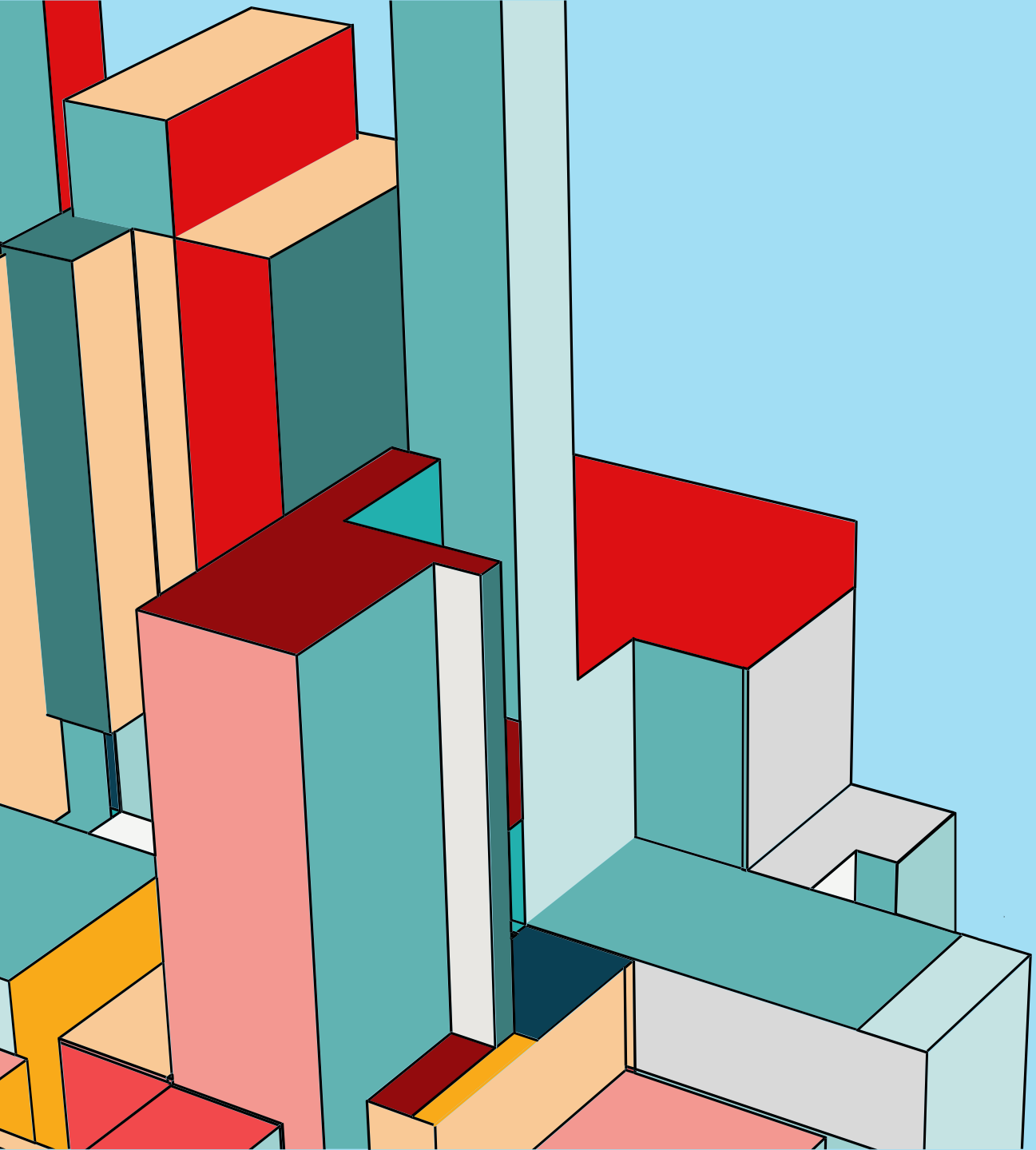


PAY TO WIN!

