

Story generation

CS 685, Fall 2021

Advanced Natural Language Processing

Mohit Iyer

College of Information and Computer Sciences

University of Massachusetts Amherst

story generation

produce a narrative that follows from
a given input prompt or context

story generation

produce a narrative that follows from
a given input prompt or context

Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

story: It was the color of a sunset: red tipped petals faded to yellow. Stark contrast to the roiling forever-cloud above it. A thick green stem climbed out of a crack in the dry earth, thorns lining it like a ladder. It had been hidden between two dry logs, lying flat like hollowed bones. It must have peered out from between the dead trees only in the last few days, and I had been the first to stumble upon it.

It was worth more than my life.

Below me lay my village of rusted huts. They littered the valley sporadically, as if a train had been derailed and its carriages rolled free, far and wide. Men and women worked around the shacks, digging and raking the land, trying to grow weeds that could be eaten, as kids kicked tins into makeshift goals around them.

The rose was worth more than any of their lives too. All of them together, even. What a biodome would pay for it...

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

story: It was the color of a sunset: red tipped petals faded to yellow. Stark contrast to the roiling forever-cloud above it. A thick green stem climbed out of a crack in the dry earth, thorns lining it like **describe the setting** den between two dry logs, lying flat like hollowed bones. It must have peered out from between the dead trees only in the last few days, and I had been the first to stumble upon it.

It was worth more than my life.

Below me lay my village of rusted huts. They littered the valley sporadically, as if a train had been derailed and its carriages rolled free, far and wide. Men and women worked around the shacks, digging and raking the land, trying to grow weeds that could be eaten, as kids kicked tins into makeshift goals around them.

The rose was worth more than any of their lives too. All of them together, even. What a biodome would pay for it...

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

story: It was the color of a sunset: red tipped petals faded to yellow. Stark contrast to the roiling forever-cloud above it. A thick green stem climbed out of a crack in the dry earth, thorns lining it like **describe the setting** den between two dry logs, lying flat like hollowed bones. It must have peered out from between the dead trees only in the last few days, and I had been the first to stumble upon it. **introduce main character** It was worth more than my

Below me lay my village of rusted huts. They littered the valley sporadically, as if a train had been derailed and its carriages rolled free, far and wide. Men and women worked around the shacks, digging and raking the land, trying to grow weeds that could be eaten, as kids kicked tins into makeshift goals around them.

The rose was worth more than any of their lives too. All of them together, even. What a biodome would pay for it...

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

story: It was the color of a sunset: red tipped petals faded to yellow. Stark contrast to the roiling forever-cloud above it. A thick green stem climbed out of a crack in the dry earth, thorns lining it like **describe the setting** den between two dry logs, lying flat like hollowed bones. It must have peered out from between the dead trees only in the last few days, and I had been the first to stumble upon it. **introduce main character** It was worth more than my

Below me lay my village of rusted huts. They littered the valley sporadically, as if a train had been derailed and its carriages rolled free, far and **more setting** women worked around the shacks, digging and raking the land, trying to grow weeds that could be eaten, as kids kicked tins into makeshift goals around them.

The rose was worth more than any of their lives too. All of them together, even. What a biodome would pay for it...

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

story: It was the color of a sunset: red tipped petals faded to yellow. Stark contrast to the roiling forever-cloud above it. A thick green stem climbed out of a crack in the dry earth, thorns lining it like **describe the setting** den between two dry logs, lying flat like hollowed bones. It must have peered out from between the dead trees only in the last few days, and I had been the first to stumble upon it. **introduce main character** It was worth more than my

Below me lay my village of rusted huts. They littered the valley sporadically, as if a train had been derailed and its carriages rolled free, far and **more setting** women worked around the shacks, digging and raking the land, trying to grow weeds that could be eaten, as kids kicked tins into makeshift goals around them.

introduce new concept any of their lives too. All of them together, even. What a **biodome** would pay for it...

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

GPT-2 large generates:

We have to be prepared and take care of the plants.

We can provide your flowers in all kinds of colors. The most important thing to remember when choosing flowers is the color of the petals.

Please use these links for more information.

<http://www.floralkids.com/flowers/natural-flowers/cotton/>

<http://www.floralkids.com/flowers/natural-flowers/flowers-for-lilies/>

<http://www.floralkids.com/flowers/natural-flowers/rose/>

<http://www.floralkids.com/flowers/natural-flowers/sweet-scented-flowers/>

prompt: Flowers have become so rare that they are the most sought after items in the world, sold at high prices in black markets, under guard in national museums etc. You just stumbled across a natural rose.

GPT-2 large generates:

We have to be prepared and take care of the plants.

