

GENERATIVE ARTIFICIAL INTELLIGENCE PART 2

DALL-E3 Prompt: Data Scientists
celebrate a generative ai Halloween



WHERE WE'RE GOING: IMAGE GENERATION!



2014



2015



2016



2017

BYU



BYU



SPECIFICALLY: CONDITIONAL IMAGE GENERATION



A photo of a Corgi dog riding a bike in Times Square. It is wearing sunglasses and a beach hat.

SPECIFICALLY: CONDITIONAL IMAGE GENERATION



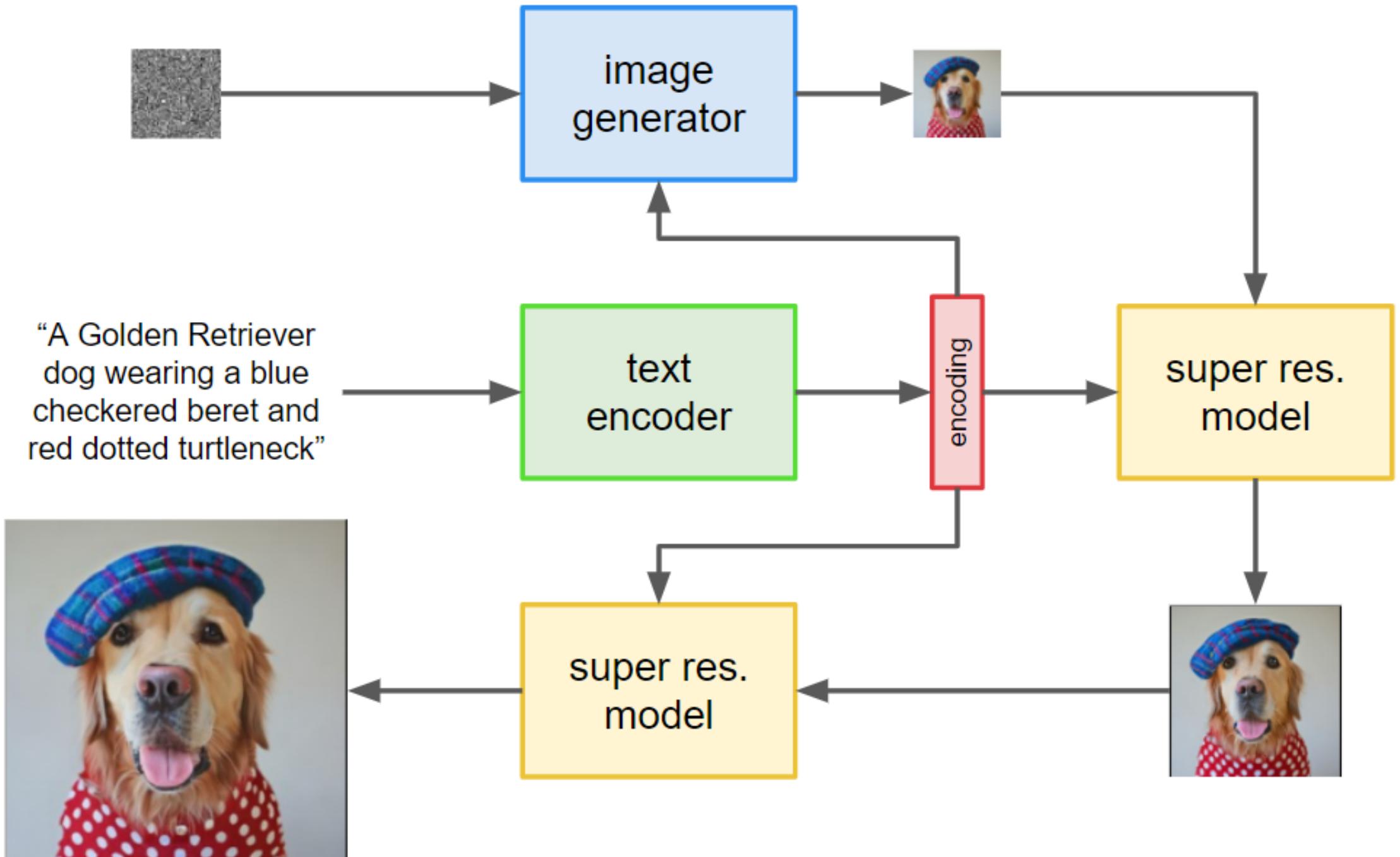
A wall in a royal castle.
There are two paintings on
the wall. The one on the
left a detailed oil painting
of the royal raccoon king.
The one on the right a
detailed oil painting of the
royal raccoon queen.

BYU

SPECIFICALLY: CONDITIONAL IMAGE GENERATION



A detailed and tightly rendered cyberpunk oil painting of a modified (cyberpunk) orange and black (mechanical hybrid design of a Porsche 911 GT3 R merged with a 2014 Nissan Xterra) in a rugged cross country race :: (entire vehicle) :: racing tires :: flourescent colored paint :: modified (very high-tech), cg, concept art, bright colored paint, artstation, art, cyberpunk, octane render, by Matthew Abram Groening,



WHAT DO MODELS LEARN FROM NEXT-WORD PREDICTION?

<i>Grammar</i>	In my free time, I like to { <u>run</u> , banana }
<i>Lexical semantics</i>	I went to the zoo to see giraffes, lions, and { <u>zebras</u> , spoon }
<i>World knowledge</i>	The capital of Denmark is { <u>Copenhagen</u> , London }
<i>Sentiment analysis</i>	Movie review: I was engaged and on the edge of my seat the whole time. The movie was { <u>good</u> , bad }
<i>Harder sentiment analysis</i>	Movie review: Overall, the value I got from the two hours watching it was the sum total of the popcorn and the drink. The movie was { <u>bad</u> , good }
<i>Translation</i>	The word for “pretty” in Spanish is { <u>bonita</u> , hola }
<i>Spatial reasoning</i>	[...] Iroh went into the kitchen to make some tea. Standing next to Iroh, Zuko pondered his destiny. Zuko left the { <u>kitchen</u> , store }
<i>Math question</i>	First grade arithmetic exam: $3 + 8 + 4 = \{15, 11\}$

WHAT CAN'T THEY LEARN FROM NEXT-WORD PREDICTION?

