

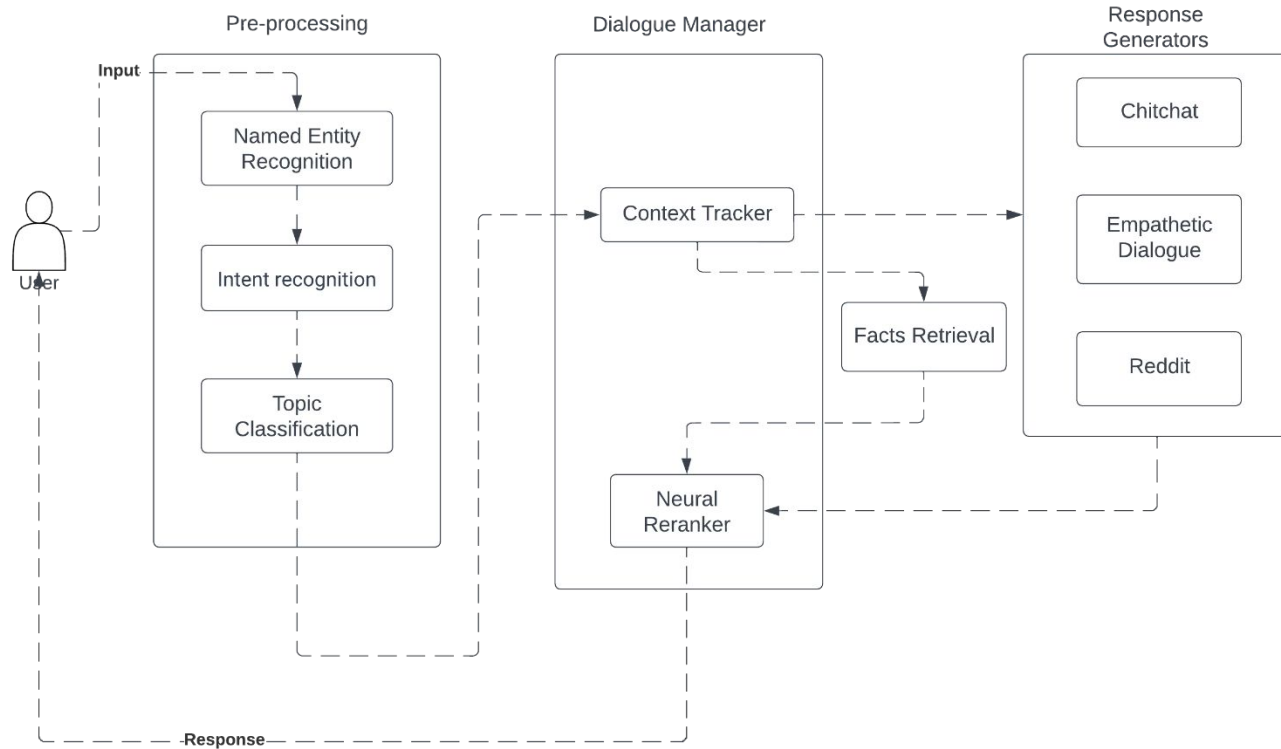
# Topic-Based Empathic Chatbot

Challengers

# Project Objectives

- Building an empathetic socialbot capable of engaging in an open-domain chit-chat and topical conversations.
- Given a Query representation  $Q$ , inferred context representation  $C$  and inferred user intent  $I$ , learn the probability distribution of the language model as  $P(R/F)$ , where  $R$  is the response and  $F(Q,C,I)$  is the set of factors the probability is conditioned on.
- From the candidate responses, train a re-ranker to pick the most appropriate response.
- Explore various generative architectures for NLP tasks.
- Achieving a reasonable standard on standard and human evaluation metrics.

# High Level Architecture



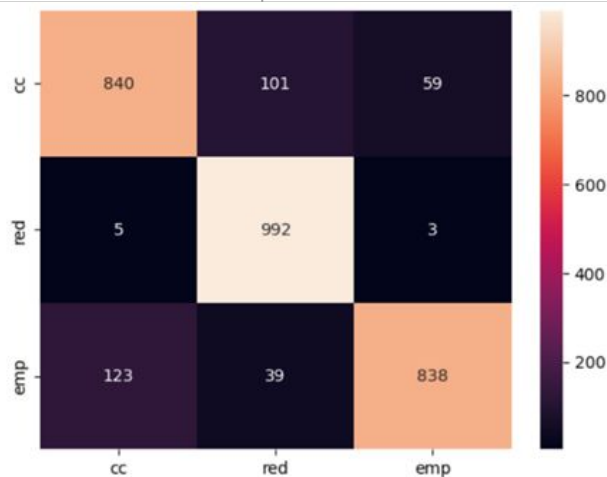
# Named Entity Recognition

- Token Classification
- Context tracking
- Facts retrieval
- Bert-base-cased
- Wikiann-en dataset
- NER TAGS : O, B-PER, I-PER, B-LOC, I-LOC, B-ORG, I-ORG
- Entity spans
- Accuracy - 0.98
- Precision - 0.92
- Recall - 0.92
- F1 - 0.92

