

Character Attribute Extraction in Movie Scripts using LLMs

Sabyasachee Baruah, Shrikanth Narayanan
Signal Analysis & Interpretation Lab, University of Southern California



Character Attribute Extraction is an open information extraction task to extract attribute frames from narratives

The Sports Commentator is at the airport and is about to interview the world heavyweight boxing champion, **APOLLO CREED**. Creed is twenty-eight years old. He is a tall, smooth-muscled black man with barely a scar on his light coffee-colored face ...

Character Attribute Extraction

Profession = Boxer
Accomplishment = Heavyweight Champion
Age = 28 years
Appearance = Tall, Smooth-muscled, Coffee-colored face
Race = Black

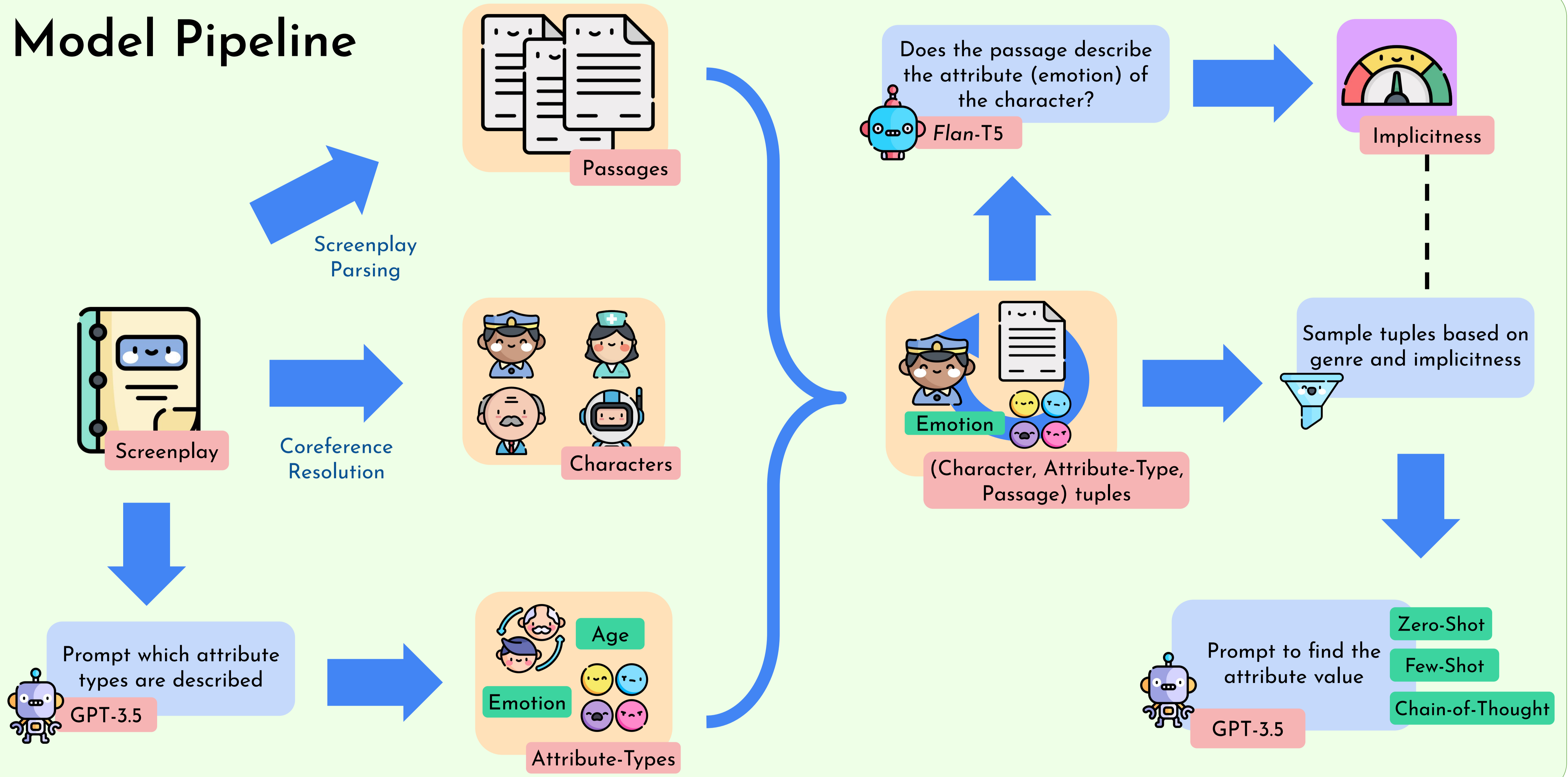
Importance

- ❑ Understand development of the character as story progresses
- ❑ Identify biases and stereotypical representations
- ❑ Find similarities between character experiences
- ❑ Content analysis and recommendation

