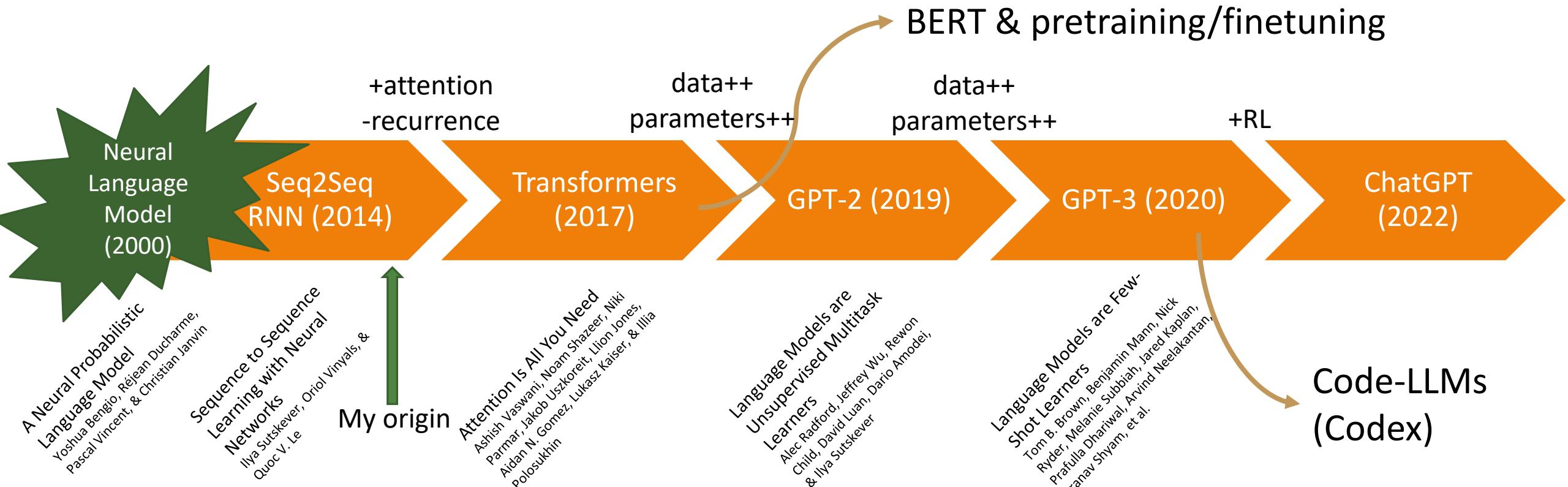


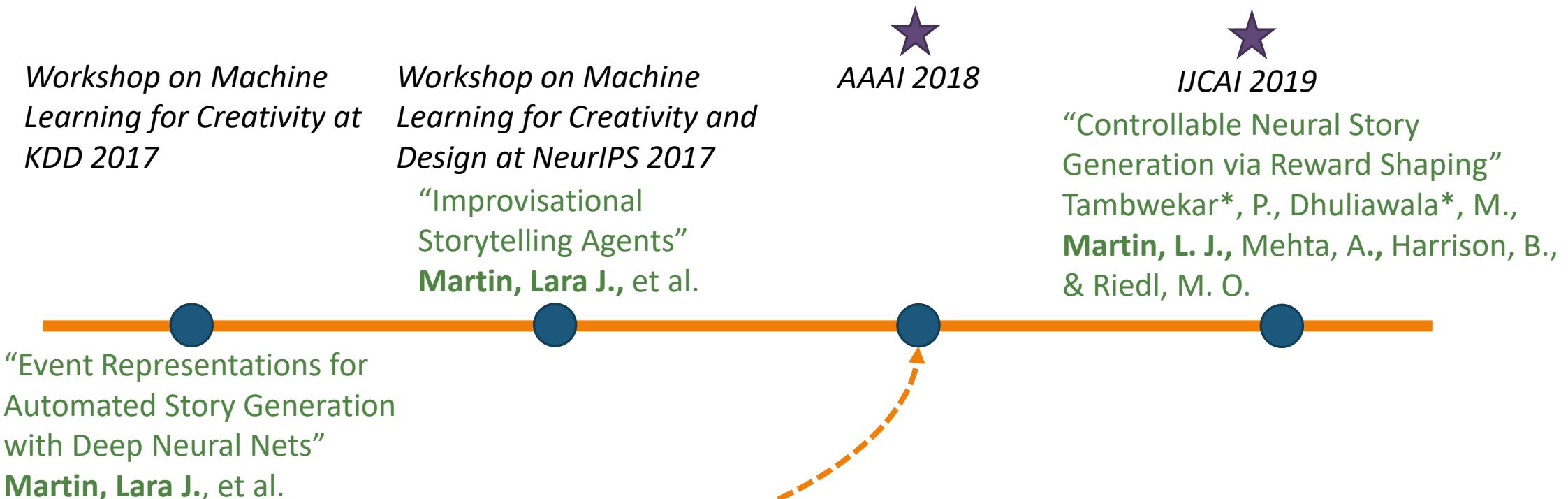
Storytelling, D&D, and AAC

CMSC 473/673 - NATURAL LANGUAGE PROCESSING

Neural Language Model Timeline



My early history with neural language models...



Seq2Seq Neural Networks (2017)

R2-D2 carrying some drinks on a tray strapped to his back passes Yoda who uses his powers to hog the drinks

Expected:

Obi Wan and Anakin are drinking happily when Chewbacca takes a Polaroid picture of Anakin and Obi Wan

Predicted:

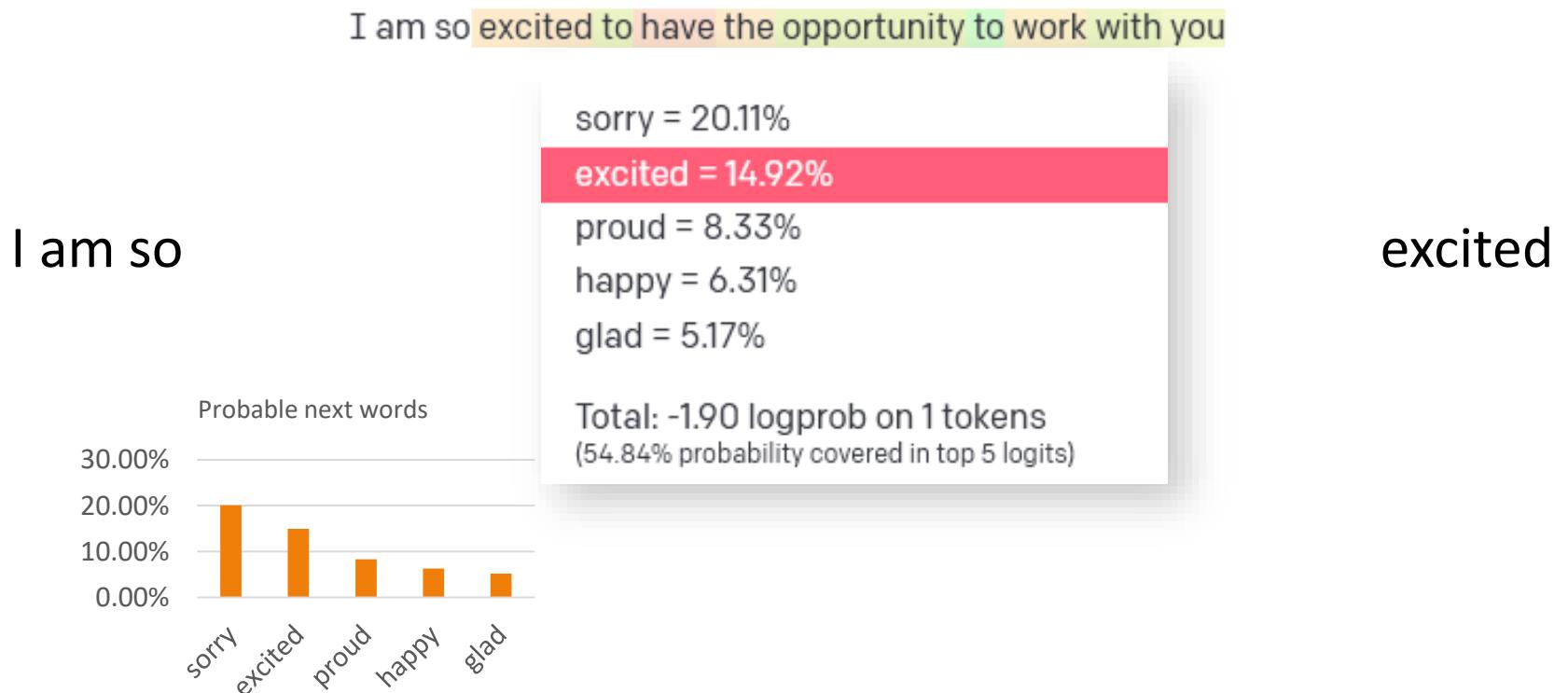
Can this block gives him the advantage to personally run around with a large stick of cheese



Summary for *Lego Star Wars: Revenge of the Brick*. From the original presentation for

Martin, Lara J., Prithviraj Ammanabrolu, William Hancock, Shruti Singh, Brent Harrison, and Mark O. Riedl. "Event Representations for Automated Story Generation with Deep Neural Nets." *KDD 2017 Workshop on Machine Learning for Creativity*, 2017.

What is a language model?



What isn't a language model?

Likelihoods → Not cause & effect

What is probable might not be possible.

GPT-3 Tells a Story

The hungry dog licked her lips as she watched her owner eat.



GPT-3 Tells a Story

The hungry dog licked her lips as she watched her owner eat.

“You’ve been a good girl,” he told her. “I think you deserve a reward.”

Once she was done, she jumped back on the couch and waited patiently.

What does this refer to?

Her owner took a piece of steak out of the fridge and gave it to her.

“Thank you,” he said. “I’m glad you’re my dog.”

She wagged her tail and ate the steak.

Forgot what has happened already

“If you’re good, you can have a treat later,” he said. “But for now, you have to sleep. I have a long day tomorrow.”

She nodded and lay down on the floor.

Lacking commonsense

Her owner got up, turned off the lights, and lay down on the bed.

But ChatGPT solves all these problems, right?

A few turns into the story:

LA

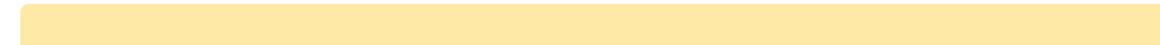
It was spying on the Hendrick's garbage can.

LA

The giant squirrel, whose name was Fwwok, decided to plan a heist and gathered their top 5 compatriots.

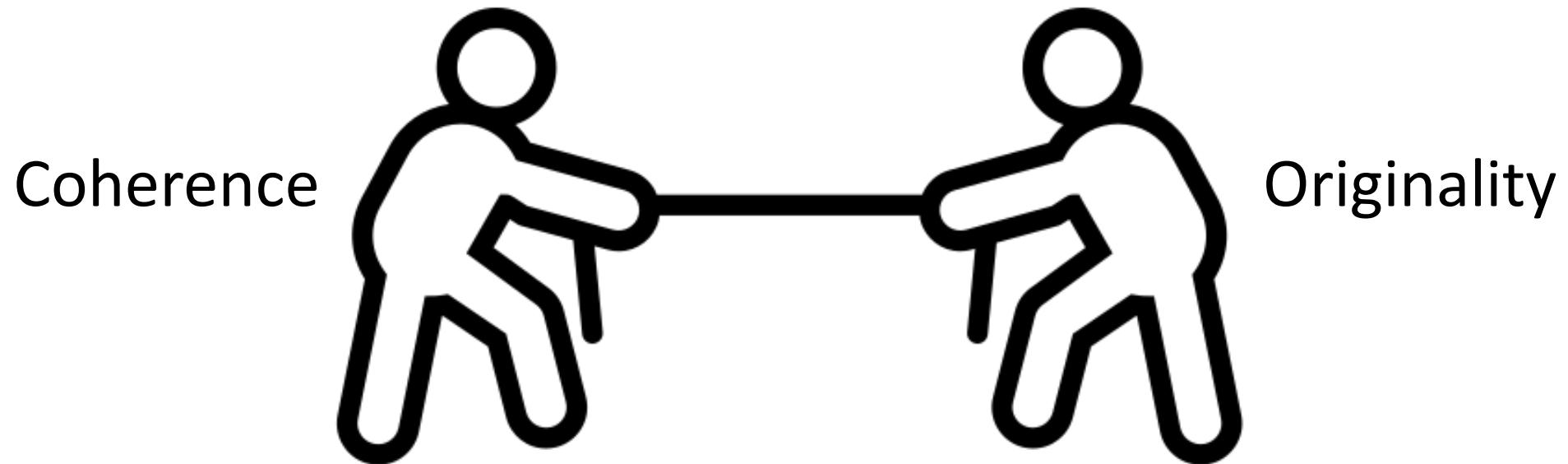


Rephrasing me



Inconsistent over time

What is a language model? (alternate version)



There's even an explicit knob in GPT-3

Playground Save View code Share ...

Mode

Model: text-curie-001

Temperature: 0.35

Does it always rain on Tuesdays?

No, it does not always rain on Tuesdays.

Does it always rain on Tuesdays?

No, Wednesday is the normal precipitation day. However, Tuesday can occasionally experience light rain or even a thunderstorm.

Mode

Model: text-curie-001

Temperature: 1

Do we really need to trade off?



https://st2.depositphotos.com/1151061/6575/v/450/depositphotos_65750413-stock-illustration-story-reading.jpg

Entertain us

Tell us a spooky
story about the
Jersey Devil.



Plan with us



Will the Ikea Klippan loveseat fit into my 2019 Honda Odyssey if I fold down the seats?



87/100

Will the Ikea Klippan loveseat fit into my 2019 Honda Odyssey if I fold down the seats?

KLIPPAN Love seat, Vissle gray - IKEA

<https://www.ikea.com/us/en/p/klippan-loveseat...>

KLIPPAN Love seat, Vissle gray We launched KLIPPAN sofa in the 1980s, and it's still a favorite. It's comfortable, fits almost everywhere and has ...