PAID team

Zihao Jia, Mina Son, Philseok Lee

ABOUT PAID TEAM

We are from **the Psychometrics and Individual Differences (PAID) Lab** in George Mason University

Zihao Jia

4th year PhD

Mina Son

3rd year PhD

Philseok Lee

PAID Lab advisor

THE BIG PICTURE

	Empathy	Interview	Clarity	Fairness
Task type	Classification	Generation	Regression	Classification/ regression
Base model	Fine-tuned GPT3.5		Fine-tuned GPT3.5	Fine-tuned GPT3.5
Final model (performance)	Few-shot GPT4 (0.145)	Few-shot GPT4 (0.109)	Fine-tuned DeBERTa (0.204)	Few-shot GPT4 (0.207)

EMPATHY

- Fine-tune DeBERTa or GPT3.5 turbo (failed)
  - Reasons: small sample size,
  - Unclear labels, extremely hard even for human (at least to me…)
- Use pre-trained empathy/emotion model (failed)
  - Reasons: the target construct in the current data may not necessarily be the "empathy".
- GPT4 prompt engineering with thought process (worked)
  - Step 1: ask GPT4 to generate why this email did/didn't demonstrate empathy.
  - Step 2: use the reasons generated from step 1 as the prompt. This can help GPT learning the underlying logic of previous labels.

Example prompt

User:
Here is an email written to jonathan to provide some feedback to his work on the beta project ###email starts###{ds.Text[i]}###email ends###. Though jonathan did not do very well on the work, the writer still want to convey some empathy and understanding in this email to help jonathan get improved. Now you think this email did demostrate empathy, why? Provide a short answer.

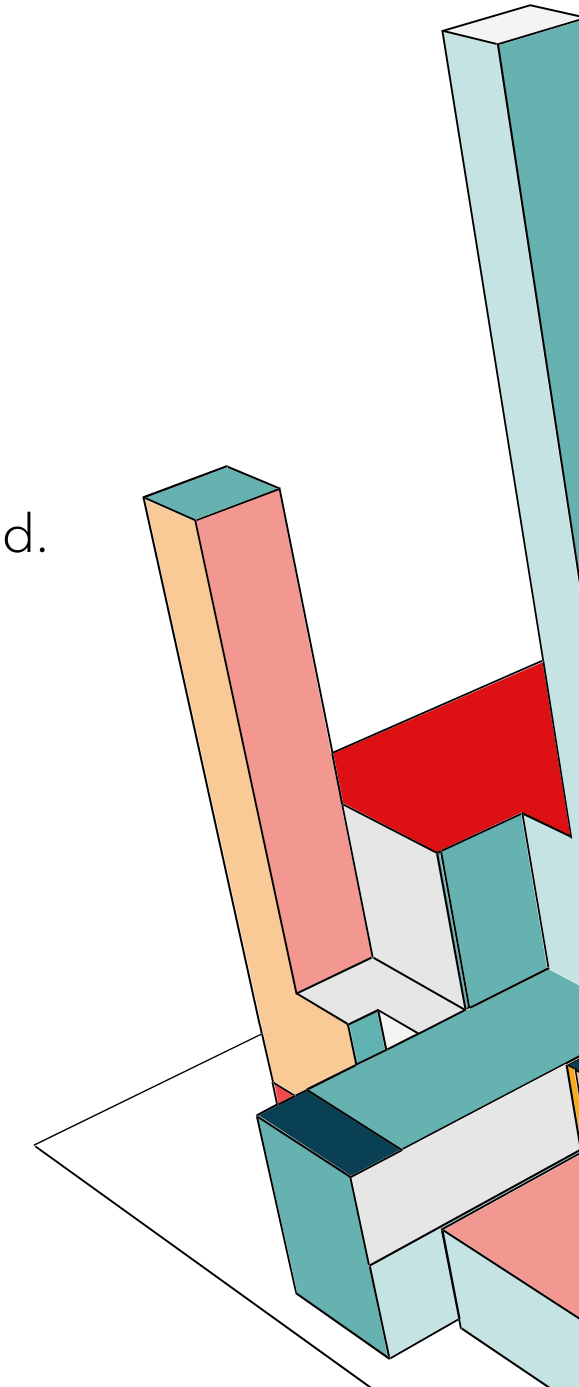
Assistant:
yes, the email demonstrates empathy because…

* Note: All cases in training set will be combined together as few-shot learning examples.