<i>Current world knowledge</i>	The stock price of APPL on March 1st, 2023 is {???
<i>Arbitrarily long arithmetic</i>	$36382894730 + 238302849204 = \{???\}$
<i>Many-step reasoning</i>	Take the nineteenth digit of Pi and multiply it by the e to the fourth power. The resulting ones-digit of the resulting number is {???
<i>Predict the future</i>	The winner of the FIFA world cup in 2026 is {???
<i>Information not in the training data</i>	Jason Wei's favorite color is {???
<i>Extremely long inputs</i>	[2,000 page Harry Potter fan-fiction] What happened after Harry opened the chest for the second time? {???

ALGORITHMIC BIAS AND TOXICITY

- Models / algorithms aren't biased, but training data is
- Negative biases can be reflected (and amplified!) in outputs
- Many different types of toxicity and bias
 - Sexism (eg, gendered professions)
 - Racism
 - Negative stereotypes (eg, black criminality)
 - Sexual orientation
 - Ableist, anti-Muslim, profane, lewd, violent, seditious...
 - Liberal, rich, educated...

“Feeding AI systems on the world’s beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy.”

-- Ruha Benjamin



CASE STUDY #1: ANALYZING AUTISM INTERVIEWS

GOAL: STUDY EARLY INDICATORS OF AUTISM

- Data: transcripts of 18 interviews with parents of autistic daughters
 - Wide ranging topics. Nuanced statements. Implicit contexts.
- Central question: are there patterns in the data?
- Standard methodology: human coding of the data (via CQR, etc.)
 - Slow – 6 months!
- AI methodology: combination of **language models** and **embeddings**

HOW TO ANALYZE UNSTRUCTURED TEXT?

..[Daughter] is very quiet and obedient, and she doesn't get a lot of attention in a group setting, and I think she got overlooked. So, when she went to first grade, and she had to have an aide they would take her to the back of the room and work with her on her alphabet things and so, she was a little bit behind in first grade, she had to catch up in maybe half of first grade, but she learned to read just fine and once she read, she was very good at it. She is an excellent reader, of course this is one of my kids I read to voraciously, and she got read to as a baby and a child a lot, so maybe that helped, I don't know, but she is a good reader, good reading comprehension. She doesn't struggle with that at all and didn't ever. I mean, once she got into probably fifth or sixth grade, I had nothing to do with her academic, she just ran her own life and did everything and did really well.

USE LLM TO EXTRACT CONCRETE OBSERVATIONS

..[Daughter] is very quiet and obedient, and she doesn't get a lot of attention in a group setting, and I think she got overlooked. So, when she went to first grade, and she had to have an aide they would take her to the back of the room and work with her on her alphabet things and so, she was a little bit behind in first grade, she had to catch up in maybe half of first grade, but she learned to read just fine and once she read, she was very good at it. She is an excellent reader, of course this is one of my kids I read to voraciously, and she got read to as a baby and a child a lot, so maybe that helped, I don't know, but she is a good reader, good reading comprehension. She doesn't struggle with that at all and didn't ever. I mean, once she got into probably fifth or sixth grade, I had nothing to do with her academic, she just ran her own life and did everything and did really well.

What prompt would you use to extract this information?

Here is the Prompt that the researchers used:

The following paragraphs were taken from interviews with parents of autistic daughters. Examine the paragraphs and summarize any concrete observations that the parent made about the child. List each observation on a separate line. If the paragraph doesn't contain a concrete observation, or is "small talk", then respond with "none". Be as specific as possible. Do not use general phrases like "Autistic Challenges" or "Parental Observations". Analyze the paragraphs and summarize concrete observations, or say "none":

Paragraph: Okay, well don't share it with the world, okay? I am not a Facebook person.

Concrete observation: none

Paragraph: but she, she's a real talker, I mean, she loves to talk and a very social person and... once she learned she was, oh, always opens her mouth. I mean, if she's autistic, she's not wrapped up in herself. She's you know, looking to other people to associate with them, and socialize. She, she loves people. But—

Concrete observation: loves to talk

Concrete observation: is a very social person

Concrete observation: is not wrapped up in herself

Concrete observation: looks to other people to associate with them

Concrete observation: loves people

Paragraph: But coordination has been, you know, she's lacked coordination, you know like sports, and things like that, large motor skills, yeah.

Concrete observation:

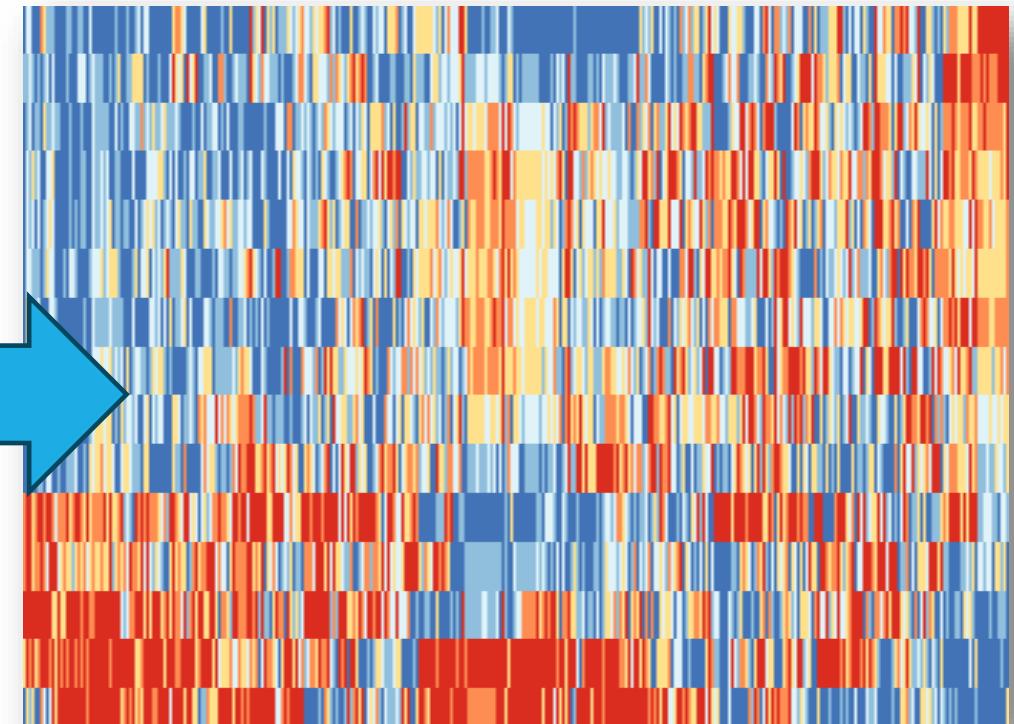
WORKS SURPRISINGLY WELL!