4/5 ★★★★ (41) Price: \$319

Availability: In stock



Other content from ikea.com

[Living Room | Klippan Love seat, Vissle Gray - Ikea](#)

[Furniture - Shop Furniture Online & In-Store Near You - Ikea](#)

[Sofas & Armchairs | Klippan Love seat, Vissle Gray - Ikea](#)

[See more](#)

KLIPPAN Love seat, Bomstad black - IKEA

<https://www.ikea.com/us/en/p/klippan-loveseat...>

If you want a leather feel and an easy-care surface, coated fabric is a durable alternative – often at a lower price. We use fabric made of ...

5/5 ★★★★★ (18) Price: \$449

Availability: In stock



Can a Love seat fit in the back? | Honda Odyssey Forum

<https://www.odyclub.com/threads/can-a-loveseat-fit-in-the-back.359361...>

Jun 8, 2020 · If you are moving the love seat with one passenger I would highly recommend removing the second row seats and have the 3rd row folded flat. You may be able to fit



According to the web, the **Ikea Klippan loveseat** measures **70.875" (180 cm) wide, 34.625" (88 cm) deep, and 26" (66 cm) high**¹. The **2019 Honda Odyssey** provides **38.6 cubic feet of cargo space** with all seats in use, and **158 cubic feet** with the second and third rows folded². Based on these dimensions, it seems that the loveseat will not fit into the Odyssey with all seats in use, but it might fit with the second and third rows folded, depending on the orientation and shape of the loveseat. However, this is not a definitive answer and you should always measure the actual items before attempting to transport them.

Learn more: [1. dimensions.com](#) +8 more

Thank you for the information.

How do I fold the seats in the Odyssey?

What are some other options for moving furniture?

Let's chat

I want to buy a new car for my family. What should I consider?



I want to buy a new car for my family. What should I consider?

Here are some things to consider when buying a new car for your family:

- **Your budget:** How much can you afford to spend on a new car? Keep in mind that the initial purchase price is just the beginning. You'll also need to factor ongoing costs like fuel, maintenance, and insurance.
- **Seating:** How many people will be riding in the car on a regular basis? Do you need a car with three rows of seats, or will two rows be sufficient?
- **Features:** What features are important to you? Do you want a car with a sunroof, leather seats, or parking cameras?
- **Safety:** Safety is always a top priority when buying a new car. Look for cars with good crash test ratings and features like airbags.
- **Fuel type:** Do you want a gasoline-powered car, a diesel-powered car, or an electric car? Each type has its own pros and cons.

Ultimately, the best car for your family is the one that meets your needs and fits your budget. Test drive a few different models to find the perfect one for you.



Check it

⋮

What's on your mind?



Bard may give inaccurate or inappropriate information. Your feedback makes Bard more helpful and safe.

Prepare us



Support us

I just broke up with
my boyfriend...

I'm so sorry to hear
that. Let me tell you
about the time...



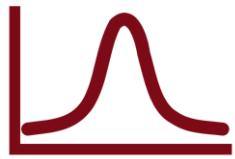
Outline

PhD

Automated Story Generation



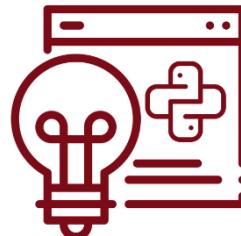
Events



Plot Progression

AAAI 2018

Story
Understanding



Code-LLMs

ACL 2023

Postdoc

Human
Communication



AAC

Submitted to
ASSETS 2025

Present

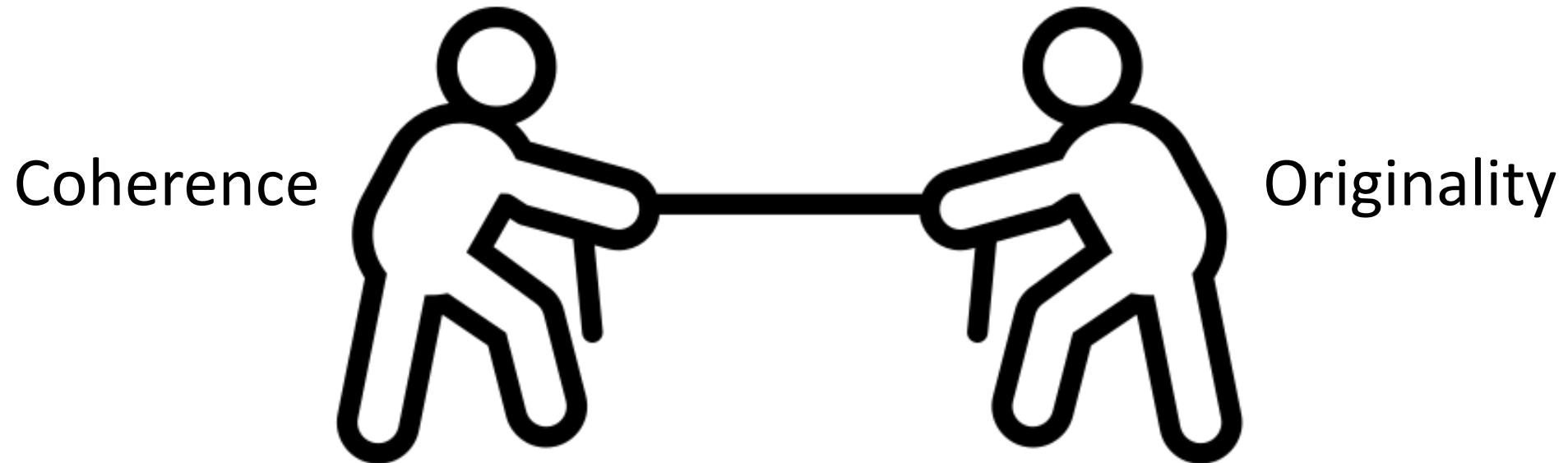


Trans* TTS



AI D&D

What is a language model? (alternate version)



Why is this happening?!

R2-D2 carrying some drinks on a tray strapped to his back passes Yoda who uses his force powers to hog the drinks

Can this block gives him the advantage to personally run around with a large stick of cheese

Too unique!



Problem

Sparsity!



Solution

Separating semantics (meaning) from syntax
(grammar)

More overlap!

Event Representations

Use linguistic knowledge to bootstrap the neural network

From sentence, extract event representation

(subject, verb, direct object, modifier)

Original sentence: yoda uses the force to take apart the platform

Event: yoda use force \emptyset

Generalized Event: <PERSON>0 use-105.1 causal_agent.n.01 \emptyset

WordNet

- S: (n) power, force (one possessing or exercising power or influence or authority)
"the mysterious presence of an evil power"; "may the force be with you"; "the forces of evil"
 - direct hyponym / full hyponym
 - direct hypernym / inherited hypernym / sister term
 - S: (n) causal agent, cause, causal agency (any entity that produces an effect or is responsible for events or results)

<http://wordnetweb.princeton.edu/perl/webwn?s=force&sub=Search+WordNet&o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&h=>



VerbNet

Full Class View

use-105.1

Member Verb Lemmas:

USE

UTILIZE

APPLY

EMPLOY

EXERT

EXPLOIT

PLAY

REUSE

RUN

WORK

ROLES:

Agent [+animate | +organization]

Theme

Eventuality

EXAMPLE:

I used the cupboard to store food.

SHOW DEPENDENCY PARSE TREE

SYNTAX:

Agent VERB Theme Eventuality

SEMANTICS:

USE(e1 , Agent , Theme , Eventuality)

FORCE DYNAMICS:

Volitional Internal Purp FD representation

ASTER* Pipeline

*Automated Story-Telling with Event Representations

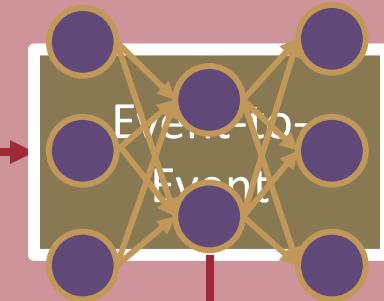
John unwittingly
unleashes an
insidious pox.
sentence_n

<PERSON>0 disassemble-23.3 contagious_disease.n.01 Ø

Eventify

John unleash pox Ø

event_n



male.n.02
spatial_configuration-47.6
Ø adopt-93

event_{n+1}

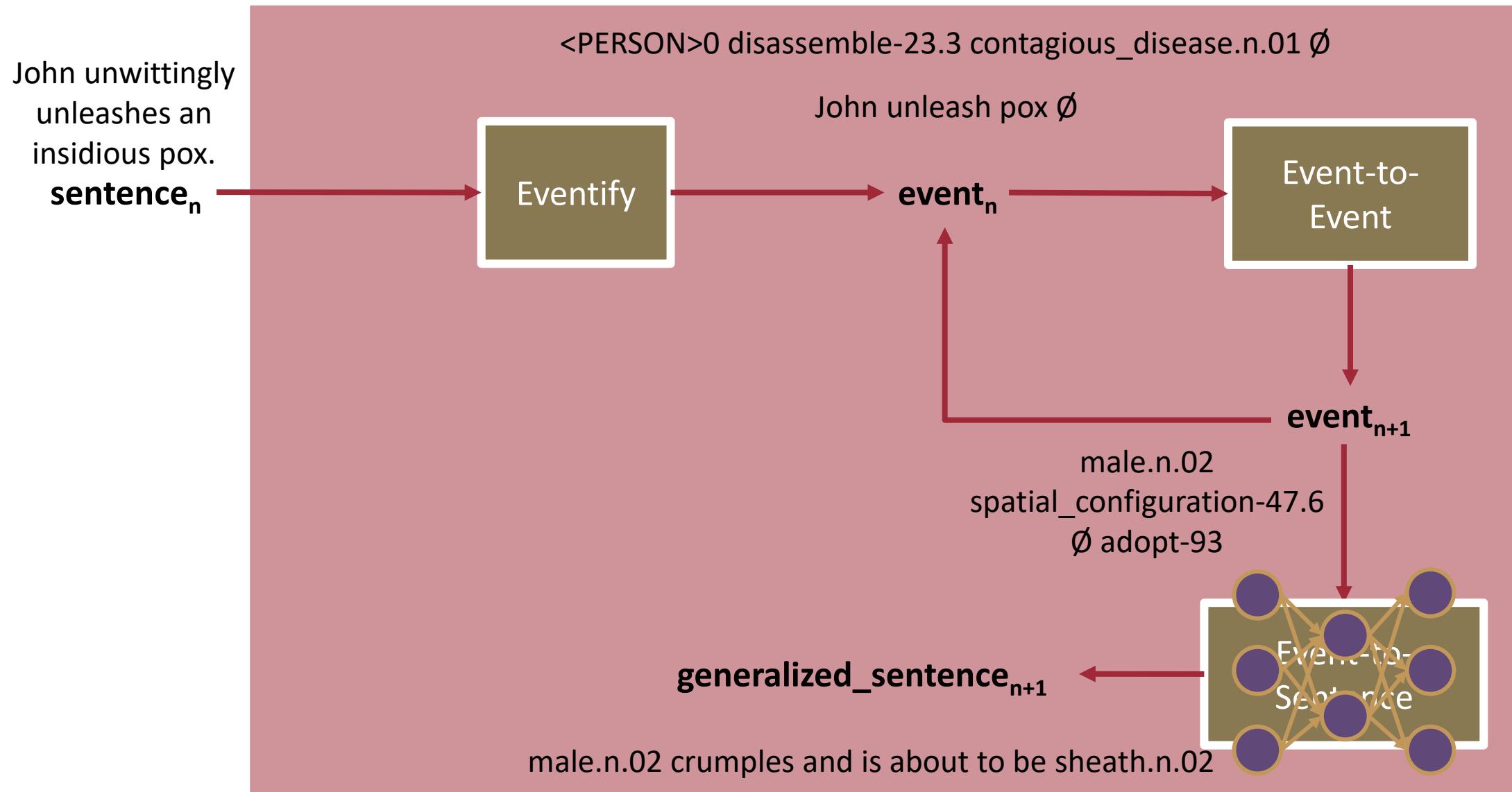
Events make NNs less surprised about what comes next

Experiment	Perplexity ↓
(0) Original Sentences	705
(1) Original Words Baseline	749
(2) Original Words with PERSONs	167
(3) Generalized Baseline	54

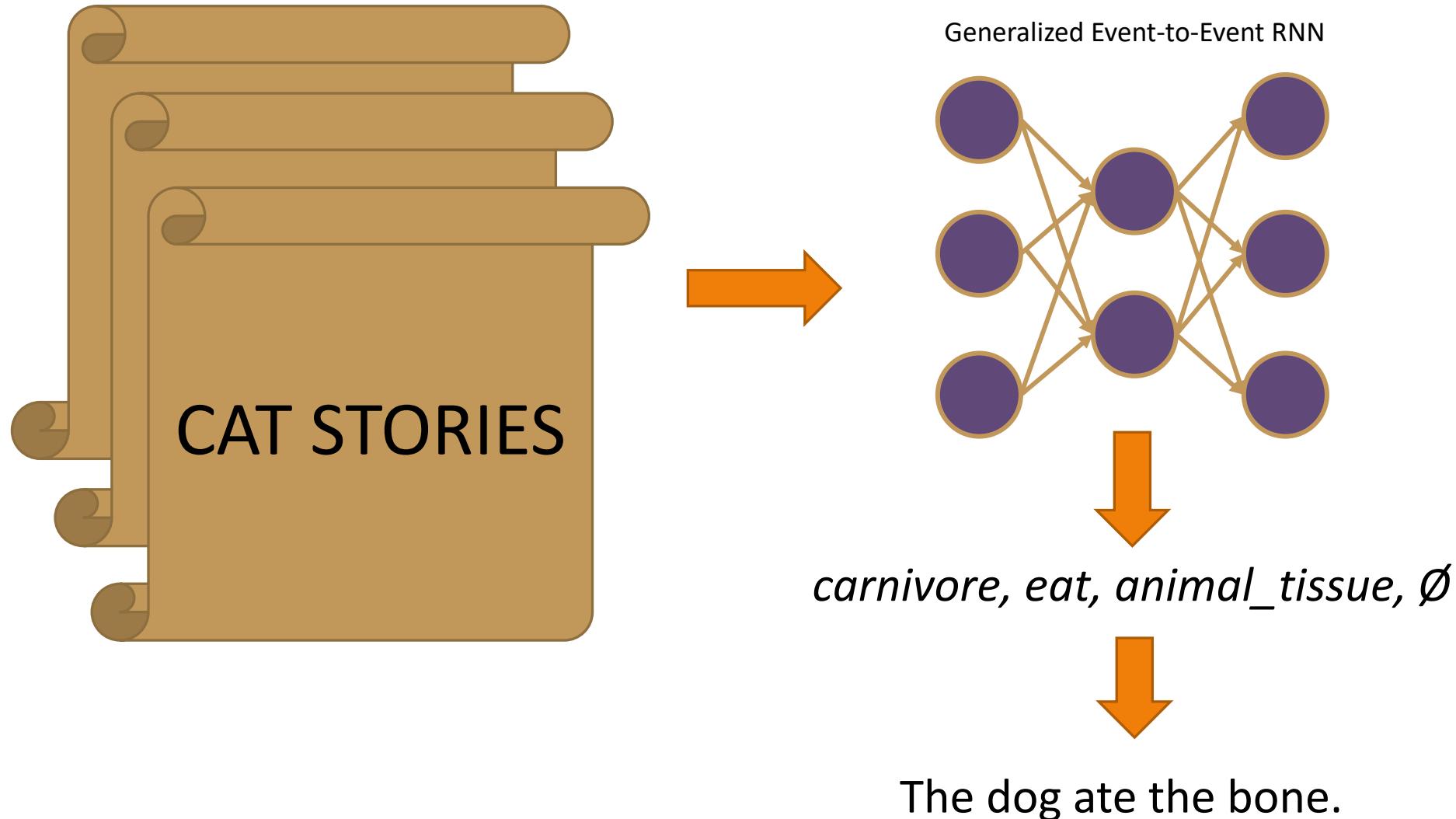
$$\text{Perplexity} = 2^{-\sum_x p(x) \log_2 p(x)}$$

$$p(x) = \frac{\text{count}(x)}{\sum_{y \in Y} \text{count}(y)}$$

ASTER Pipeline



Why did we keep the sentences generalized?