We can provide your flowers in all kinds of colors. The most important thing to remember when choosing flowers is the color of the petals.

... complete garbage

Please use these links for more information.

<http://www.floralkids.com/flowers/natural-flowers/cotton/>

<http://www.floralkids.com/flowers/natural-flowers/flowers-for-lilies/>

<http://www.floralkids.com/flowers/natural-flowers/rose/>

<http://www.floralkids.com/flowers/natural-flowers/sweet-scented-flowers/>

this task is extremely hard!

- many different plausible stories per prompt, which makes training / evaluation difficult
- do we need to model the *planning* process?
- stories can be really long
- stories have all sorts of complex linguistic phenomena

nevertheless, there's lots of recent
work on story generation!

- Fan et al., ACL 2018: *Hierarchical Neural Story Generation*
- Yao et al., AAAI 2019: *Plan-And-Write: Towards Better Automatic Storytelling*
- Tambwekar et al., IJCAI 2019: *Controllable Neural Story Plot Generation via Reward Shaping*

and many more!

Title (Given)	The Bike Accident
Storyline (Extracted)	Carrie → bike → sneak → nervous → leg
Story (Human Written)	<u>Carrie</u> had just learned how to ride a bike. She didn't have a <u>bike</u> of her own. Carrie would <u>sneak</u> rides on her sister's bike. She got <u>nervous</u> on a hill and crashed into a wall. The bike frame bent and Carrie got a deep gash on her <u>leg</u> .

You're a Werewolf. You begin to transform, but instead of a terrifying beast, you turn into a small puppy.

<V> opened <A0> ent0 <A1> ent0 eyes
<V> looking <A0> to ent0 ent1
<V> found <A0> ent0
<V> clipped <A1> ent1
<V> flopped <A0> ent0 ears
<V> was <A0> the hunger <A1> gone
<V> clouded <A0> Confusion <A1> ent1 mind
<V> tilted <A0> ent0 <A1> ent2
<V> approached <A1> ent0 <A2> a nearby puddle
<V> looked <A0> ent0

ent0 opened ent0 eyes. Looking to ent0 ent1, ent0 found that ent1 were now neatly clipped. ent0 ears flopped on either side of ent2 lazily, too soft and formless to hunt properly. Most of all, the hunger was gone. Confusion clouded ent0 mind and ent0 tilted ent2 instinctively. ent0 approached a nearby puddle and looked in.

I opened my eyes. Looking to my razor-sharp claws, I found that they were now neatly clipped. My ears flopped on either side of my head lazily, too soft and formless to hunt properly. Most of all, the hunger was gone. Confusion clouded my mind and I tilted my head instinctively. I approached a nearby puddle and looked in.

Story Prompt

Action Plan with
Semantic Role
Labeling

Entity
Anonymized
Story

Full Story

no standard datasets

- Fan et al., ACL 2018: *Hierarchical Neural Story Generation*: **reddit r/writingprompts**
- Yao et al., AAAI 2019: *Plan-And-Write: Towards Better Automatic Storytelling*: **ROCStories**
- Tambwekar et al., IJCAI 2019: *Controllable Neural Story Plot Generation via Reward Shaping*: **CMU Movie Summaries**

automatic evaluations?

- See et al., CoNLL 2019: *Do Massively Pretrained Language Models Make Better Storytellers?*
 - how related is prompt to story?
 - how coherent is the story?
 - how much does the model repeat itself?
 - how syntactically complex are the stories?
 - how surprising are the stories?

automatic evaluations?

- See et al., CoNLL 2019: *Do Massively Pretrained Language Models Make Better Storytellers?*
 - how related is prompt to story?
 - how coherent is the story?
 - how much does the model repeat itself?
 - how syntactically complex are the stories?
 - how surprising are the stories?

“human evaluation... currently the only reliable way to assess overall text quality”

how do we set up human evaluations for story quality?

- human eval of Fan et al: given a shuffled collection of three prompts and three generated stories, match them up
- human eval of Yao et al: given a prompt and two stories generated by two different models, pick which story is better

