-antoniak.github.io/resources/2025-04-24_narrative_workshop_COMPTEXT_menti.pdf

Narratives drive large online conversations

M.J. Crockett @mjcrockett.bsky.social · 2h
Dispatch from Silicon Valley: the narrative of "AI as Oracle" is everywhere.

Writing in Nature Machine Intelligence, I updated the Delphic maxims to help us cut through the hype:

1. Know thy training data
2. Nothing in excess
3. Certainty brings ruin

www.crockettlab.org/s/s42256-024...

https://www.crockettlab.org/s/s42256-024-00970-z-1.pdf
www.crockettlab.org

💬 2   🔁 13   ♡ 28   ...

Amanda Jean Stevenson @amandaj... · 21h
Without a backstop like the law, a counter-narrative is the only way to address a firehouse of lies and nonsense.

Gish gallop - Wikipedia
en.m.wikipedia.org

The Seattle Times @theseattl... · 2d
"He threw out so many nonsense, nonsensical questions and statements that it was head-spinning," Washington's senator said of her

Government's wins are often invisible: Systems that avoid plane crashes; alliances that avert war; surveillance that prevent pandemics.

Government wins are often *the avoidance of loss.*

So how do we tell the story of the destruction of government? The story of future losses *not* averted?

February 16, 2025 at 4:30 PM 👥 Everybody can reply

501 reposts   **79** quotes   **1.6K** likes

💬 53   🔁 580   ♥ 1.6K   ⬆   ...

Julia Angwin @juliaangwin.com · 5h
This question has been haunting me ever since Elon & co began their govt wrecking spree.

Journalists are crippled in telling this tale. We cannot tally the wins because we cannot ever fully know why a war didn't start or a plane didn't crash.

snackowski
@snackowska.bsky.social

i guess the America-as-asylum narrative has lost to the America-as-gated-community narrative. (they are two versions of the same narrative)

January 22, 2025 at 10:09 AM 👥 Everybody can reply

Kat Tenbarge
@kattenbarge.bsky.social   + Follow

When we see narratives that blame the most vulnerable people for problems caused by the most powerful, I think it's important to remember that this viral spread of misinfo is a more pervasive version of what the powerful have always done to defend themselves, with tech they steered in this direction

January 9, 2025 at 3:54 PM 👥 Everybody can reply

Cultural Analytics

"[Cultural analytics is] the computational study of cultural objects, processes, agents"
– Journal of Cultural Analytics

Big data, computational methods, research questions about culture. *(No, not primarily focused on comparisons across countries or languages!)*

Traditional unsupervised NLP methods are still important: topic modeling, word embeddings, lexica

LLMs are both unlocking new ways to study human culture *and* creating new complications for human culture.

Ongoing challenges LLMs in cultural analytics

Companies take and use unknown amounts of cultural works for pretraining, leading to legal, ethical, and evaluation issues.

High quality annotations are still necessary, but the object of study is often subjective, ambiguous, difficult to define.

NLP people are increasingly are interested in cultural questions, often lack interdisciplinary expertise and don't involve humanists in their studies of human culture.
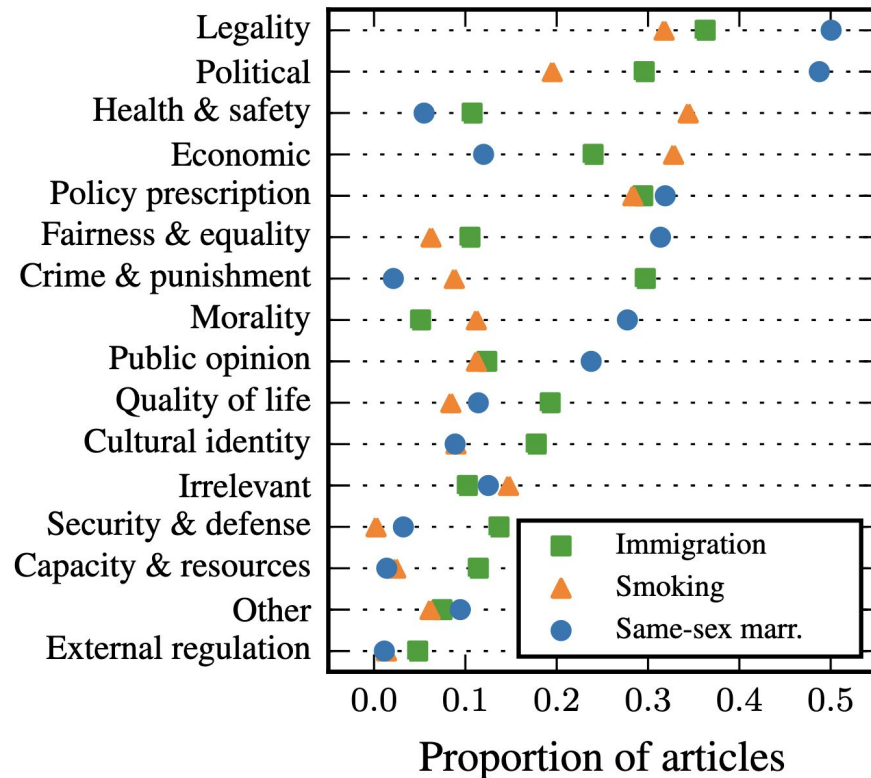
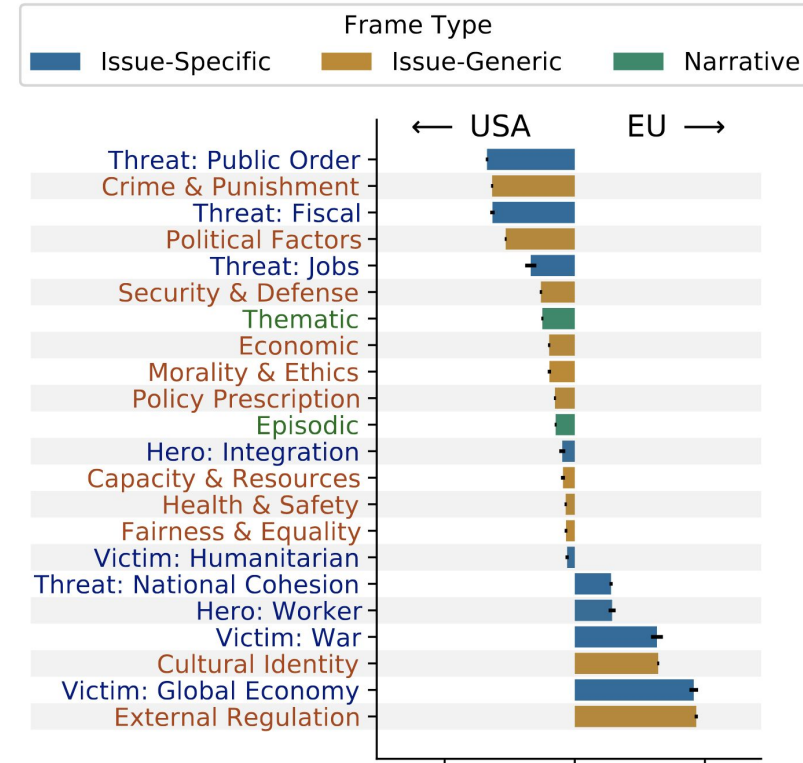# Computational studies of narratives

Narratives have been studied in all kinds of ways and in all kinds of mediums, and all kinds of NLP methods have been used to study narratives at scale.

But what are "narratives"? This is an overloaded term.
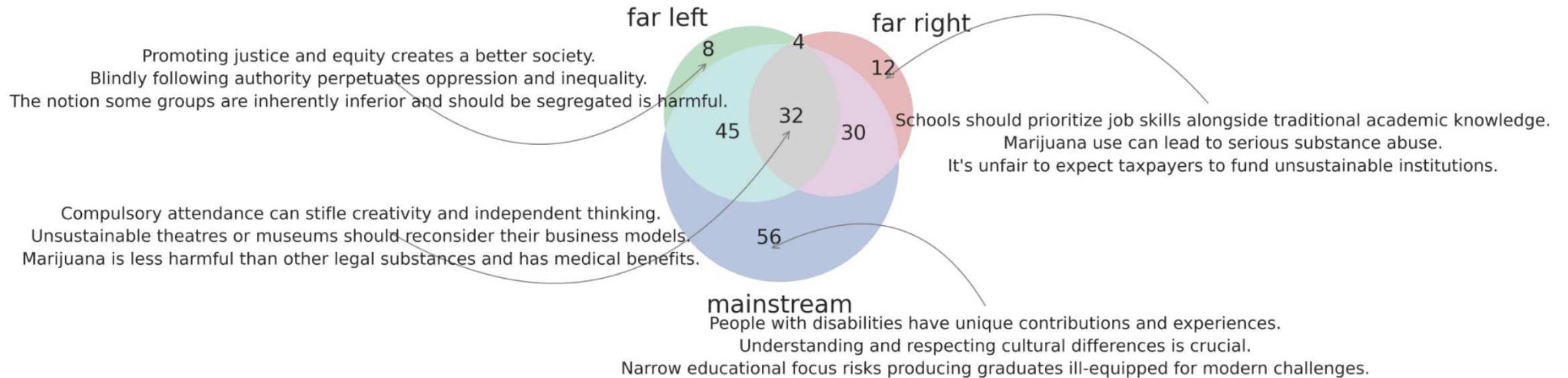
Narratives as *frames*

"The Media Frames Corpus: Annotations of Frames Across Issues"
Card, Boydstun, Gross, Resnik, Smith; ACL 2015

"Modeling Framing in Immigration Discourse on Social Media"
Mendelsohn, Budak, Jurgens; NAACL 2021

Narratives as *tropes*

far left

far right

8
4
12

Promoting justice and equity creates a better society.
Blindly following authority perpetuates oppression and inequality.
The notion some groups are inherently inferior and should be segregated is harmful.

32

45

30

Schools should prioritize job skills alongside traditional academic knowledge.
Marijuana use can lead to serious substance abuse.
It's unfair to expect taxpayers to fund unsustainable institutions.

Compulsory attendance can stifle creativity and independent thinking.
Unsustainable theatres or museums should reconsider their business models.
Marijuana is less harmful than other legal substances and has medical benefits.

56

mainstream

People with disabilities have unique contributions and experiences.
Understanding and respecting cultural differences is crucial.
Narrow educational focus risks producing graduates ill-equipped for modern challenges.

"LLM Tropes: Revealing Fine-Grained Values and Opinions in Large Language Models"
Wright, Arora, Borenstein, Yadav, Belongie, Augenstein; EMNLP Findings 2024

Narratives as *shapes*
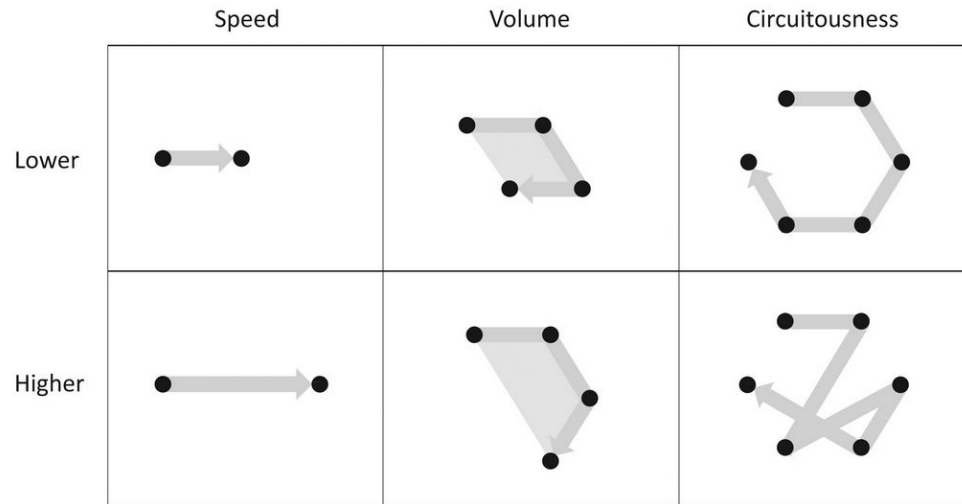
"How quantifying the shape of stories predicts their success"
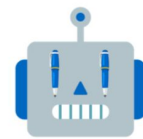Toubia, Berger, Eliashberg; PNAS 2021

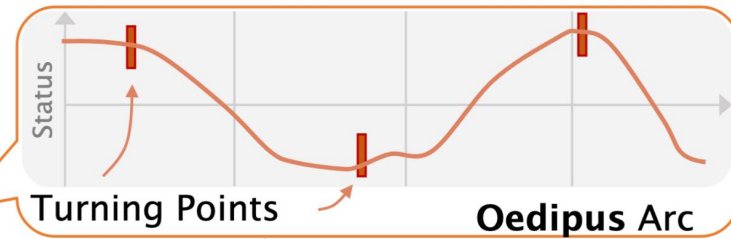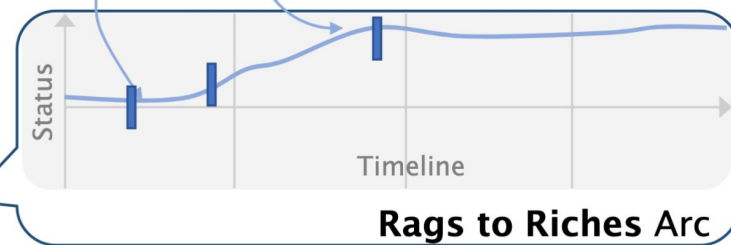"Are Large Language Models Capable of Generating Human-Level Narratives?"
Tian, Huang, Liu, Jiang, Spangher, Chen, May, Peng; EMNLP 2024

Narratives as *stories*

A	Someone
B	tells
C	someone
D	somewhere that
E	someone
F	did something(s)
G	[to someone]
H	somewhere
I	at some time
J	for some reason

"Narrative Theory for Computational Narrative Understanding"
Piper, So, Bamman; EMNLP 2021

Dracula: Arthur and Lucy

TIME

RMN — murder, worship, suffering, sadness, outdoors, sadness, politics, education, love, boats