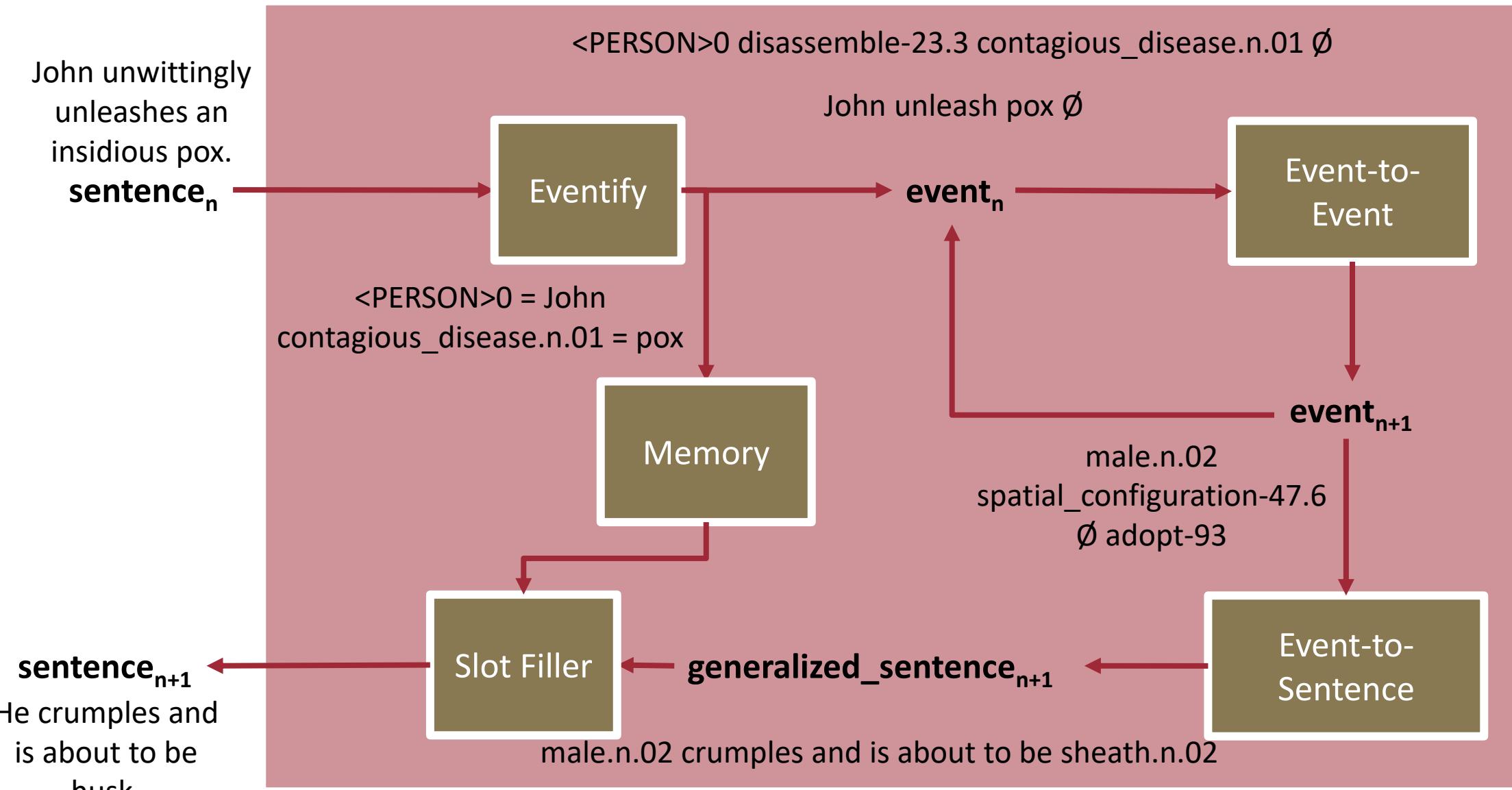


Generalizing helps generating sentences

Experiment	Perplexity ↓	BLEU ↑
Original Words Event → Original Sentence	1585	0.0016
Generalized Event → Generalized Sentence	57	0.0331
All Generalized Events → Generalized Sentence	59	0.0366