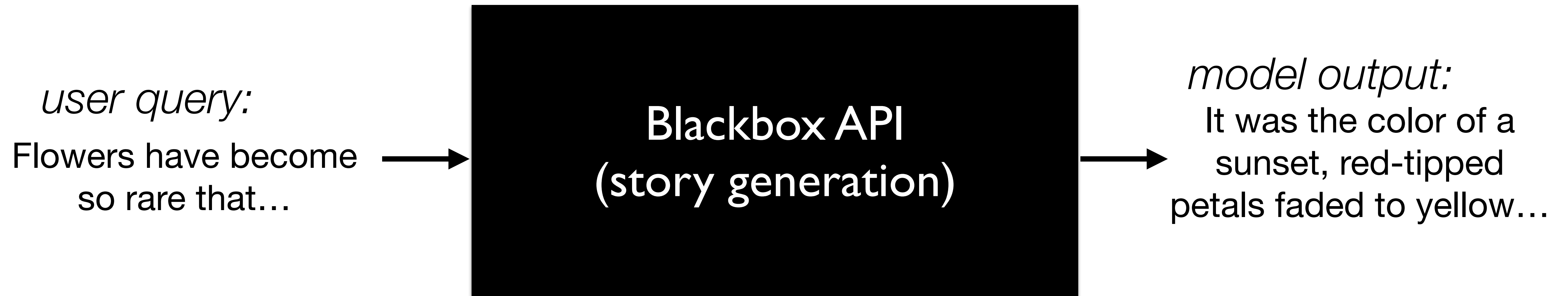
machine-in-the-loop storytelling

- How do we get evaluators to read and engage with long, complex stories?
- What if the people evaluating stories were also *authors* of the stories?
- **machine-in-the-loop** storytelling: a model provides *suggestions* (e.g., of scenes) to real authors, who can then use the generated text directly or as inspiration in their final story

Roemelle et al., ICIDS 2015, Clark et al., IUI 2018

machine-in-the-loop storytelling

- A text generation model provides *suggestions* (e.g., of scenes) to real authors, who can then use them as inspiration for their final story