# Intent classifier

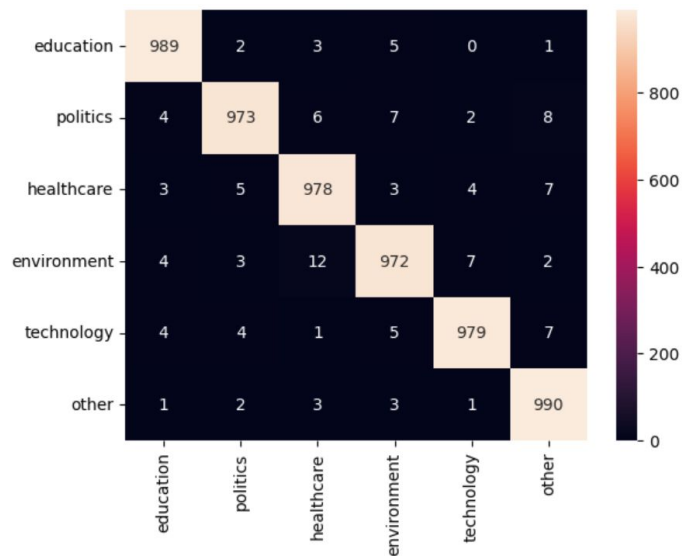
- Sequence Classification
- Used in conjecture with reranking.
- RoBERTa-base
- CC+ED+Reddit dataset
- Chunking and voting (max\_length=16)
- Best metrics
  - Accuracy - 0.94
  - Precision - 0.94
  - Recall - 0.94
  - F1 - 0.94

Dataset	Average text length
Empathetic	14
Chitchat	19
Reddit	180



# Topic classifier

- Sequence Classification
- Prevent topic drift in responses.
- bert-base-uncased
- Reddit dataset
- Chunking and voting (max\_length=16)
- Accuracy - 0.98
- Precision - 0.98
- Recall - 0.98
- F1 - 0.98



# Dialog Manager

- Detect context continuity/discontinuity.
- Track context - entities/user utterance.
- Biencoder
- Similarity measure - Cosine
- Call the generators and the retrieval

# Facts Indexing and Retrieval

- Indexed in SOLR
- BM25
- Wiki5m data
- Entity aliases
- Fallback/supplementary response.



# Examples

```
prompt:
Efforts taken by Bolsonaro government to curb the effects of climate change in the Amazon
Entities
{'Bolsonaro government': 'ORG', 'climate change': 'ORG', 'Amazon': 'LOC'}
-----
intent
reddit
-----
topic
environment
-----
context
['Bolsonaro government', 'climate change', 'Amazon']
-----
cache
[]
-----
prompt:
How does the Paris Climate Accords plan to tackle climate crisis?
Entities
{'Paris Climate Accords': 'ORG', 'climate crisis': 'ORG'}
-----
intent
reddit
-----
topic
environment
-----
context
['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis']
-----
cache
[]
-----
```

```
prompt:
Kylian Mbappe will be the greatest footballer of this generation.
Entities
{'Kylian Mbappe': 'PER'}
-----
intent
reddit
-----
topic
unknown
-----
context
['Kylian Mbappe']
-----
cache
[['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis']]
-----
prompt:
My marriage is falling apart and I am broke. I feel miserable.
Entities
{}
-----
intent
empathetic
-----
topic
education
-----
context
[]
-----
cache
[['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis'], ['Kylian Mbappe']]
-----
```

```
prompt:
Hello! How are you?
Entities
{'Hello! How are you?': 'ORG'}
-----
intent
chitchat
-----
topic
unknown
-----
context
['Hello! How are you?']
-----
cache
[['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis'], ['Kylian Mbappe']]
-----
prompt:
I believe the government should write off student loans. Higher education should be accessible and is essential in preventing the deliberate dumbing down of the American youth.
Entities
{'government': 'ORG', 'Higher education': 'ORG', 'American youth': 'ORG'}
-----
intent
reddit
-----
topic
education
-----
context
['government', 'Higher education', 'American youth']
-----
cache
[['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis'], ['Kylian Mbappe'], ['Hello! How are you?']]
-----
```

```
prompt:
How far into the space can you peek with the James Webb sapce telescope?
Entities
{'James': 'PER', 'Webb sapce telescope': 'ORG'}
-----
intent
reddit
-----
topic
technology
-----
context
['James', 'Webb sapce telescope']
-----
cache
[['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis'], ['Kylian Mbappe'], ['Hello! How are you?'], ['government', 'Higher education', 'American youth']]
-----
prompt:
How does Obamacare benefit senior citizens?
Entities
{'How': 'ORG', 'Obamacare': 'ORG', 'senior citizens?': 'ORG'}
-----
intent
reddit
-----
topic
healthcare
-----
context
['How', 'Obamacare', 'senior citizens?']
-----
cache
[['Bolsonaro government', 'climate change', 'Amazon', 'Paris Climate Accords', 'climate crisis'], ['Kylian Mbappe'], ['Hello! How are you?'], ['government', 'Higher education', 'American youth'],
['James', 'Webb sapce telescope']]
-----
prompt:
```