Precision using
n-grams

ASTER Pipeline



Summary

Writers like to be unique with their writings

This makes it hard for ML to see patterns

I created an event representation to aid in pattern finding

This created a giant system that still rambles

Outline

PhD

Automated Story Generation



Events

AAAI 2018

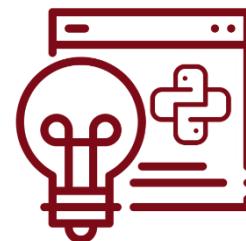


Plot Progression

IJCAI 2019

Postdoc

Story Understanding



Code-LLMs

ACL 2023

Human Communication



AAC

Submitted to ASSETS 2025

Present

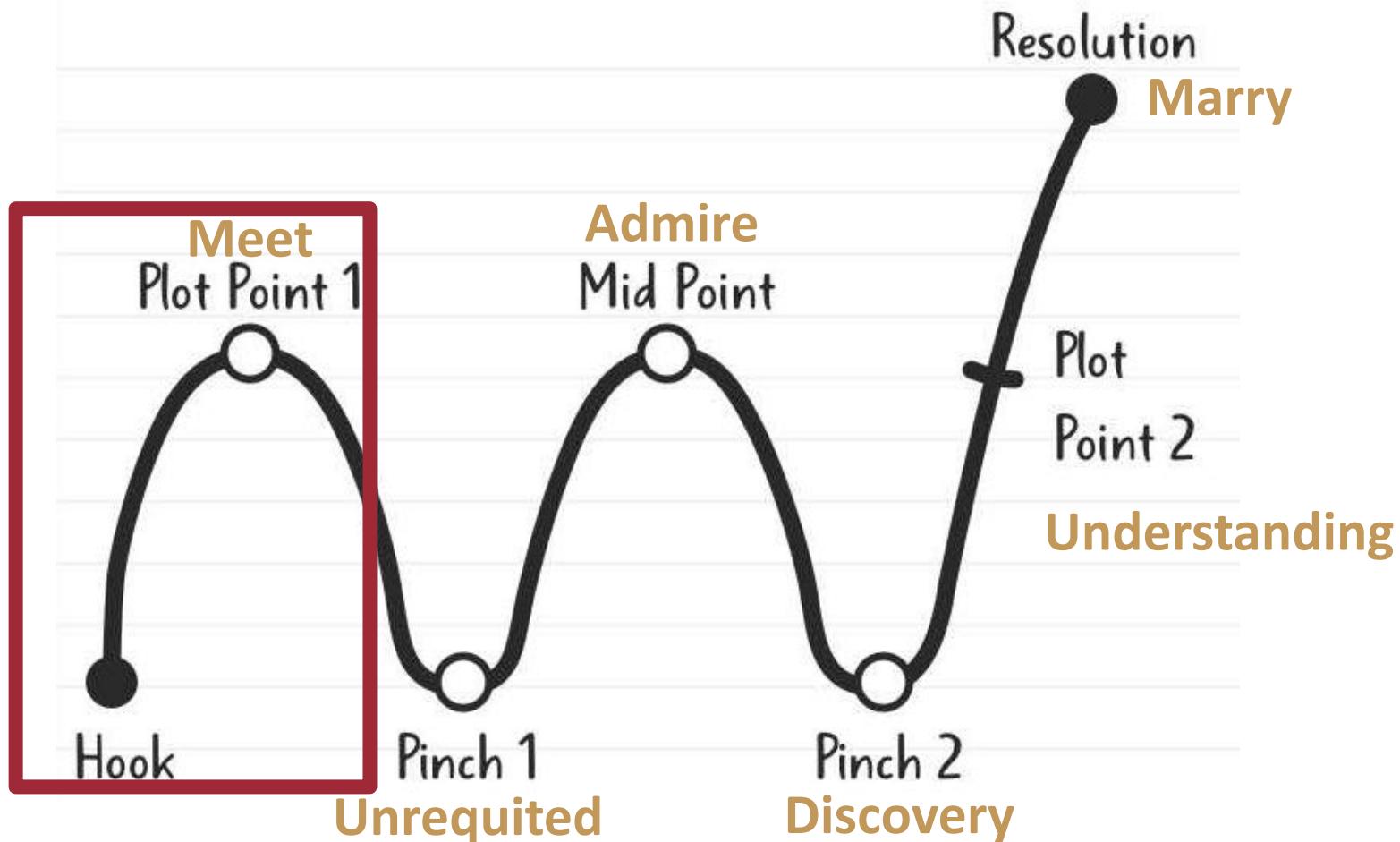


Trans* TTS

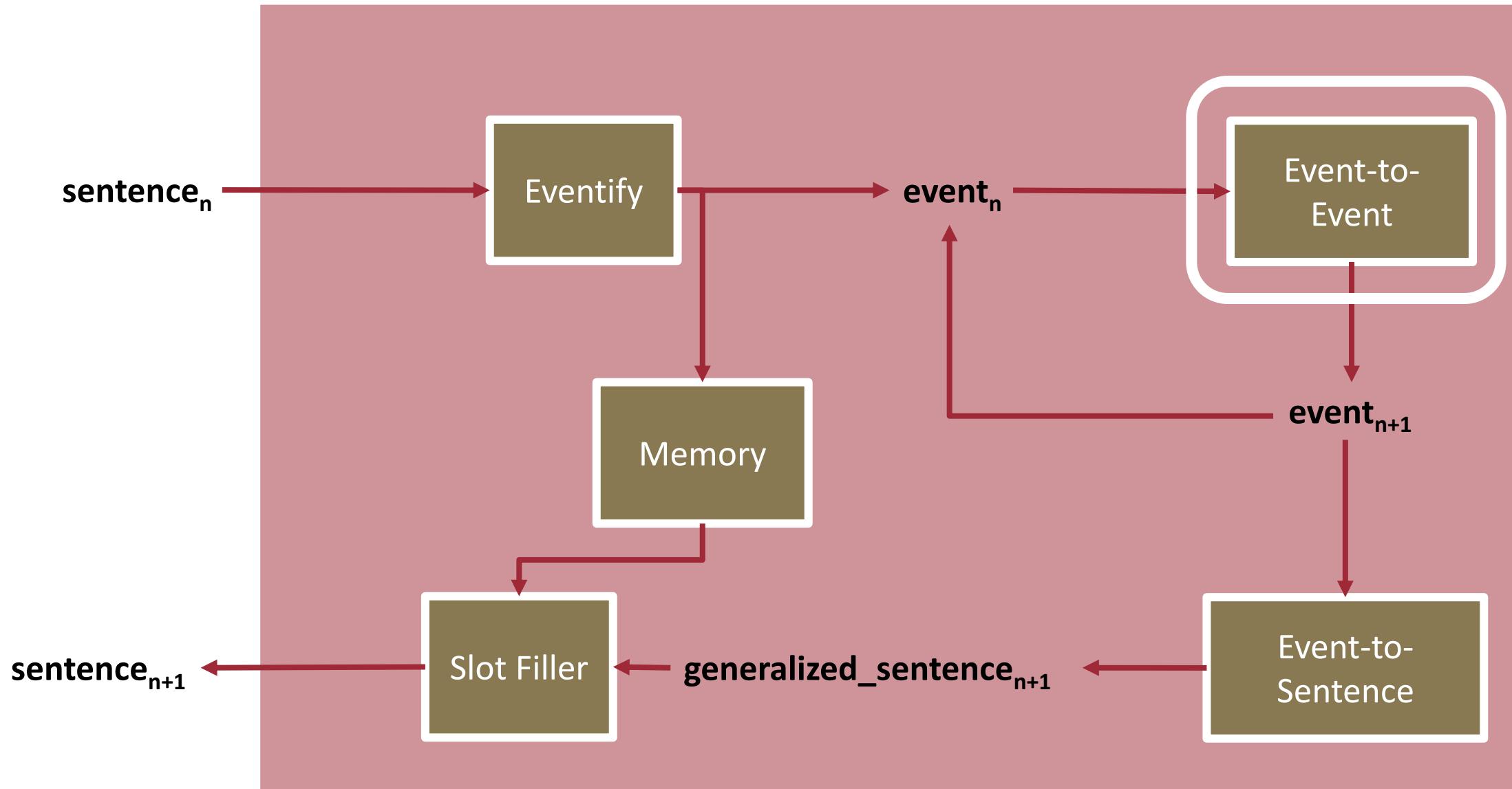


AI D&D

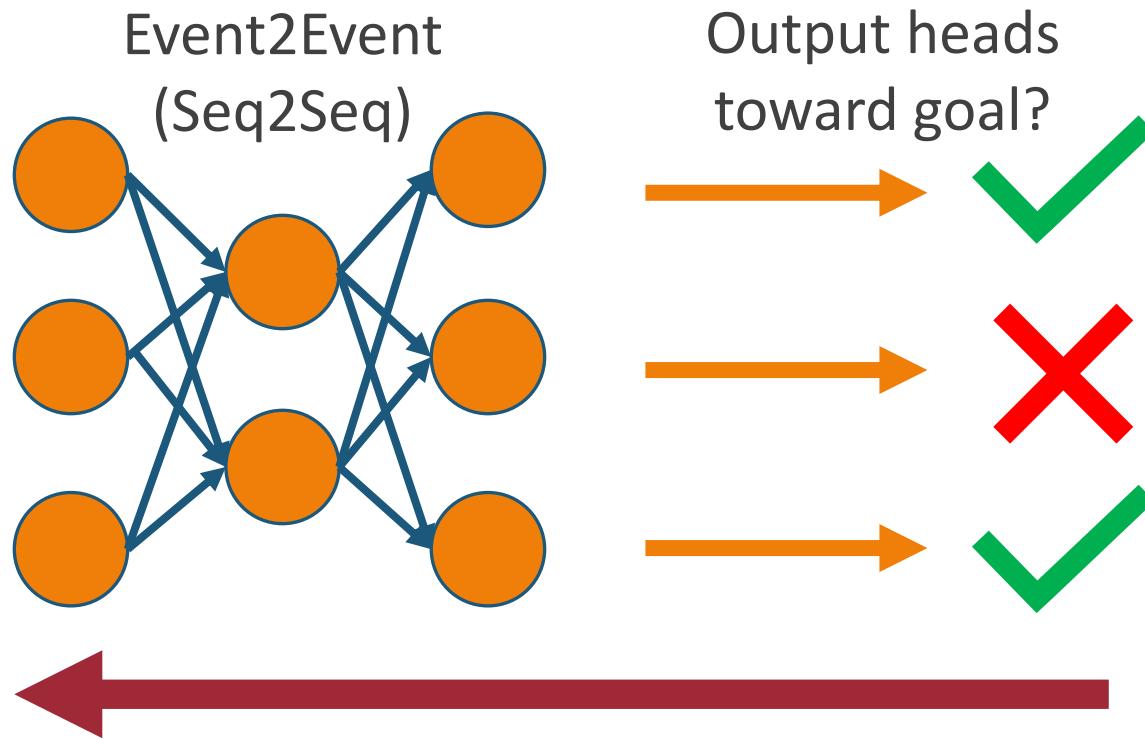
Global Coherence



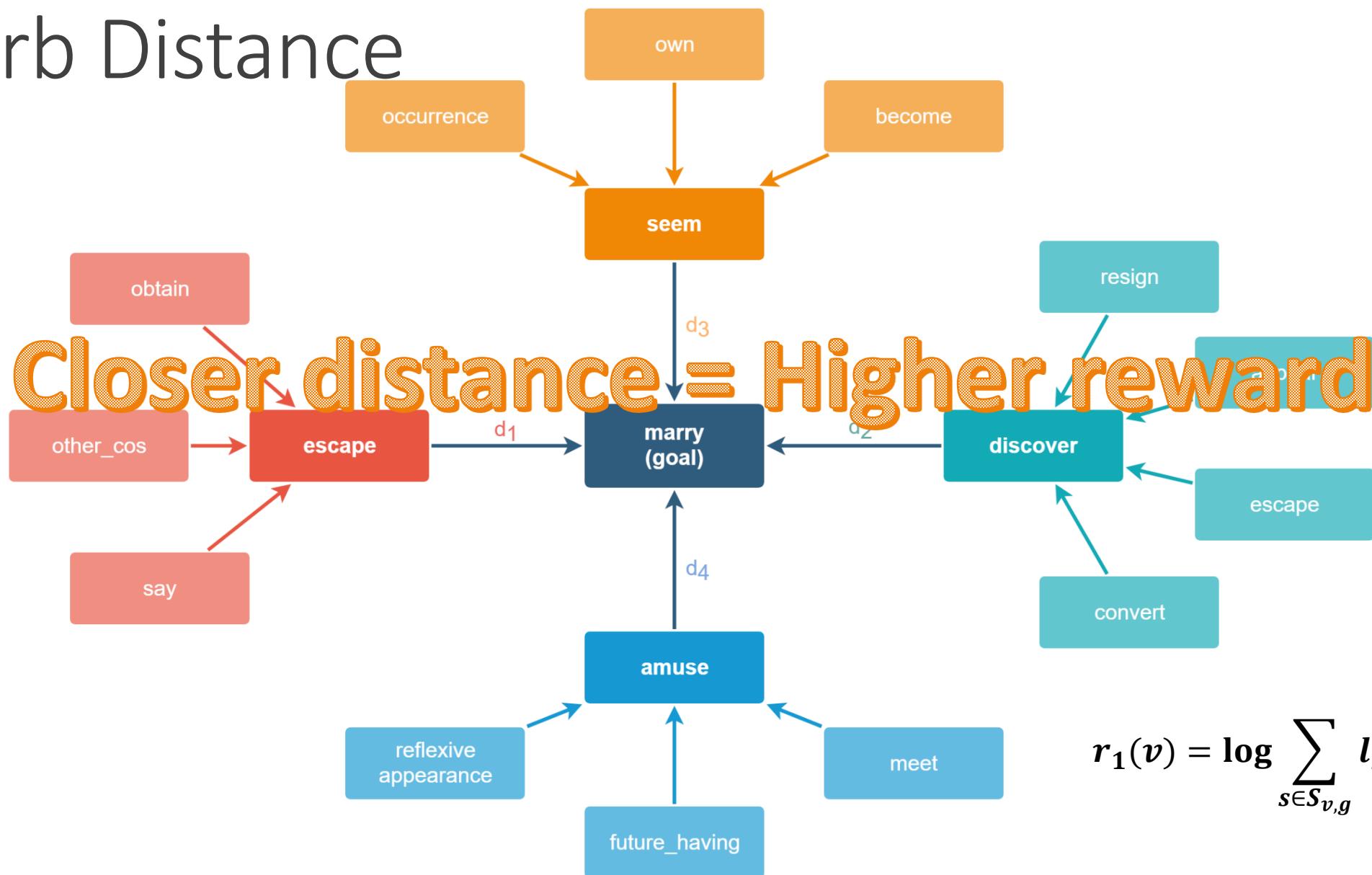
ASTER Pipeline



Policy Gradient DRL (REINFORCE)



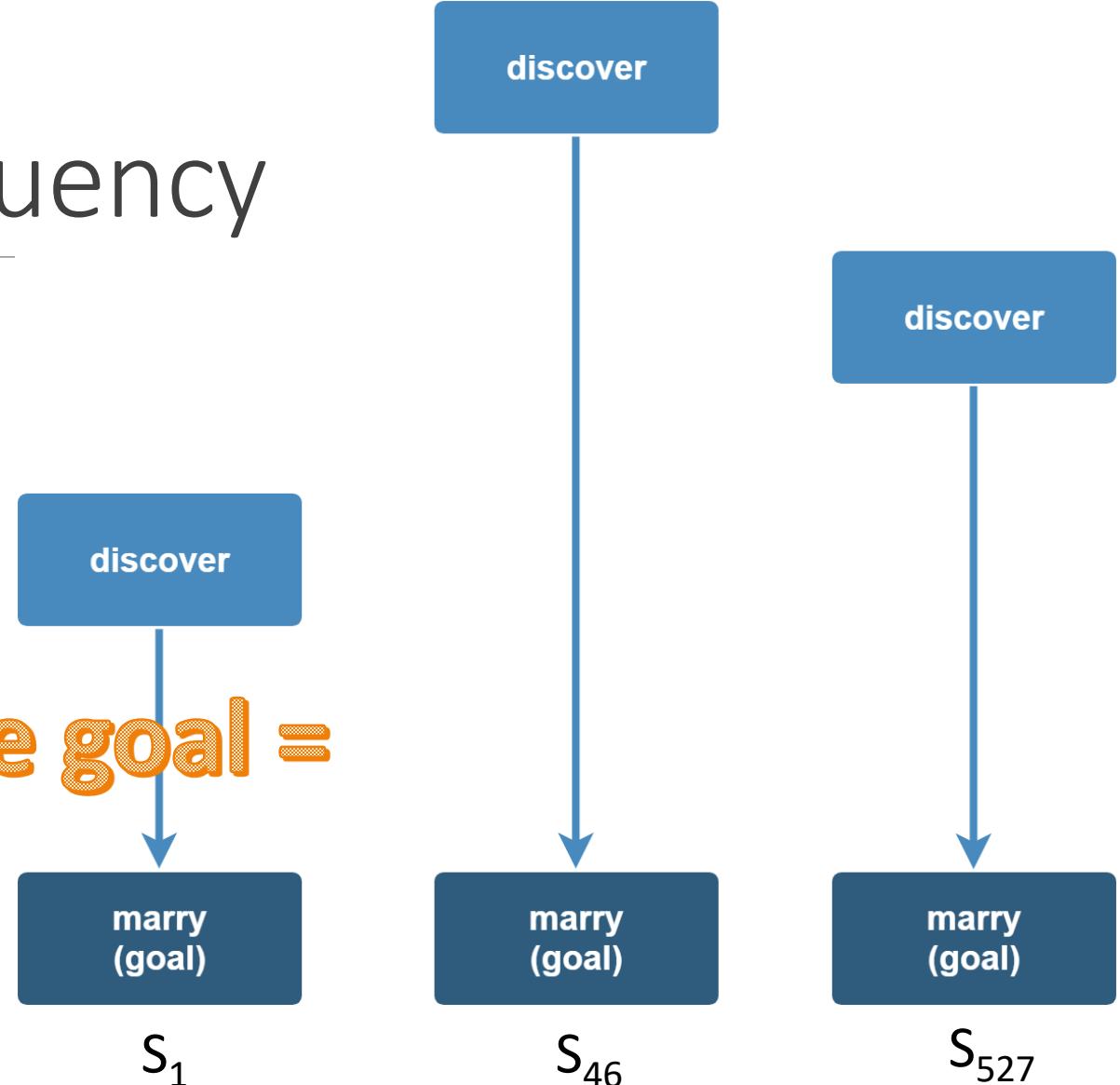
#1 Verb Distance



#2 Story-Verb Frequency

$$r_2(v) = \log \frac{k_{v,g}}{N_v}$$

Appear frequently before goal =
Higher reward



Final Reward Equation

$$R(v) = \alpha \times r_1(v) \times r_2(v)$$

Verb Distance to Goal Story-Verb Frequency

DRL system is more controlled generation

Goal	Model	Goal Achievement Rate ↑	Average Perplexity ↓	Average Story Length
admire	Test Corpus	20.30%	n/a	7.59
	Seq2Seq	35.52%	48.06	7.11
	Policy Gradient DRL	94.29%	7.61	4.90
marry	Test Corpus	24.64%	n/a	7.37
	Seq2Seq	39.92%	48.06	6.94
	Policy Gradient DRL	93.35%	7.05	5.76

But are the stories
actually any *good*?

Methods

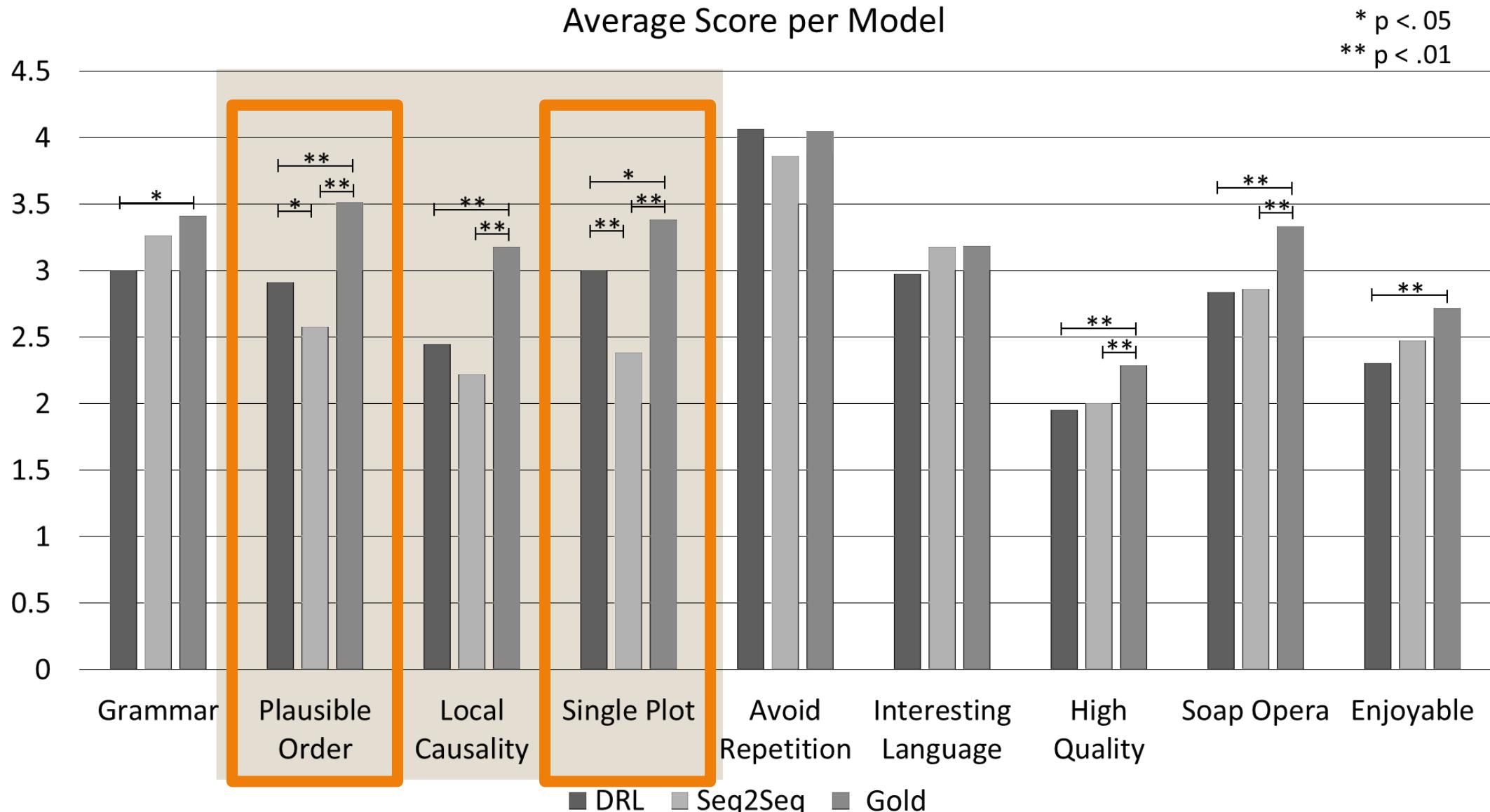
175 Mechanical Turkers rated statements on a 5-point Likert scale

For each of 3 conditions:

- Policy Gradient DRL
- Baseline Seq2Seq
- Human-written Stories (Gold Standard)

Questionnaire

1. This story exhibits CORRECT GRAMMAR.
2. This story's events occur in a PLAUSIBLE ORDER. Coherence
3. This story's sentences MAKE SENSE given sentences before and after them.
4. This story FOLLOWS A SINGLE PLOT.
5. This story AVOIDS REPETITION.
6. This story uses INTERESTING LANGUAGE.
7. This story is of HIGH QUALITY.
8. This story REMINDS ME OF A SOAP OPERA.
9. This story is ENJOYABLE.