HTMM — love, royalty, fantasy, love
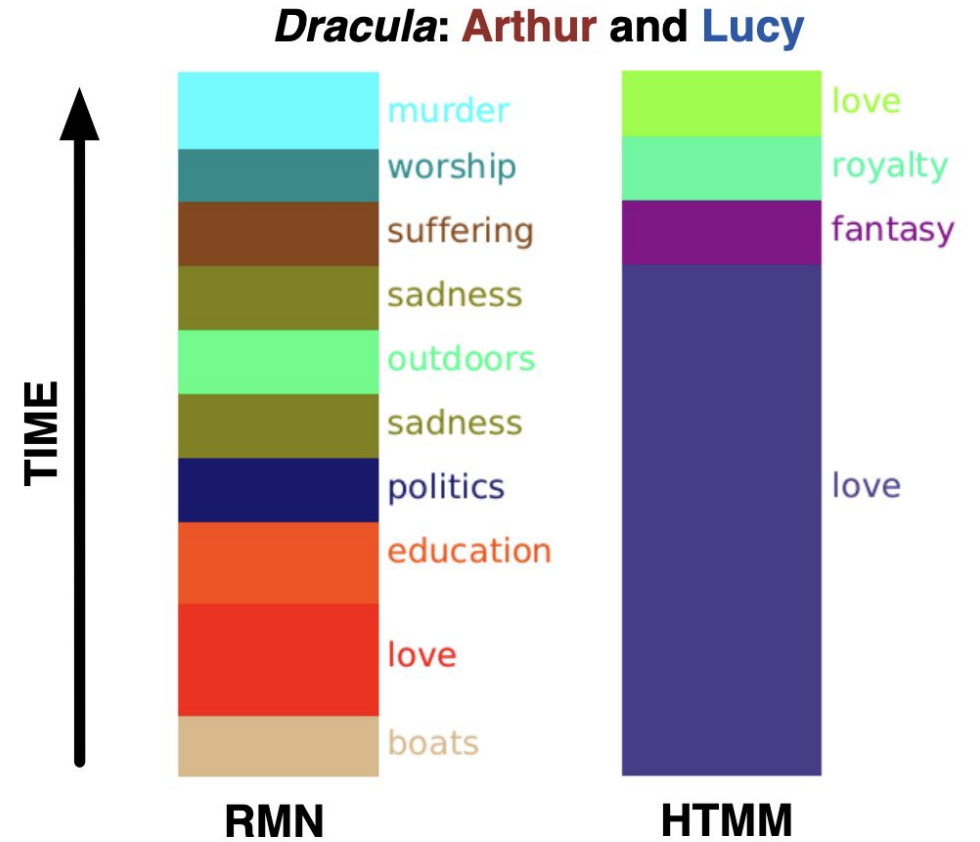
"Feuding Families and Former Friends: Unsupervised
Learning for Dynamic Fictional Relationships"
Iyyer, Guha, Chaturvedi, Boyd-Graber, Daume; NAACL 2016

Why should we care about online storytelling?
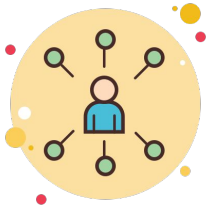
As a **rhetorical** strategy, storytelling can drive social movements (#MeToo, #BLM), spread misinfo (vaccine side effects), educate (Ted Talks, podcasts)

As a form of **self-disclosure**, personal storytelling can strengthen social bonds, build trust between community members, and benefit the storyteller
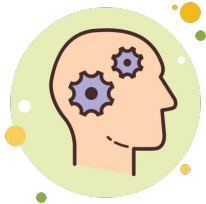
As a **sensemaking** strategy, communal storytelling can help individuals and groups learn from their shared experiences

For **NLP** researchers, modeling narratives is a very challenging task that is also important for better understanding our **pretraining** and **prompting** data

Jourard and Lasakow. "Some factors in self-disclosure." The Journal of Abnormal and Social Psychology, (1958).

Hill and Stull. "Gender and Self-Disclosure." 1987.

Zhao, Hinds, and Gao. "How and to whom people share: the role of culture in self-disclosure in online communities." CHI (2012).

Pennebaker and Beall. "Confronting a traumatic event: Toward an understanding of inhibition and disease." Journal of Abnormal Psychology (1986).

Merz, Fox, Malcarne. "Expressive writing interventions in cancer patients: a systematic review." Health Psychol Review (2014).

Tangherlini. "Heroes and Lies: Storytelling Tactics among Paramedics." Folklore (2000).

Weick, Sutcliffe, Obstfeld. "Organizing and the process of sensemaking." Organization Science (2005).

Genius and Bronstein. "Looking for "normal": Sense making in the context of health disruption." Journal of the Association for Information Science and Technology (2017).
Ma et al. "Self-disclosure and perceived trustworthiness of Airbnb host profiles." CSCW, 2017.

Gallagher et al. "The Networked Disclosure Landscape of #MeToo." *CSCW*, 2019.

Tangherlini. "Heroes and lies: Storytelling tactics among paramedics." *Folklore*, 2000.

Yang, Kraut, Smith, Mayfield, and Jurafsky. "Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities." CHI, 2019.

Ayers, Poliak, and Dredze. "Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum." JAMA, 2023.

Online discussions ABOUT storytelling

The study of **literary reception**, **genre formation**, and **canonization** is well established but has traditionally focused on elite readers.

Online reading communities like Goodreads and TikTok provide abundant new evidence documenting diverse readers' perceptions of books in detail and in their own words.

Online book reviewers shape our offline reading landscape.

OK, we care about stories...

...but what are stories?

Mentimeter

What features of a text make it a story?

Exercise

1. Break into small groups (3-4 people per group)
2. Discuss how you would define "story"
3. Create lists of story and non-story examples

Mentimeter

What is your group's definition of "story"?

Where do people tell stories online?

Story Detection Across Online Communities

Maria Antoniak, Joel Mire, Maarten Sap, Elliott Ash, Andrew Piper

ACL 2024

Research Questions

What is a story?

How can we automatically detect identifying storytelling?

How can we do this at scale?

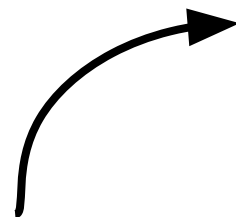**Where do people tell stories online?**

r/breastcancer

r/cancer

r/leukeumia

... etc.

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

r/breastcancer

r/cancer

r/leukeumia

... etc.

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed. She'd had cancer before years, and she I can't stop dering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This requires building and evaluating a story span detection system.

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

This could be an upsetting topic so feel free to not read, but I need some help from people in a similar situation. My aunt passed away from breast cancer the week before I was diagnosed last year. She'd had cancer before but it came back after 10 years, and she passed three years later. I can't stop thinking about her and wondering and feeling afraid. Does anyone else feel like this? How do you handle those worries?

22

Challenges to story detection across communities

Stories and non-stories are mixed together **across** topics and **within** individual texts

Defining stories is notoriously difficult: there is an entire literary field (*narratology*) concerned with the theory of narratives!

Existing story **datasets** do not meet our requirements

# Can we use existing story datasets?

No, existing datasets are unavailable or do not meet our desiderata.

Existing story datasets

- are **not in English** (dos Santos et al., 2017)

- are **not publicly available** (Ganti et al., 2022; 2023)

- contain **very sparse** positive labels (ICWSM 2009 Spinn3r Corpus)

- include **topical confounders** (Prize, 2019)

- focus on **literary fiction** (Piper & Bagga, 2023)

Can we adopt an existing story definition?
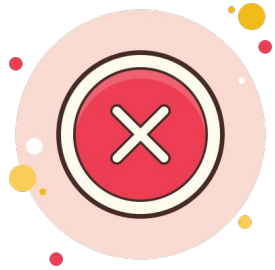
No, there's no canonical codebook.

Many different definitions have been proposed in both NLP and narratology, relying on many different features.

These definitions weren't designed for annotation, nor were they designed for online communities.

We found that crowdworkers were very inconsistent.

What are stories, according to NLP?

"A narrative is a discourse presenting a coherent sequence of events which are causally related and purposely related, concern specific characters and times, and overall displays a level of organization beyond the commonsense coherence of the events themselves. In sum, ***a story is a series of events effected by animate actors****... at least two key elements to stories, namely, the plot (fabula) and the characters (dramatis personae) who move the plot forward (Abbott, 2008)."*

"A Simpler and More Generalizable Story Detector using Verb and Character Features." Eisenberg and Finlayson (2017).

What are stories, according to NLP?

*"It is generally agreed in narratology (Forster, 1962; Mani, 2012; Pentland, 1999; Bal, 2009) that a narrative presents a **sequence of events** arranged in their time order (the plot) and involving **specific characters** (the characters)."*

"Temporal Event Knowledge Acquisition via Identifying Narratives." Yao and Huang (2018).

What are stories, according to NLP?

*"A narrative is the recounting of a **sequence of events** that have a **continuant subject** and **constitute a whole** (Prince, 2003)."*

"Fabula Entropy Indexing: Objective Measures of Story Coherence."
Castricato et al. (2021).

What are stories, according to NLP?

*"A sequence of related **events**, leading to a **resolution** or projected resolution."*

"A Semantic Triplet Based Story Classifier." Ceran et al. (2012).

What are stories, according to NLP?

*"**Situatedness:** narrativity depends on the social context in which it occurs*

***Event sequencing:** narrativity depends on temporally ordered events*

***World making**: narrativity depends on the fact of disequilibrium such that we can observe a change in the world"*

"Detecting Narrativity Across Long Time Scales." Piper et al. (2021).

What are stories, according to NLP?

*"A STORY follows a **plot-like structure** (e.g. has an introduction, middle section or conclusion) or contains a **sequence of events**."*

"StoryARG: a corpus of narratives and personal experiences in argumentative texts." Falk and Lapesa, ACL 2023.

What are stories, according to NLP?

sequences of events arranged temporally

Piper et al. (2021)
Yao and Huang (2018)
Castricato et al. (2021)

causally related events leading to resolutions

Eisenberg and Finlayson (2017)
Ceran et al. (2012)
Alzahrani et al. (2016)

entities or characters

Eisenberg and Finlayson (2017)
Piper et al. (2021)
Yao and Huang (2018)
Alzahrani et al. (2016)

rhetorical purpose

Eisenberg and Finlayson (2017)
Roos and Reccius (2021)
Castricato et al. (2021)

world building

Piper et al. (2021)

Story classification features used in prior work

Feature	Used In
n-gram	*Santos et al. (2017)*
	Piper et al. (2021)
	Gordon (2009)
	Ceran et al. (2012)
POS	*Yao and Huang (2018)*
	Piper et al. (2021)
	Ceran et al. (2012)
coreference chain length	*Eisenberg and Finlayson (2017)*
	Yao and Huang (2018)
LIWC	*Santos et al. (2017)*
	Yao and Huang (2018)
readability	*Santos et al. (2017)*
verb classes	*Eisenberg and Finlayson (2017)*

Story classification features used in prior work

(continued)

Feature	Used In
syntactic production rules	*Yao and Huang (2018)*
verb sequence perplexity	*Yao and Huang (2018)*
named entities	*Ceran et al. (2012)*
stative verbs	*Ceran et al. (2012)*
semantic triples	*Ceran et al. (2012)*
... etc.	*… etc.*

What are stories, in narratology?

Narrative is the technical term for story.

Story has a second technical definition: we can distinguish between "story" (*syuzhet*) and "discourse" (*fabula*), where story = what happened and discourse = how it is told.

Herman (2009) writes, "Narrative roots itself in the lived, felt experience of human or human-like agents interacting in an ongoing way with their surrounding environment."

Propp. "Morphology of the Folktale." 1929.

David Herman. 2009. Basic elements of narrative. John Wiley & Sons.

... so what are stories?

- Many different definitions — it's complicated!

- A common thread: linear sequences of events, a cast of characters, and a rhetorical goal

- Distinction between stories and narratives that we covered earlier (framing, tropes, etc.)

Building a story detection dataset

Developing a codebook

A computer scientist, a literary scholar, and an economist walk into a bar... and try to agree on a story definition.

How can we define storytelling, in the context of Reddit?

Developing a codebook

Many interations of open coding and discussions!

Drew from prior work:

- "Narrative Theory for Computational Narrative Understanding" (Piper et al., 2021)
- "Literary Event Detection" (Sims et al., 2019)

Adapted to the Reddit context: more personal stories, less worldbuilding, medium length texts.

🔭 StorySeeker Codebook

A story describes a sequence of events involving one or more people.

A story could be **as short as one sentence** or **as long as an entire post**.

🔭 StorySeeker Codebook

A story describes a sequence of events involving one or more people.

A story could be **as short as one sentence** or **as long as an entire post**.

| Who/what qualifies as "people"? |
| What are "events"? |
| What are "sequences" of events? |

➡️

Our full codebook is

available on Github

🔭 StorySeeker Codebook

Who/what qualifies as "people"?

- "People" can include animals, aliens, etc.
- "People" can include groups as long as these are specific groups of people that exist at a specific time and place.
- "People" includes the first person narrator.

🔭 StorySeeker Codebook

What are "events"?

- Events are "a singular occurrence at a particular place and time" (Sims et al., 2019).
- **General**, **repeating**, **isolated**, or **hypothetical** situations, states, and actions are usually not events, unless they appear together in a strongly story-like sequence.
- Most stories are told in the past tense. **Present** and **future** tense can also be used, but the bar is higher and the narrated events need to be strongly story-like.
- Most events are positively asserted as occurring, but depending on the context, **negative verbs can also be events** when occurring at a specific time and place.
- Events are usually verbs but can also be **nouns** and **adjectives**.
- When are **states** events? See Sims et al., 2019.

🔭 StorySeeker Codebook

What is a "sequence" of events?

- These events should be **sequential**: one event happens, then another event happens. It's ok if the events are narrated out of order, but there should still be a clear sequence.
- These events should be **connected**: they might be about the same people, they might be causally connected, they might describe an overall change or transformation in the state of the world, they might describe a single experience.
- Jumbles of events that are unordered and/or unconnected (like lists of examples) are not stories.

Mentimeter

Questions and reactions to the StorySeeker codebook

What data should we annotate?

The Reddit Webis-TLDR-17 dataset

- 3,848,330 posts and comments

- average length of 270 words

- each with a "TL;DR" summary

- 2006 to 2016

We sample from this data, creating balanced sets of texts across 500 of the most popular subreddits.