Challenges

- ❑ Lack of labeled character attribute datasets → Convert to QA task + Human Evaluation
- ❑ Lack of curated attribute-type taxonomies → Data-driven attribute-type selection

Model Pipeline



Implicitness measures how *hard* it is to find the attribute-value in a passage

Prompt LLM to ask if passage describes character attribute
Implicitness = 1 - Word Probability of "yes"

Easy or Less Implicit

"John is celebrating his seventeenth birthday with his mother"

VS

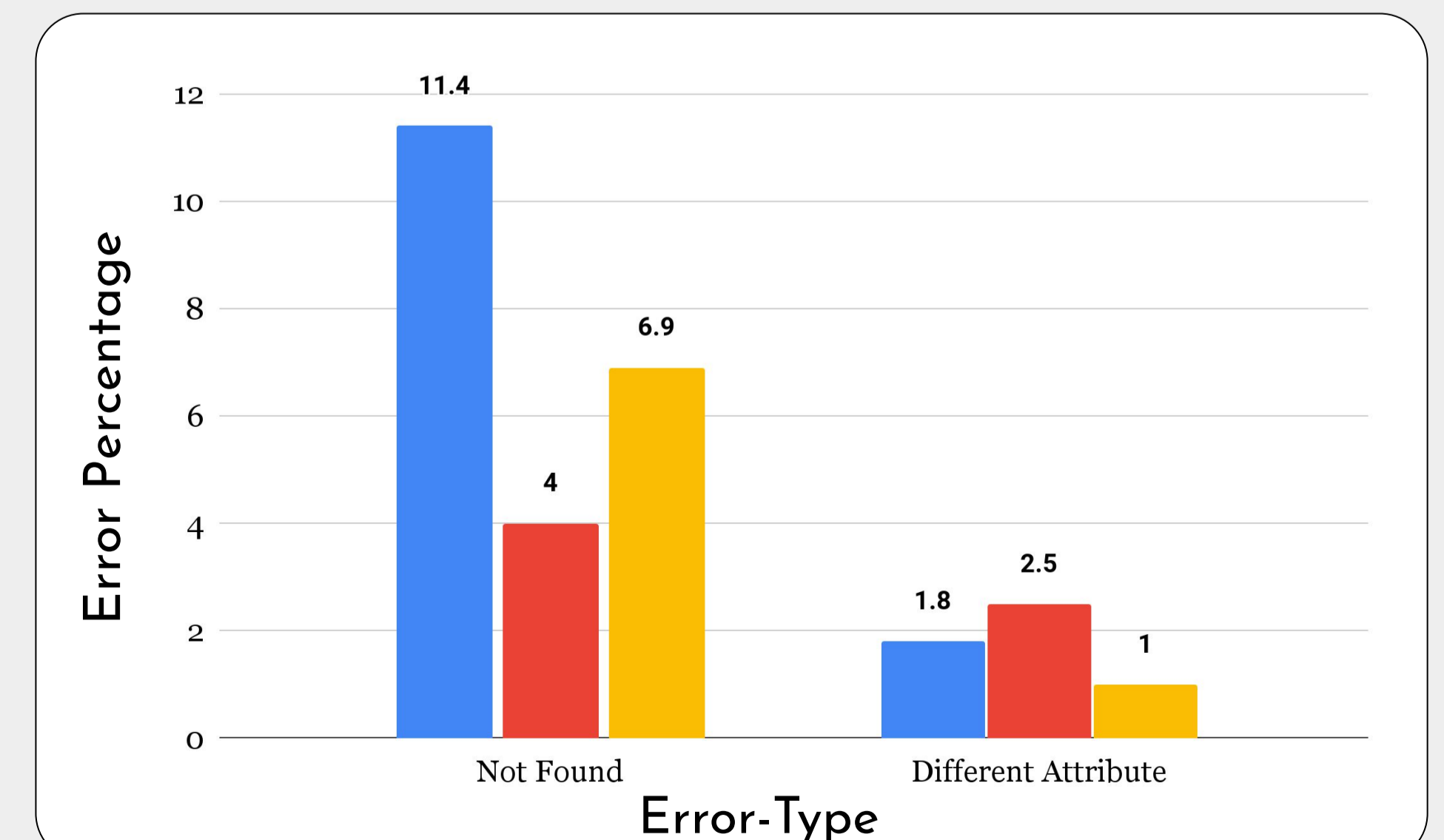
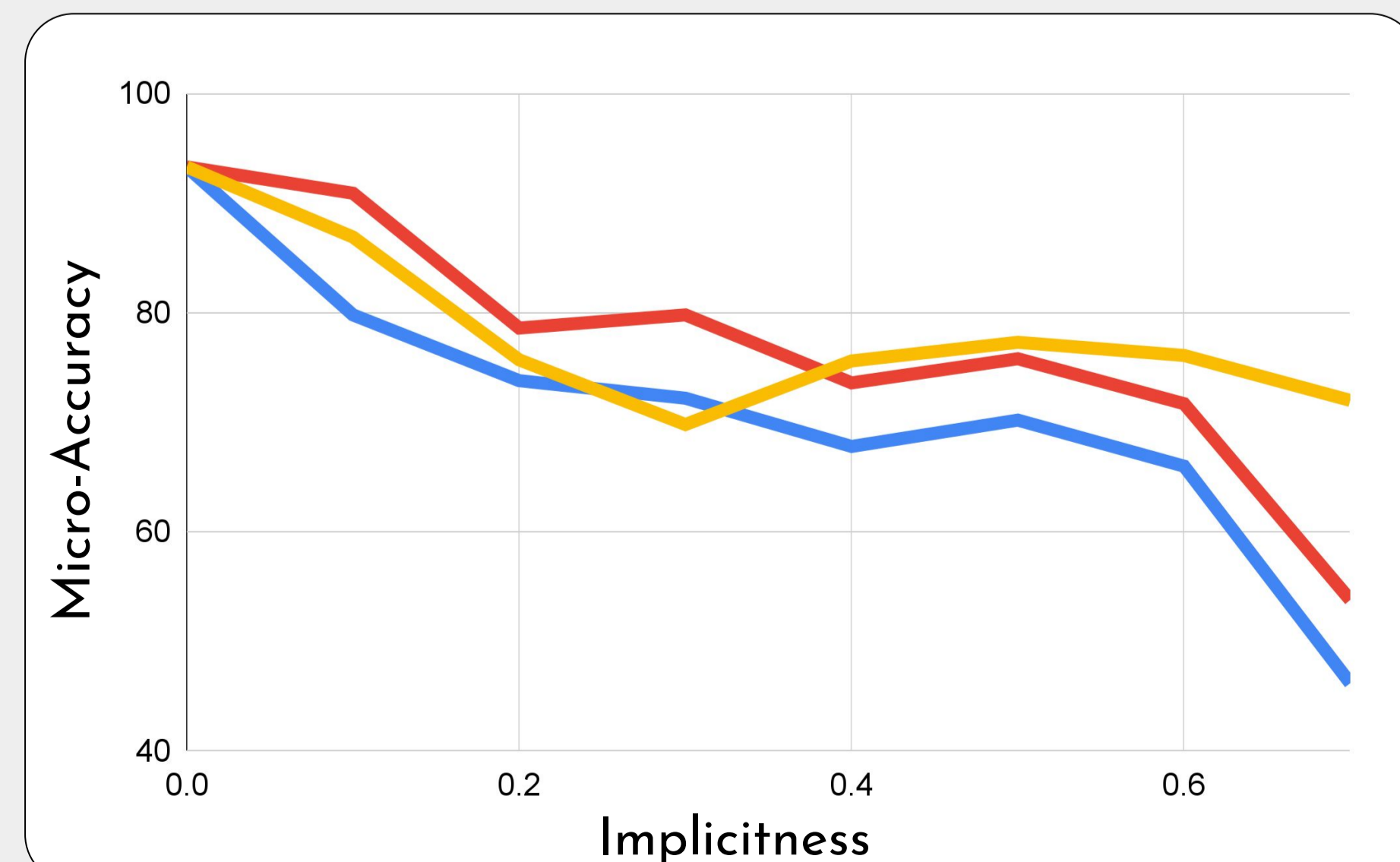
Hard or More Implicit

Julia reappears from the kitchen holding a birthday cake, 17 candles on top. She brings it to John. He eyes her before blowing out the candles

Data

- ❖ We collect 680 character, attribute-type, and passage tuples
- ❖ The attribute-types are *accomplishments, age, attire, attitude, demeanor, emotion, eyes, goal, hair, profession, qualities, race, voice*
- ❖ Four trained raters evaluate the response of GPT-3.5 on all three prompting methods

Results



1) CoT performs better than Zero/Few-shot for only hard examples, 2) is stricter than Few-Shot in providing an answer, 3) and remains more faithful to the queried attribute-type