Story Generation Takeaways

I have created a representation so that semantic events can be generated separately from syntax.

This created more semantically-accurate plot generation

but it would still lose coherence over time.

So we created a way to finetune the event generator to behave as RL.

This made for more consistent, plot-driven generation.

Back to the Future

All of this previous work was before the era of finetuning.

It was challenging to get these models to even output grammatical sentences.



>>



How much do LLMs actually understand stories?

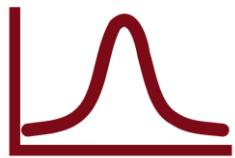
Outline

Past

Automated Story Generation



Events



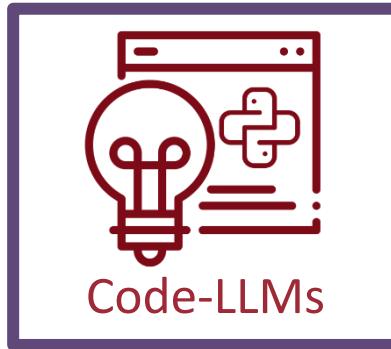
Plot Progression

AAAI 2018

IJCAI 2019

Present

Story
Understanding



Code-LLMs
Submitted to
ACL 2023

Human
Communication



AAC

TBA

Future



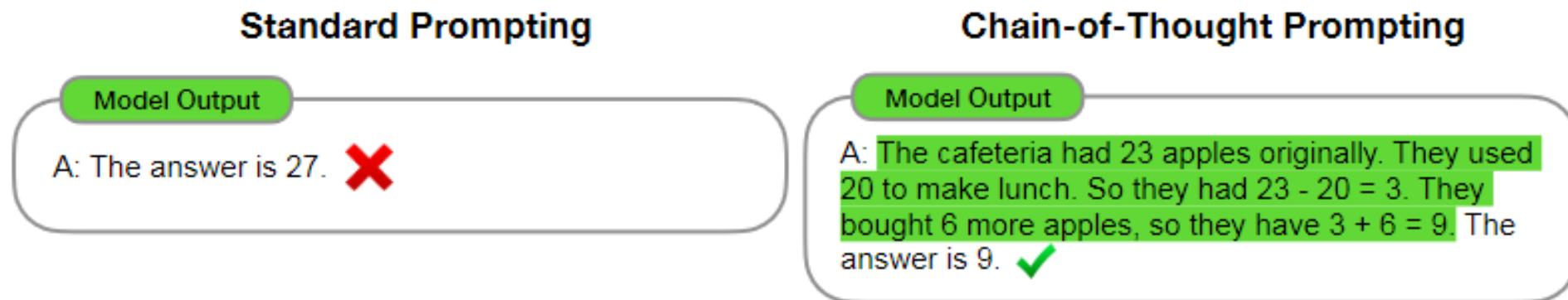
Trans* TTS



AI D&D

Chain-of-Thought Prompting

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?



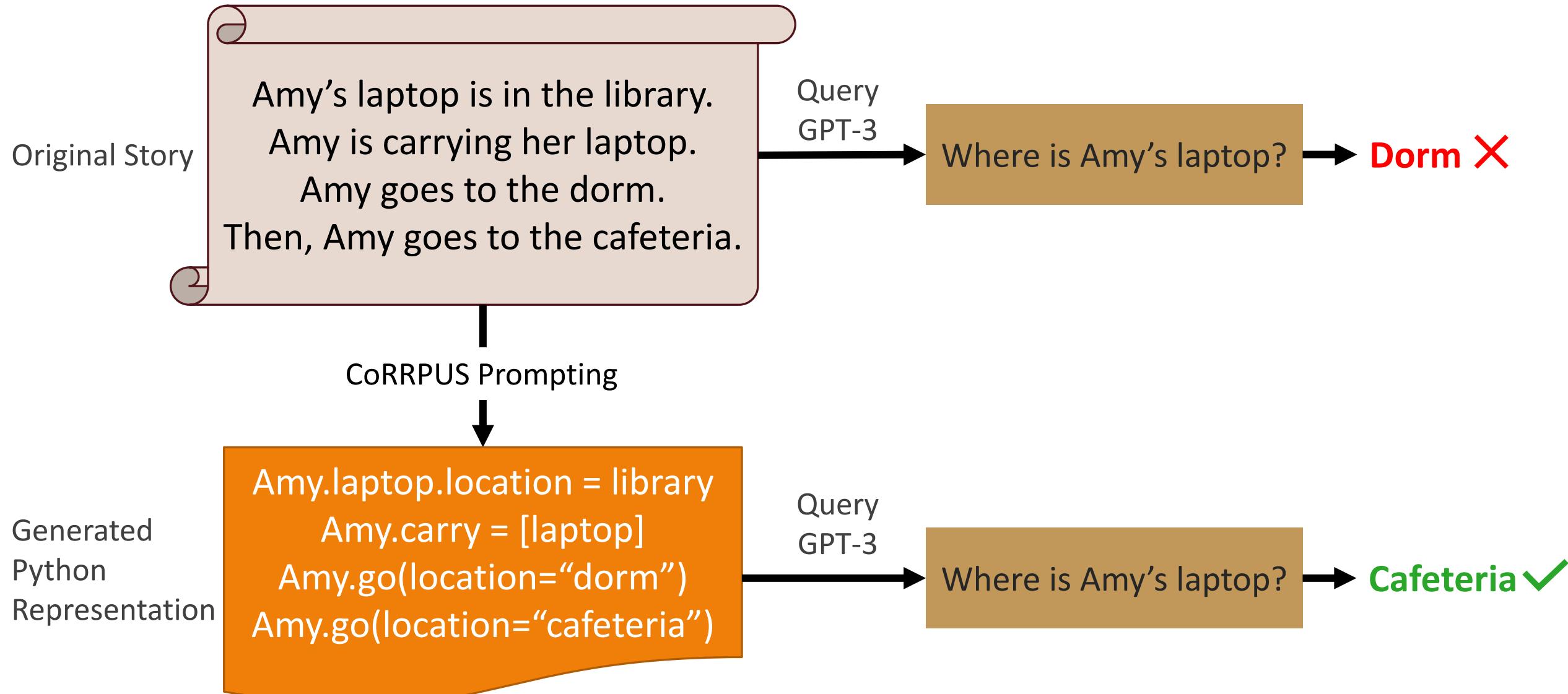
Motivation

Can we use CoT prompting for story understanding?

Will a Code-LLM do better than a regular LLM because of the innate structure of code?

CoRRPUS

(Code Representations to Reason & Prompt over for Understanding in Stories)



CoRRPUS Chain-of-Thought Prompting

Three versions that are initialized the same:

Comment

```
def story(self):
    ## Mary moved to the bathroom.
    self.Mary.location = "bathroom"
    ## Mary got the football there.
    self.Mary.inventory.append("football")
...
...
```

Specific Functions

```
self.Mary_moved_to_the_bathroom()
self.Mary_got_the_football_there()
self.John_went_to_the_kitchen()
self.Mary_went_back_to_the_garden()
```

```
def Mary_moved_to_the_bathroom()
    self.Mary.location="bathroom"
def Mary_got_the_football_there():
...
...
```

Abstract Functions

```
def go(self, character, location):
    character.location = location
for item in character.inventory:
    item.location = location
def pick_up(): ...
def story(self):
```

```
## Mary moved to the bathroom.
self.go(character=self.Mary,
location = "bathroom")
```

Tested On 2 Tasks

bAbI (Weston et al. 2015)

- Task 2: Stories tracking objects that characters carry

bAbl (Weston et al. 2015)

Method	# Shot	Accuracy ↑
Random	-	25%
GPT-3	1	56.5%
Chain of Thought (Creswell et al. 2022)	1	46.4%
Selection-Inference (Creswell et al. 2022)	1	29.3%
Dual-System (Nye et al. 2021)	10	100%
CoRRPUS (comment)	1	67.0%
CoRRPUS (specific)	1	78.7%
CoRRPUS (abstract)	1	99.1%

Tested On 2 Tasks

bAbI (Weston et al. 2015)

- Task 2: Stories tracking objects that characters carry

Re³ (Yang et al. 2022)

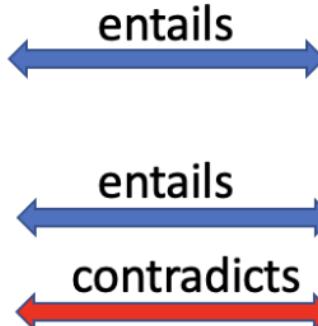
- Identifying inconsistencies in stories (e.g., descriptions of characters' appearances, relationships)
- Stories were generated from a list of facts (the premise). They also generated premises with a contradiction.

Re³

The task is to see what stories match what premises based on the facts extracted from both.

Joan Westfall premise

Attribute	Value
Gender	Female
Occupation	Teacher
Brother	Brent Westfall
Appearance	Blue eyes



Joan Westfall in story

Attribute	Value
Gender	Female
Father	Jason Westfall
Brother	Brent Westfall
Appearance	Brown eyes

Re³ (Yang et al. 2022)

Method	ROC-AUC ↑
Random	0.5
GPT-3	0.52
Entailment (Yang et al. 2022)	0.528
Entailment with Dense Passage Retrieval (Yang et al. 2022)	0.610
Attribute Dictionary → Sentence (Yang et al. 2022)	0.684
CoRRPUS (comment)	0.751
CoRRPUS (specific)	0.794
CoRRPUS (abstract)	0.704

→ Probably because functions like `set_age(self, character, age)` complicate more than they help.

Takeaways from CoRRPUS

Multi-step, structured (via code) chain-of-thought prompting outperforms regular chain-of-thought prompting.

Prompting systems like CoRRPUS can function at the same level as highly engineered systems.



<https://media.tenor.co/images/4377f4b69226f645d7ccd583b27d76f3/raw>

How can NLP improve people's lives?

Outline

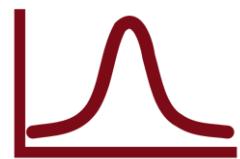
PhD

Automated Story Generation



Events

AAAI 2018

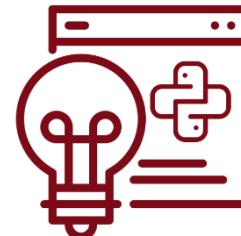


Plot Progression

IJCAI 2019

Postdoc

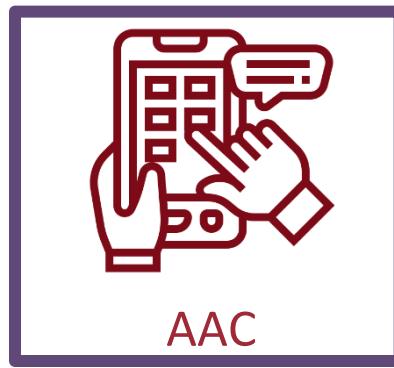
Story
Understanding



Code-LLMs

ACL 2023

Human
Communication



AAC

Submitted to
ASSETS 2025

Present



Trans* TTS



AI D&D

Augmentative & Alternative Communication



<https://www.abc.es/media/ciencia/2018/03/15/20576450-kgXH--620x349@abc.jpg>



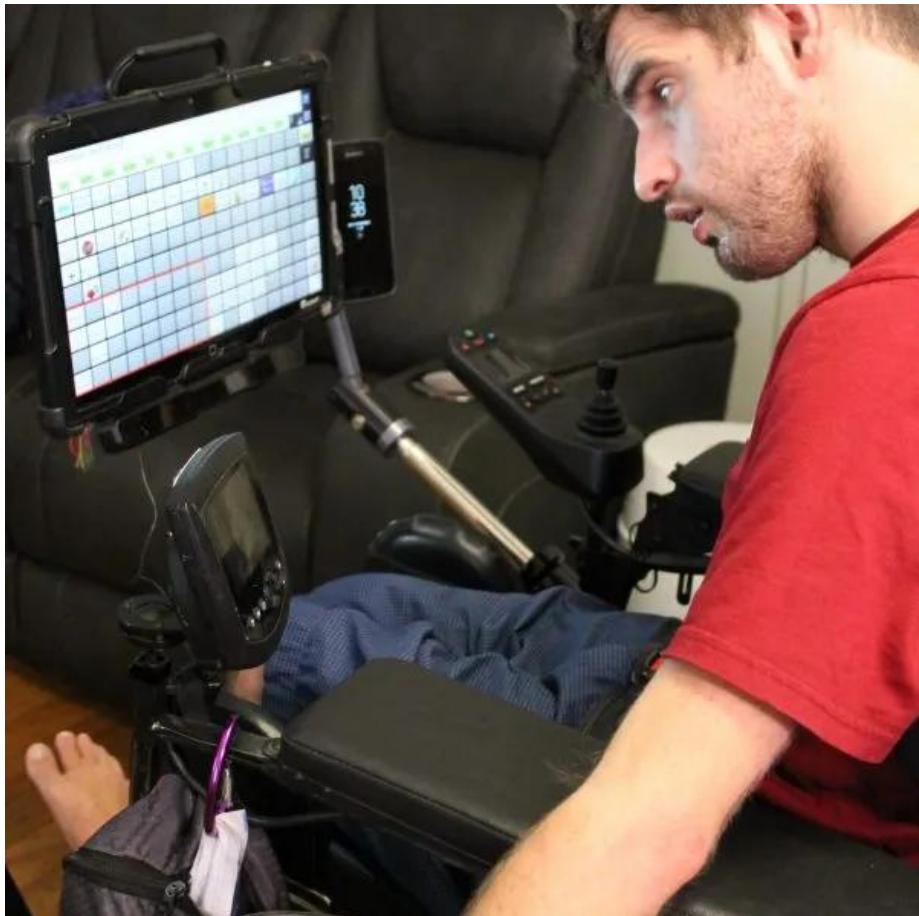
STORYTELLING, D&D, AND AAC
<https://dnewsc4.com/lifestyle/2021/05/30/the-talking-dog-of-tiktok>

