

Machine-in-the-Loop Rewriting for Creative Image Captioning

Vishakh Padmakumar, He He

New York University



Machine-In-The-Loop Creative Writing

- Creative writing tasks can be challenging for both humans and machines.

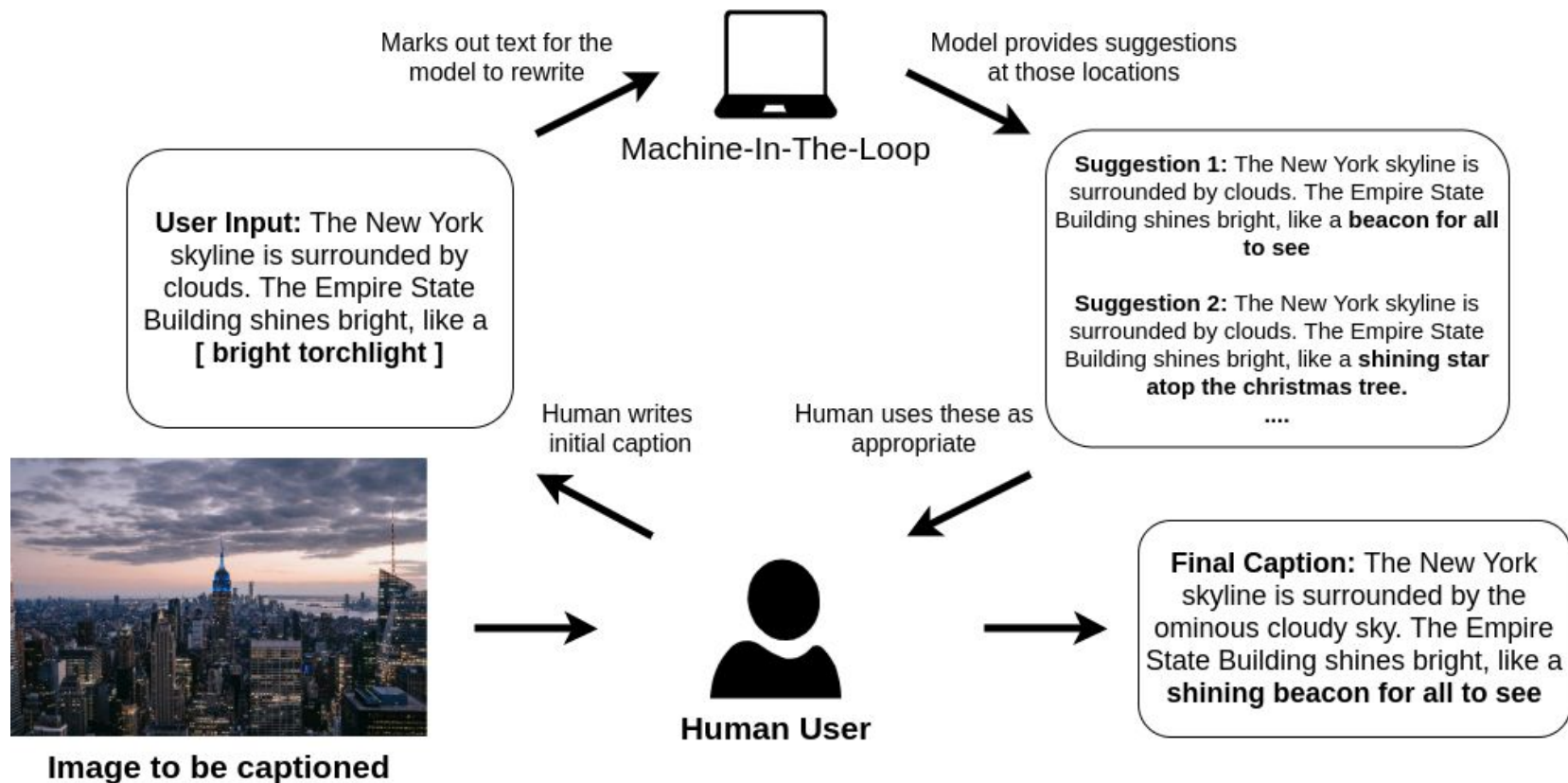
Machine-In-The-Loop Creative Writing

- Creative writing tasks can be challenging for both humans and machines.
 - Humans would benefit from suggestions on wording and framing their ideas
 - Models are able to rewrite spans of text but struggle with global coherence