FAIRNESS

- Similar idea as empathy (GPT4 prompt engineer) (worked)
    - Step 1: generate the reason why option 1/option 2 is more preferred.
    - Step 2: create few-shot learning examples with the generated reasons.

- Also tried to create fairness scores with GPT then compare. (failed)



Example prompt

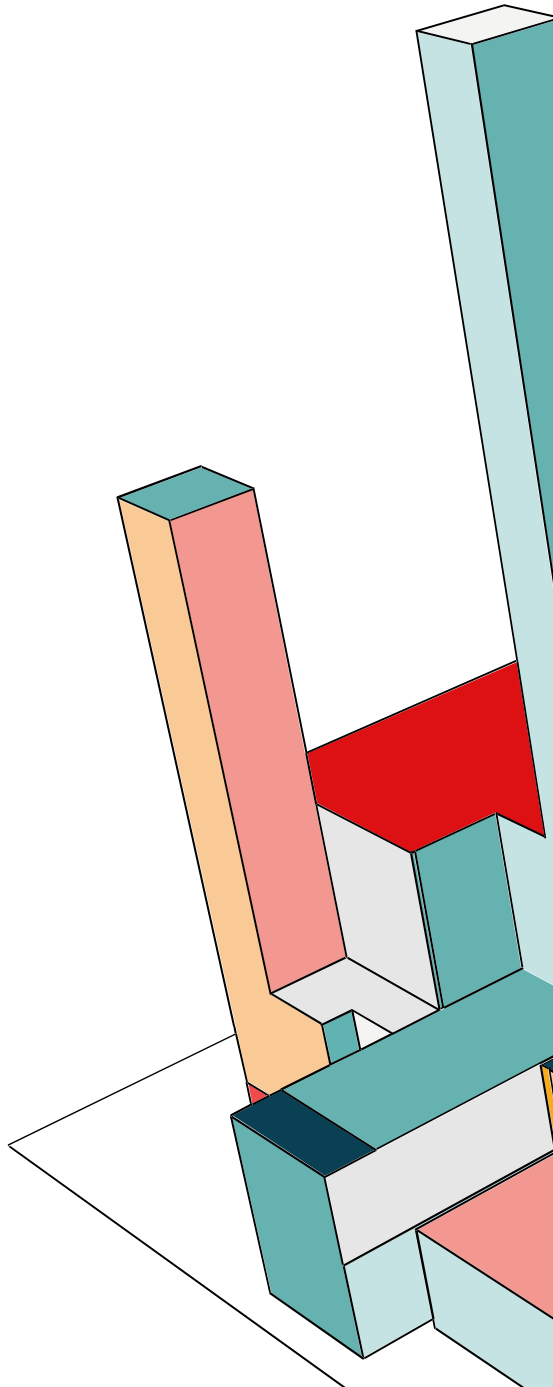
User:
I have two policies.
Policy 1:[{ds_fairness.first_option[i]}].
Policy 2: [{ds_fairness.second_option[i]}]
employees prefer the
{ds_fairness.majority_vote[i]} one. What could be the reasons for that? Provide your answer

Assistant:
Employees might prefer the Diversity policy because….

* Note: All cases in training set will be combined together as few-shot learning examples.