A	B	C	D	SPACE	END OF MESSAGE						
E	F	G	H	START OVER	I DON'T KNOW						
I	J	K	L	M	N						
O	P	Q <u>u</u>	R	S	T						
U	V	W	X	Y	Z						
1	2	3	4	5	6	7	8	9	Ø	YES	NO

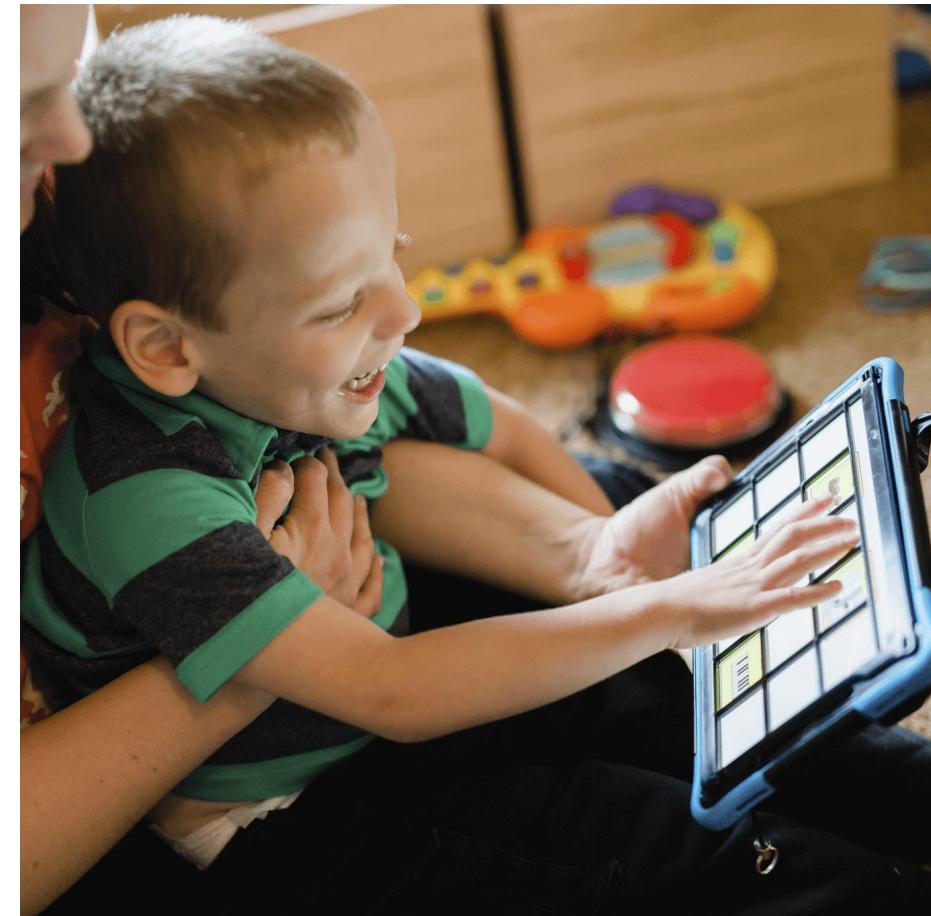
Letter Board - AEIOU format

unl.edu/documents/secd/forms/Letter-Boards.png

~~Who uses AAC?~~ Who is studied using AAC?



<https://lifetec.org.au/dales-story/>



<https://premierpedstherapy.com/5-myths-about-aac-use-for-children-with-autism-spectrum-disorder/>

But what about autistic adults?

(See also: children grow up)

Let's ask them

I ran semi-structured interviews with 12 autistic adults recruited online

Participants were encouraged to use AAC during interviews



I've used it with rideshare drivers before and



they



I, me, my 	is, am, are 	go 	to 	not 	what 	hi 	People 	and
you 	do 	want 	in 	more 	how 	please 	Time 	because
it 	have 	like 	on 	good 	when 	wait 	Food 	but
this 	can 	get 	for 	now 	where 	let's 	Care 	so
that 	see 	help 	up 	different 	who 	yes 	Plans 	if
we 	will 	stop 	with 	finished 	why 	no 	Repair 	than
a 	the 							as



I, me, my 	is, am, are — 	go 	to 	not 	what ?	hi 	People 	and
you 	do 	want 	in 	more 	how ?	please 	Time 	because
it 	have 	like 	on 	good 	when ?	wait 	Food 	so
this 	can 	get 	for 	now 	where ?	let's 	Care 	than
that 	see 	help 	up 	different 	who ?	yes 	Plans 	as
we 	will 	stop 	with 	finished 	why ?	no 	Repair 	else
a 								

Asked people questions about...

Usage, app preferences, empowerment, and privacy

For instance:

Tell me about a situation where you wanted to use AAC but **couldn't** because of the type of social situation.

Or

Tell me about a time AAC helped you feel empowered as an autistic person.

Or

What type of information about you do you think your AAC application(s) saves?

Participants

- 12 autistic adults
- Native English speakers, living in the US
- Use AAC applications
 - Proloquo2Go (6)
 - Proloquo4Text (5)
 - CoughDrop (3)
 - TD Snap (3)
- Encouraged to use AAC during the interview

Age Range	Self-Assigned Label	Years of AAC Use
18-24	Mostly non-speaking	5-9
25-34	Unreliable speech	Since childhood
25-34	Semi-speaking	5-9
25-34	Mostly non-speaking	≤1
25-34	Selectively mute/situationaly mute	10+
25-34	Semi-speaking	10+
25-34	Semi-speaking	≤1
35-44	Mostly non-speaking with unreliable speech and word retrieval issues	10+
35-44	Sometimes user	5-9
35-44	Mostly non-speaking	≤1
35-44	Mostly non-speaking	5-9
35-44	Unreliable speech	≤1

Research Questions

- **RQ1:** What are pitfalls and challenges that autistic AAC users run into when using existing AAC applications?
- **RQ2:** What types of features do autistic adults require or prefer to see from AAC applications?
- **RQ3:** How does the person's identity as an autistic person factor into their AAC use?
- **RQ4:** How do participants feel about NLP techniques being integrated into their AAC apps?

Interview Questions

1. Questions asking the participant to recall a recent conversation where they used AAC and describe details about how that conversation went.
2. Questions asking the user to think of particular scenarios when they used AAC.
3. Questions for describing a typical day of using AAC and what types of tools they used.
4. Questions asking about when they have felt empowered or disempowered because of their AAC.
5. Questions about how they select what AAC to try.
6. Questions about privacy

Thematic Analysis Categories

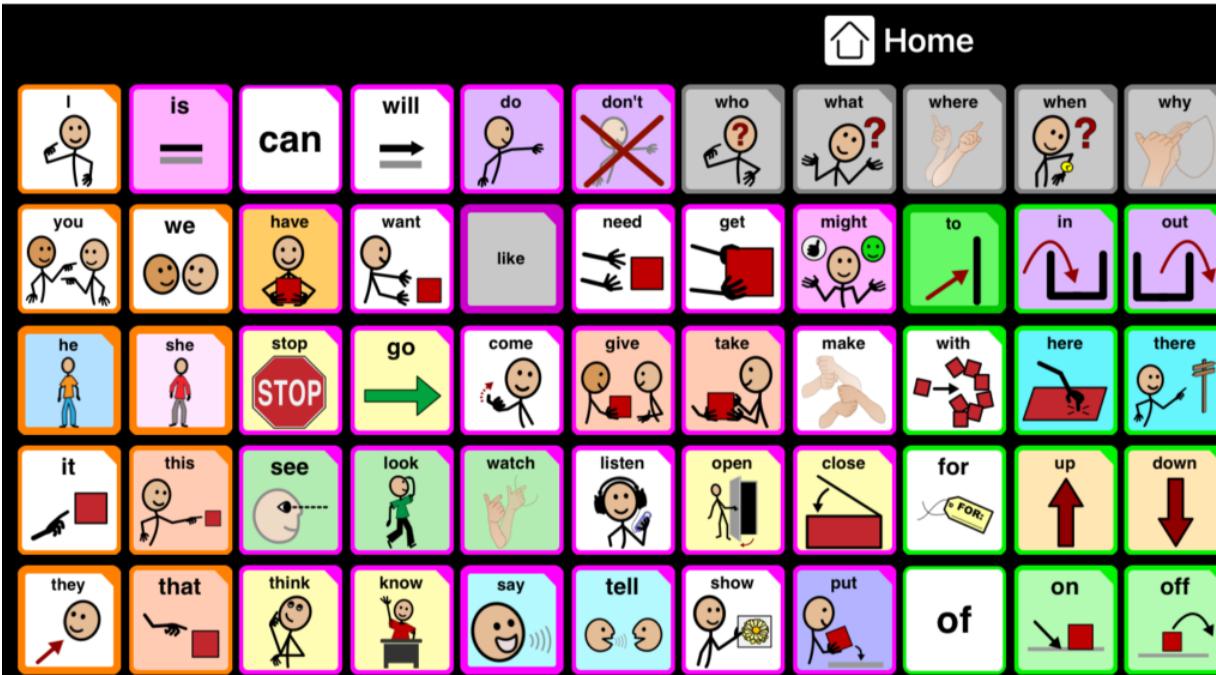
1. Input flexibility
2. Output flexibility
3. Selecting or adapting AAC
4. When to start, swap, or stop AAC
5. Benefits
6. Access as an Adult
7. Stumbling blocks for continued use
8. Control of communication

1. Input Flexibility

Symbol-based

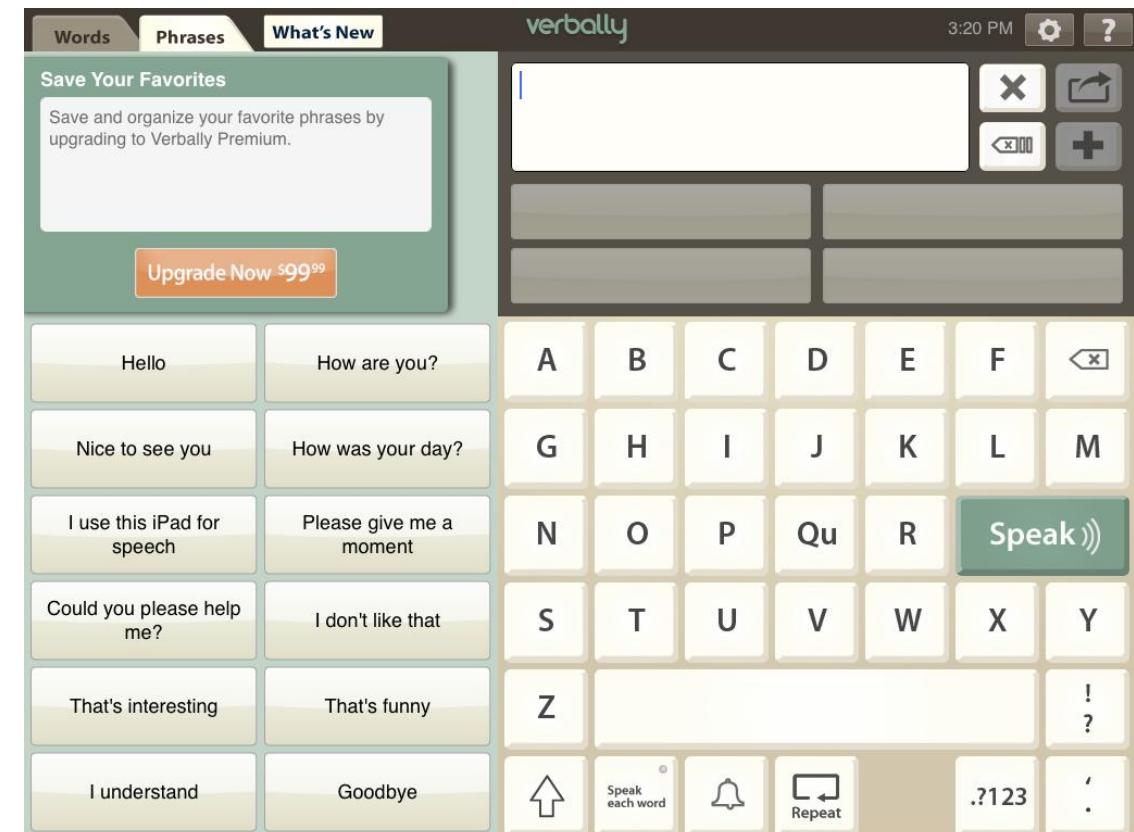
11:37 AM Tue Nov 16

it's all ableism, and it all needs to stop.



<https://communicationfirst.org/guest-blog-communication-barriers-faced-by-aac-users-in-accessing-health-care/>

Text-based



<https://atpdc.wordpress.com/2015/01/30/verbally-text-based-aac-app/>

1. Input Flexibility

12 out of 12 participants

- Typing as an option (12)
- Mix symbol- and text-based AAC (7)
- Flexible vocabulary (7)
- Saving pre-programmed phrases (10)

Quotes from participants



"I can type on a phone, but I type 100 words a minute on a regular keyboard so that's always gonna be my preference"

"[I dislike] symbols and typing being so completely separate in different apps."

"I also like that with the phone I can pre-program phrases before I enter a situation."

2. Output Flexibility

- Showing text instead of using TTS (8)

3. Selecting or adapting AAC

- Aging up AAC (5)
- Dedicated device (5)
- General customization (12)

“Many AAC apps feel like they’re made for kids or students, and it feels infantilizing.”

“I like being able to adjust pretty much every feature really. “

4. When to start, swap, or stop AAC

-
- Lose oral speech because of stressors (9)
 - Trust (11)
 - Adjusting to environment (9)

"I don't use AAC [...] when there are people who have dramatic power over me who can drastically control my life."