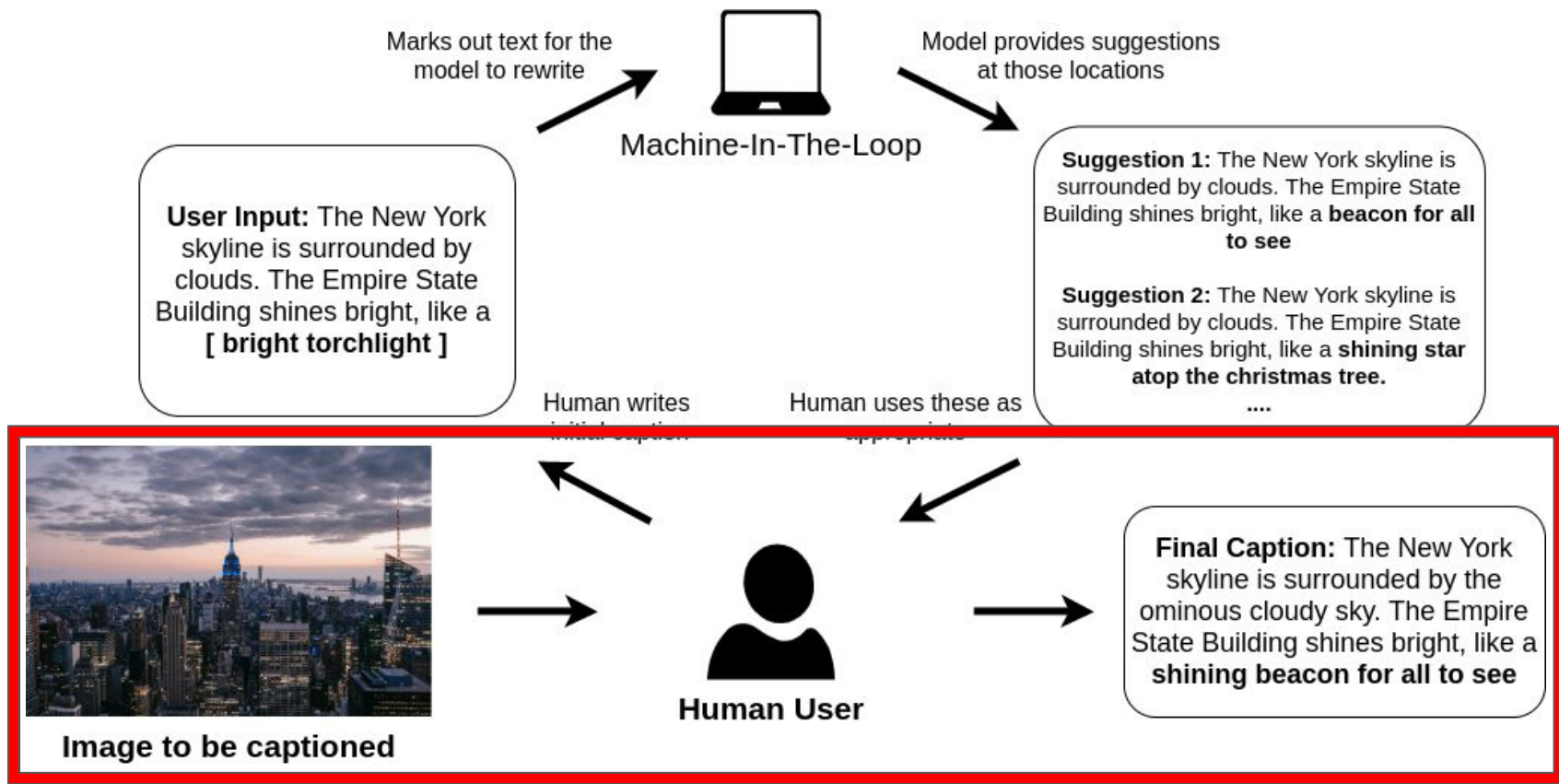
Machine-In-The-Loop Creative Writing

- Creative writing tasks can be challenging for both humans and machines.
 - Humans would benefit from suggestions on wording and framing their ideas
 - Models are able to rewrite spans of text but struggle with global coherence
- Motivates a cooperative setting: Can a model help the author improve their creative output?