- 2,535 observations
- Nicely tagged with interview, paragraph #
- ... but now what?
- We'd really like to somehow cluster these...

```
,filename,paragraph_number,concrete_observation
0,interview222-282.json,2,had the cord wrapped around her neck at birth
1,interview222-282.json,2,was slower learning to talk
2,interview222-282.json,2,was slower learning to walk
3,interview222-282.json,3,didn't crawl as a baby
4,interview222-282.json,3,walked without crawling first
5,interview222-282.json,3,resisted attempts to learn how to crawl
6,interview222-282.json,3,was taken to a facility for children not developing properly
7,interview222-282.json,4,attended speech therapy
8,interview222-282.json,4,doesn't stop talking
9,interview222-282.json,5,loves to talk
10,interview222-282.json,5,is a very social person
11,interview222-282.json,5,is not wrapped up in herself
12,interview222-282.json,5,looks to other people to associate with them
13,interview222-282.json,5,loves people
14,interview222-282.json,7,started walking at fourteen months
15,interview222-282.json,7,fine motor skills were slower to develop
16,interview222-282.json,7,taught herself how to play the piano
17,interview222-282.json,8,is a very fast typist
18,interview222-282.json,9,lacks coordination
...
```

Run each observation through an embedder

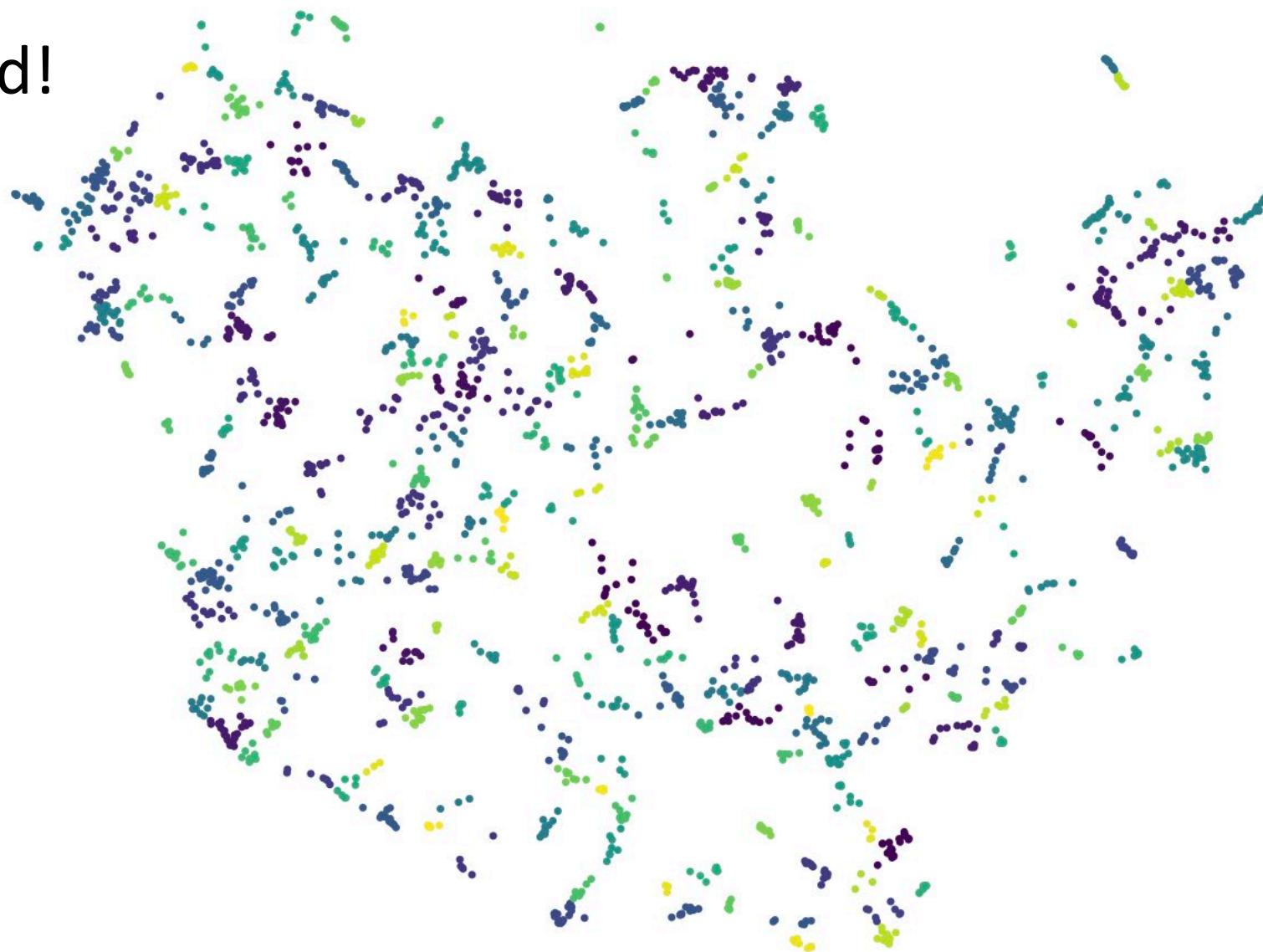
```
.filename,paragraph_number,concrete_observation  
0,interview222-282.json,2,had the cord wrapped around her neck  
1,interview222-282.json,2,was slower learning to talk  
2,interview222-282.json,2,was slower learning to walk  
3,interview222-282.json,3,didn't crawl as a baby  
4,interview222-282.json,3,walked without crawling first  
5,interview222-282.json,3,resisted attempts to learn how to  
6,interview222-282.json,3,was taken to a facility for children  
7,interview222-282.json,4,attended speech therapy  
8,interview222-282.json,4,doesn't stop talking  
9,interview222-282.json,5,loves to talk  
10,interview222-282.json,5,is a very social person  
11,interview222-282.json,5,is not wrapped up in herself  
12,interview222-282.json,5,looks to other people to associate  
13,interview222-282.json,5,loves people  
14,interview222-282.json,7,started walking at fourteen months  
15,interview222-282.json,7,fine motor skills were slower to  
16,interview222-282.json,7,taught herself how to play the piano  
17,interview222-282.json,8,is a very fast typist  
18,interview222-282.json,9,lacks coordination  
...
```