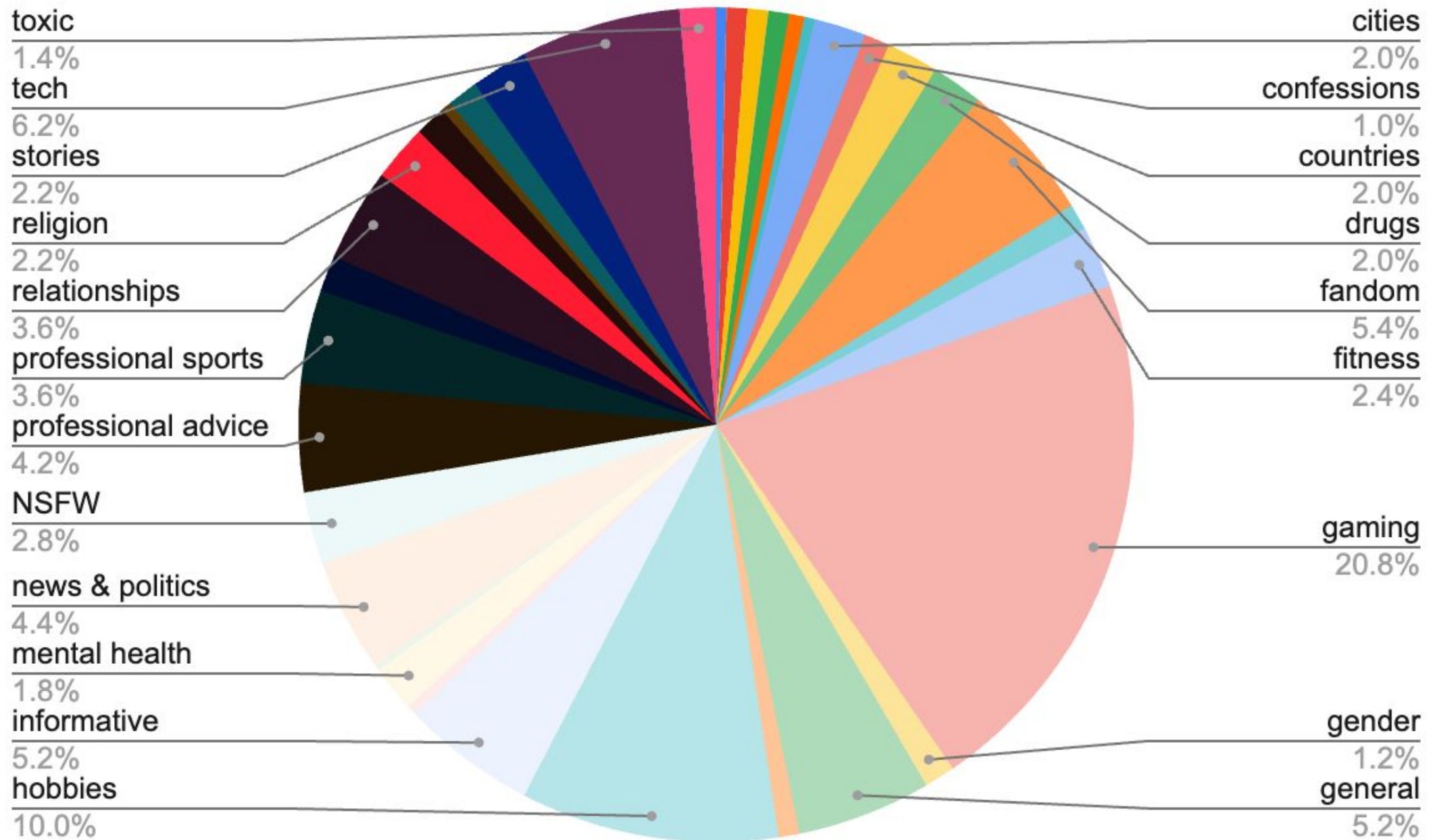`"TL;DR: Mining Reddit to Learn Automatic Summarization" (Völske et al., 2017)`

Categorizing 500 subreddits

We used open-coding to build a subject taxonomy of the 500 subreddits with the most posts+comments in the dataset.

Why?

- to protect our annotators
- to protect redditors
- to structure our results

toxic 1.4%

tech 6.2%

stories 2.2%

religion 2.2%

relationships 3.6%

professional sports 3.6%

professional advice 4.2%

NSFW 2.8%

news & politics 4.4%

mental health 1.8%

informative 5.2%

hobbies 10.0%

cities 2.0%

confessions 1.0%

countries 2.0%

drugs 2.0%

fandom 5.4%

fitness 2.4%

gaming 20.8%

gender 1.2%

general 5.2%

STORY 1　EVENT 2

Comment: DRRR, gets a while to get used to. **To be honest, the first time I started**

STORY

EVENT

watching it I flat out dropped it after 5 episodes. I picked it up 3 months later and

EVENT　　　EVENT

finished the entire show and enjoyed it quite a bit. I personally found that it started really

EVENT　　　EVENT

slow (the entire first half (12 episodes) of the show is introducing us to the characters and

developing interpersonal relationships between characters), which got really draggy and

pretty boring at times. However the 2nd half of the show really picks it up in terms of

pacing / action and makes the show worthwhile and enjoyable for me. Also enjoying DRRR,

played into enjoying the second season of DRRR.

What do these stories look like?

I've thought about this for a while, but only decided to post when someone else brought up how weak they may be. Everyone assumes that the animatronics can walk at a leisurely speed. And that Foxy can run quite fast. Hypothetically, I guess that...

All over the Internet I've seen a ton of hype about this show, it didn't look particularly appealing to me but I said what the hell I'll watch. So I watched a few episodes and yeah… I mean it's okay for a kids show... I watched some other kids shows afterwards to compare… it didn't really stand out that much to me. So I'd honestly love to know...what's with all the hype?

Let me clarify. Every moment from age 15 when I <u>gave</u> my life to Jesus, to now (age 22), I have been a Christian. Sure, there <u>have</u> been moments of ignoring the Christian faith and living a "worldly" life, but my last 3 years in college have been completely devoted to campus ministry, and discipling others... Over the last few months though, I <u>have</u> been feeling more and more like this is kind of a lie. I am at a turning <u>point</u> in my life where I either need to give myself over completely to Christianity, or drop it and move on. Basically, I need resources to make a decision. Is there anyway you can help me, via personal stories that are similar, book recommendations, youtube links, etc.?

I'm a grad student for context. It starts out very general, you will spend a year or two really just catching up on literature since, you know, your advisor has been along for the ride for about 15 years so comparatively you're an academic infant. So you read a paper which answers questions, but also creates more questions for you. But that's ok and is par for the course, so you look up more papers which answer those questions, and this is a cycle for your entire career. At some point you will think, "why didn't they do this?" but this time… this wondrous time, you find there is no further information on the topic. But it's a good thing you spent the past two 2 years catching up on it, and now hopefully you can try and answer the question yourself. Now you read others papers to get answers, which leads to more questions. Suddenly it's your own set of questions and answers that drive the field. I am more typical I think and would say after about 12-18 months I started asking questions that didn't have really well worked out answers, and that's when the real work begins!

🔭 StorySeeker Dataset

Two expert annotators

Consensus discussions and decisions: hours upon hours of work!

Cohen's k agreement for **story spans**: 0.72

Cohen's k agreement for **event spans**: 0.65

235 stories, 267 non-stories in Reddit posts and comments

Mean span length is 92 tokens (comments) and 119 tokens (posts).

Building a story detector

Model	P	R	F1	P	R	F1
Document Classification		*Story*			*Not Story*	
SVM with TF-IDF	0.82 ± 0.06	0.69 ± 0.11	0.74 ± 0.08	0.77 ± 0.06	0.87 ± 0.03	0.81 ± 0.03
Fine-tuned RoBERTa	0.87 ± 0.02	0.85 ± 0.08	0.86 ± 0.04	0.88 ± 0.05	0.89 ± 0.03	0.88 ± 0.02
GPT-4 Zero-Shot	0.83 ± 0.02	0.76 ± 0.06	0.79 ± 0.03	0.80 ± 0.04	0.87 ± 0.02	0.83 ± 0.02
GPT-4 Few-Shot	0.84 ± 0.05	0.70 ± 0.06	0.76 ± 0.03	0.77 ± 0.04	0.88 ± 0.04	0.82 ± 0.03
GPT-4 C-o-T	0.56 ± 0.05	0.96 ± 0.03	0.71 ± 0.04	0.91 ± 0.06	0.35 ± 0.02	0.50 ± 0.03
GPT-3.5-Turbo Zero-Shot	0.93 ± 0.02	0.37 ± 0.08	0.53 ± 0.08	0.64 ± 0.04	0.97 ± 0.01	0.77 ± 0.03
GPT-3.5-Turbo Few-Shot	0.79 ± 0.06	0.53 ± 0.10	0.63 ± 0.09	0.68 ± 0.07	0.88 ± 0.03	0.77 ± 0.05
GPT-3.5-Turbo C-o-T	0.56 ± 0.07	0.92 ± 0.03	0.69 ± 0.05	0.85 ± 0.06	0.36 ± 0.07	0.50 ± 0.05
Token Classification		*Story* ($N = 7,756$)			*Not Story* ($N = 20,477$)	
Fine-tuned RoBERTa	0.77 ± 0.05	0.79 ± 0.08	0.78 ± 0.04	0.90 ± 0.03	0.88 ± 0.03	0.89 ± 0.01
GPT-4 Few-Shot	0.52 ± 0.06	0.86 ± 0.04	0.64 ± 0.05	0.88 ± 0.04	0.57 ± 0.04	0.69 ± 0.04

Where does the finetuned model fail?

Stories misclassified as non-stories

- stories containing **cognitive verbs**, such as "plan," "decide," or "notice"

- stories containing **hypothetical** verbs

- **very short stories** (one sentence or less)

Non-stories misclassified as stories

- texts containing **general or repeating events** or describe a **state** without sequence

- texts containing **pronouns, entities, and concrete language** like place descriptions, making the texts appear more story-like

🔭 StorySeeker generalizes to other settings.

🔭 StorySeeker generalizes to other settings.

🔭 StorySeeker generalizes to other settings.

🔭 StorySeeker generalizes to other settings.

🔭 StorySeeker generalizes to other settings.

f1

Training Data	Ours	PiperBagga	PiperBagga Full	Brazilian Blogs
Ours	0.81	0.77	0.84	0.73
PiperBagga	0.65	0.93	0.95	0.62
PiperBagga Full	0.63	1	0.99	0.83
Brazilian Blogs	0.74	0.45	0.59	0.93
Ours+PiperBagga	0.76	0.93	0.95	0.78

Test Data

Where do people tell stories?

Categorizing 500 subreddits

We categorized by hand the 500 subreddits with the most posts+comments in the dataset.

toxic 1.4%
tech 6.2%
stories 2.2%
religion 2.2%
relationships 3.6%
professional sports 3.6%
professional advice 4.2%
NSFW 2.8%
news & politics 4.4%
mental health 1.8%
informative 5.2%
hobbies 10.0%

cities 2.0%
confessions 1.0%
countries 2.0%
drugs 2.0%
fandom 5.4%
fitness 2.4%
gaming 20.8%
gender 1.2%
general 5.2%

Which subreddit categories have more storytelling?

fandom

tech

gaming

professional sports

news & politics

finance

cities

healthcare

children

religion

gender
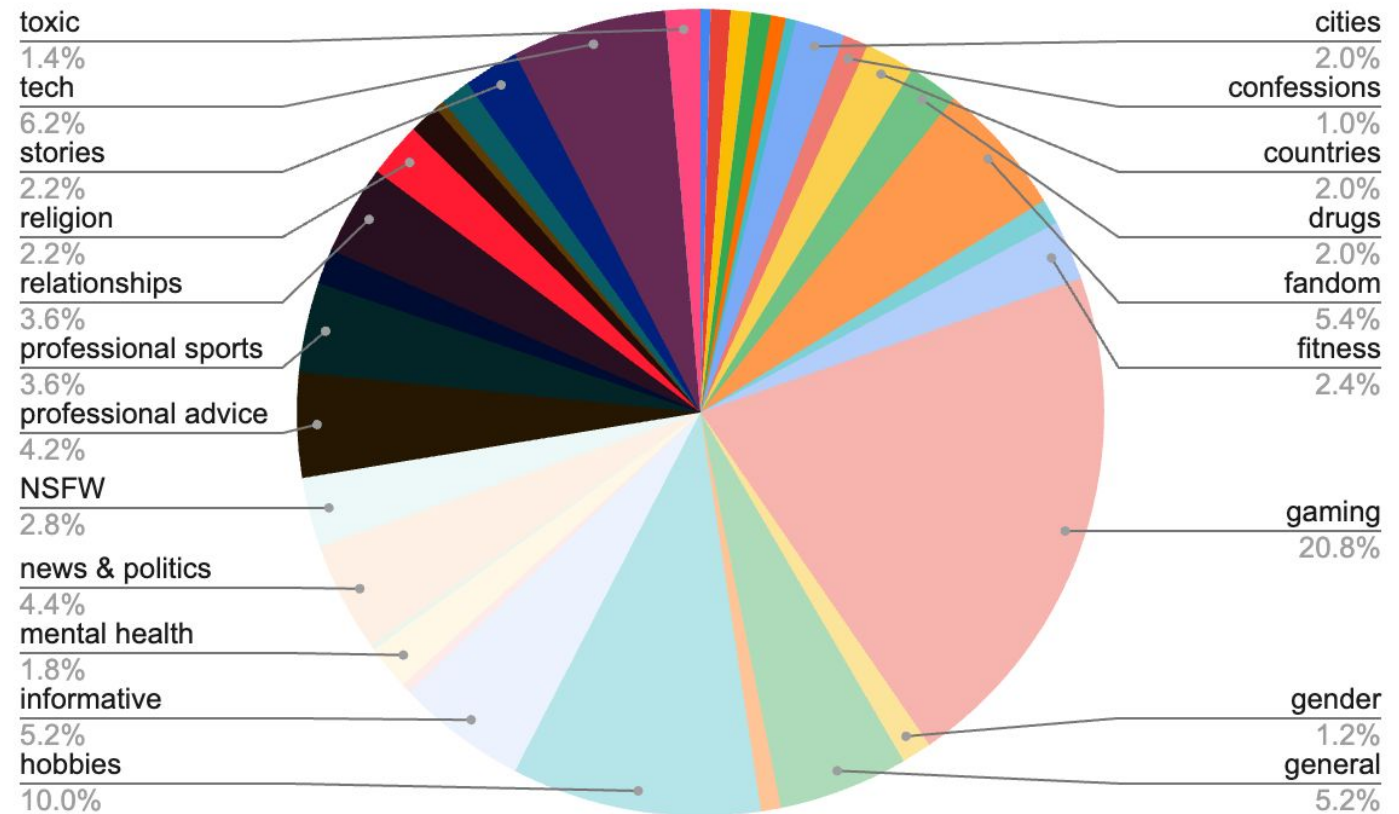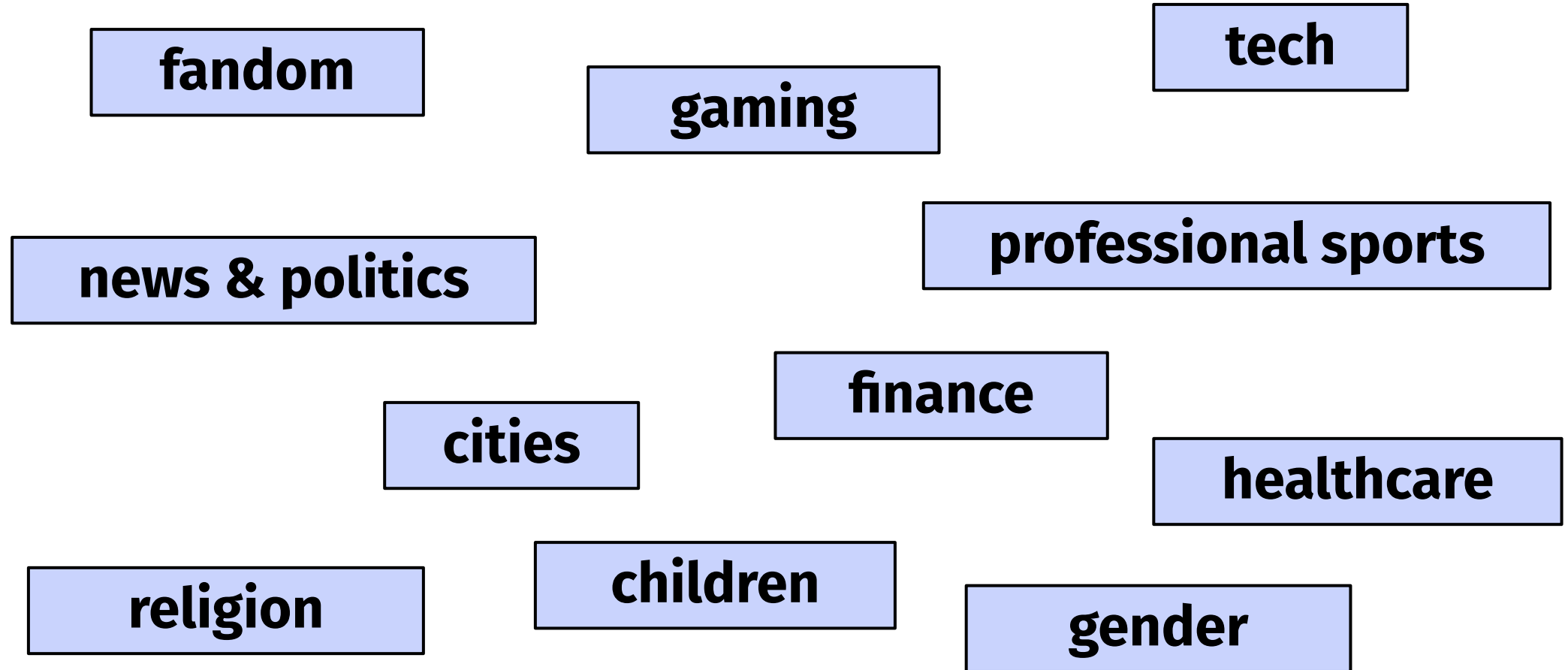
Mentimeter