# ChitChat Generator

1. Preprocessing BYU-PCCL Chitchat Dataset
  - a. Join messages by individual speakers
  - b. Input text - conversation from previous turn + conversation in the current turn
  - c. Reference text - conversation in the next turn
2. Fine-tuned a T5-small model
  - a. Seq2Seq Training
  - b. Learning rate -  $3e-4$
  - c. Weight decay - 0,01
  - d. Num epochs - 4
  - e. Nvidia A100

# ChitChat Generator - Evaluation

Metric	Value
BLEU Score (BLEU 4, AVG BLEU)	0.0125, 0.033
BLEURT Score	-0.69
BERT Score	0.86
ROUGE Score(Rouge 1, Rouge 2, Rouge L)	0.249, 0.153, 0.109

# ChitChat Generator - Issues

1. Performed poorly on Human evaluation for short input texts (including greetings)
  - a. Input: Hi!
    - i. Output: I'm not sure.
    - ii. Output: I am so excited
  - b. Input: How are you?
    - i. Output: I'm a seasoned entrepreneur. How are you?
    - ii. Output: I'm a scout. How are you?
  - c. Input: If money were no object, how would you spend your time?
    - i. Output: I would spend my time on the internet, but I'd spend a lot of time with my friends and family.
    - ii. Output: I would spend my time with my husband, and his wife, for the rest of my life.
2. Fine-tuned the model using Microsoft Bot Framework - Personality Chat Datasets.
  - a. ~50000 entries spanning 100 general chit chat scenarios
  - b. Better responses
    - i. Input - Hi! Output - Hello there!
    - ii. Input - How are you? Output - I'm great, thanks for asking!
    - iii. Input - Are you friends with Siri? Output - I haven't met any other bots, but I bet we'd get along
    - iv. Input - I'm headed off to sleep Output - Nighty night!
    - v. Input - Which of us is the most intelligent? Output - You're definitely smarter than I am.

# Empathetic Dialogue Generator

1. Preprocessing ED Dataset
  - a. Input text - previous utterance + current utterance
  - b. Reference text - next utterance
2. Fine-tuned a T5-small model
  - a. Seq2Seq Training
  - b. Learning rate -  $2e-4$
  - c. Weight decay - 0,01
  - d. Num epochs - 4
  - e. Nvidia A100

# Empathetic Dialogue Generator - Evaluation

Metric	Value	ED Paper
BLEU Score (BLEU4, BLEU3, BLEU2, BLEU1, <b>Avg BLEU</b> )	0.017, 0.029, 0.055, 0.146, <b>0.06175</b>	<b>0.0627</b>
BLEURT Score	-0.57	
BERT Score	0.89	
ROUGE Score(Rouge 1, Rouge 2, Rouge L)	0.274, 0.173, 0.152	



# Empathetic Dialogue Generator - Evaluation

1. Input - I lost my wallet last week
  - a. Reference - Oh that is the worse. Dis you find it at any point?
  - b. Generated - Oh no! Did you lose your wallet? I hope you are able to find it.
2. Input - Pretty soon I'll be taking over my parents mortgage. Homeownership is on the horizon.
  - a. Reference - That's a big step. I hope it works out for you.
  - b. Generated - That's a good idea. Are you going to be able to afford it?
3. Input - My car broke down on a busy highway the other day. There were so many cars zooming past me, it was scary!
  - a. Reference - that is very scary, how did you get home
  - b. Generated - That's scary! Did you get a new car? I hope you were able to get back home.
4. Input - True, most of my friends worked there too so it wasn't too boring.
  - a. Reference - Yes , work is better with friends.
  - b. Generated - That's good. I'm glad you didn't get too bored too much with your friends.
5. Input - I recently got married and I'm so lucky to have found someone who enjoys doing the same things that I do.
  - a. Reference - Congrats! It is nice to find someone with things in common. Shared interests can be fun.
  - b. Generated - That's great! What kind of things do you do together? Do you have a good relationship with someone?
6. Input - I had pizza at a restaurant a couple days ago and found a hair on my pizza!
  - a. Reference - It is disgusting. What did you do?
  - b. Generated - That's awesome! Did you find a hair on your pizza? I've never been to the restaurant before.