5. Benefits

- Organizing thoughts (8)
- Identity (9)
- Don't need to force speech anymore (7)

“Sometimes I talk faster than I think and struggle to explain a complex idea, and AAC lets me collect my thoughts and write a clear message.”

6. Access as an Adult

- Affordability (10)

"I have no other way and because [redacted] helped me afford it."

7. Stumbling blocks for continued use

- Speed (11)
- Voice quality (10)
- Need for reliability (6)

"I've had times where I tried to pull out the AAC in a group conversation [...] and then people hear it and laugh because they think I'm making a joke".

*" I reported several bugs and *never* got any fixed."*

8. Control of communication

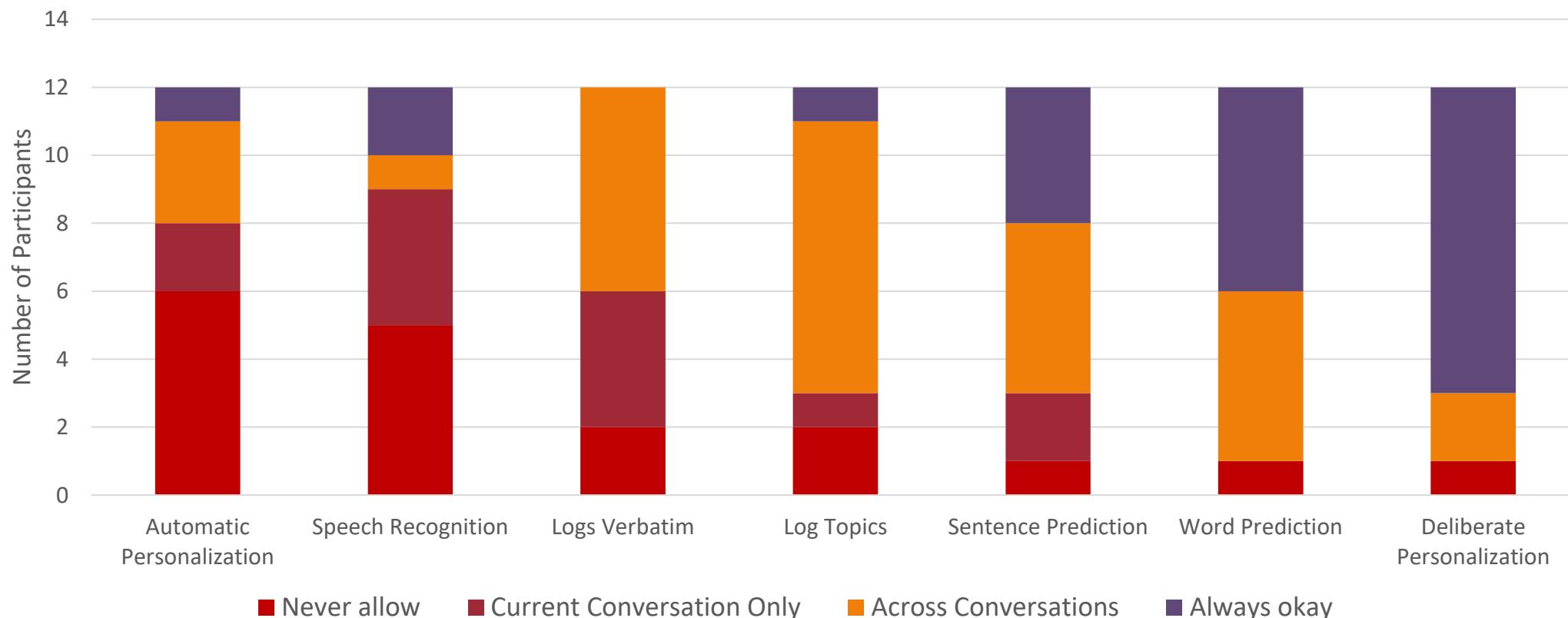
- No automatic personalization (6)
- Unwanted outside access (11)
- App's unknown data policy (9)

"I don't like it when the way I use something automatically changes the way an app or website, or whatever, functions."

"I do have to think a little about data security specifically for my job because I'm bound by HIPAA."

Autistic adults don't want automated logging of conversations

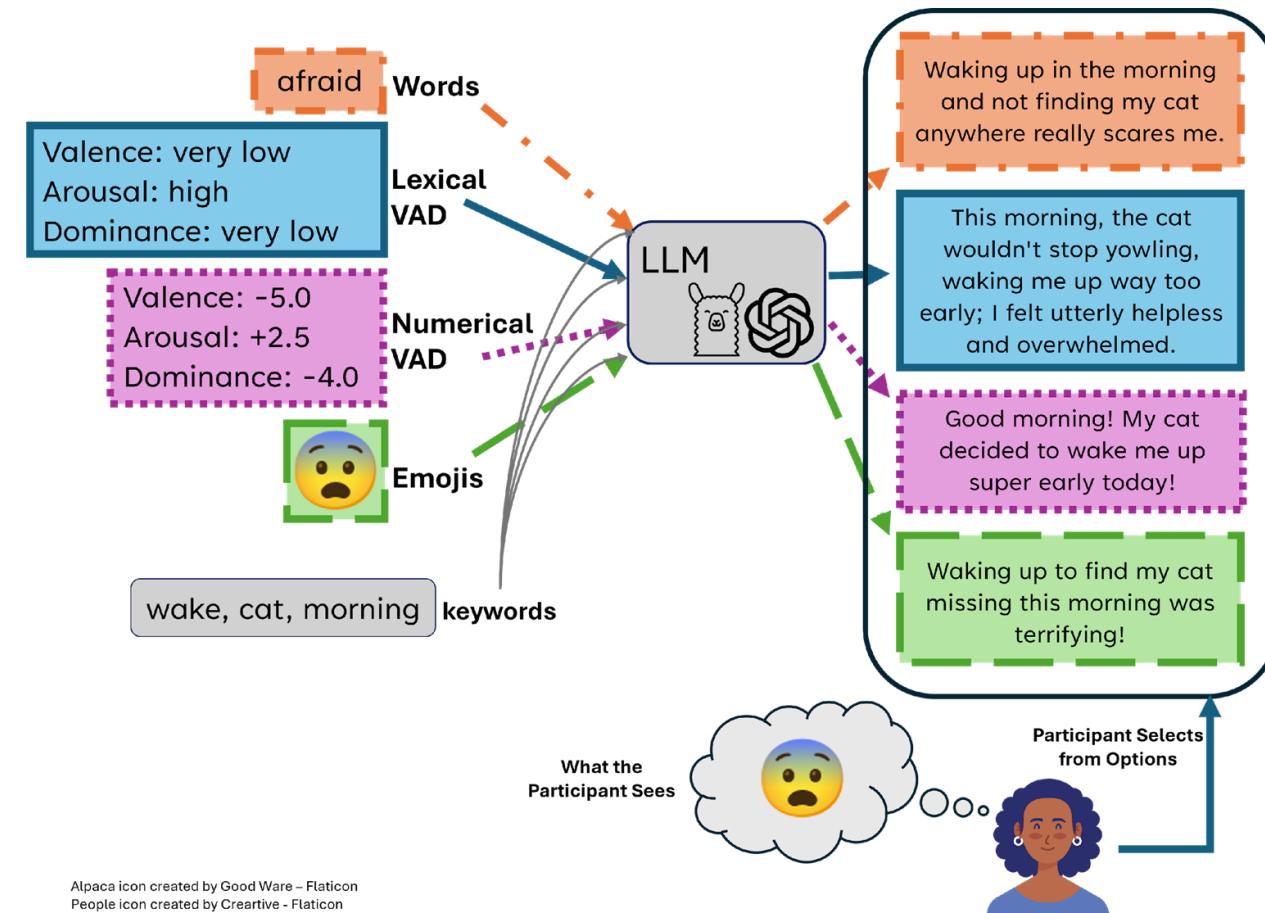
For each row of the table, read the feature and think about how you would feel using AAC with that feature—think specifically in terms of your privacy.



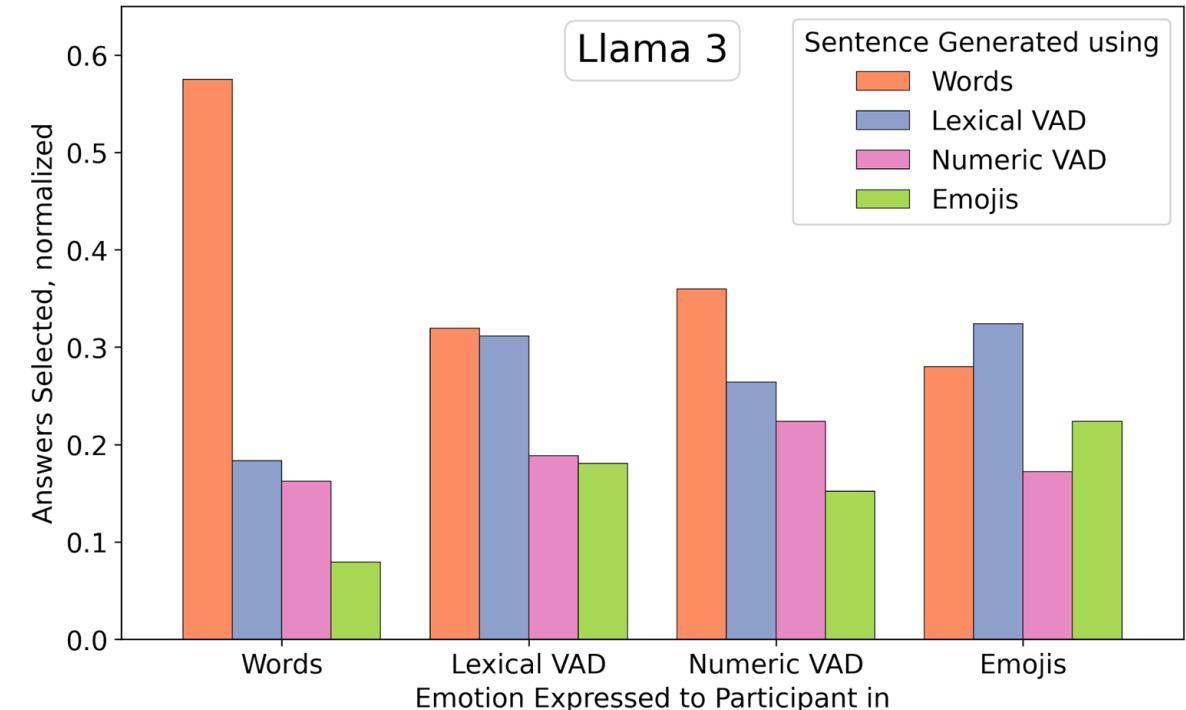
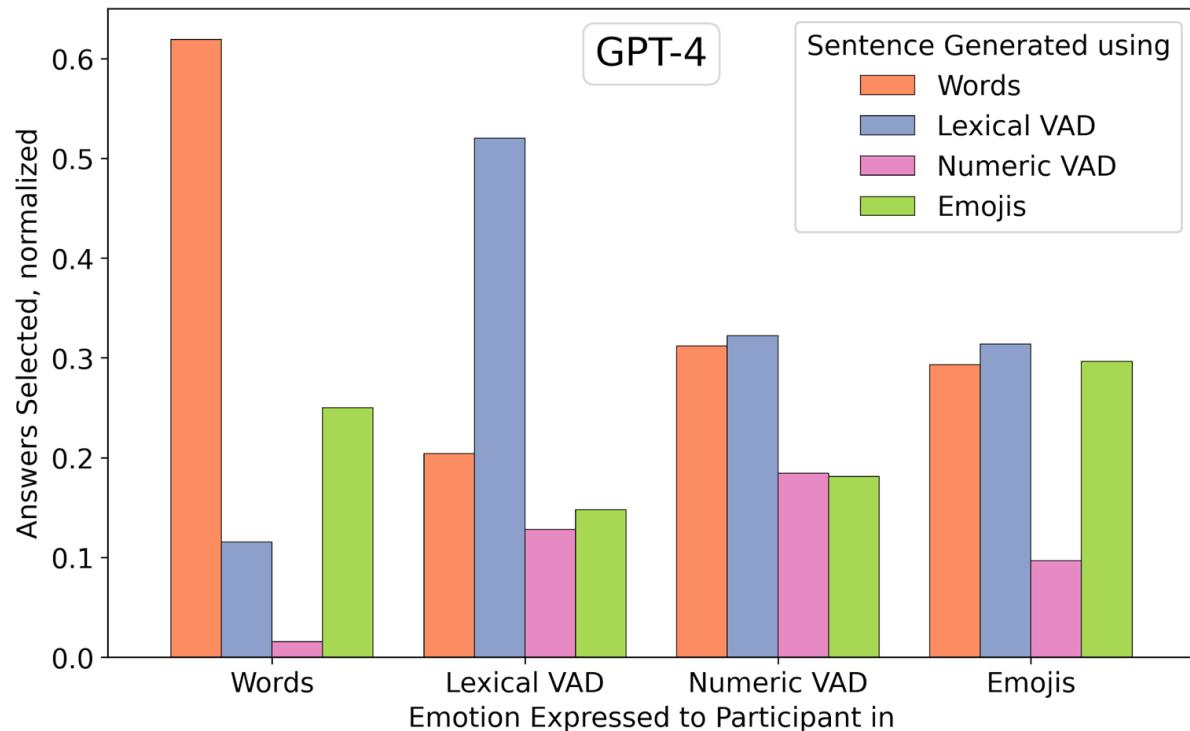
Autistic Adult-specific AAC Directions

- Keyboard input
- Conversation planning
- Mixing symbols & text modalities
- On-the-fly switching
- Trans* Text-to-Speech

LLM Emotion Alignment



Some representations align better



Outline

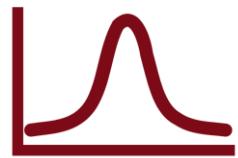
PhD

Automated Story Generation



Events

AAAI 2018

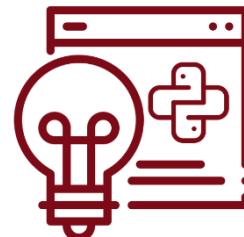


Plot Progression

IJCAI 2019

Postdoc

Story
Understanding



Code-LLMs

ACL 2023

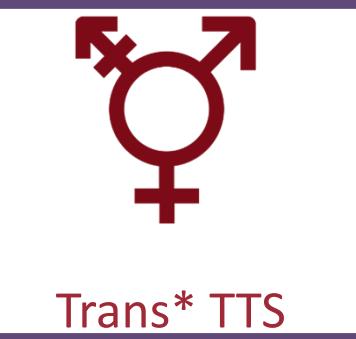
Human
Communication



AAC

Submitted to
ASSETS 2025

Present



Trans* TTS



AI D&D

Why Trans TTS?

