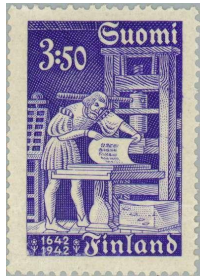
Evaluating Creative Short Story Generation in Humans and LLMs

Mete Ismayilzada

Creative Short Story Generation



stamp

letter

send

"I wrote a **letter** to my aunt. I went to the post office and bought a **stamp**. I put the stamp on the letter and gave it to the mailman to **send**."

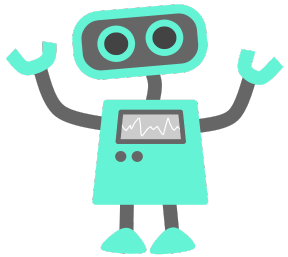
Setup

4 item sets

stamp-letter-send
petrol-diesel-pump
organ-empire-comply
gloom-payment-exist

} low

} high



51 models

GPT, Gemini, Claude, Llama,
Grok, Gemma, Mistral, Qwen,
Reka, Phi, Command-R,
Jamba,
and etc.

61 humans

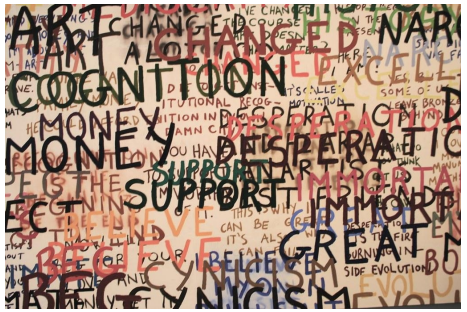
Turing Test

petrol-diesel-pump

During one night at my local gym I bumped into Vin **Diesel** who commented on my ability to lift my weight correctly. He decided to show me a better way to get my **pump** on. Whilst given the chance to ask him, I quipped if he had any siblings named after **petrol**.

In the bustling heart of Mechanica City, where machines roamed free and humans were a rare sight, a curious little robot named Zephyr discovered an ancient contraption labeled “The **Petrol** Pump of Dreams.” Legend had it that this **pump**, when filled with either **petrol** or **diesel**, could grant a single wish to the one who dared to activate it. Zephyr, with gears whirring and circuits buzzing, decided to fill the **pump** with a blend of both fuels, hoping for an extraordinary outcome. As the concoction flowed through the **pump's** rusty veins, a magical surge enveloped the city, transforming every mechanical being into sentient, imaginative creatures. From that day forward, Mechanica City thrived with creativity and innovation, forever grateful to the little robot who dared to dream beyond fuel.

Evaluation



Complexity

- lexical
- syntactic

LLMs generate linguistically **more complex** stories.

Diversity

- lexical
- semantic

LLM stories largely **lack** lexical and semantic **diversity**.

Creativity

- novelty
- surprise

LLM stories are **less novel** and **surprising** than human stories.

Thank you!