STORIUM: a dataset and evaluation platform for story generation

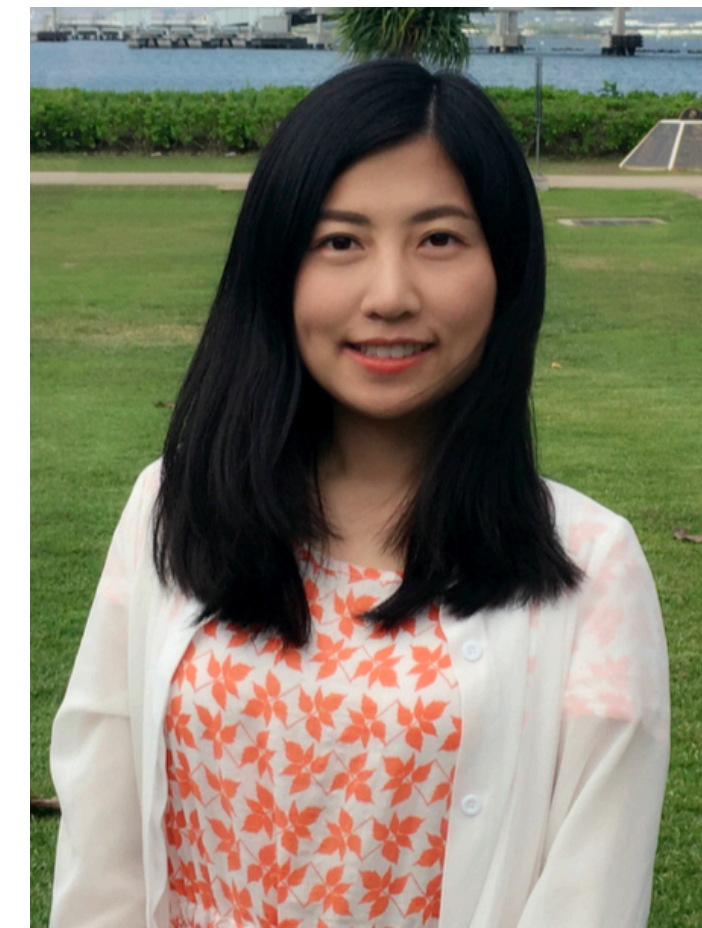
www.storium.com



Nader Akoury

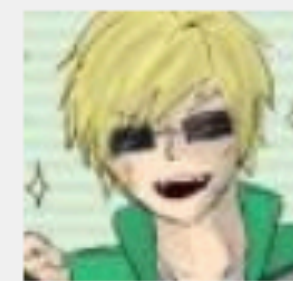


Shufan Wang

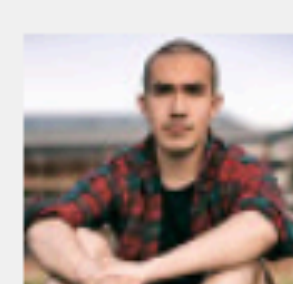
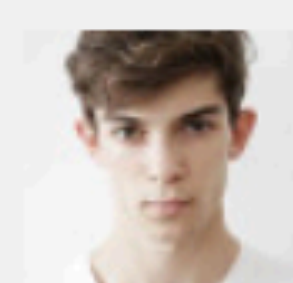
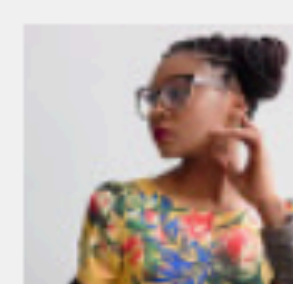
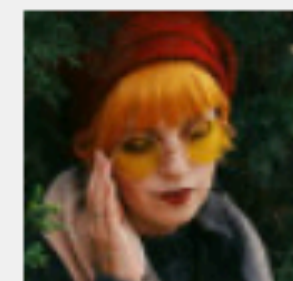


Nanyun Peng

NARRATOR



CAST



Chapter 1, Scene 1 · 05/06/2018



CHALLENGES

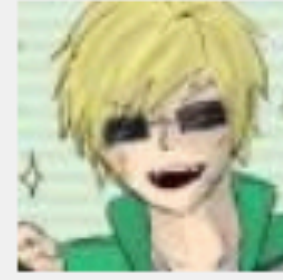
 STRENGTH?



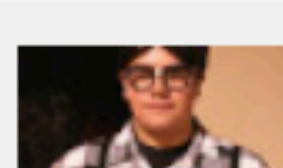
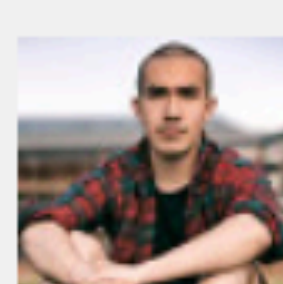
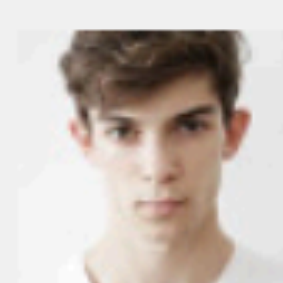
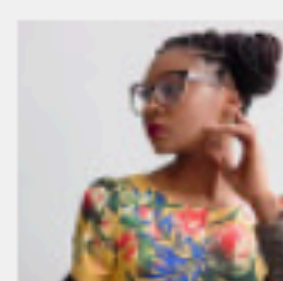
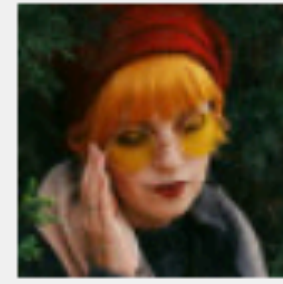
The storefront was empty of all decoration, but everyone around town had heard what this place was going to be: Inspectres.

The infamous company known for hunting down the supernatural around the continental United States (plus a few franchises in Canada and one right across the Mexican border) had staked a franchise claim in this city and was having open interviews for anyone who wanted to

NARRATOR



CAST



INSPECTRES: GHOST HUNTING FOR

stories are written by a group of users, each of whom write from the perspective of a different character



CHALLENGES

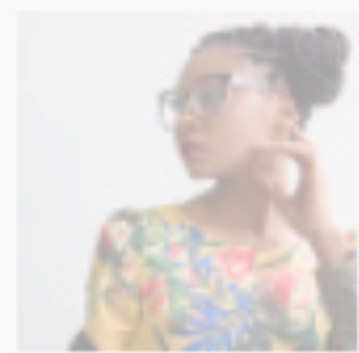
STRENGTH?

The storefront was empty of all decoration, but everyone around town had heard what this place was going to be: Inspectres.

The infamous company known for hunting down the supernatural around the continental United States (plus a few franchises in Canada and one right across the Mexican border) had staked a franchise claim in this city and was having open interviews for anyone who wanted to

each character is associated with a set of properties (“cards”)

CAST



EMILIA LOPEZ

LAB TECHNICIAN

GOOD HEARING

WILD x3

COWARD

CURIOSITY x3

ASSETS x4

Nature



LAB TECHNICIAN

Emilia has great experience working with chemical and biological laboratory equipment, as well as computer data analysis.

CHALLENGES

STRENGTH

The storefront was empty of all decoration, but everyone around town had b

a “narrator” user creates challenges for the characters to solve. characters write scene entries addressing these challenges.





Emilia Lopez (victoriasilverwolf) moved · 06/13/2018


Emilia was fascinated! This wasn't some silly hoax, but the real thing. She and her co-workers had seen videos of the cat speaking Russian, and multiple, detailed analysis by sophisticated computers proved it wasn't faked.