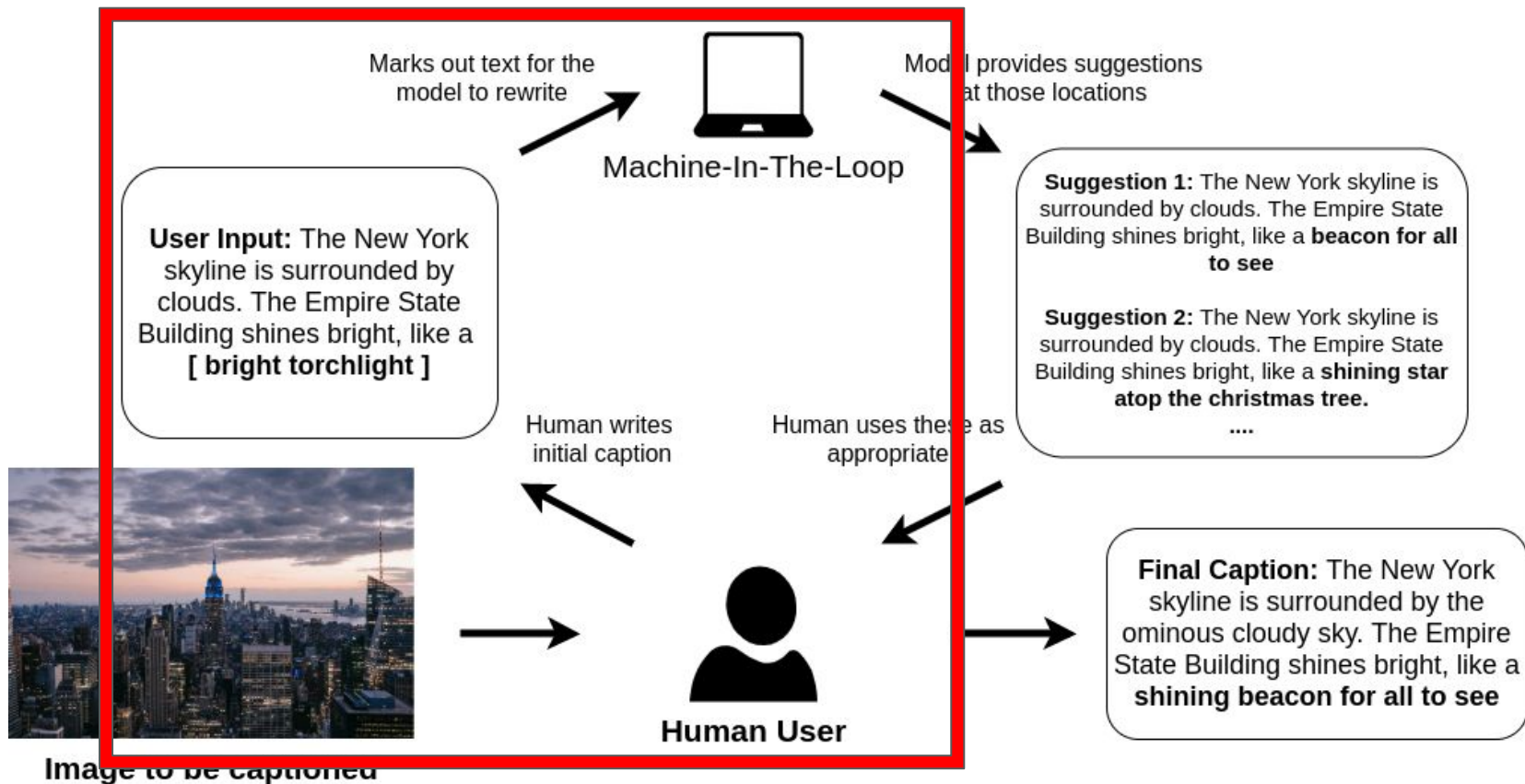
Task Setup - Creative Image Captioning



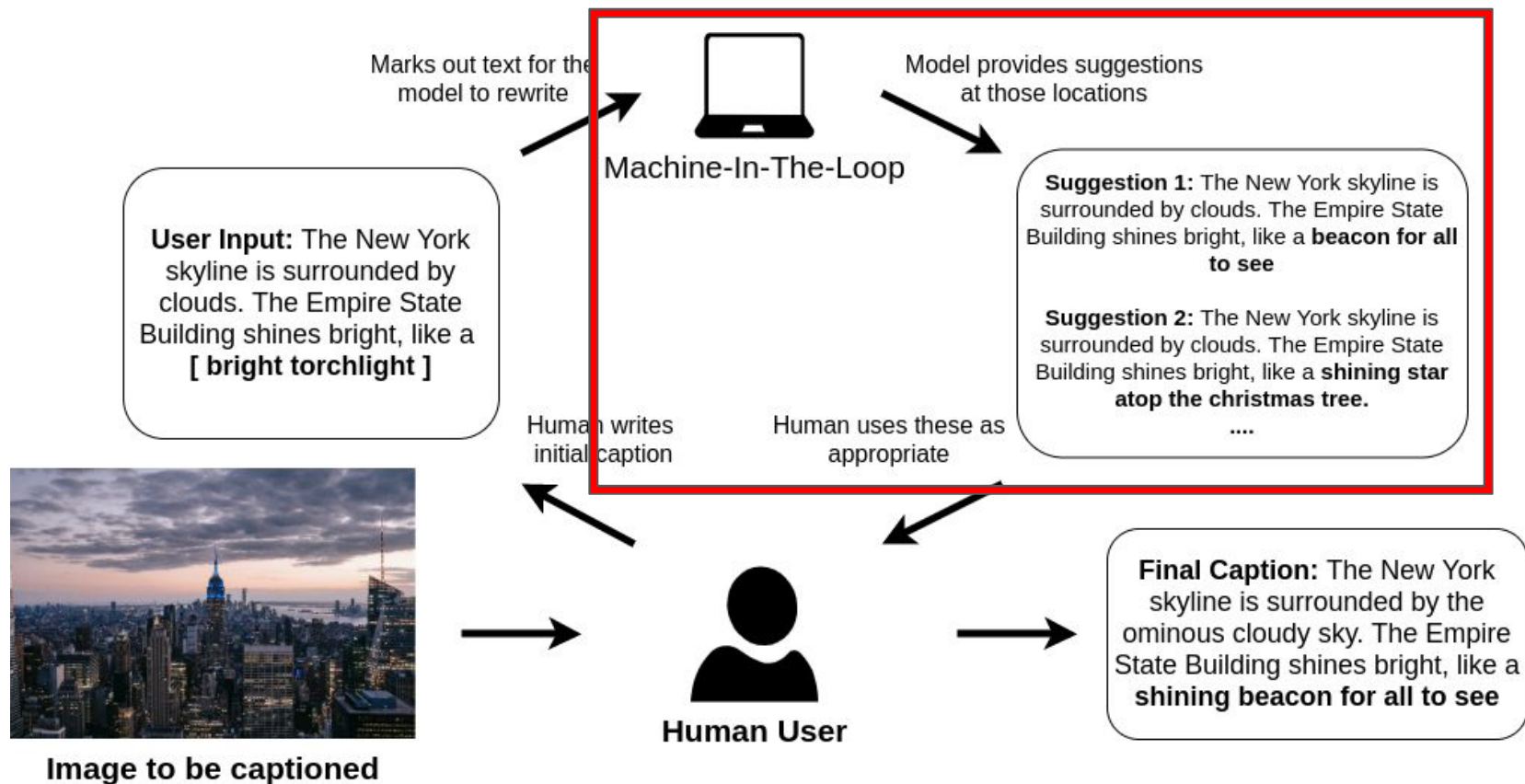
Task Setup - Creative Image Captioning



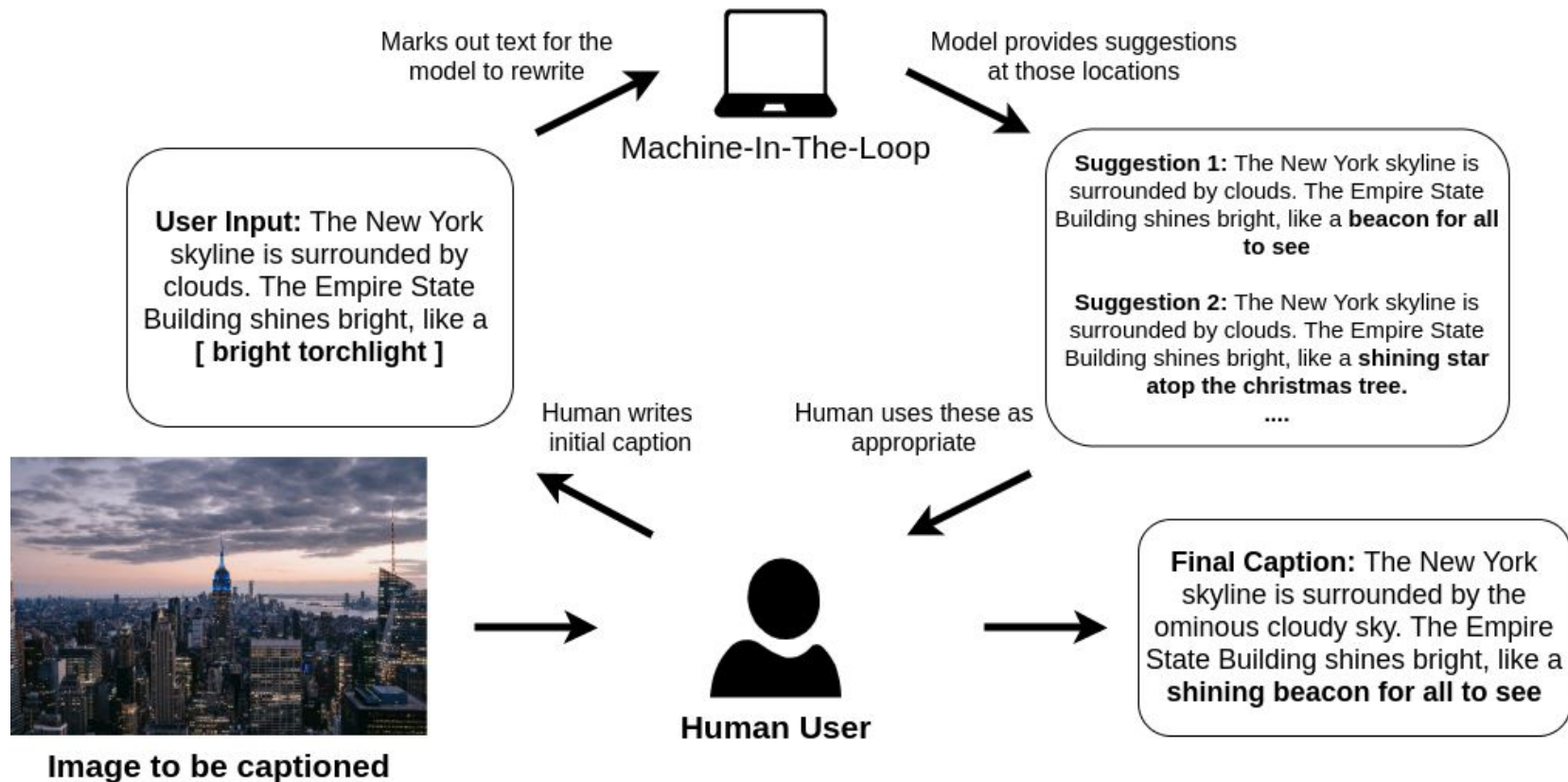