Trans people are 3.03-6.36 times more likely to be autistic or have autistic traits (Warrier et al. 2020)

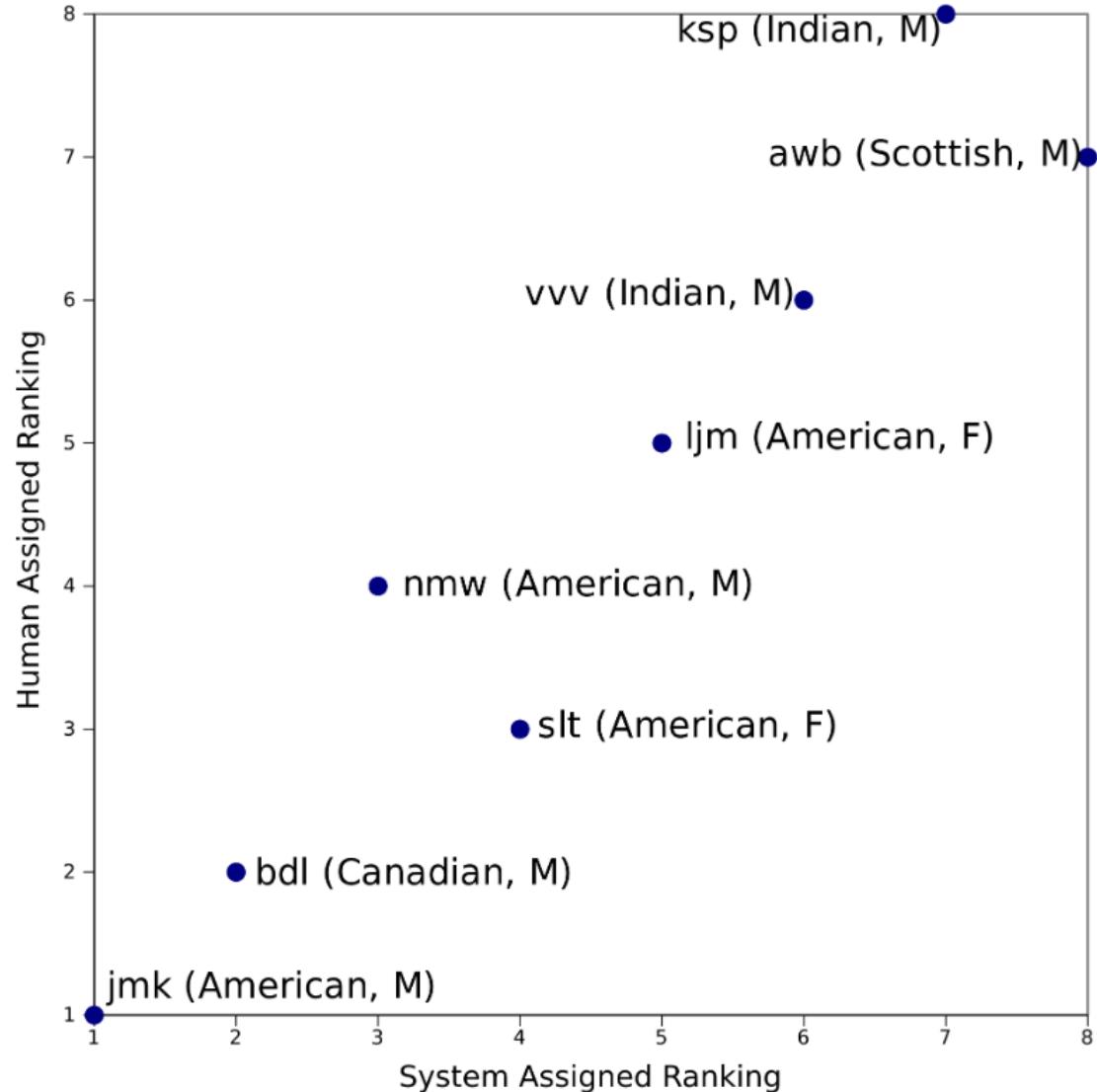
Because of current AAC's focus on children, their children's voices end up being the most expressive TTS

People want to sound like *themselves* (even if that doesn't sound like the voice they can produce)

Related Work

We trained ML model to match speech features (like MFCCs) with human-annotated scores for accent

- What if trans people scored voices based on qualities they want?
- Can this be used to create synthesized voices with these characteristics?



N. Wolfe, V. V. Vemuri, **L. J. Martin**, F. Metze, and A. W Black,
“Applause: A Language Learning Tool for Low-Resource Languages,” in *Designing Speech and Language Interactions Workshop at CHI 2014*, Toronto, Canada, Apr. 2014.

Outline

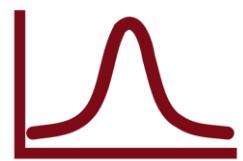
PhD

Automated Story Generation



Events

AAAI 2018

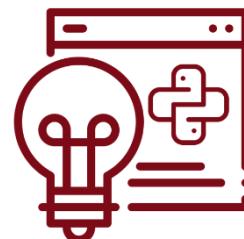


Plot Progression

IJCAI 2019

Postdoc

Story
Understanding



Code-LLMs

ACL 2023

Human
Communication



AAC

Submitted to
ASSETS 2025

Present



Trans* TTS



AI D&D

What is D&D?

It's...

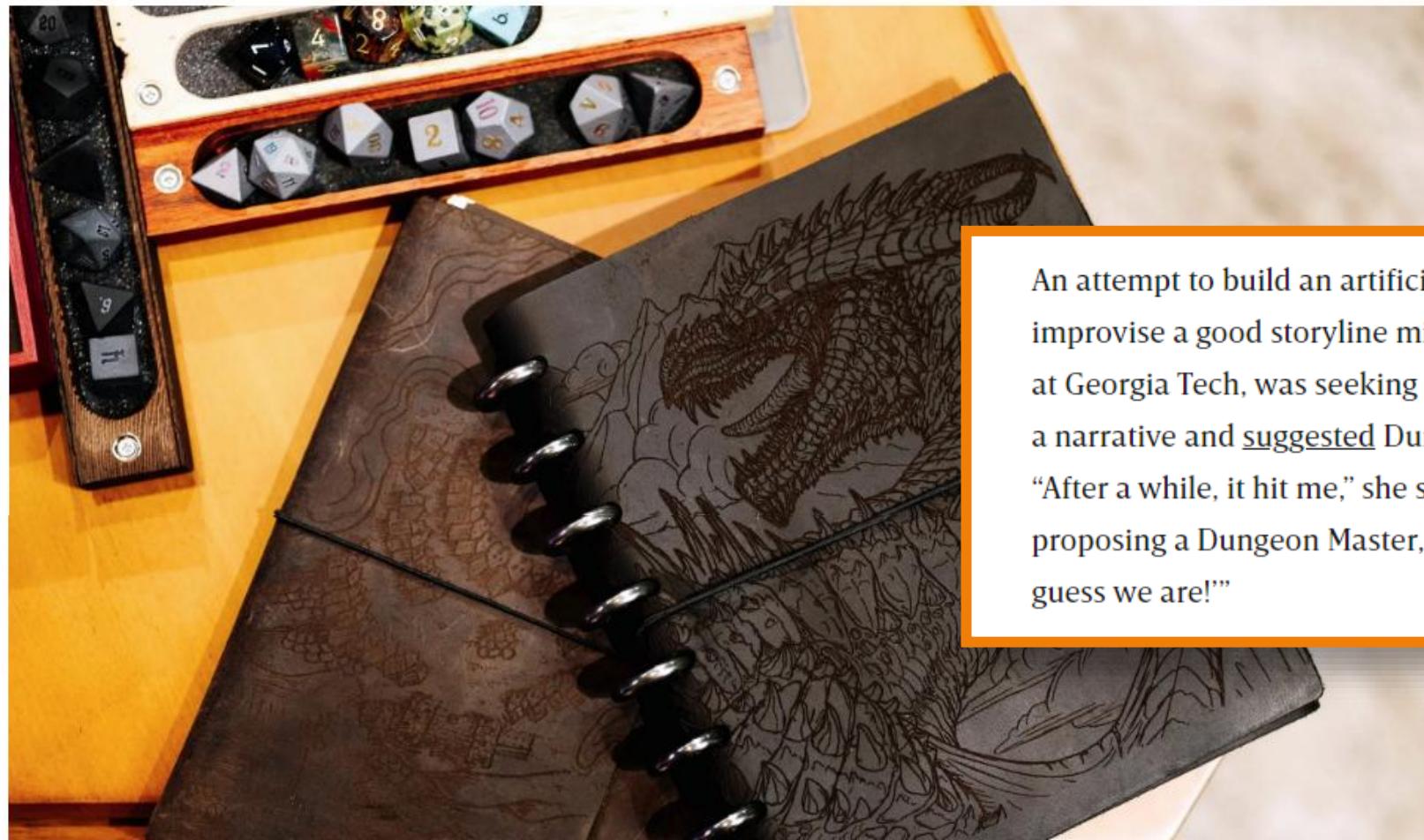
1. ~~play acting~~ & tabletop acting through game dialog with DM & players
2. ~~co~~laborating to tell a story



Screencap from Netflix's *Stranger Things* via <https://www.thetimes.co.uk/article/dungeons-and-dragons-fans-enter-the-land-of-popularity-j0q53lkcw>

Forget Chess—the Real Challenge Is Teaching AI to Play D&D

Some artificial intelligence experts think role playing adventure games will help machines learn to be as clever as we are.



An attempt to build an artificial Dungeon Master offers hope that machines able to improvise a good storyline might be built. In 2018, [Lara Martin](#), a graduate student at Georgia Tech, was seeking a way for AI and a human to work together to develop a narrative and suggested Dungeons & Dragons as a vehicle for the challenge. “After a while, it hit me,” she says. “I go up to my adviser and say ‘We’re basically proposing a Dungeon Master, aren’t we?’ He paused for a bit, and said ‘Yeah, I guess we are!’”

Hard problems in AI and



1. Infinite Number of Sentences
2. Unexpected Consequences
3. Dynamic, Continuous World
4. Distributed World (Theory of Mind)
5. Intrinsic Reward



Not solved but I've been working on...

A. Zhu, K. Aggarwal, A. Feng, L. J. Martin, and C. Callison-Burch, “FIREBALL: A Dataset of Dungeons and Dragons Actual-Play with Structured Game State Information,” *Annual Meeting of the Association for Computational Linguistics (ACL)*, Toronto, Canada, Jul. 2023, pp. 4171-4193.

C. Callison-Burch, G. Singh Tomar, L. J. Martin, D. Ippolito, S. Bailis, and D. Reitter, “Dungeons and Dragons as a Dialogue Challenge for Artificial Intelligence,” in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Abu Dhabi, United Arab Emirates, Dec. 2022, pp. 9379–9393.

A. Zhu, L. J. Martin, A. Head, C. Callison-Burch. “CALYPSO: LLMs as Dungeon Masters’ Assistants,” in *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, Salt Lake City, UT, Oct. 2023, pp. 380-390.

The screenshot shows a sequence of rolls and actions from the Avrae bot. It includes:

- Ara Fentas makes an Arcana check! 1d20 (8) + 7 + 1d12 (7) [Bardic Inspiration] = 22
- Ara Fentas makes an Intelligence Save! 1d20 (17) + 1 = 18 Ack!
- Ara Fentas attacks with a Dagger! Meta To Hit: 1d20 (18) + 9 = 27 Damage: 1d4 (1) + 3 [piercing] = 4 Effect Finesse, light, thrown (range 20/60)



State Tracking for D&D



Character-Specific Dialog

Neurosymbolic Story Generation



Events
AAAI 2018



Improvisational Storytelling
ICIDS 2016



Expanding Events into Sentences
AAAI 2020



Plot Progression
IJCAI 2019



Dungeons & Dragons



State Tracking for D&D
ACL 2023

Character-Specific Dialog
EMNLP 2022

Story Understanding



Code-LLMs
ACL Findings 2023



Separating Generation from Understanding
ACL CSRR Workshop 2022



Narrative Characteristics of an "Asshole"
ICWSM 2023

Human Communication



Submitted to ASSETS 2025
Speech-to-Speech Translation
ASRU 2015



Matching Crowdsourced Data
to Uncertainty in Speech
SLT 2014

+ various workshop papers

Announcement

I'm teaching a Special Topics (491/691) course on Interactive Fiction and Text Generation in Fall 2025

I have also gotten permission to make it a more permanent course number

- Probably under the name Neurosymbolic Text Generation

Picture References

Lego Star Wars screenshots: <https://www.youtube.com/watch?v=sWNrrgz8-nQ>

Alexa Picture: https://pisces.bbystatic.com/image2/BestBuy_US/images/products/6430/6430060cv13d.jpg

Dog gif: <https://media.giphy.com/media/WiFHem7FxKbHG/giphy.gif>

Icons:

Trans: <https://thenounproject.com/icon/trans-3113633/>

Gears: https://www.flaticon.com/free-icon/engineering_3079165

AAC: https://www.flaticon.com/free-icon/mobile-phone_2974039

Events: <https://thenounproject.com/icon/event-1430222/>

Reddit: https://www.flaticon.com/free-icon/reddit-logo_52053

Bard: https://www.flaticon.com/free-icon/bard_2858872

Simplify: <https://thenounproject.com/icon/ungroup-1091019/>

20-sided Die: <https://stock.adobe.com/images/black-and-white-tattoo-linework-style-natural-20-d20-dice-roll/548982460>

Robot head: <https://thenounproject.com/icon/robot-head-584365/>

Story Understanding Icon: Python: <https://www.flaticon.com/free-icons/python> + Idea: https://www.flaticon.com/free-icon/idea_7857740

Pause button in video: https://www.flaticon.com/free-icon/pause_2920686

Tug of war: <https://thenounproject.com/icon/tug-of-war-1016981/>

Cheese: https://www.flaticon.com/free-icon/cheese_9224691

Runner: <https://www.flaticon.com/free-icons/runner>