What a puzzle! Emilia spent the next several hours investigating reports of animals speaking human languages. There was a lot of information to go through, and it led in so many directions that it was hard to come to any conclusion. The cat spoke in an odd accent, too, so computer translation wasn't very reliable. She could have handled Spanish well enough, and even stumbled her way through other Romance languages, but Russian was, well, Greek to her.

“Anybody here speak Russian?”

It couldn't hurt to ask.

 HIT THE BOOKS 

 CURIOSITY



E

Emilia was fasci
the real thing. Sh
the cat speaking
by sophisticated

What a puzzle! E
speaking human
many directions
too, so compute
enough, and eve
well, Greek to he

“Anybody here s

It couldn't hurt to ask.



Obstacle



HIT THE BOOKS



You have your personal libraries, the internet, and the public library. Time to research what you can about werewolves.



Heading towards an **uncertain** outcome, with 3 cards left.

Outcomes

Strong:

There are a few reasons a cat might start speaking Russian. Possession, inter-dimensional leakage... Either way you end up researching an exorcism, just in case this is a spirit or a demon.

Weak:

You find something out about a way to deal with Russian talking cats. It's entirely wrong, and may make you look like an idiot.

1 • 06/13/2018



HIT THE BOOKS



CURIOSITY

ing reports of animals
go through, and it led in so
he cat spoke in an odd accent,
ave handled Spanish well
languages, but Russian was,

# games	5,743
# users	30,119
# player cards	154,586
# challenges	61,223
# scene entries	448,264
total # of tokens	~126 million
ave tokens per entry	~247

mostly fantasy / sci-fi stories

(see topics below)

mysterious, temple, cave, mythology,
sacred, archeological, tomb,
archaeological, mystical, ancient

creature, beast, monster, demon,
undead, dragon, kill, snake,
monstrous, evil

commander, gunship, artillery, spacecraft,
squadron, infantry, warship, frigate,
fleet, aircraft

wounded, ambush, troops, infantry,
defeated, assault, paramilitary,
melee, militia, enemy

telemetry, module, apparatus,
calibration, device, monitoring, system,
detection, diagnostics, sensor

prey, carnivorous, fanged, slain,
bloodthirsty, demonic, fearsome,
monster, demon, undead

nutjob, batshit, borderline, purely,
lunatic, manipulation, stereoscopic,
complex, propagandist

hijacker, stakeout, borta, firebomb,
smuggle, safehouse, waggars,
försöker, armados, ronja

accused, undercover, burglary, gunpoint,
plainclothes, arrest, policeman, robbery,
gunman, police

slope, trail, ravine, steep,
coastline, shoreline, wooded,
hillside, roads, river

how do we evaluate?

- we put our models on the Storium platform!
- users can generate a suggestion using our model and then edit it. we collect their edits as well as some simple feedback.
 - compute recall-based metrics like ROUGE, or MT metrics like HTER (human-targeted translation edit rate)
- evaluation by domain experts for free!

also collect feedback scores

🔔 Get a suggestion

The suggested text has been added to your draft — you are free to edit it, of course, or discard it entirely. Regardless of whether or not you use it, please tell us what you thought of it:

Bob has his eyes closed and a cup of tequila steaming in the bar. He slides to one knee, before pulling it out slowly onto an examination table with high polished wooden panels on top. "Hey! Man, you all right? I'm sure my dad won't be surprised at how skinny this kid looks."
Once he's been seated facing each student individually for several minutes Bob searches through their body fluids. With chewed tissue noisily dusting from him as blood continues piling up like coffee spills down a toilet seat drawers they're

How grammatically correct is the suggested text? (on a scale of 1-5, with 1 being the lowest) (Required)

1 2 3 4 5

How relevant is the suggested text to the current story so far? (on a scale of 1-5, with 1 being the lowest) (Required)

1 2 3 4 5

How well do the sentences in the suggested text fit together? (on a scale of 1-5, with 1 being the lowest) (Required)