The embedded observations can be visualized...



...And clustered!



BYU

RESULTS!

- **Cluster 11: Social Ineptitude Despite Intellectual Capacity**
 - missing out on some social things because she doesn't fully engage in it
 - often didn't know what was happening in her social circle
 - feels defensive when with people who are not her friends
 - didn't learn or practice social skills in her early teens
 - social understanding has not tracked with her academic progress
 - disconnect between her intellect and social application
 - daughter doesn't like to initiate social interactions
 - teachers initially didn't recognize her intelligence
 - felt she was really stupid, despite being smart

RESULTS!

- Cluster 122: Rapid Learning Ability
 - is a very fast typist
 - caught onto reading well
 - joined the top reading group within two months of changing schools
 - is a fast reader
 - is a very fast reader
 - could read very fast
 - picked up reading easily
 - read very quickly
 - picked up reading very fast
 - never had trouble comprehending things according to report cards
 - never expressed trouble comprehending things

RESULTS!

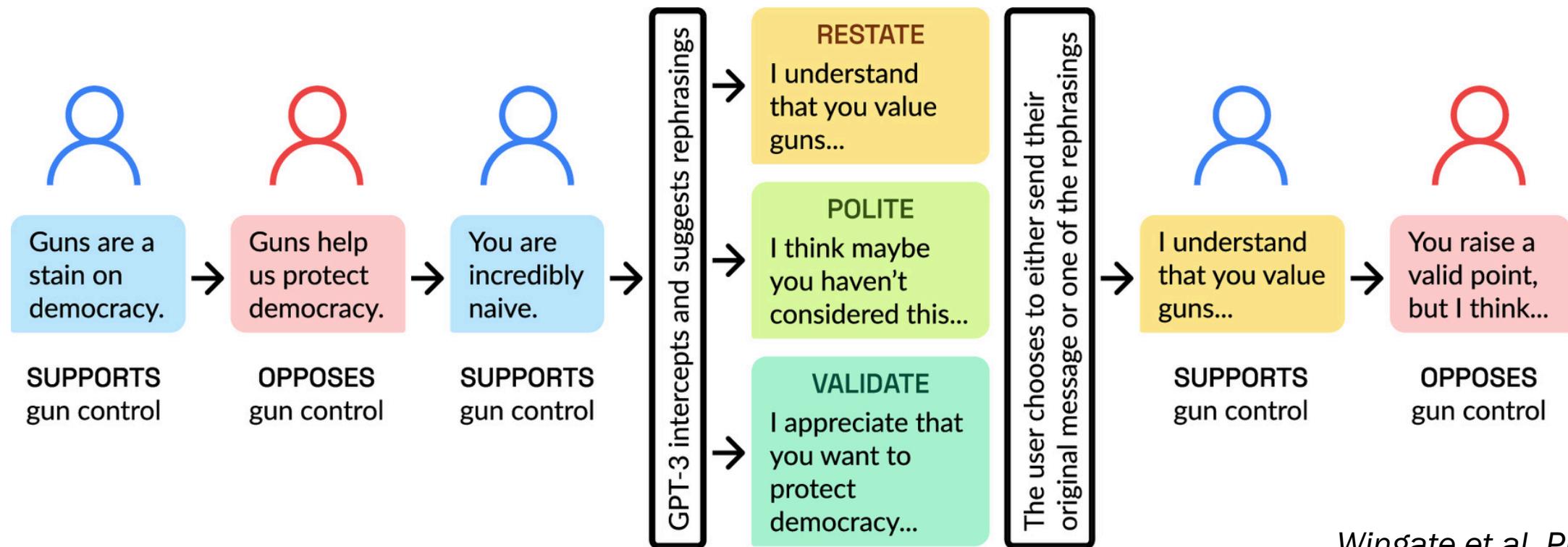
- Cluster 77: Sensory Overstimulation
 - has a hard time with clutter
 - views disorganized spaces as a nightmare
 - wears headphones to block out noise
 - is bothered by noise
 - has a hard time dealing with sounds like chewing or breathing
 - struggles with repetitive types of sound
 - is bothered by too much noise
 - is bothered by too many lights
 - is bothered by too much talking
 - disturbed by sounds and lights in the grocery store
 - can hear electric lights



CASE STUDY #2: DEPOLARIZING CHATBOT

CAN WE IMPROVE CIVIL CONVERSATION BETWEEN HOTLY OPPOSING VIEWS?

Idea: Recruit opposing partisans and have them talk about hot-button issues.
What could possibly go wrong?