Task Setup - Creative Image Captioning



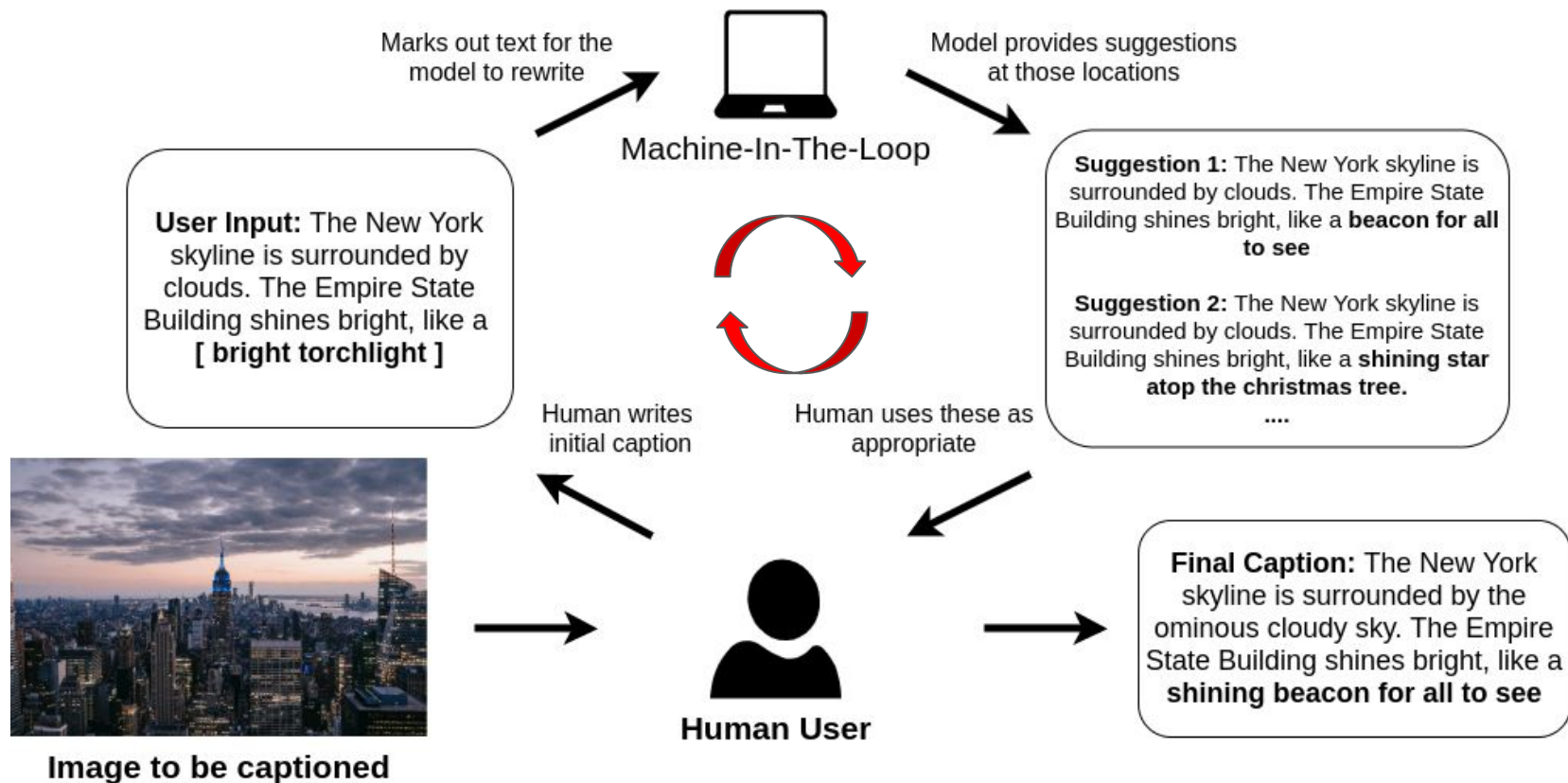
Task Setup - Creative Image Captioning



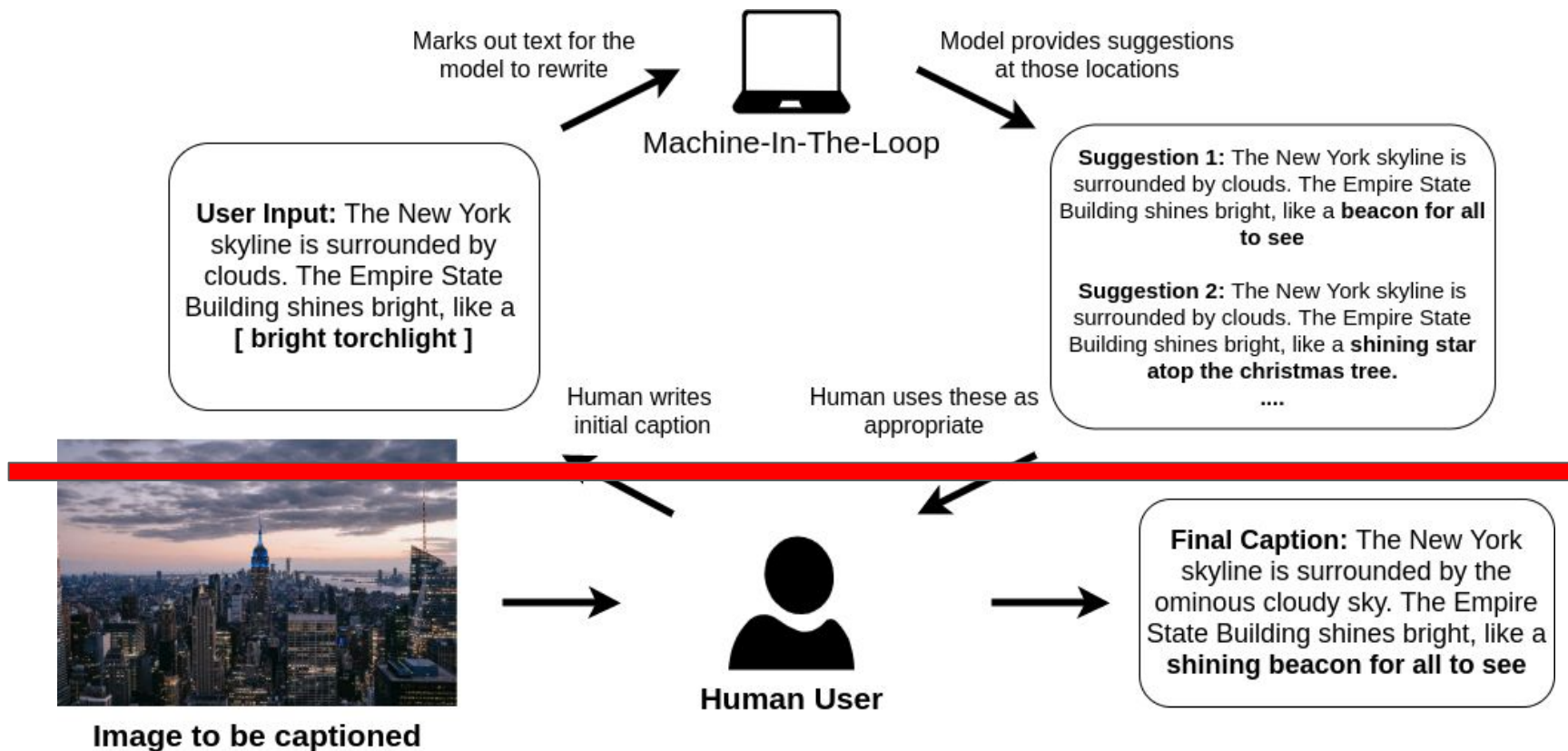
Task Setup - Creative Image Captioning



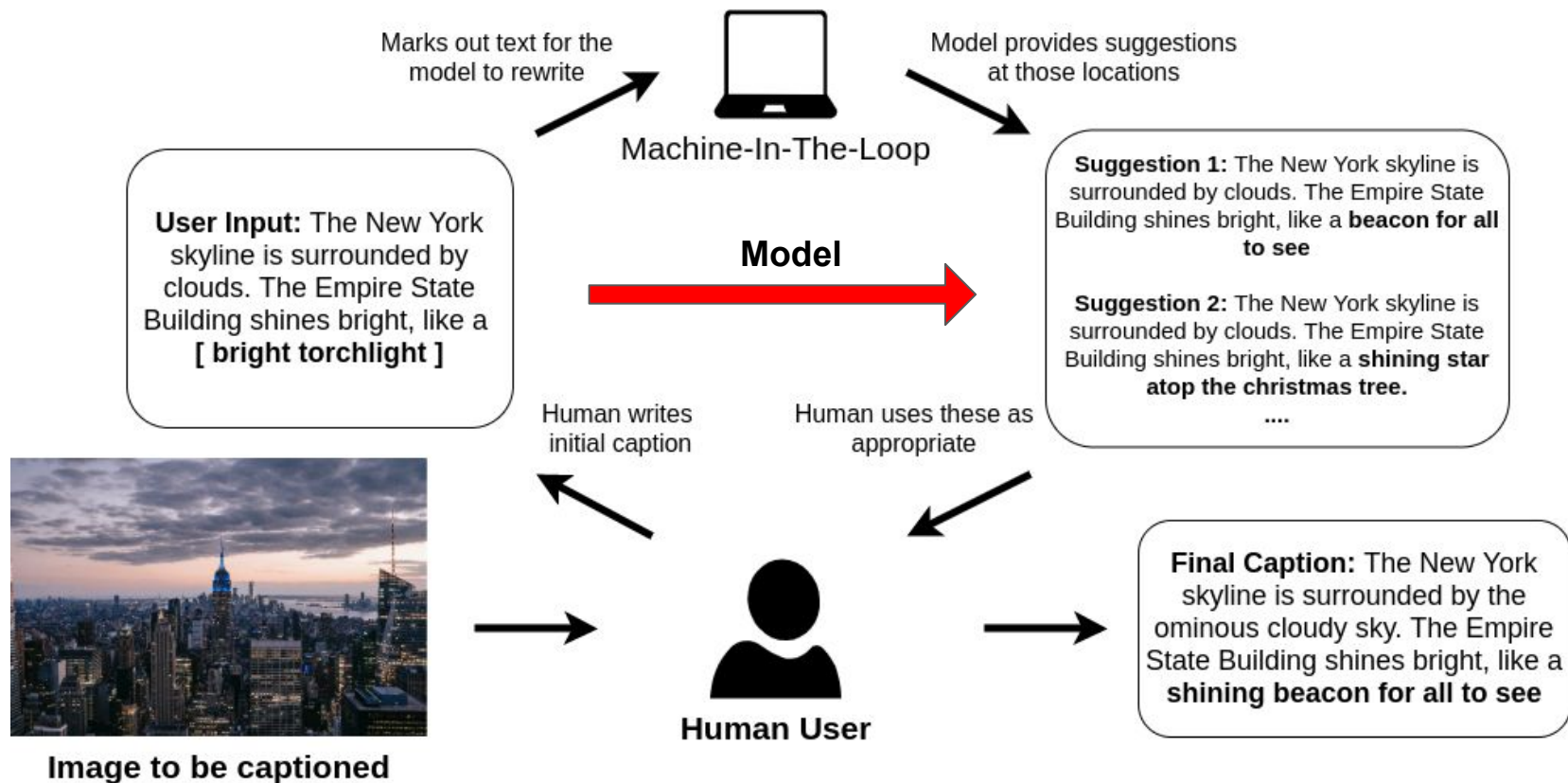