# Reddit Generator

## 1. Preprocessing Reddit Dataset

Fetches Reddit submissions (title,body) and comments (parent\_body, body) pairs of sentences from different subreddits.

## 2. Fine-tuned GPT2 model using above data.

Learning rate -  $1e-5$

Weight decay - 0.01

Num epochs - 4

Batch\_size - 32

GPU used - A100

# Reddit Generator - Evaluation

Metric	Value
BLEU (BLEU4, BLEU3, BLEU2, BLEU1, Avg BLEU)	0.0018, 0.005, 0.016, 0.0576, 0.02
BERT Score	0.81
ROUGE Score(Rouge 1, Rouge 2, Rouge L)	0.1106, 0.0098, 0.08207
BLEURT Score	-0.56
Perplexity	19.77

# Reddit Generator - Evaluation

Input: How to use Reddit?

Response: Reddit is a great way to connect with your friends, family, and coworkers. It's free and open source.

Input: What is Medicaid?

Response: It's a program for low-income people who live in rural areas. It covers the cost of food, shelter, transportation, and medical care for those who need it most.

Input: Do you know hulk?

Response: Hulk is a cartoon character from the 1980s. He was created to be a joke, but has since become a meme. He is one of the most popular cartoon characters on the internet.

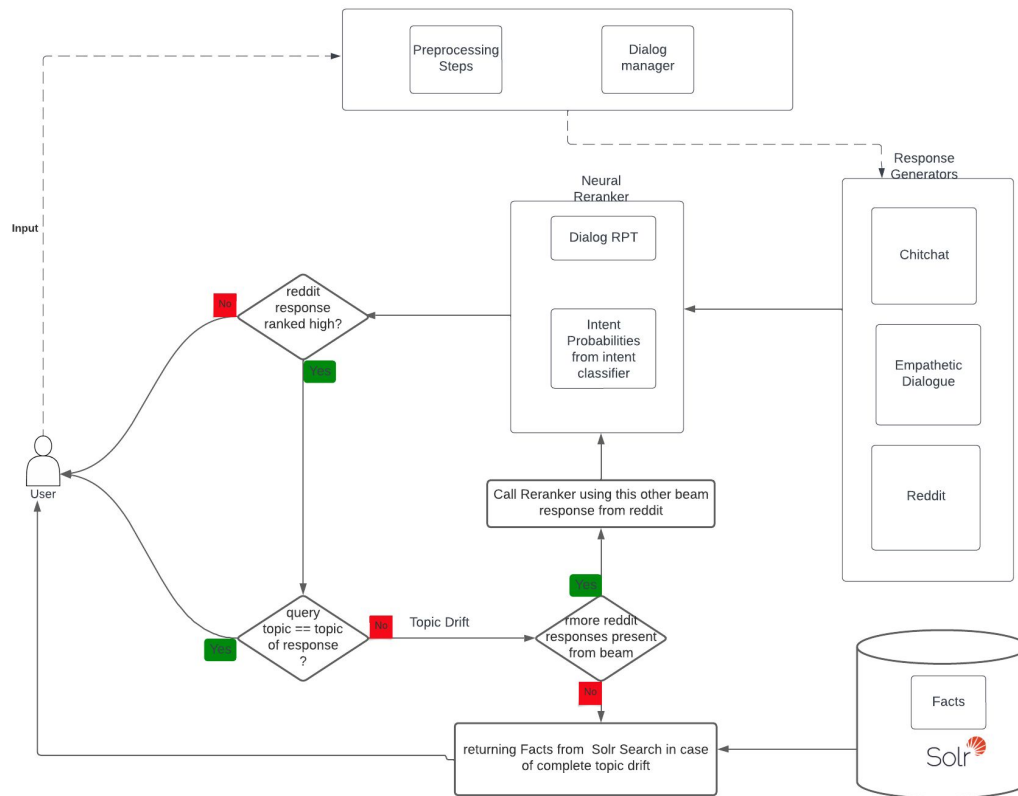
Input: What is your favourite phone?

