# MAUVE: Measuring the Gap Between Neural Text and Human Text

**Krishna Pillutla**, Swabha Swayamdipta, Rowan Zellers, John Thickstun, Sean Welleck, Yejin Choi, Zaid Harchaoui

## Motivation

Enormous language models can now write high quality text. But how close is it to human text?


Brown et al. (2020), Devlin et al. (2018)

### Open Ended Text Generation

**>> prompt:** In a shocking finding, scientists discovered a herd of unicorns living in a remote, previously, unexplored valley, in the Andes Mountains.


**Continuation**. The scientists named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously….

**Desiderata**

- Coherent
- Creative
- Fluent

### How good is neural text?

**Continuation 2**. This discovery has kicked off an all-out search for other mythical creatures from the frozen reaches of the Antarctic to the tropical islands of the Pacific

**Continuation 3**. Perhaps most astonishingly, these unicorns have developed their own artificial general intelligence named Yuyaysapa …

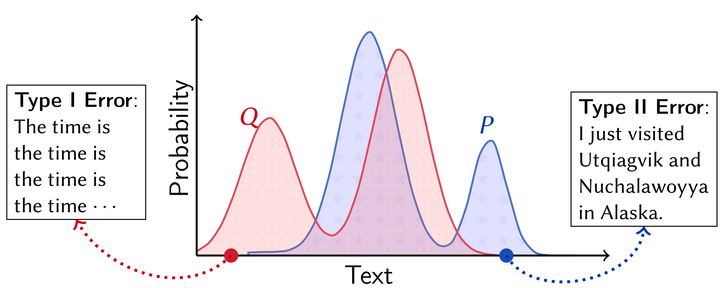
Multiple Correct Generations!

**Our Approach:** Directly compare distributions!

## Errors In Text Generation

$Q$: machine text distribution

$P$: human text distribution




e.g. **repetitions**

Type I Error = $\mathrm{KL}(Q|P)$

e.g. **truncation**

Type II Error = $\mathrm{KL}(P|Q)$

## MAUVE

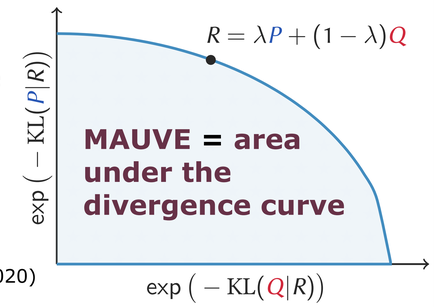
$P$: human distribution

$Q$: machine distribution

i. KL can be infinite

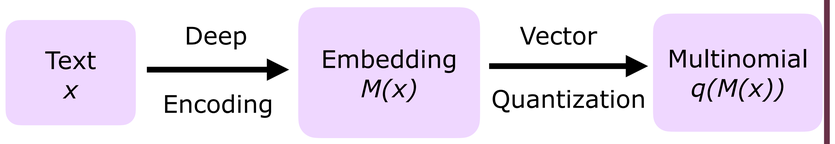
ii. Smooth KL with mixture distribution

iii. Varying mixture weight $\implies$ **divergence curve**




Sajjadi et al. (2018); Djolonga et al. (2020)

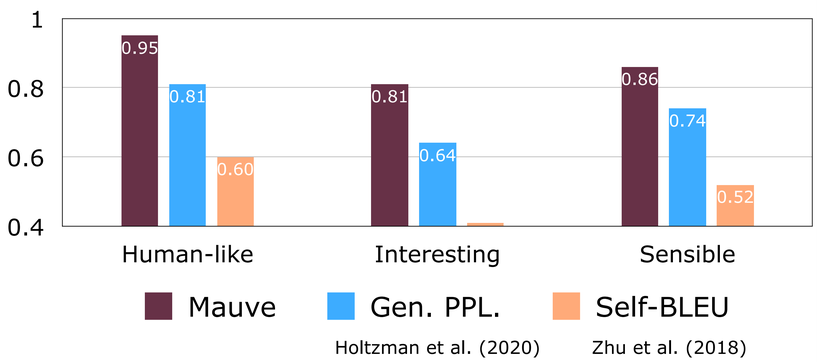
### Computing MAUVE: KLs are intractable




## Experiments

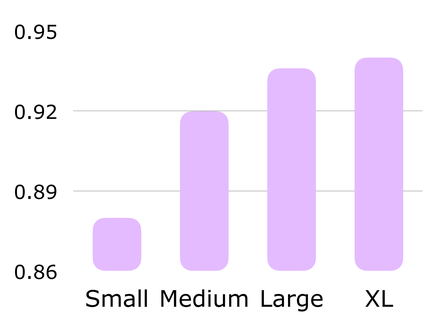
### MAUVE correlates with human judgements

Spearman Correlation w/ human eval (↑)

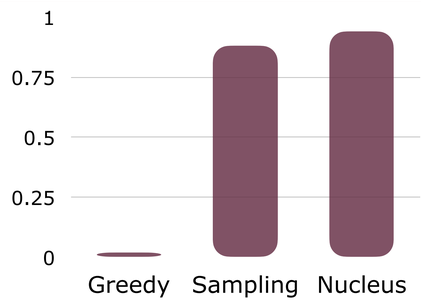



Holtzman et al. (2020) Zhu et al. (2018)

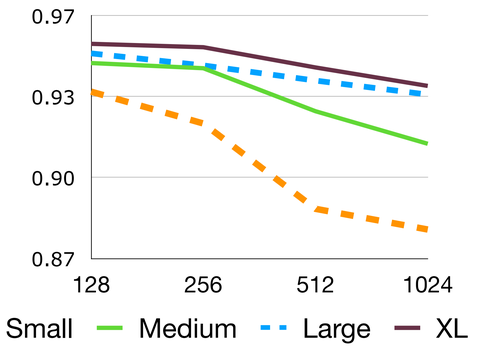
**MAUVE captures the effect of scale**




**MAUVE captures the effect of decoding**




**MAUVE captures the effect of text length**




## Software

**pip install mauve-text**

```
from mauve import compute_mauve
out = compute_mauve(p_text=p_text, q_text=q_text)

print(f'Mauve(P, Q) = {out.mauve}')
```

## Conclusion

**MAUVE can accurately measure the gap between neural text and human text!**

Theory of **MAUVE**: See our other paper at **NeurIPS 2021**

***Liu, Pillutla, et al***. *Divergence Frontiers for Generative Models: Sample Complexity, Quantization Level, and Frontier Integral.*

krishnap25

KrishnaPillutla

krishnap25.github.io

**Software**

SCAN ME