1 2 3 4 5

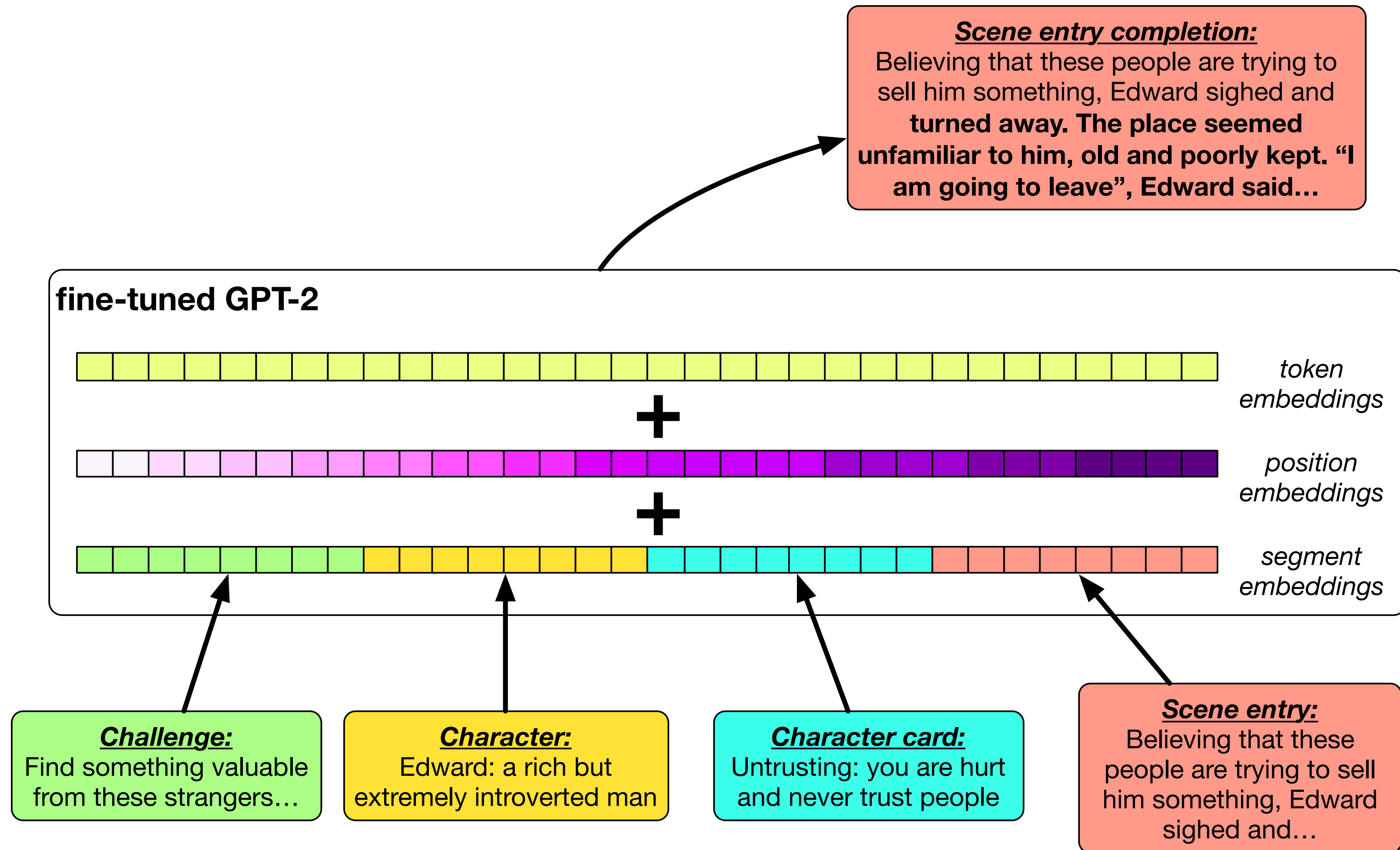
How enjoyable do you find the suggested text? (on a scale of 1-5, with 1 being the lowest) (Required)