Task Setup - Creative Image Captioning



Task Setup - Creative Image Captioning



Task Setup - Creative Image Captioning



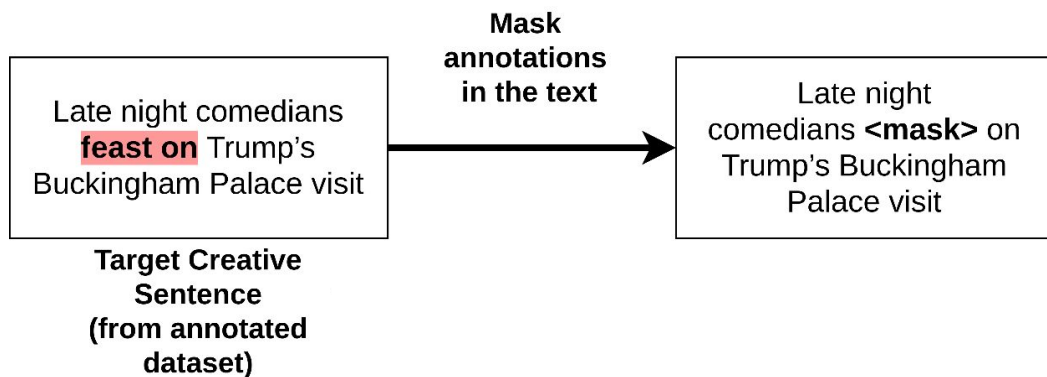
Training the Creative Rewriting Assistant (CRA) Model