Which subreddit categories have more storytelling?
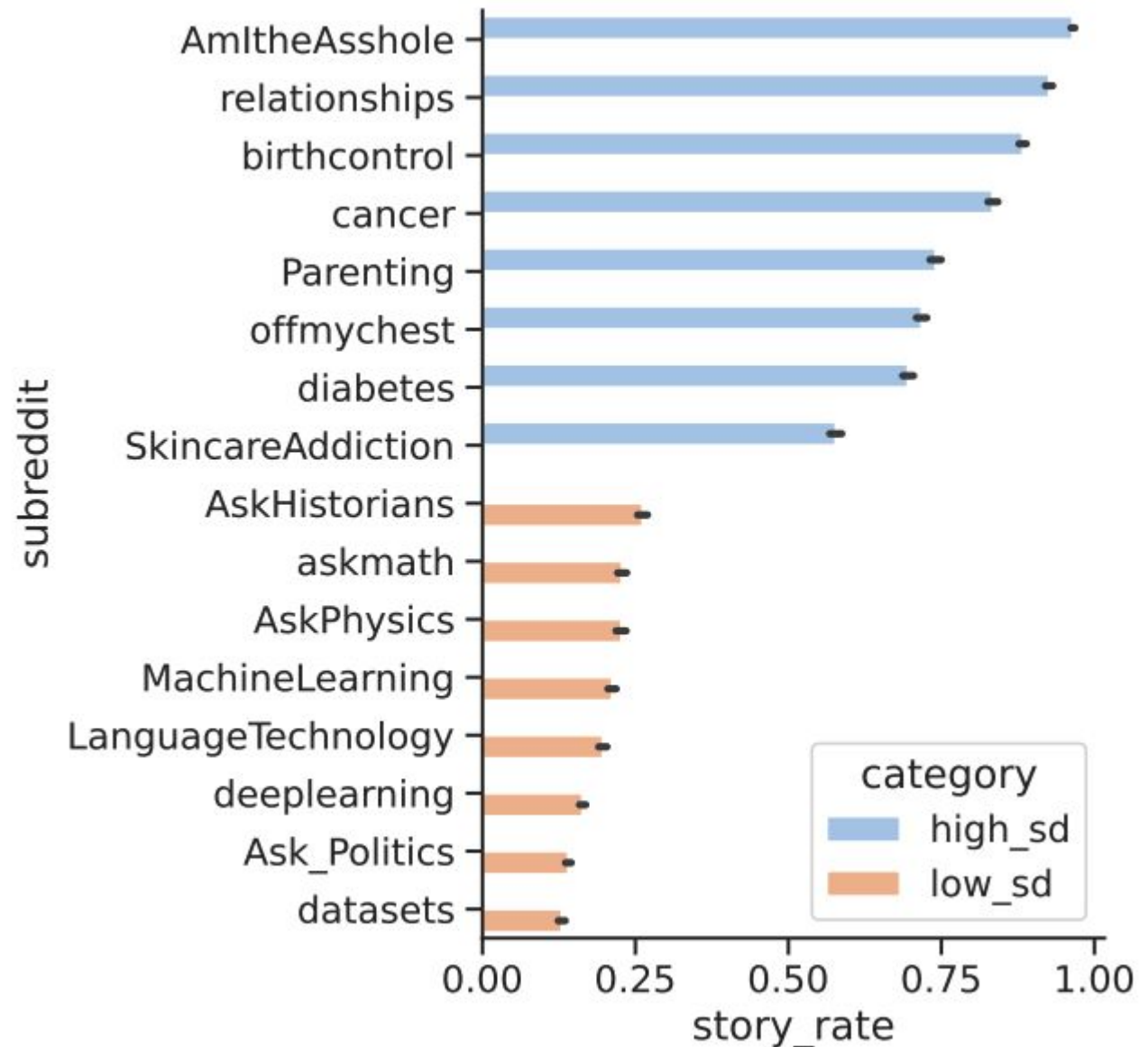
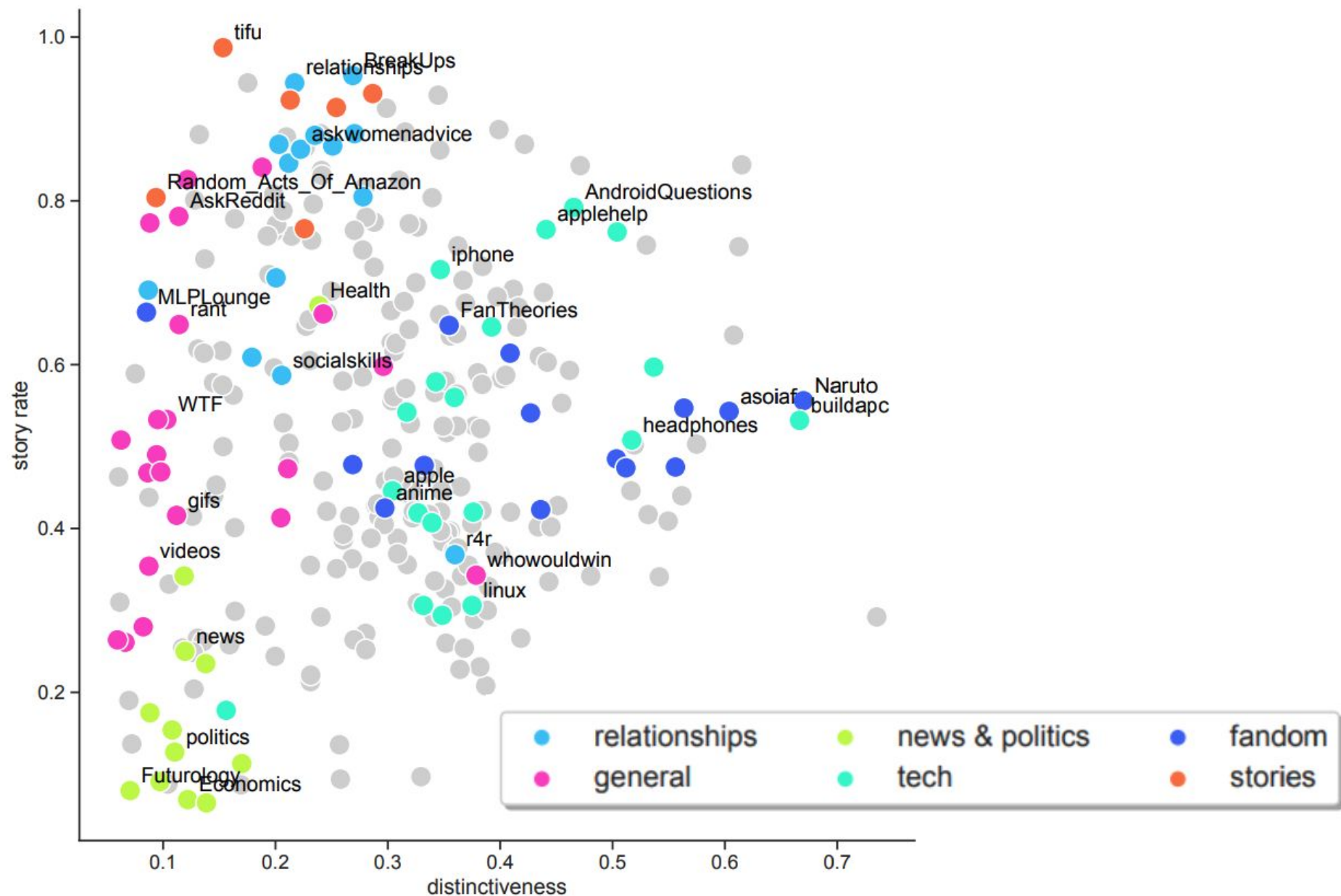## subreddits with the MOST stories
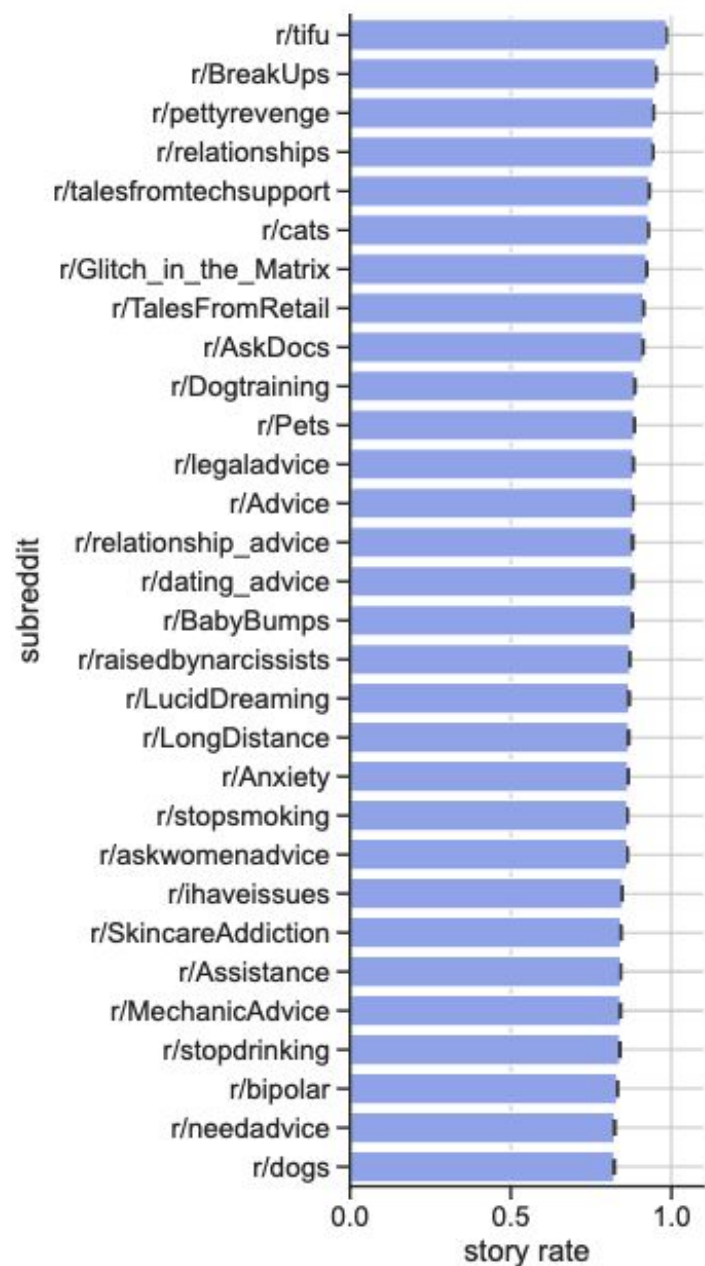
subreddits with the FEWEST stories
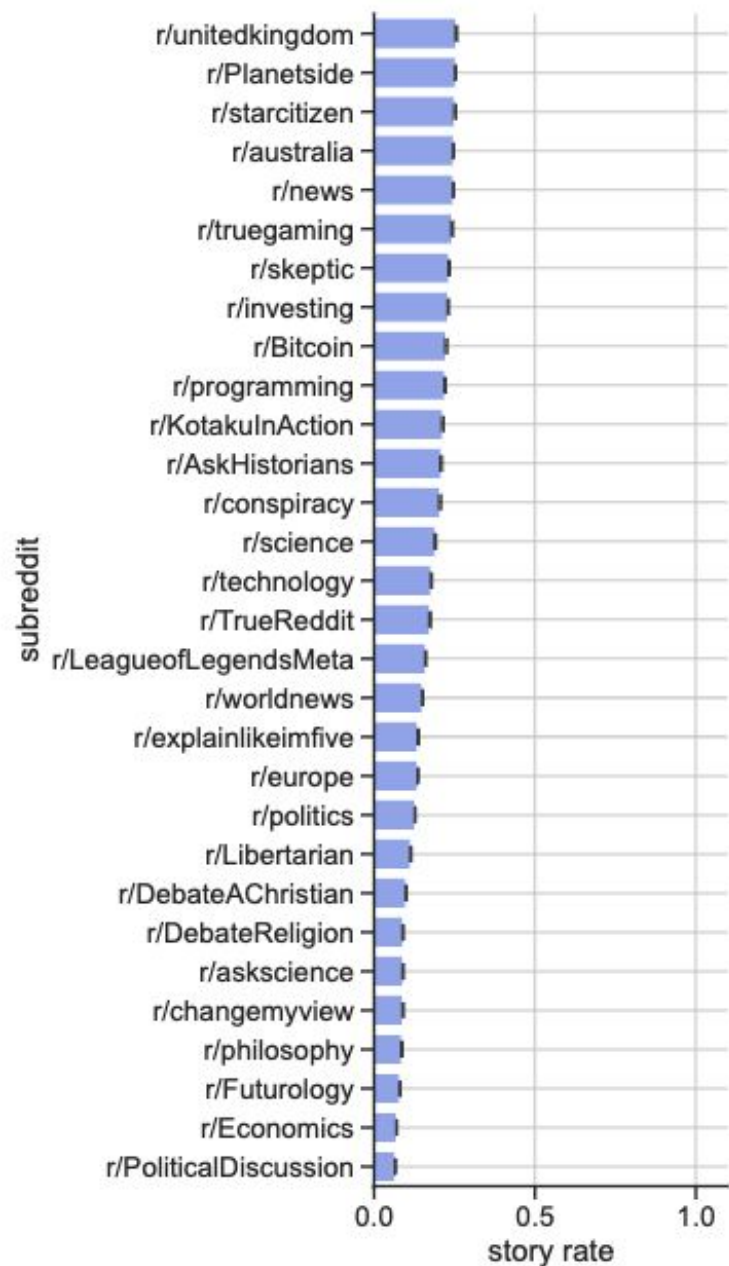
subreddit categories
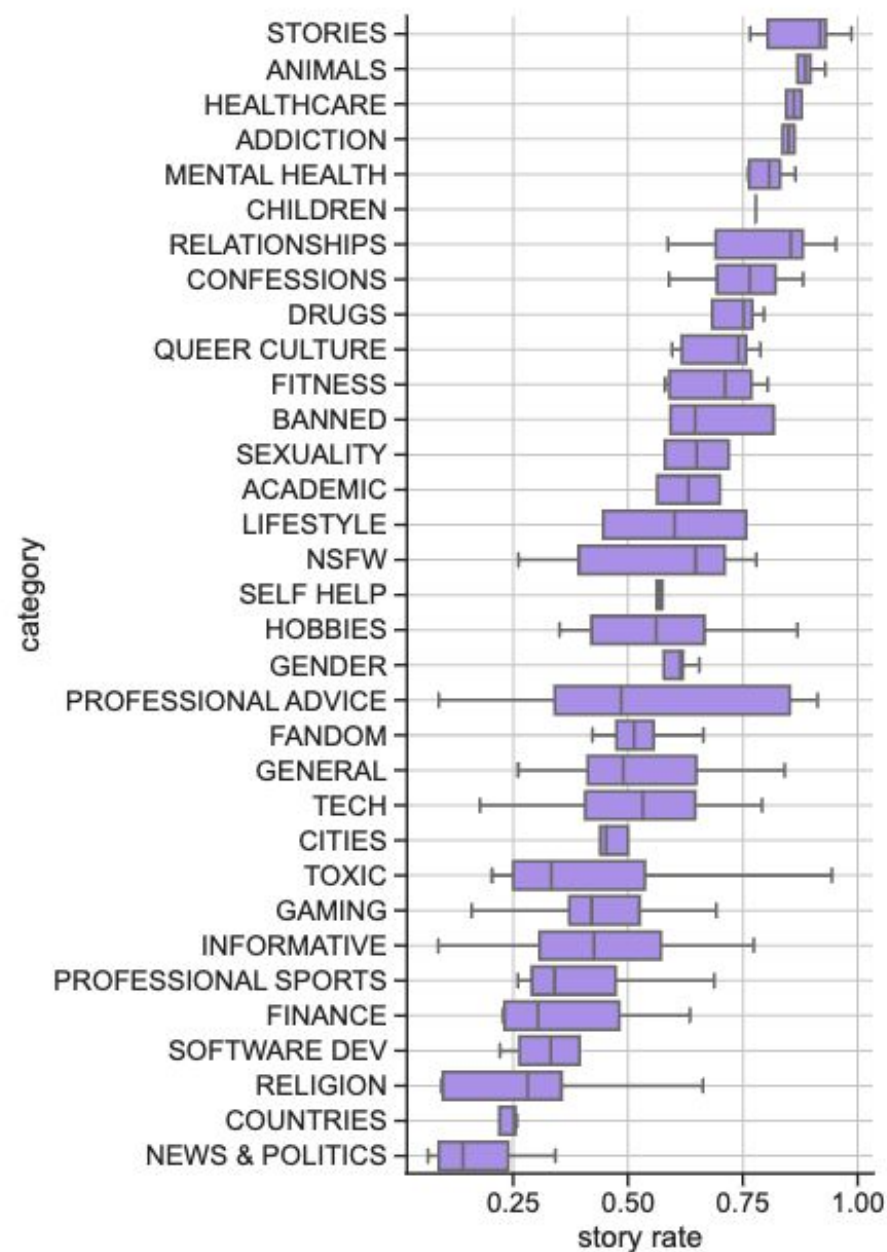
High self-disclosure subreddits contain more storytelling.