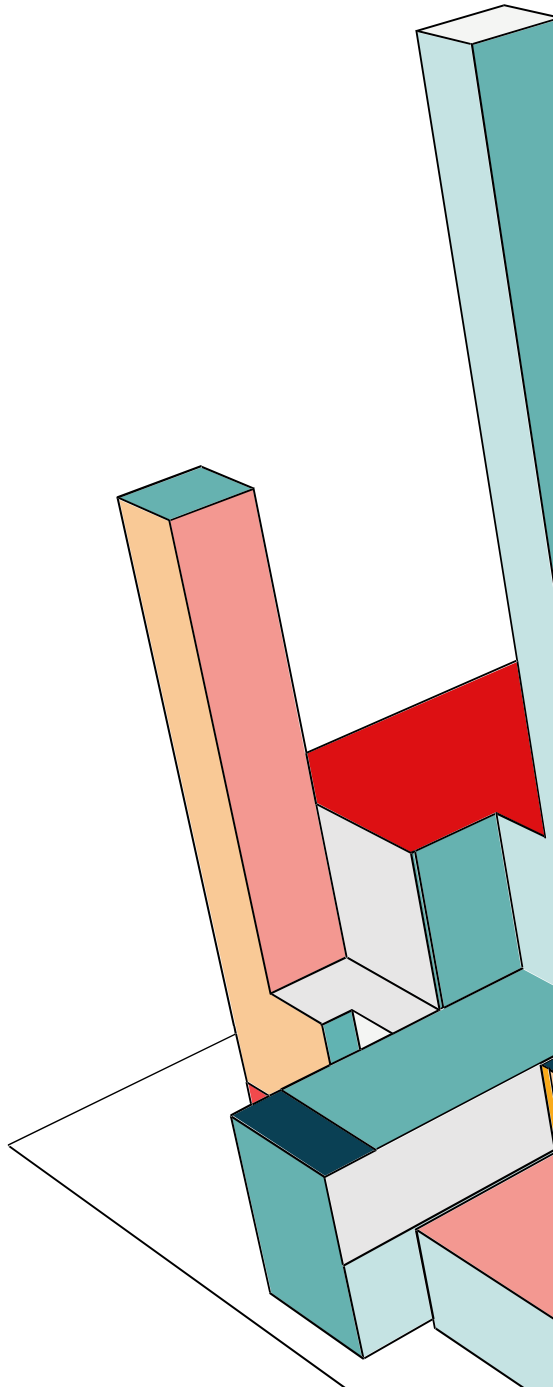
CLARITY

- Fine-tuned DeBERTa-V3-base (worked)
    - Can use R or MSE as loss during training
    - Weight_decay = 0.1
    - Train/validation split on original training set.
- Tried GPT4 (failed)
- Tried fine-tuned GPT3.5 (failed)



INTERVIEW

- GPT 4 – person centered (worked)
- GPT 4 – question centered (worked)
    - Step 1: collect all questions from train/validation/testing set.
    - Step 2: collect all response to each question from all datasets.
    - Step 3: Use step 2 response as few-shot learning examples and generate the response to that question.
- Difficult to predict job candidate's specific experience
    - e.g., I met *a women in business panel conference*



Example prompt

User:
Respond to the last question using no more than 120 words. Keep the tone and personality reflected in the provided previous responses.

Question: {ds_test['Q1'][i]}
Response: {ds_test['R1'][i]}
Question: {ds_test['Q2'][i]}
Response: {ds_test['R2'][i]}
Question: {ds_test['Q3'][i]}
Response: {ds_test['R3'][i]}
Question: {ds_test['Q4'][i]}

Assistant:
In a group project for my final year at university, I was tasked with leading the research component….

DISCUSSION

- Is personality important in the current data?
  - Meh, too hard for the model

- Some takeaways
  - Pay to win (GPT)
  - Well, sometimes we can save some money (open-resource)
    - Avoid overfitting, try more parameter tunning
  - Think about difficulty, sample size before choosing a model.
  - Help the model to think as a human. You are ALWAYS smarter than LLM.

THANK YOU

For questions: zjia2@gmu.edu

My NLP journey all begins with Phil, I wouldn't be here without him

Mina is the best team member

We are looking for a new member(s)!
For details, check out our lab website! (QR)