Training the Creative Rewriting Assistant (CRA) Model

Late night comedians
feast on Trump's
Buckingham Palace visit

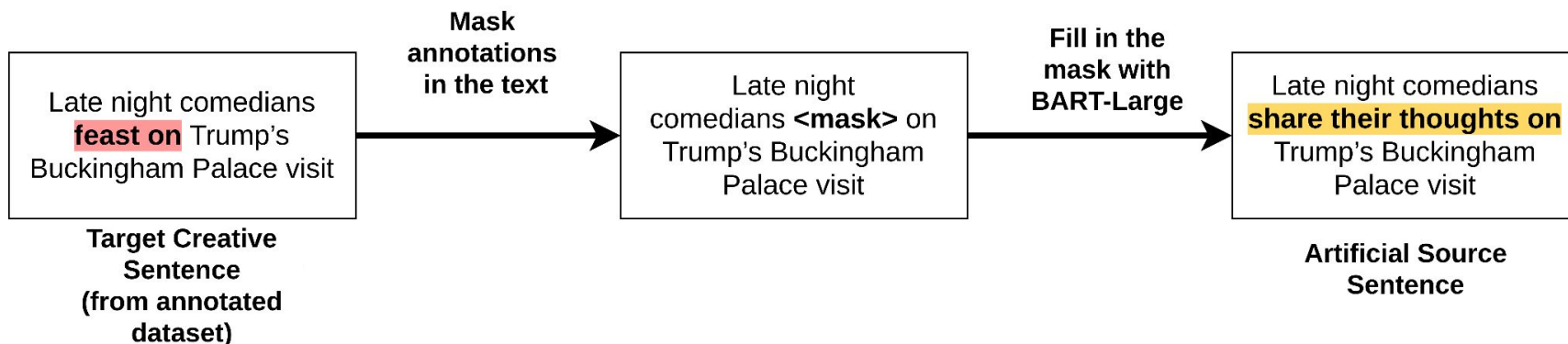
**Target Creative
Sentence
(from annotated
dataset)**

Training the Creative Rewriting Assistant (CRA) Model



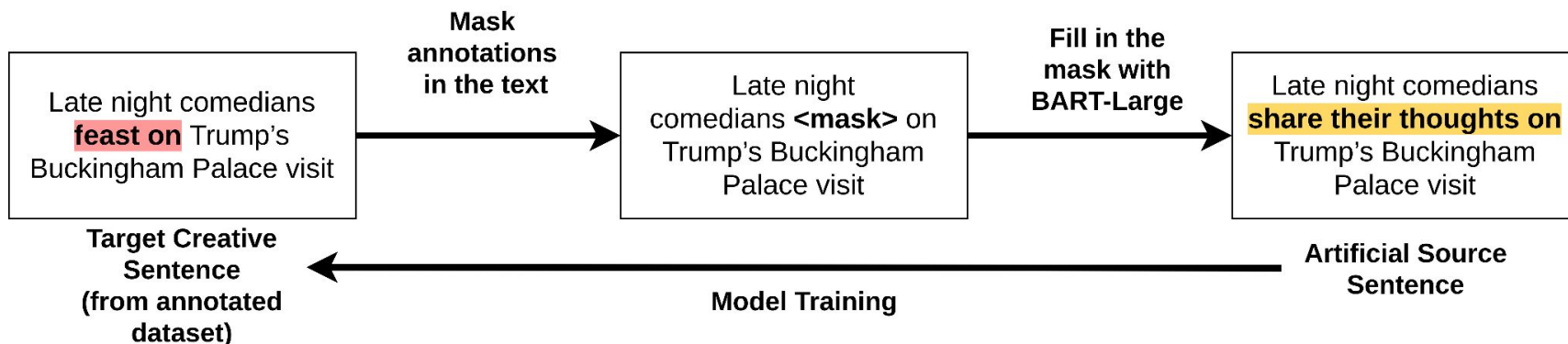
Training the Creative Rewriting Assistant (CRA) Model

- Fine-tuning Data: We create a pseudo-parallel corpus of creative sentences (annotated for literary devices) and corresponding generic sentences



Training the Creative Rewriting Assistant (CRA) Model

- Fine-tuning Data: We create a pseudo-parallel corpus of creative sentences (annotated for literary devices) and corresponding generic sentences
- **CRA** is a fine-tuned BART-Large model

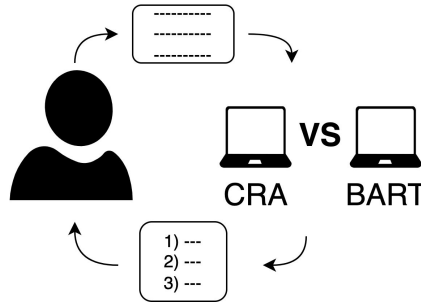


Research Questions

- Do users find CRA model suggestions helpful?
- Are users more effective at creative image captioning with model help?
- How does collaboration with the model impact different users?

Research Questions

- **Do users find CRA model suggestions helpful?**

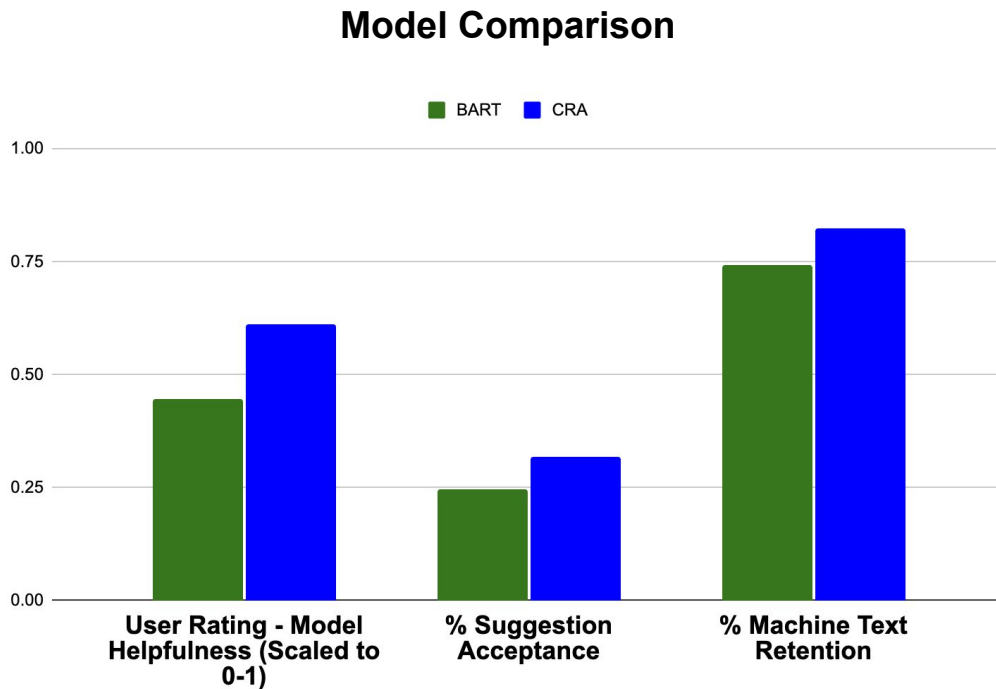


- Are users more effective at creative image captioning with model help?
- How does collaboration with the model impact different users?