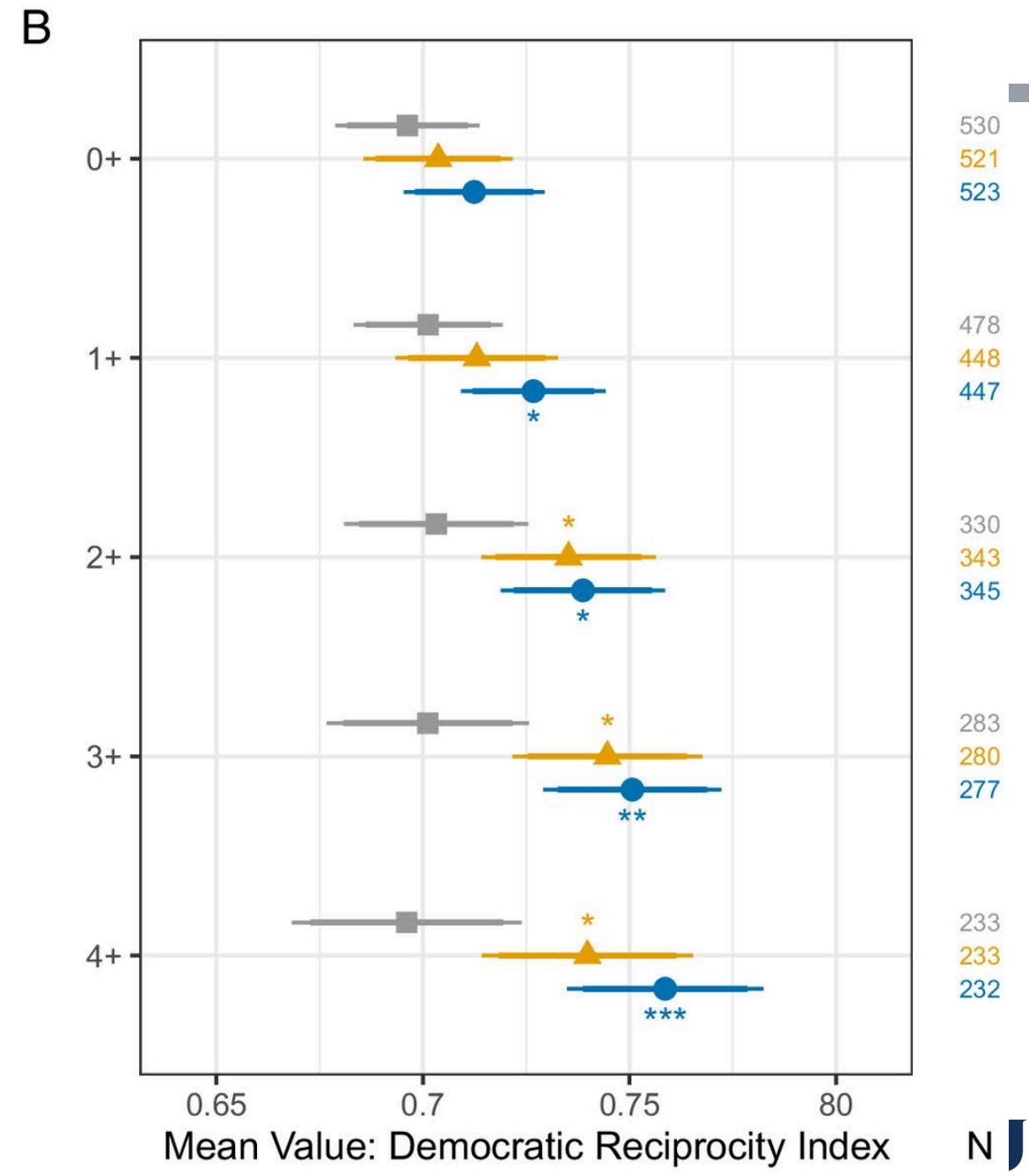
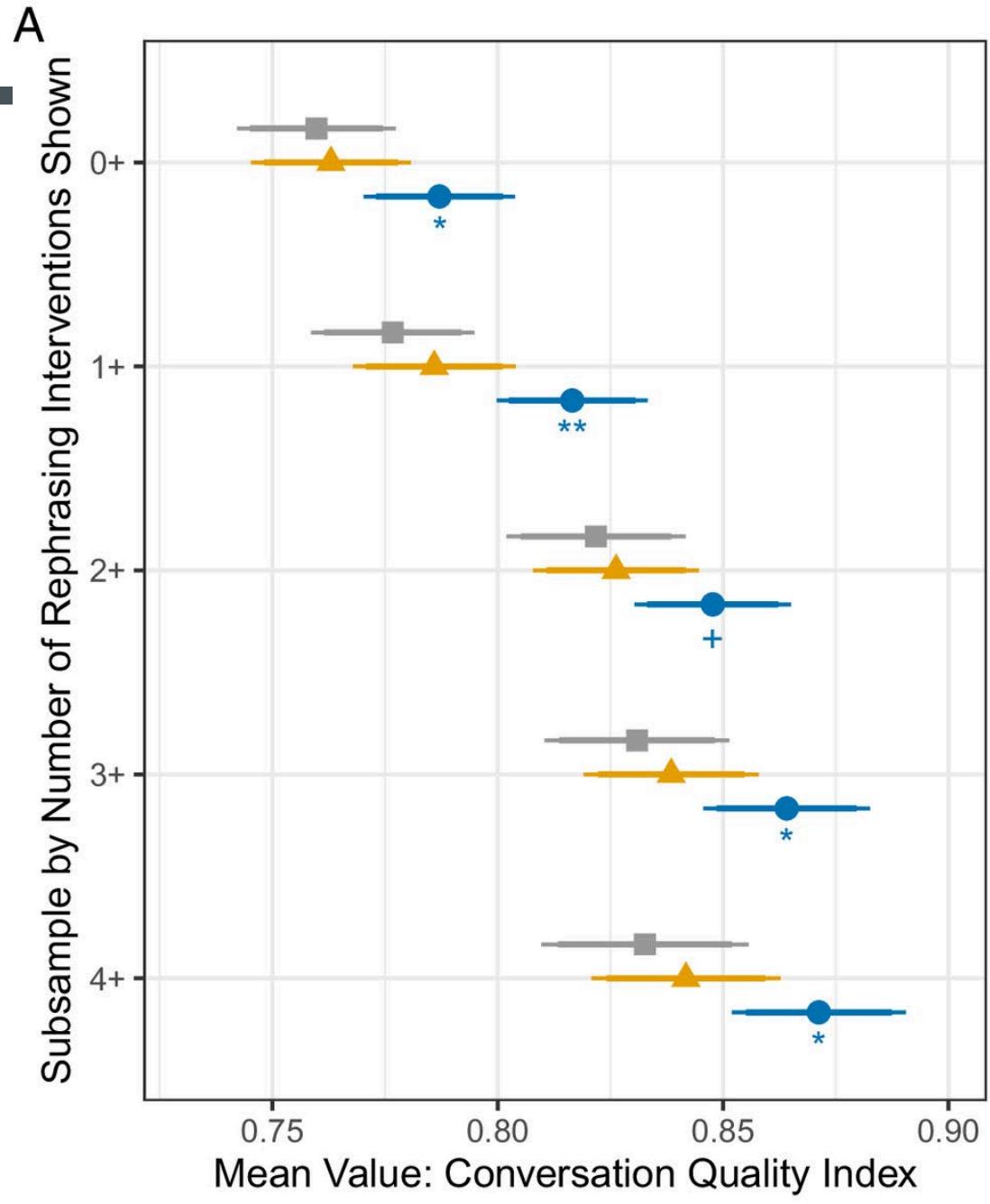