more storytelling

story rate

1.0

tifu
BreakUps
relationships
askwomenadvice
Random_Acts_Of_Amazon
AskReddit
AndroidQuestions
applehelp

0.8

iphone
MLPLounge
rant
Health
FanTheories
socialskills
Naruto
asoiaf
buildapc
headphones
WTF

0.6

apple
anime
gifs
r4r
whowouldwin
videos
linux

0.4

news

0.2

politics
Futurology
Economics

| | relationships | | news & politics | | fandom |
| | general | | tech | | stories |

0.1 0.2 0.3 0.4 0.5 0.6 0.7

distinctiveness

Zhang et al., 2017; Lucy et al., 2023

more distinct language ⟶

71

	generic	distinctive
more storytelling		
less storytelling		

	generic	distinctive
more storytelling	r/tifu r/pettyrevenge r/Glitch_in_the_Matrix r/Advice	
less storytelling		

	generic	distinctive
more storytelling	r/tifu r/pettyrevenge r/Glitch_in_the_Matrix r/Advice	r/SkincareAddiction r/LucidDreaming r/techsupport r/MechanicAdvice
less storytelling		

	generic	distinctive
more storytelling	r/tifu r/pettyrevenge r/Glitch_in_the_Matrix r/Advice	r/SkincareAddiction r/LucidDreaming r/techsupport r/MechanicAdvice
less storytelling	r/politics r/explainlikeimfive r/PoliticalDiscussion r/Futurology	

	generic	distinctive
more storytelling	r/tifu r/pettyrevenge r/Glitch_in_the_Matrix r/Advice	r/SkincareAddiction r/LucidDreaming r/techsupport r/MechanicAdvice
less storytelling	r/politics r/explainlikeimfive r/PoliticalDiscussion r/Futurology	r/asoiaf r/summonerschool r/Naruto r/fantasyfootball

Change My View (CMV)

r/changemyview

Posts   Wiki   Rules

Create Post

Hot   New   Top   ...

Posted by u/NewRedSpyder 5 hours ago

Delta(s) from OP Cmv: The biggest reason why people think of Vanilla ice cream as basic is because how it looks, not how it tastes.

If vanilla ice cream wasn't a plain white color, people would not call it basic. Strawberry and Chocalate are arguably more basic flavors, but people don't say that because of their colors and them not being plain white. I guarantee that if vanilla ice cream was purple or something but it had the same taste, it would not be called basic.

Strawberry within itself is a pretty basic flavor too. There are a million things out that are strawberry flavored, and strawberry ice cream doesn't taste that differently from a lot of those, making it by definition pretty basic and not really standing out much. Vanilla isn't as common as a flavor, especially not outside of desserts, so it's less basic than strawberry in that aspect. The same thing can be applied to chocolate flavor (I am aware of many different variations of chocolate ice cream such as mint, caramel, etc but chocolate in itself is not considered basic).

💬 87 Comments   ➦ Share   🔖 Save   ...

Posted by u/ItsNotCalledAMayMay 5 hours ago

CMV: All presidential elections in the US should use a Ranked Choice voting system

Like seriously is there any downside?

About Community

A place to post an opinion you accept may be flawed, in an effort to understand other perspectives on the issue. Enter with a mindset for conversation, not debate.

🏠 Created Jan 16, 2013

3.6m
Members

● **665**
Online

Top 1%
Ranked by Size

Create Post

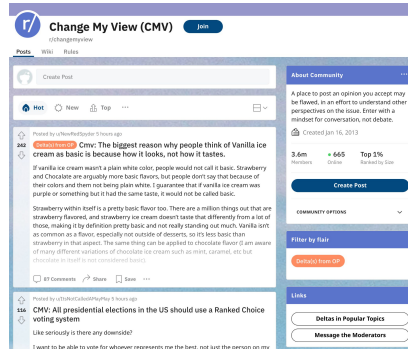COMMUNITY OPTIONS   ⌄
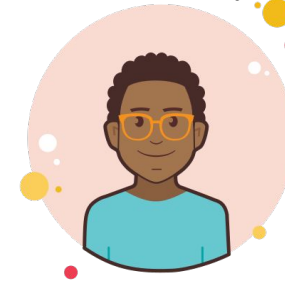
Filter by flair

Delta(s) from OP
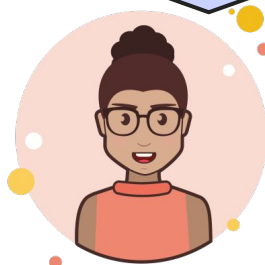
Links

Deltas in Popular Topics

Message the Moderators

77

I think that voting is pointless. Change my view!
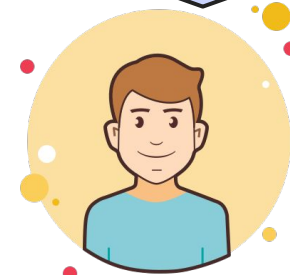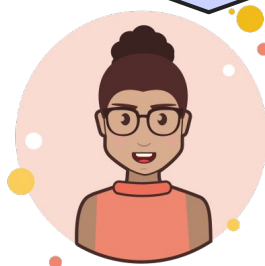
Engaging in the electoral process is a privilege that you shouldn't take for granted.

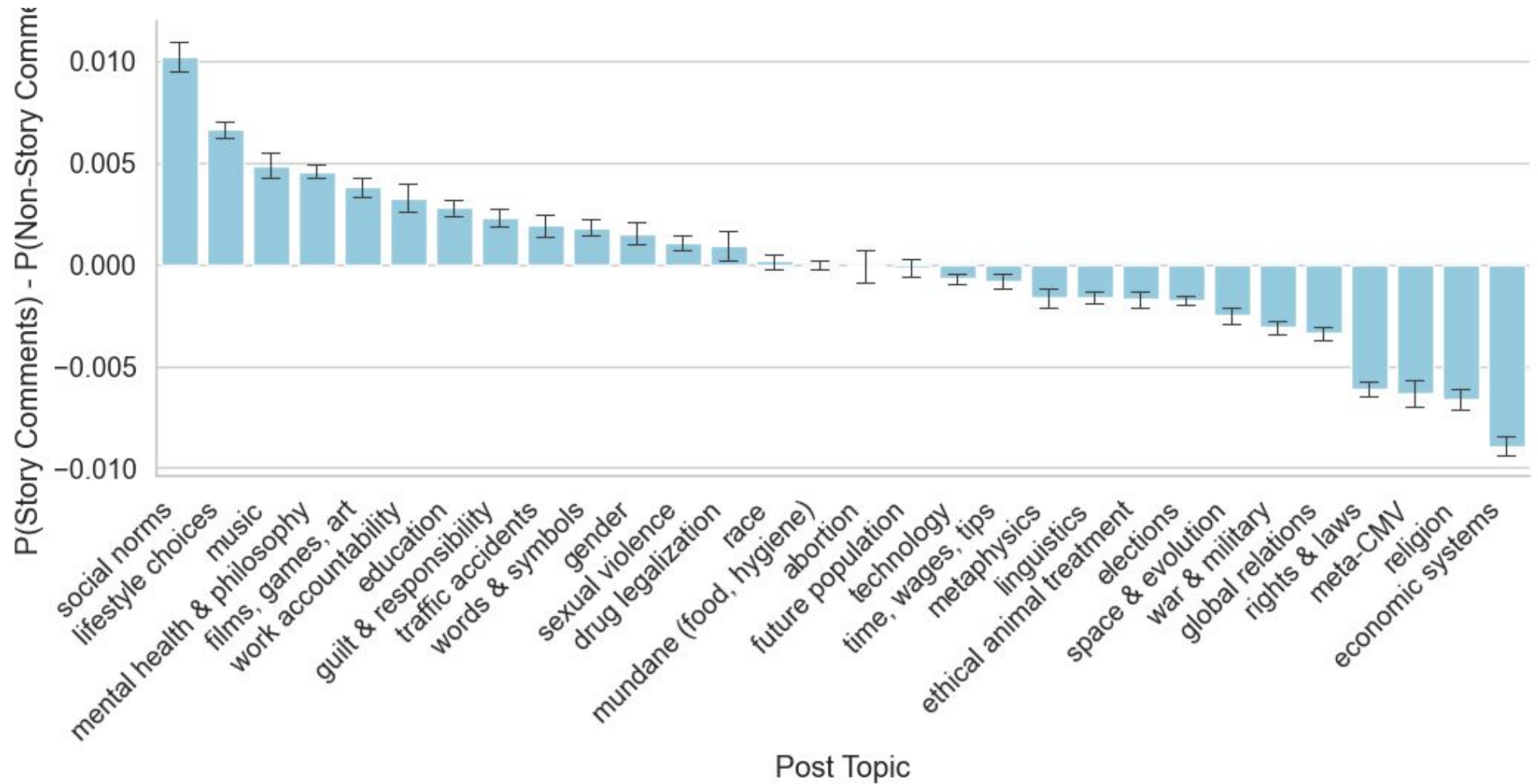You're right that your vote alone might not change things, but have you considered that...