1 2 3 4 5

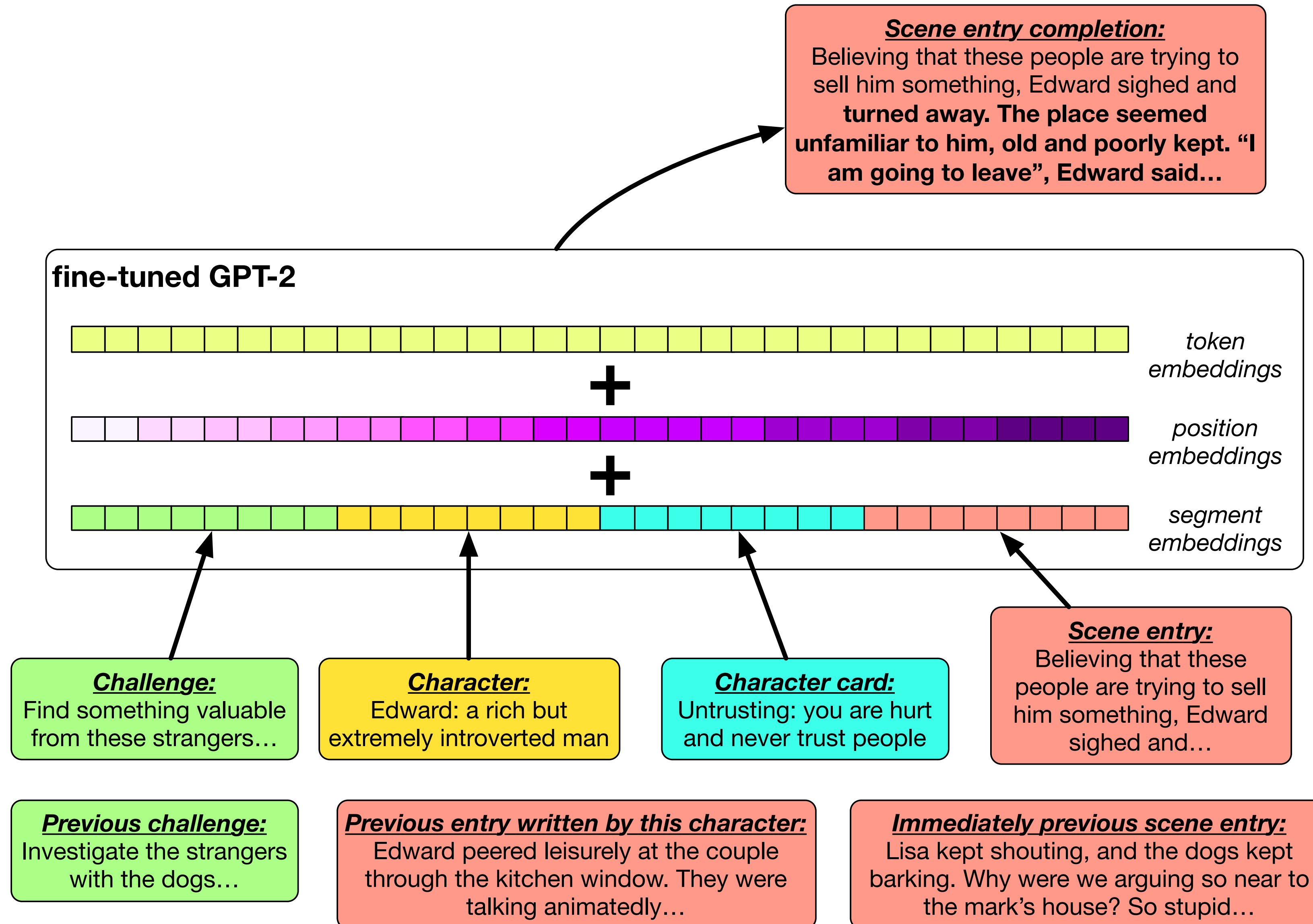
Please provide any additional comments you have about the suggested text. (Optional)

Done

We've built models that look like this:



adding more context...



adding more context improves feedback scores

Model	Fluency	Relevance	Coherence	Likability	Perplexity
Base (fine-tuned GPT-2 medium)	3.6	2.8	3.6	2.4	22.9
Base + history	3.7	3.2	4.1	2.7	21.0

adding more context improves feedback scores

Model	Fluency	Relevance	Coherence	Likability	Perplexity
Base (fine-tuned GPT-2 medium)	3.6	2.8	3.6	2.4	22.9
Base + history	3.7	3.2	4.1	2.7	21.0

A metric over user edits

- Intuitively, edits with fewer changes (additions or deletions) should receive a **higher** score, as this indicates that the generated text was useful to the Storium author.
- Our **USER** metric is a modification of **ROUGE-LCS** that measures the *longest common subsequence* between the edited text and the generated text

top- k sampling performs better than nucleus sampling

Decoding algorithm	Fluency	Relevance	Coherence	Likability	USER
top- k	4.0	2.6	3.4	3.3	15.6
nucleus	3.8	2.4	2.9	3.2	9.9

we collected >300 edits for each decoding algorithm

top- k sampling performs better than nucleus sampling

Decoding algorithm	Fluency	Relevance	Coherence	Likability	USER
top- k	4.0	2.6	3.4	3.3	15.6
nucleus	3.8	2.4	2.9	3.2	9.9

we collected >300 edits for each decoding algorithm

correlations between Likert scores and **USER**

	likability	fluency	coherence	USER
relevance	0.51	0.28	0.55	0.51
likability		0.28	0.35	0.34
fluency			0.54	0.13
coherence				0.25

correlations between feedback and **USER**

	likability	fluency	coherence	USER
relevance	0.51	0.28	0.55	0.51
likability		0.28	0.35	0.34
fluency			0.54	0.13
coherence				0.25

relevance correlates most strongly with **USER**

correlations between feedback and **USER**

	likability	fluency	coherence	USER
relevance	0.51	0.28	0.55	0.51
likability		0.28	0.35	0.34
fluency			0.54	0.13
coherence				0.25

fluency does not really correlate with **USER**

how do users modify the generated suggestions?