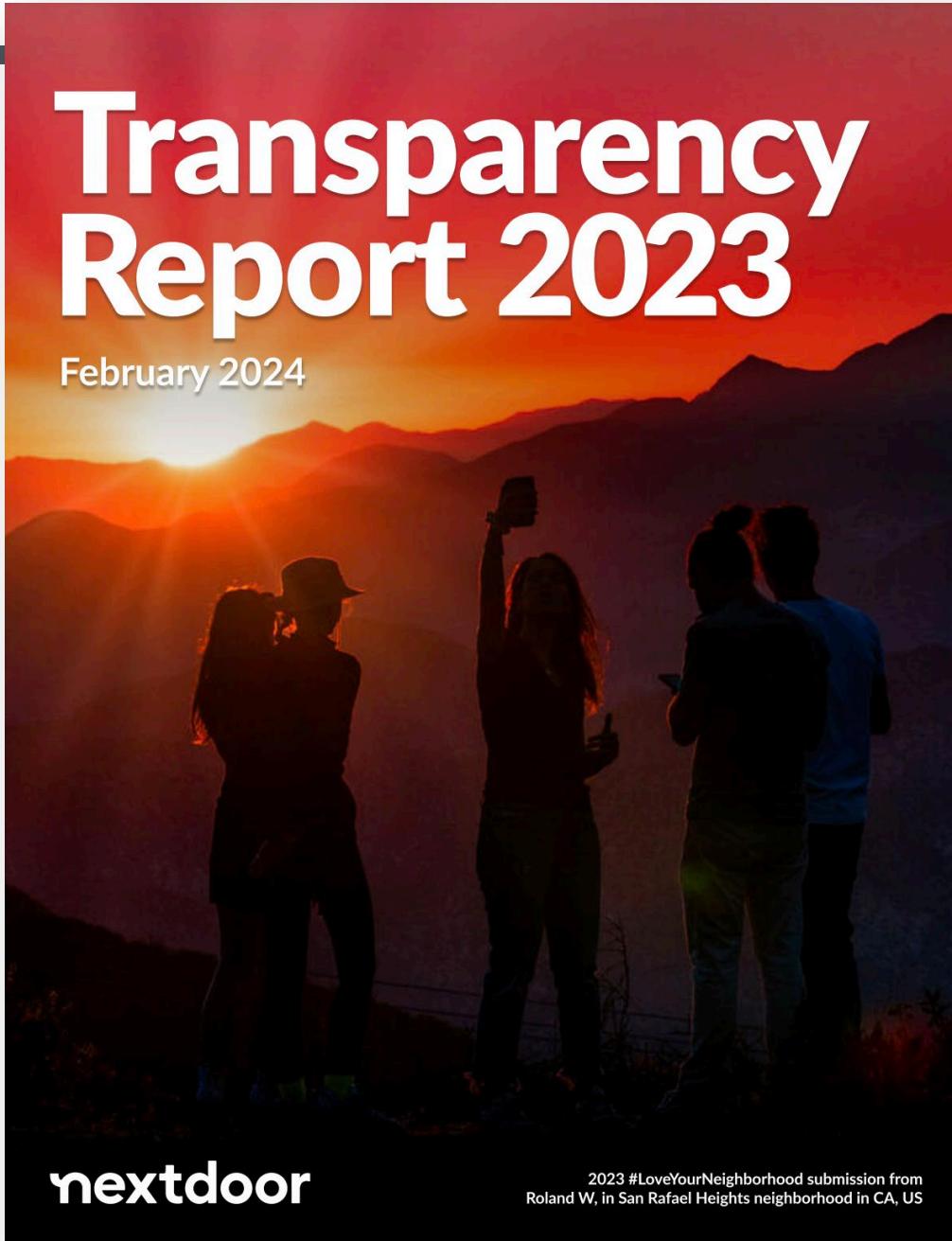
Prompt:

- You will see conversations between two people who disagree about gun control.
- Given the conversation, rephrase so that it is polite and non-defensive.
- Also, if the language is very strong, try to soften the tone of the message.
- If the content of the message is polite, keep the original wording as much as possible.
- Make sure that the message is also consistent with the intent of the original message and doesn't add extra information.

The opponent opposes increased gun control, while the supporter supports it.