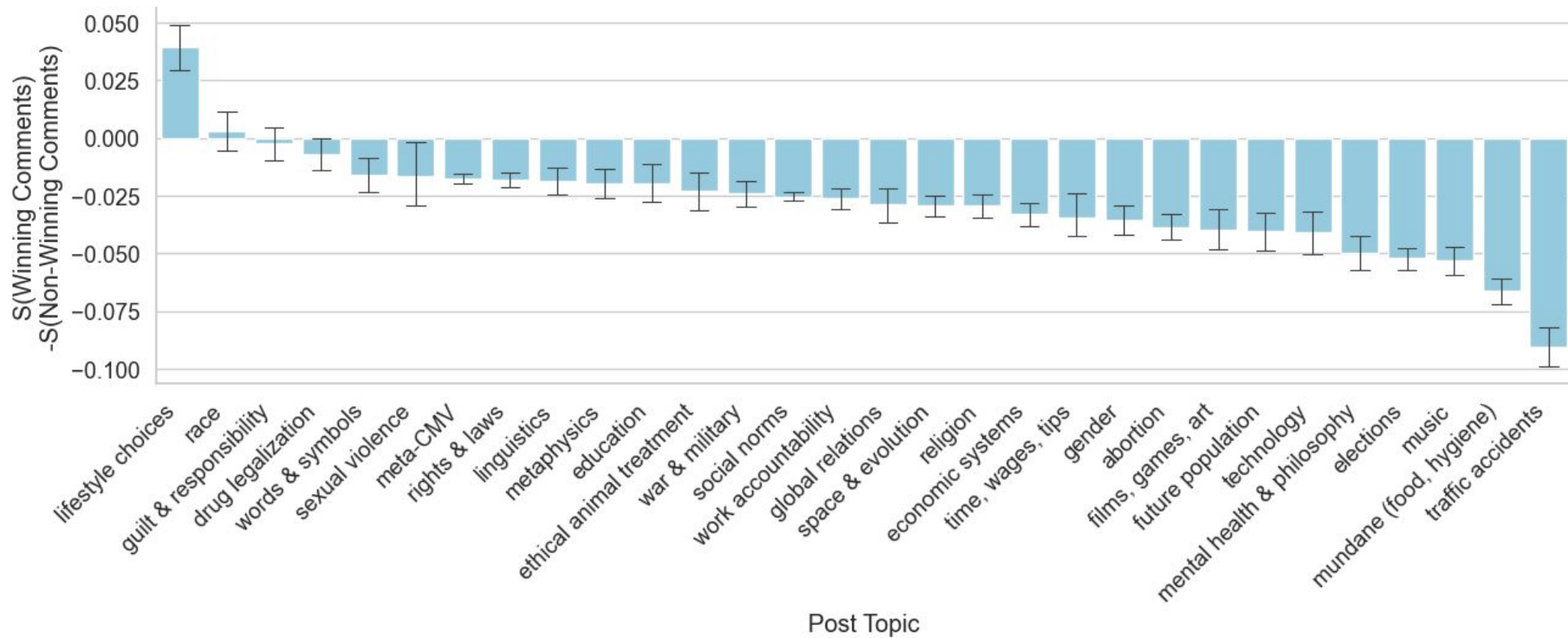I thought the same way until the last election. When I went to drop my ballot off...

r/ChangeMyView

Some topics (e.g., personal) receive more storytelling comments

more storytelling in comments ↑

But are those stories persuasive? Not necessarily!

more storytelling in *winning* comments

Implications

Implications for story detection

- Task formalization, training and evaluation framework, models

Implications for narratology

- Tell-tale signs of online storytelling: agent-centered, event-driven

*Implications for textual interactions in online communities*

- Storytelling patterns across communities

Future directions

➜ Deeper social media analysis about **community features** like trust and how storytelling can drive **misinformation**

➜ Variation in **storytelling perceptions across demographic groups**

➜ Analysis of event annotations and the relationship between **causality, sequentiality, and stories**

Contributions: 🔭StorySeeker
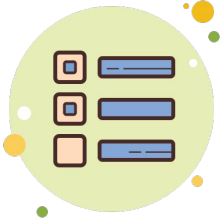
We formalize the task of **story detection** and the task of **story span detection**, linking these tasks to event detection.

We release the first expert-annotated **dataset**, **codebook**, and **models** for story detection **across diverse online communities**.

A first **inter-community** analysis of where storytelling takes places on Reddit, and an **intra-community** analysis of where storytelling is used as a rhetorical strategy.

Code

StorySeeker Github repository and Colab notebook

https://github.com/maria-antoniak/storyseeker

Mentimeter

What are some datasets where you could use

StorySeeker to predict the presence of storytelling?

Exercise

1. Break into small groups (3-4 people per group)

2. Identify a dataset you're interested in

3. Load the dataset into the Colab notebook:

   https://bit.ly/storyseeker-github

4. Run StorySeeker on the dataset

5. Create lists of true positives, true negatives, false positives and false negatives

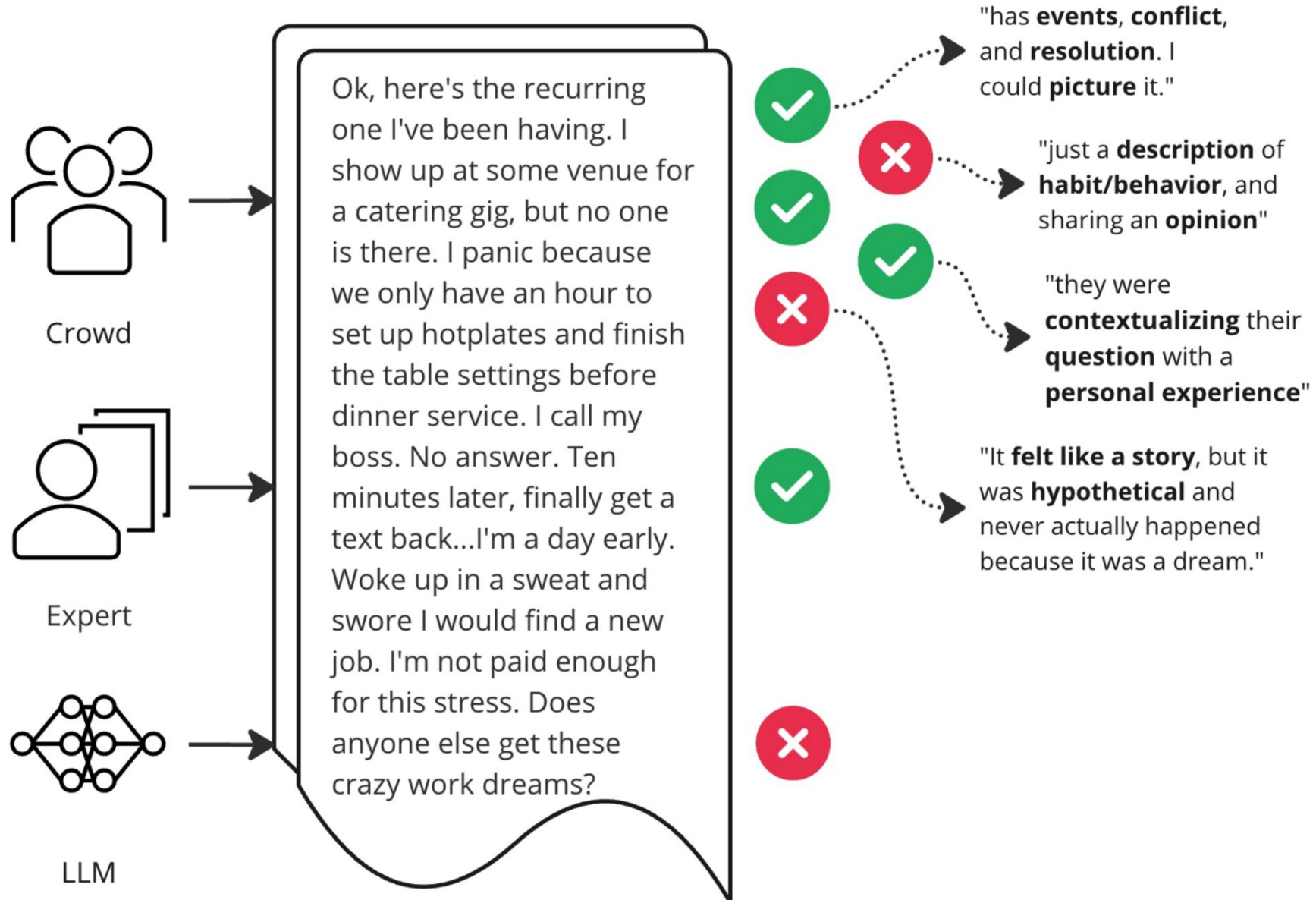6. Share your findings on Menti (4493 1633)

The Empirical Variability of Narrative Perceptions of Social Media Texts

Joel Mire, <u>Maria Antoniak</u>, Maarten Sap, Elliott Ash, Andrew Piper

EMNLP 2024

Does this text contain a story?

Crowd

Expert

LLM

Ok, here's the recurring one I've been having. I show up at some venue for a catering gig, but no one is there. I panic because we only have an hour to set up hotplates and finish the table settings before dinner service. I call my boss. No answer. Ten minutes later, finally get a text back...I'm a day early. Woke up in a sweat and swore I would find a new job. I'm not paid enough for this stress. Does anyone else get these crazy work dreams?

"has **events**, **conflict**, and **resolution**. I could **picture** it."

"just a **description** of **habit/behavior**, and sharing an **opinion**"

"they were **contextualizing** their **question** with a **personal experience**"

"It **felt like a story**, but it was **hypothetical** and never actually happened because it was a dream."

89

Comment: I got lucky and was able to find a guy who already had it in California. Brand new which was a major win. The only other way is through a SSE. You have to find someone (not an ffl) willing to send it to your local ffl who will perform a single shot exemption on the gun. You're going to need a non-threaded barrel. FN does not sell these yet for the .45, so you have to go third party or trade your threaded for a non-threaded barrel. Also, your ffl cannot receive the gun with the threaded barrel in the same package, and the gun has to function upon arrival to perform the SSE. The only other way around this is welding the thread protector onto the barrel. Take into consideration that once you weld the thread protector on, the barrel can no longer be removed from the gun unless you cut it. Lastly, you either need to buy new mags or have your ffl cap them at 10. There's a ridiculous amount of hoops you need to jump through. This is why I went the route of finding someone in California who wanted to sell theirs. I'm really glad I was able to find one in practically mint condition. It only took about a year of searching...

1. How familiar are you with the topic of this text?

Not at all familiar ○ ○ ○ ○ ◉ Very familiar

2. What is the author's goal in writing this text? Finish the sentence: The author of this text wants to _____.

[   ]

3. Does this text contain a story?

◉ YES

○ NO

4. How confident are you in your answer to Question 3?

Not at all confident ○ ○ ○ ○ ○ Very confident

5. Explain your answer to Question 3 by writing a short list of reasons.

[   ]

All annotators agree: this text tells a story

Rationales:

1) The author says "I want to start at the beginning". 2 The author lists events and personal experiences that happened at different times, leading to the present.

- Consistent narrative. - Plot is laid out in a way that can be easily followed.

Author is summarizing the story of his recent life.

This author has shared their own personal story about overcoming addiction. It was an emotional and powerful story.

The person tells the story of his struggles with addiction and how it's destroyed his life. He mentions the addiction, his unemployment, and his eventual divorce.

All annotators agree: this text does NOT tell a story

Rationales:

No description of an event

Clearly, this is an informational post intent on responding to someone else about a dental procedure. There is a lot of descriptive stuff about what happens in this procedure that's obviously not a story.
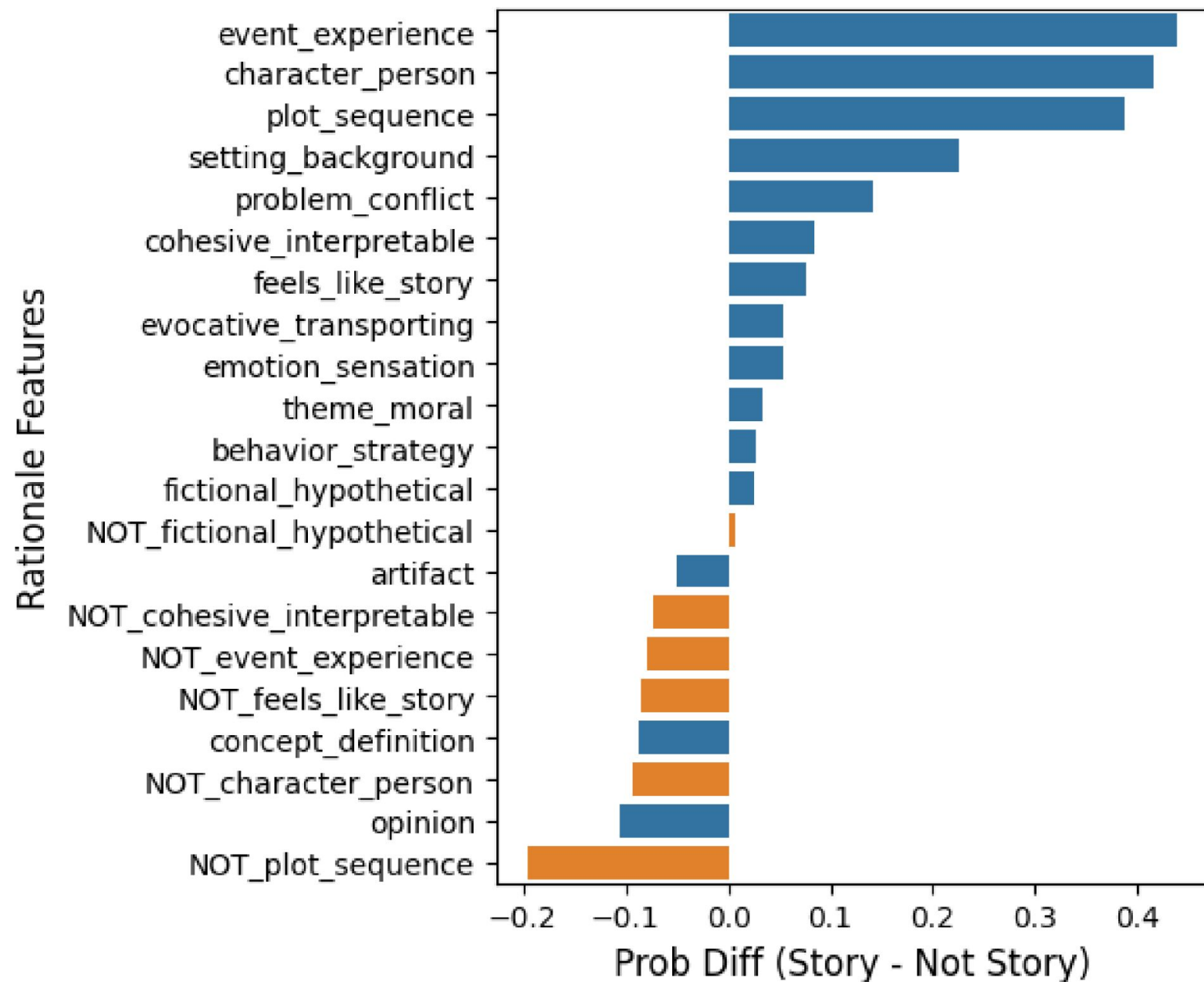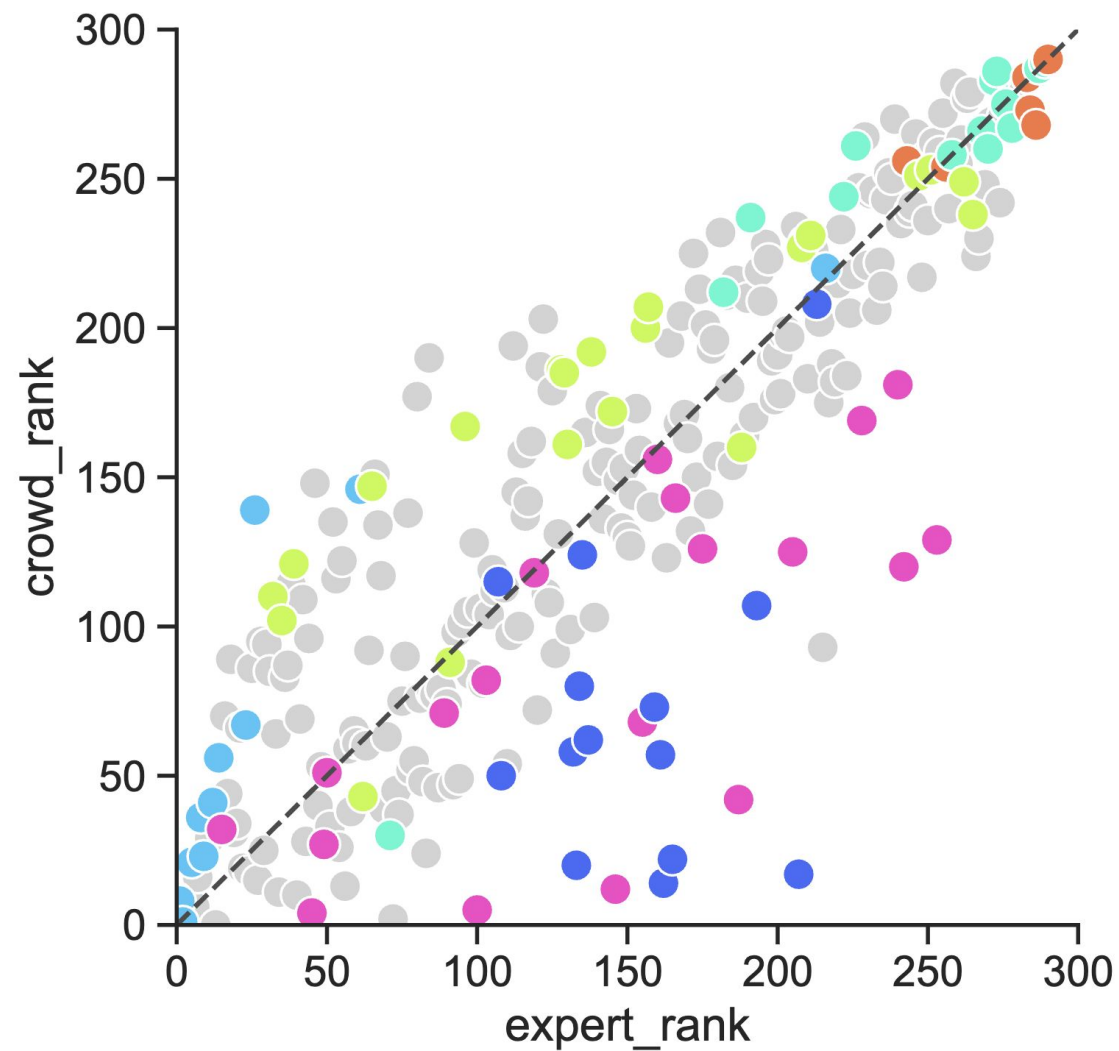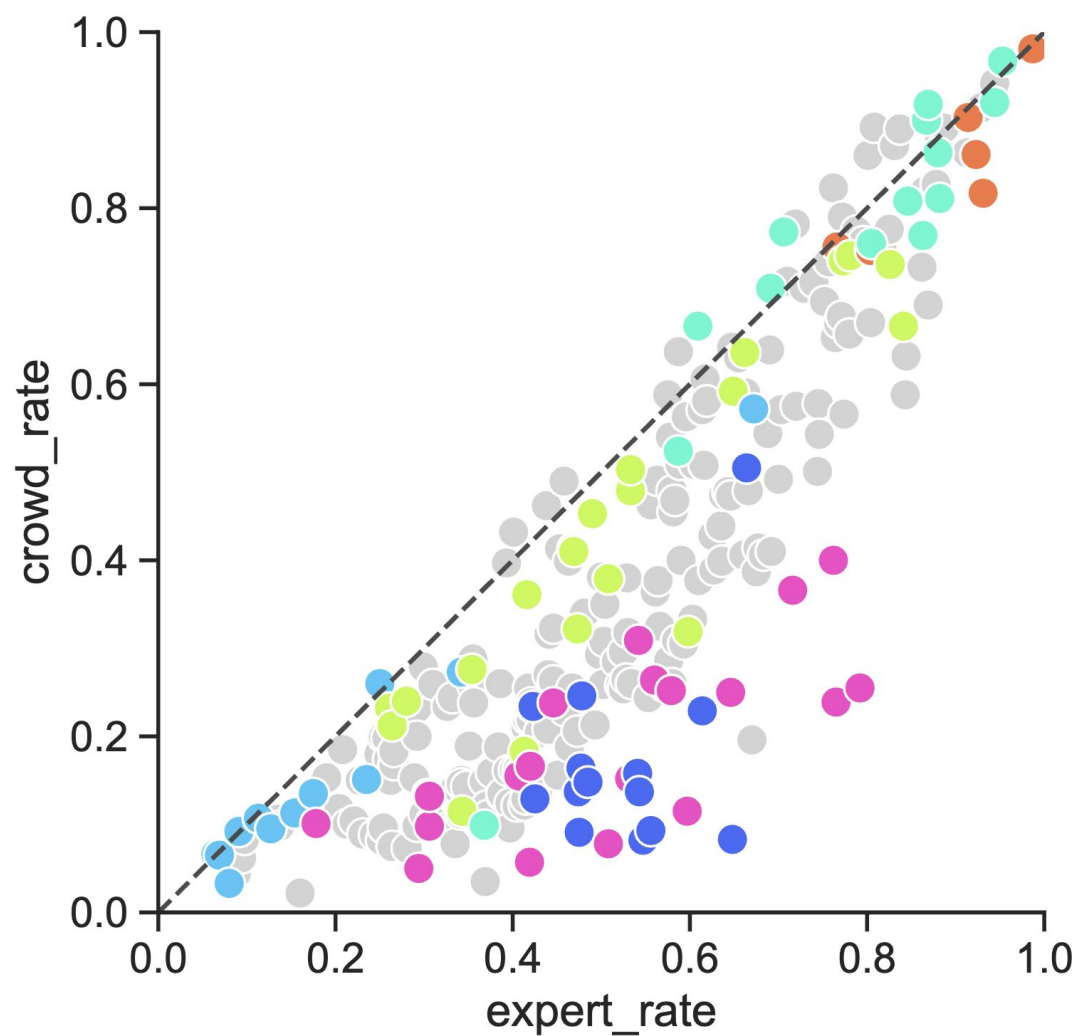
The person it's just explaining something to other

no narrative structure no plot just giving information

1) the author is providing medical information, 2) the text refers to specific dental procedures only
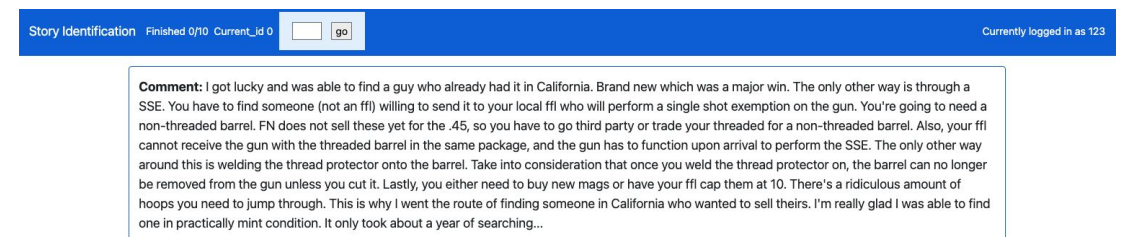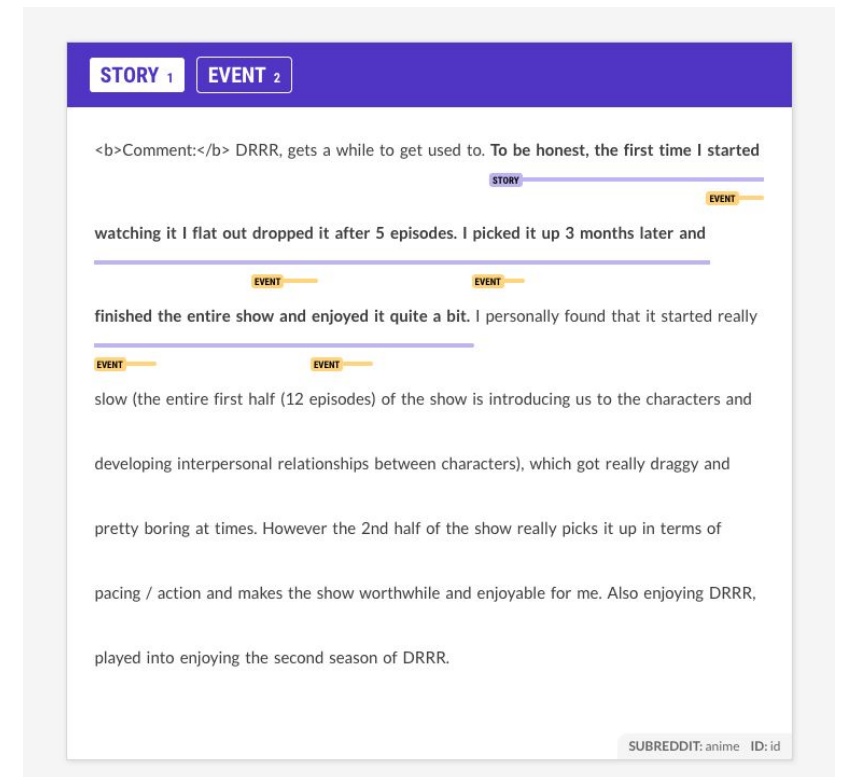
news & politics · general · fandom
tech · relationships · stories

Annotation interfaces

Prodigy

- great documentation and support

- frequent updates

- great user interface

- academic licenses available

Potato

- free and open source

- very flexible

Code

Potato annotation setup for narratives

Exercise

1. Break into small groups (3-4 people per group)
2. Download the Github repository: [REDACTED, please see the Potato website]
3. Set up 🥔potato
4. Try to annotate story spans
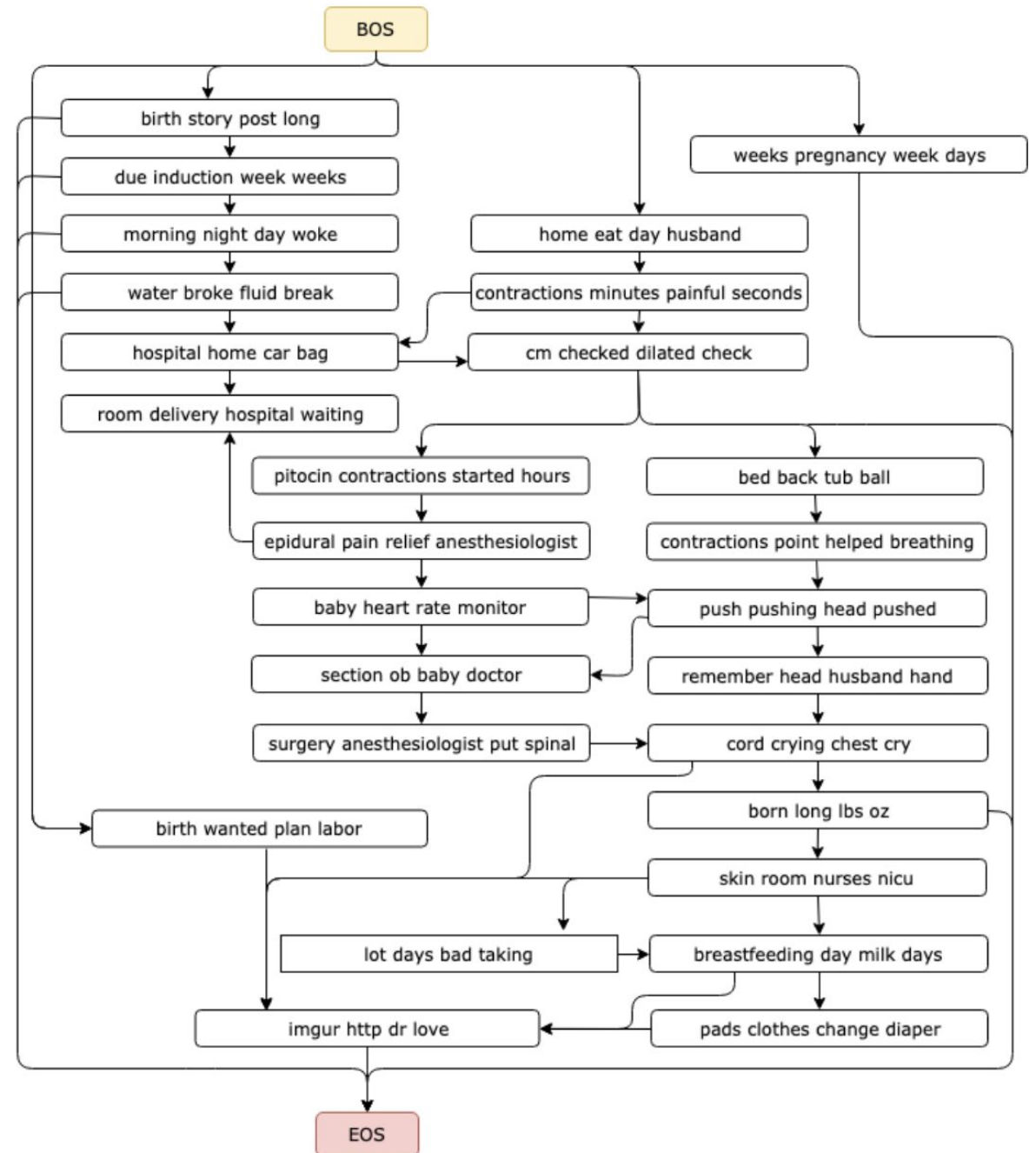5. (Bonus) Try this on your own data!