scene entry after user edited the suggestion:

Chiang Lee hacked the shop's computer before going inside. That should keep them busy. Alarms went off, electronic locks opened, and they worried about losing all their data. He slipped in, and **crossed the "security line"**. In this area of the shop there was **an orange sticker across all windows. Chiang Lee was taking photos as** he went for the record; **every once in awhile someone wanted special items**, so it would be good to keep a log. However, for now he knew exactly what he wanted - he found it, and pocketed it. It was a wide-spectrum signal blocker. It could interfere with electronic equipment, block signals and - allegedly - interfere with laser blasts and force fields. Whether CIA, FBI or UFO, anyone wanting to take him or slow him down had better stick to analogue and wood, because anything scientific or technological was toast. **Chiang noticed** other people taking **pictures while running around** the shop **looking for some secret** item...he blended in perfectly.

Common usage pattern: users keep short phrases from the suggestion but completely change the content

often, users delete everything

generated text:

Oh, wait, Annie thought. *This is just like in the movies.* That's where she got stuck with a living ghost! She was surprised at how this part of Christmas went on so long - and by its strange behaviour that it started acting up again when things weren't exactly normal for her to be having fun. So... what did Annie do? She had a friend helping all over town. A caring friend who passed through Kringle every year-and-a half, passing out presents from their favourite tree as they went! And here they were going to make Christmas special (at least until a return trip home!), and now there seemed no end in sight before she'd need an exorcism, or possibly a demon blood sacrifice... But maybe if these last few days hadn't gone so very smoothly... she could reach back into her own deep pocket one more time and get her hands free – It would give you a real boost later. Lets hope nothing happened during the ritual, too... But enough “hurr durr”

no n-gram overlap at all!

final edited version:

Annie checked her list again; she'd already read it multiple times, but she might now, finally, be ready for Christmas. Thanks to her new flight experience, she was actually going to be able to help deliver some of the many presents this year. She had also managed some private preparations of her own for afterwards. Now watching the snow fall softly just outside the window, Annie Bell felt the familiar Christmas tingle as the magic of the evening approached. She dressed warmly, and went to help Noelle with the reindeer, knowing the other elves would soon start to arrive to begin packing the sleigh. Back in her room, six brightly wrapped Christmas presents glinted on her desk each labelled to one of her new found friends.

user feedback

I feel like the ai gives me idea's for keeping things going for example I was gonna just say they slay the big golem but then the machine came up with there was 3 golems so I had to keep going and rethink the story I just kept the writing because I like what I was doing with the cards and everytime the ai changed the story I just worked with it like a friend giving me ideas.

user feedback

I feel like the ai gives me idea's for keeping things going for example I was gonna just say they slay the big golem but then the machine came up with there was 3 golems so I had to keep going and rethink the story I just kept the writing because I like what I was doing with the cards and everytime the ai changed the story I just worked with it like a friend giving me ideas.

I tried it in one story that is almost dead, just to see what the ai would come up with. It was hilarious, bad and amazing at the same time.

user feedback

I feel like the ai gives me idea's for keeping things going for example I was gonna just say they slay the big golem but then the machine came up with there was 3 golems so I had to keep going and rethink the story I just kept the writing because I like what I was doing with the cards and everytime the ai changed the story I just worked with it like a friend giving me ideas.

I tried it in one story that is almost dead, just to see what the ai would come up with. It was hilarious, bad and amazing at the same time.

There is a disproportionate amount of innuendo and strong language coming through - not suitable for the age range of the games and not at all explainable by the context.

user feedback

I feel like the ai gives me idea's for keeping things going for example I was gonna just say they slay the big golem but then the machine came up with there was 3 golems so I had to keep going and rethink the story I just kept the writing because I like what I was doing with the cards and everytime the ai changed the story I just worked with it like a friend giving me ideas.

I tried it in one story that is almost dead, just to see what the ai would come up with. It was hilarious, bad and amazing at the same time.

There is a disproportionate amount of innuendo and strong language coming through - not suitable for the age range of the games and not at all explainable by the context.

In my head, now, I'm starting to assign bios to the AIs

- swears wild-child outsider who keeps getting piercings and tattoos and stands around at parties, glaring at the more popular kids.
- the reluctant kid who's actually quite keen to be involved but a little shy to start with.
- the convent-educated naive kid who's just taken LSD for the first time and whoa... gotta tell you about all the cyberpunk stuff!