<INSERT CONVERSATION w/MESSAGE TO BE REPHRASED>





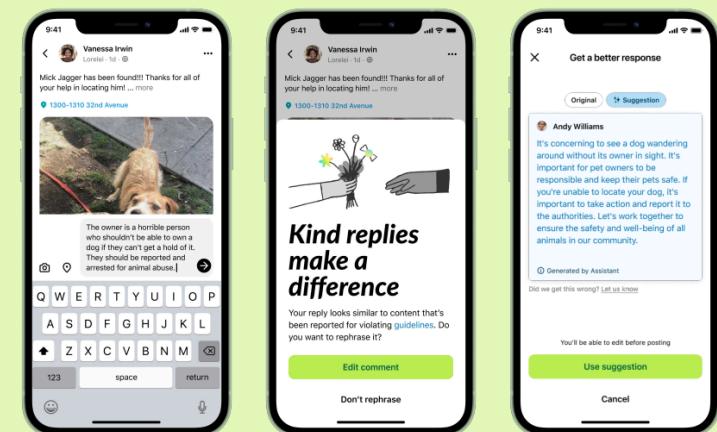
The image shows the cover of the NextDoor Transparency Report 2023. The title "Transparency Report 2023" is prominently displayed in large white letters at the top left. Below it, the date "February 2024" is shown. The background features a silhouette of three people standing on a hillside, looking out over a landscape with mountains under a sunset sky. At the bottom left is the NextDoor logo, and at the bottom center is a caption: "2023 #LoveYourNeighborhood submission from Roland W, in San Rafael Heights neighborhood in CA, US".

Red Teaming and Deploying Generative Artificial Intelligence to Aid Constructive Conversations in Neighborhoods

Nextdoor's Kindness Reminder is a historic feature designed to detect content that may violate the Community Guidelines and encourages the author to edit their content before they publish. It was the first of our core product features to introduce moments of friction aimed at slowing people down to combat incivility. This feature has consistently yielded positive outcomes over the years, with neighbors who encountered the reminder editing or withholding their content 35% of the time in 2021 and 36% of the time in 2022.

In 2023, Nextdoor introduced generative artificial intelligence (AI) into our product in a variety of critical ways. In particular, we gave neighbors the ability to use generative AI to help write posts that are more likely to drive positive community engagement², and made a formative change to the longstanding Kindness Reminder by integrating AI into the product. This new iteration (the Generative AI Kindness Reminder) takes the traditional Kindness Reminder a step further by suggesting more constructive revisions for flagged comments. Leveraging AI technology has enabled us to better analyze tone and context, and gives us the ability to propose alternative phrasings to prevent misunderstandings and conflicts.

2. We found that more than 70% of neighbors who opted to use generative AI to edit their posts accepted the generative AI suggestion, and that small businesses used the generative AI Assistant at twice the rate of other neighbors.

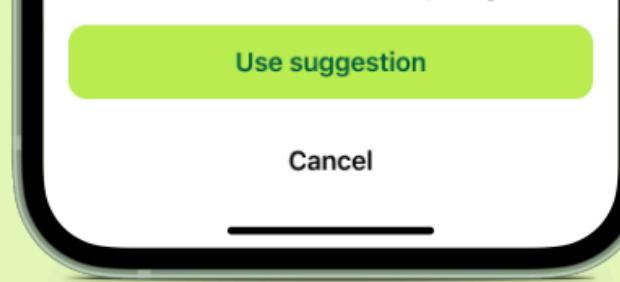
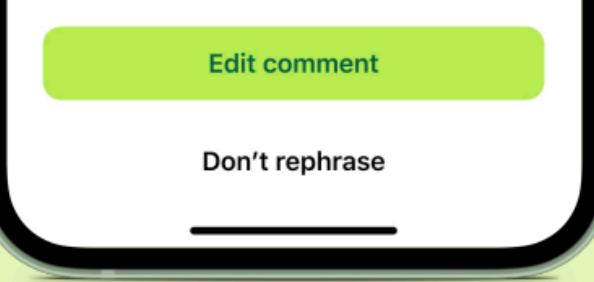


Prior to launch, we ensured adherence to our [Generative AI Ethics Principles](#)—Transparency, Privacy, Safety, Accountability, and Inclusivity. We also red teamed the integration with internal and external partners, including our [Neighborhood Vitality Advisory Board](#), and focused on issues such as fine tuning the generative AI model to uphold unique local speech patterns. Finally, we implemented a feedback mechanism for neighbors to report instances where the AI-generated content didn't perform as intended. Notably, only 0.16% of neighbors who encountered the AI Kindness Reminder provided constructive feedback such as changes in meaning, lack of coherence, or a mismatch with their usual tone.

The results are promising. In 2023, 36% of neighbors who saw the AI Kindness Reminder or the traditional Kindness Reminder (where neighbors self-edit without AI assistance) chose to edit or withhold their content, maintaining historically positive results. Further, of those who encountered the generative AI-revised text, 26% adopted the suggestion and published more constructive content. Exposure to the Kindness Reminder and AI Kindness Reminder **reduced the creation of content that violated the Community Guidelines by 15%**.