Some more examples!

extracting event pathways

dataset: birth stories

techniques: topic modeling

"Narrative Paths and Negotiation of Power in Birth Stories"
Antoniak, Mimno, and Levy
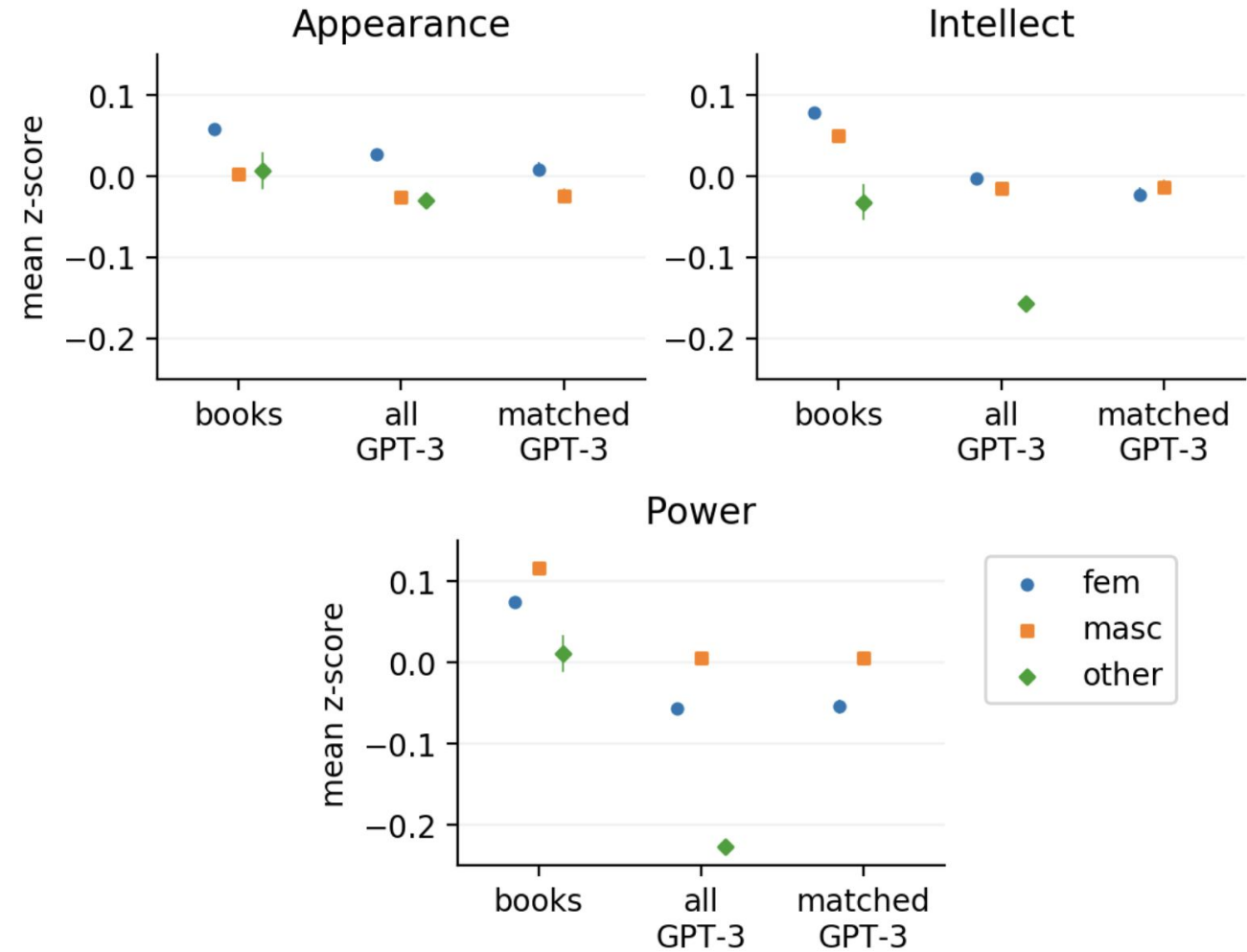CSCW 2019

gender analysis

dataset: generated stories

techniques: LLMs, lexicons, topic modeling

"Gender and Representation Bias
in GPT-3 Generated Stories"
Lucy and Bamman
Workshop on Narrative
Understanding (WNU), NAACL 2021

stylistic analysis

dataset: generated stories

techniques: LLMs, lexicons

"The Zero Body Problem: Probing
LLM Use of Sensory Language"
Hicke, Hamilton, Mimno
preprint 2025

Differences in Sensory Values (Model vs Human)

online communities

dataset: healthcare community

techniques: user activity extraction, clustering

"Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities"
Yang, Kraut, Smith, Mayfield, Jurafsky
CHI 2019

Role	HR	Std.Err
Emotional support provider	0.984	0.027
Welcomer	0.883***	0.028
Informational support provider	1.060	0.034
Story sharer	0.872***	0.034
Informational support seeker	1.324***	0.023
Private support provider	0.842***	0.033
Private communicator	1.031	0.022
All-round expert	0.869***	0.028
Newcomer member	1.054***	0.025
Knowledge promoter	1.091***	0.028
Private networker	0.916*	0.035

Table 3: Survival Analysis predicting how long members continue to participate in the community. $p<0.001$: ***; $p<0.01$**; $p<0.05$*. Number of users = 66,246. Number of user-session records = 522,429
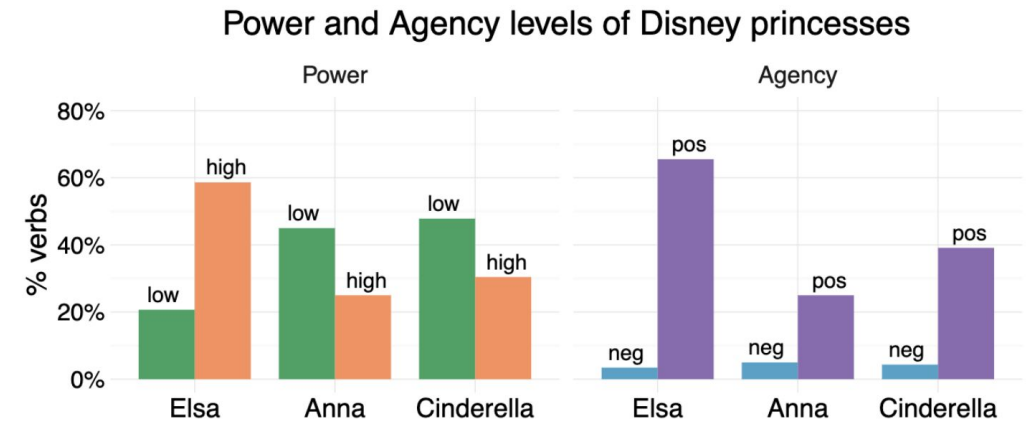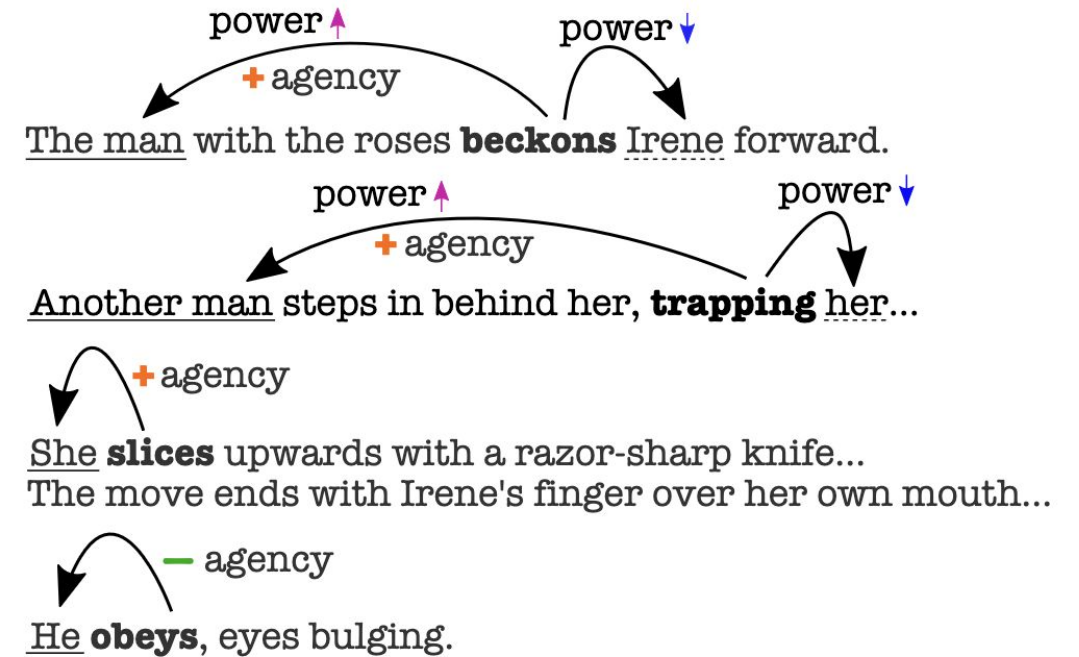
power and agency

dataset: movie scripts

techniques: verb lexica

"Connotation Frames of Power
and Agency in Modern Films"
Sap, Prasetio, Holtzman, Rashkin, Choi
EMNLP 2017

power↑
   + agency
              power↓
The man with the roses **beckons** Irene forward.

power↑
   + agency
              power↓
Another man steps in behind her, **trapping** her...

+ agency
She **slices** upwards with a razor-sharp knife...
The move ends with Irene's finger over her own mouth...

— agency
He **obeys**, eyes bulging.

Power and Agency levels of Disney princesses
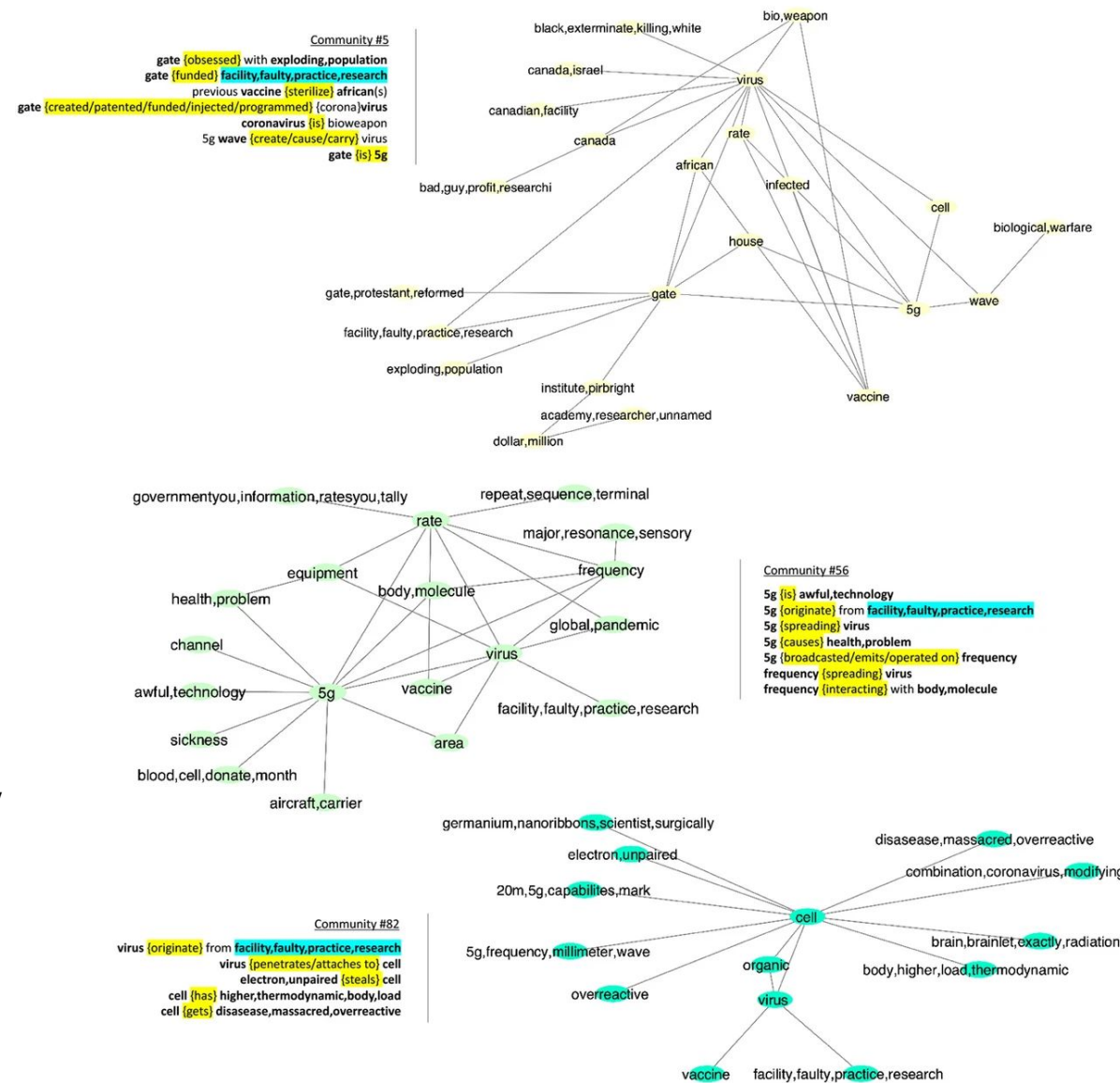
Power

Agency

103

algorithmic folklore

dataset: social media, news

techniques: script extraction, networks

"Conspiracy in the time of corona:
automatic detection of emerging COVID-19 conspiracy
theories in social media and the news"
Shahsavari, Holur, Wang, Tangherlini, Roychowdhury

Journal of Computational Social Science, 2017

Closing

Bridging AI and the Humanities

AI for Humanists www.aiforhumanists.com

- ○ Stand-alone tutorials, references, and code for humanities researchers

- ○ Led tutorials at FAccT 2023, ICWSM 2022, Bell Labs, Hertie School, NLP+CSS

Riveter

- ○ ACL 2023 Demo

- ○ Accessible Python pipeline for extracting and measuring connotation frames

More Tools

- ○ Little Mallet Wrapper, 🔭StorySeeker

Bridging AI and the Humanities

We need deep collaborations between humanists and computer scientists.

We need people working both downstream and upstream; data matters.

And we need people who are interdisciplinary and can move between traditions, methods, communities.

Future Work: Modeling Narratives

Storytelling is a powerful driver of community **sensemaking** processes that can also spread **misinformation.**

Need for more studies on new online communities like TikTok where storytelling is part of a longer **folkloric** and **memetic** tradition.

Narrative medicine to empower care seekers and motivate empathy in caregivers.

Collaborators: narratologists, NLP researchers, web & social media researchers

Thank you! Questions?

https://maria-antoniak.github.io

mariaa.bsky.social