Do Users Find Model Suggestions Helpful?

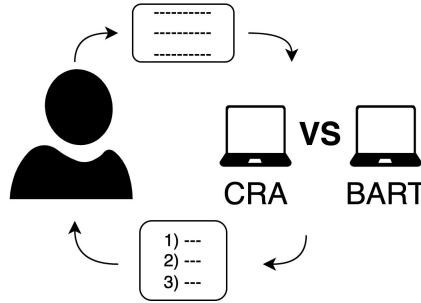
We compare the **CRA model** to a **baseline BART model** with an A/B user study (n=50)

On average, users find the **CRA model** to be more helpful than **BART** by a statistically significant margin



Research Questions

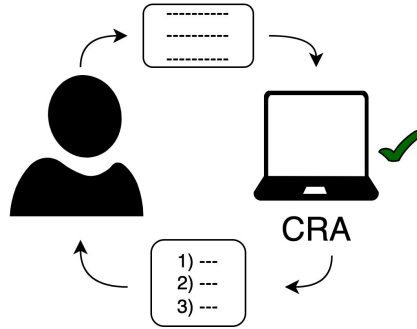
- Do users find CRA model suggestions helpful?



- Are users more effective at creative image captioning with model help?
- How does collaboration with the model impact different users?

Research Questions

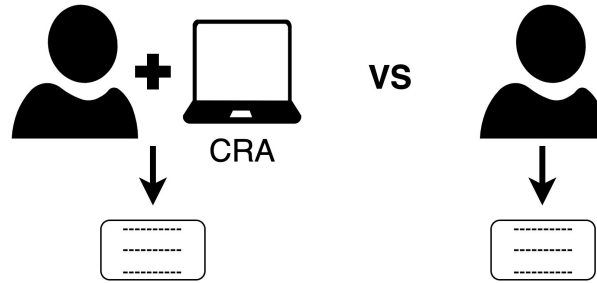
- Users find **CRA suggestions more helpful** than a baseline BART model



- Are users more effective at creative image captioning with model help?
- How does collaboration with the model impact different users?

Research Questions

- Users find **CRA suggestions more helpful** than a baseline model
- Are users more effective at creative image captioning with model help?

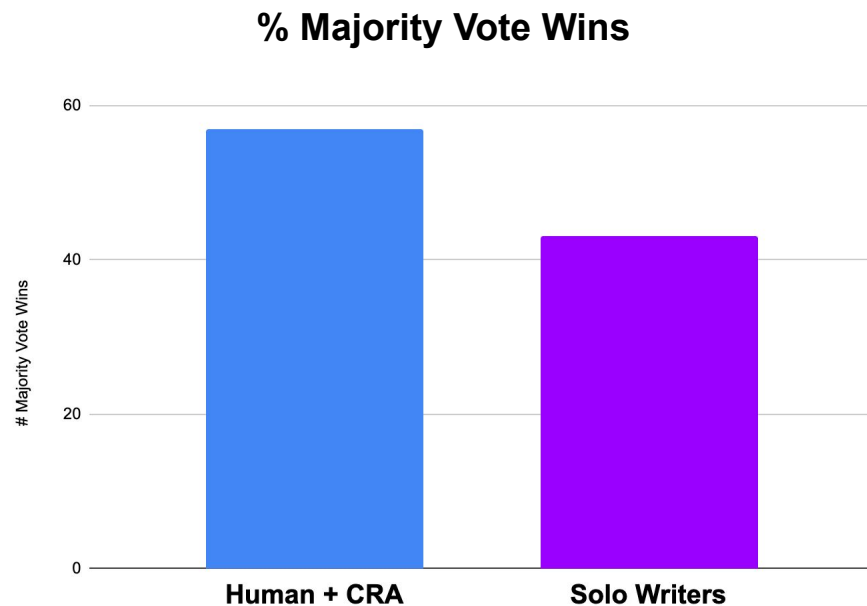


- How does collaboration with the model impact different users?

Are Users More Effective With Model Help?

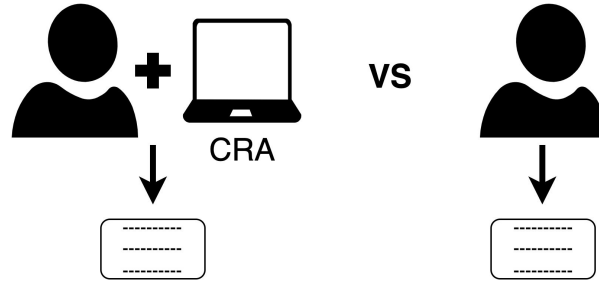
We collected captions for 100 images from **solo writers**, from users collaborating with the **CRA model** and collected 3 annotations for which is the better caption.

Collaborative users are judged to write better captions by a 57-43 margin on the majority vote.



Research Questions

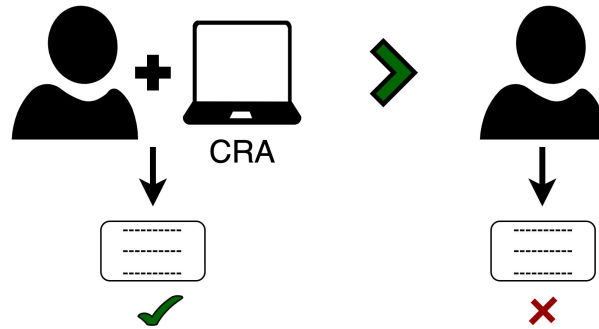
- Users find **CRA suggestions more helpful** than a baseline model
- Are users more effective at creative image captioning with model help?



- How does collaboration with the model impact different users?