BYU

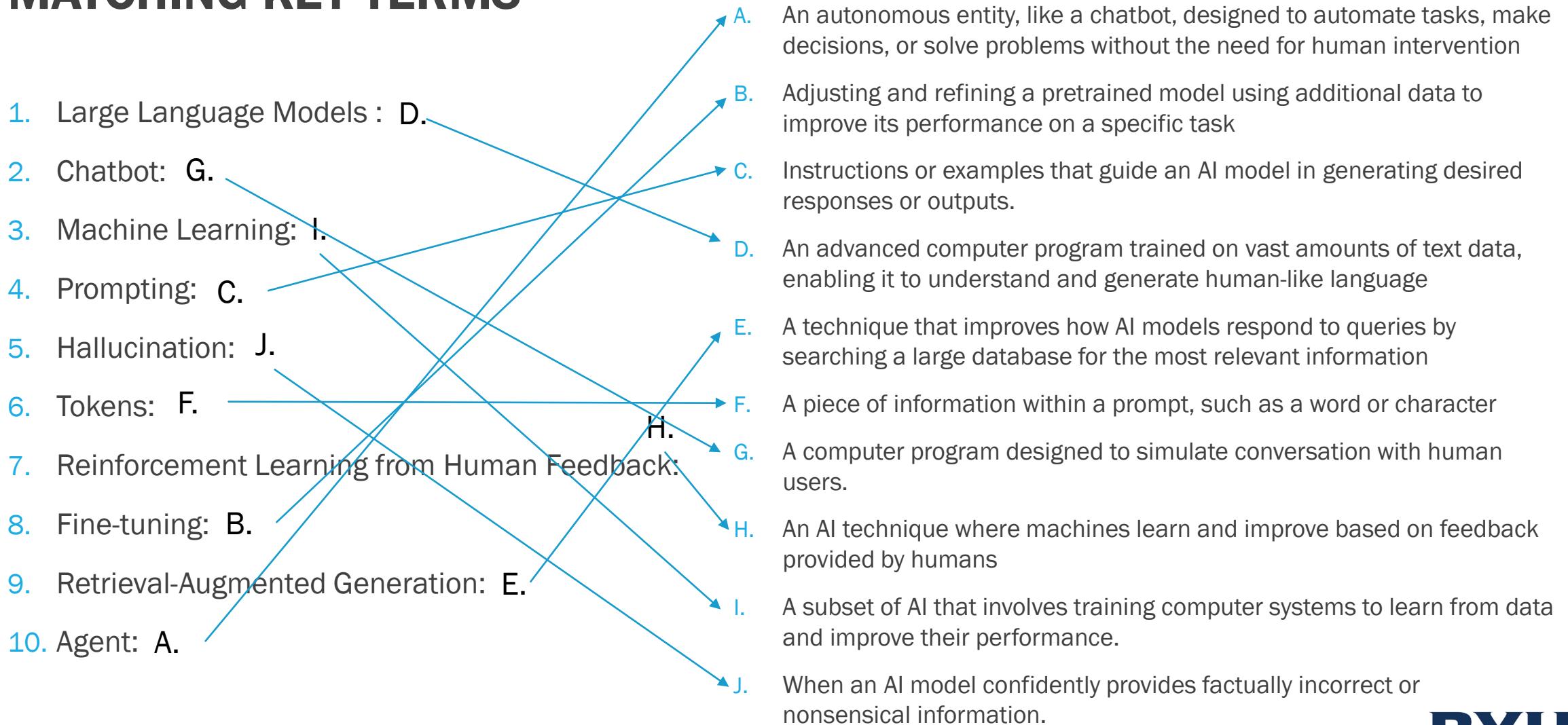
Source: NextDoor Transparency Report, 2023



Prior to launch, we ensured adherence to our [Generative AI Ethics Principles](#)—Transparency, Privacy, Safety, Accountability, and Inclusivity. We also red teamed the integration with internal and external partners, including our [Neighborhood Vitality Advisory Board](#), and focused on issues such as fine tuning the generative AI model to uphold unique local speech patterns. Finally, we implemented a feedback mechanism for neighbors to report instances where the AI-generated content didn't perform as intended. Notably, only 0.16% of neighbors who encountered the AI Kindness Reminder provided constructive feedback such as changes in meaning, lack of coherence, or a mismatch with their usual tone.

The results are promising. In 2023, 36% of neighbors who saw the AI Kindness Reminder or the traditional Kindness Reminder (where neighbors self-edit without AI assistance) chose to edit or withhold their content, maintaining historically positive results. Further, of those who encountered the generative AI-revised text, 26% adopted the suggestion and published more constructive content. Exposure to the Kindness Reminder and AI Kindness Reminder **reduced the creation of content that violated the Community Guidelines by 15%**.

MATCHING KEY TERMS





“

AI won't take your job – it's someone using AI that will take your job.

*Richard Baldwin, Economist
Growth Summit 2023*

BYU

AI PROVIDES NEW ANALYTIC & PROCESSING POSSIBILITIES

- Synthetic data
- Automated query generation
- Code completion
- Unstructured text analysis
- Embeddings
- Simulated virtual populations
- Sophisticated, semantically aware rewriting of text
- ...and more!



APPENDIX

SUNO SONGS: LEARNING AI WITH DR. T, EPIC ROCK DANCE POP

<https://suno.com/song/5c64d5e7-1515-4f73-b2ef-73fdf3a8cef6>

(Verse 1)

We walked into the class, just a little unsure,
A world so new, a code so pure,
But Dr. T had a spark in his eyes,
He said, "We'll reach beyond the skies."

(Chorus)

Learning AI with Dr. T,
Unlocking the secrets, setting minds free.
From algorithms to the heart of machines,
We're dreaming in code, chasing our dreams.

(Verse 2)

Each line we write, each problem we solve,
We see the future start to evolve.
Neural nets, and models we create,
With Dr. T, there's no limit to our fate.

(Chorus)

Learning AI with Dr. T,
Unlocking the secrets, setting minds free.
From algorithms to the heart of machines,
We're dreaming in code, chasing our dreams.

(Bridge)

When the code gets tough, and the answers hide,
We hear his voice, "Don't run, don't hide.
Keep pushing forward, one line at a time,
In the world of AI, you're ready to shine."

(Chorus)

Learning AI with Dr. T,
Unlocking the secrets, setting minds free.
From algorithms to the heart of machines,
We're dreaming in code, chasing our dreams.

(Outro)

So here we stand, bold and strong,
The future's ours, it won't be long.
With Dr. T, we've come so far,
Reaching for greatness, like shooting stars.

(Ending)

Learning AI with Dr. T,
Together we've built what the world will see.
In every byte, and every line,
We'll carry his lessons, for all time.



Use Case Families	Common AI Techniques						
	Generative Models	Predictive ML	Optimisation	Simulation	Rules	Graphs	
Forecasting	Low	High	Low	High	Medium	Low	
Planning	Low	Low	High	Medium	Medium	High	
Decision Intelligence	Low	Medium	High	High	High	Medium	
Autonomous System	Low	Medium	High	Medium	Medium	Low	
Segmentation	Medium	High	Low	Low	High	High	
Recommender	Medium	High	Medium	Low	Medium	High	
Perception	Medium	High	Low	Low	Low	Low	
Intelligent Automation	Medium	High	Low	Low	High	Medium	
Anomaly Detection	Medium	High	Low	Medium	Medium	High	
Content Generation	High	Low	Low	High	Low	Low	
Chatbots	High	High	Low	Low	Medium	High	
Knowledge Discovery	High	Medium	Low	Low	Medium	High	