Research Questions

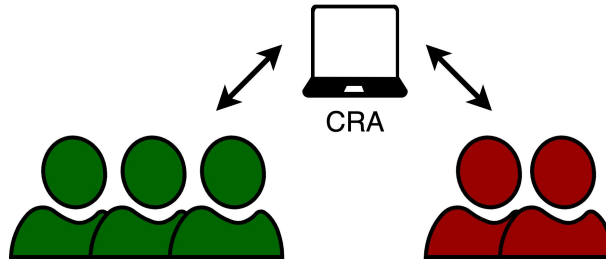
- Users find **CRA suggestions more helpful** than a baseline model
- **Collaborative users are more effective** at the creative writing task



- How does collaboration with the model impact different users?

Research Questions

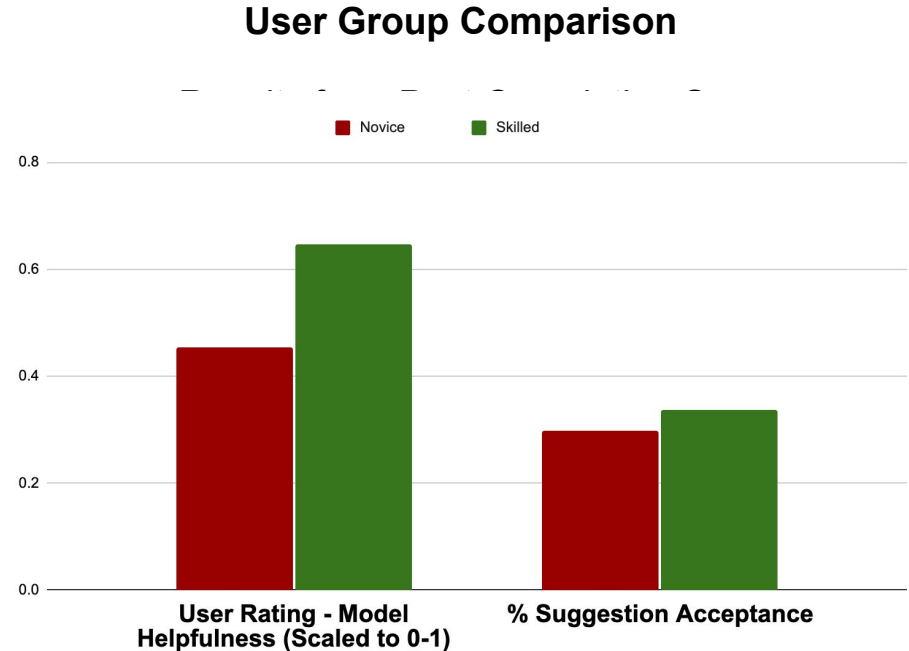
- Users find **CRA suggestions more helpful** than a baseline model
- **Collaborative users are more effective** at the creative writing task
- **How does collaboration with the model impact different users?**



How Does CRA Impact Different Users?

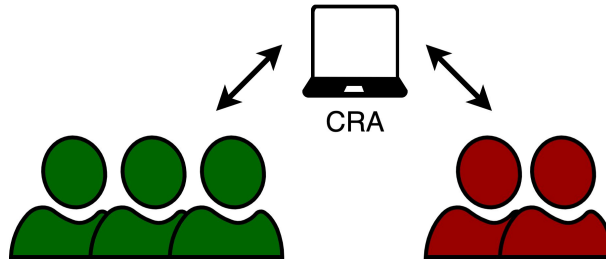
Effect of User Skill Level: We divide users into two groups, **novice** and **skilled**, based on their self-rated writing skill.

Takeaway: **Skilled** users find the CRA model to be significantly more helpful



Research Questions

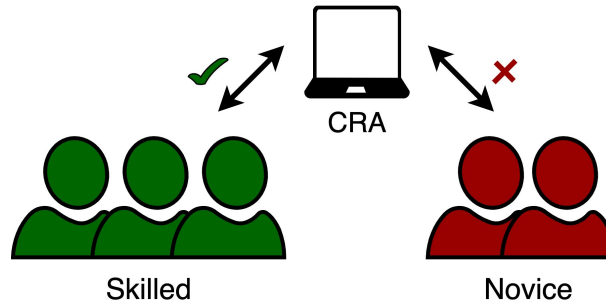
- Users find **CRA suggestions more helpful** than a baseline model
- **Collaborative users are more effective** at the creative writing task
- **How does collaboration with the model impact different users?**



Takeaways

- Users find **CRA suggestions more helpful** than a baseline model
- **Collaborative users are more effective** at the creative writing task
- **Model helps skilled writers more** potentially widening the gap in

performance



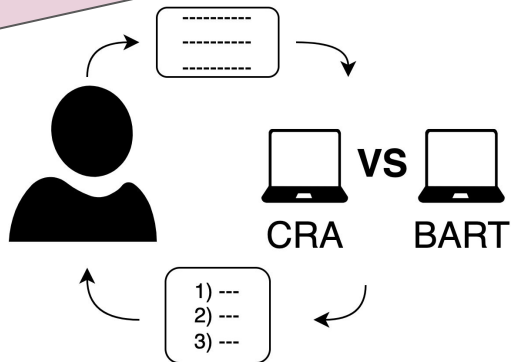
For more details, stop by
1F : Poster Session 1 on 7/11 at 10:45am :)

Machine-in-the-Loop Rewriting for Creative Image Captioning

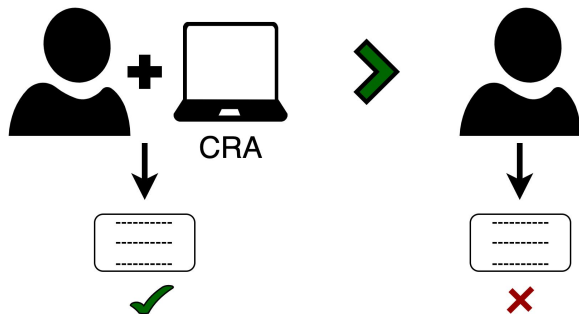
Vishakh Padmakumar, He He



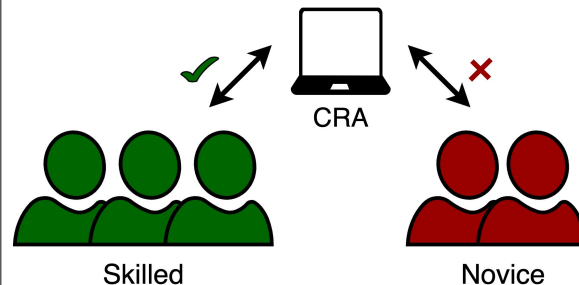
1F: Poster Session 1



Users find our **CRA** model suggestions **more helpful** than a baseline



Collaborative users outperform solo writers at creative image captioning



Models help **skilled users more**, potentially **increasing the performance gap!**