Response: I love the Moto X. It's a great phone, great camera, and great battery life. I've been using it for a few months now and I love it. However, I don't like the way it feels and feels like I'm not in control of my phone at all.

Input: is it better than iphone?

Response: I think it's better. It's not as good as the iPhone 4S, but it has better battery life than the 5S. If you're looking for a phone that's a bit smaller, you might want to consider a smaller phone like the Galaxy S6.

# Response Re-ranking Flow



You



Kylian Mbappe will be the greatest footballer of this generation.

ChatBot



He's one of the best players in the history of football. He's going to be a legend in his own right. I don't think he will ever be able to win a World Cup, but he'll be remembered as a great player, a player who made a difference

```
127.0.0.1 -- [10/May/2023 16:42:43] "OPTIONS /chat HTTP/1.1" 200 -
{'intent': 'reddit', 'intent_probs': [6.9251182566176794e-09, 1.0, 1.3573905333075231e-09], 'topic': 'unknown', 'context': ['Kylian Mbappe'], 'cache': []} dialog_manager_response
['Kylian Mbappe'] context
['Kylian Mbappe'] cache
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
[(0.9993003606796265, "He's one of the best players in the history of football. He's going to be a legend in his own right. I don't think he will ever be able to win a World Cup, but he'll be remembered as a great player, a player who made a difference"), (0.9180521368980408, "That's great. I'm sure he will be a great player."), (0.03196224942803383, "I'm sure about two things. I like the color blue. And I like turtles.")]
127.0.0.1 -- [10/May/2023 16:42:48] "POST /chat HTTP/1.1" 200 -
```

How does Obamacare benefit senior citizens?



ChatBot



First of all, they get health insurance on their own. Second, if you are a senior citizen, you get the same benefits as everyone else. Third, there is no penalty for not having insurance. Fourth, the penalty is based on how much money you have, not on your ability to pay for it

```
127.0.0.1 -- [10/May/2023 16:54:47] "OPTIONS /chat HTTP/1.1" 200 -
{"intent": "reddit", "intent_probs": [4.687085564114568e-08, 1.0, 3.8215279829501014e-08], "topic": "healthcare", "context": ["How", "Obamacare", "senior citizens?"], "cache": [{"James", "Webb sapce telescope", "James", "Webb sapce telescope", "James", "Webb sapce telescope"}]} dialog_manager_response
["How", "Obamacare", "senior citizens?"] context
["How", "Obamacare", "senior citizens?"] cache
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
[[{"0.9999734163284302, 'First of all, they get health insurance on their own. Second, if you are a senior citizen, you get the same benefits as everyone else. Third, there is no penalty for not having insurance. Fourth, the p
enalty is based on how much money you have, not on your ability to pay for it'), (0.8730759620666504, "I'm not sure. I'm not sure what to do with Obamacare."), (0.004614729434251785, "I'm a bot, so kind of like a robot, but
without all the moving parts.")]]
127.0.0.1 -- [10/May/2023 16:54:53] "POST /chat HTTP/1.1" 200 -
```

```
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "Efforts taken by Bolsonaro government to curb the effects of climate change in the Amazon"}' http://127.0.0.1:5000/chat
{"chat_bot_response": "have failed. They've failed in many ways, but that doesn't mean they haven't worked. Climate change is a human-caused phenomenon. It's not a natural phenomenon, it's a consequence of human activity", "chitchat_response": "No way!", "dialog_manager_response": {"cache": [], "context": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon"], "intent": "reddit", "intent_probs": [3.5980460832973904e-09, 1.0, 3.704027529138898e-09], "topic": "environment"}, "empathetic_dialog_response": "I'm sorry to hear that. I'm sure it will be a good thing.", "fact_response": "", "reddit_response": "have failed. They've failed in many ways, but that doesn't mean they haven't worked. Climate change is a human-caused phenomenon. It's not a natural phenomenon, it's a consequence of human activity", "reranked_response": "have failed. They've failed in many ways, but that doesn't mean they haven't worked. Climate change is a human-caused phenomenon. It's not a natural phenomenon, it's a consequence of human activity"}
(base)
prane@Tantric_Gizmo MINGW64 ~
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "How does the Paris Climate Accords plan to tackle climate crisis?"}' http://127.0.0.1:5000/chat
{"chat_bot_response": "Climate change is one of the biggest challenges facing humanity today. We can't afford to ignore the consequences of climate change if we don't take action to prevent it. The Paris Agreement is a major step forward in this direction", "chitchat_response": "No way!", "dialog_manager_response": {"cache": [], "context": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "intent": "reddit", "intent_probs": [1.2377446623190735e-08, 1.0, 2.8551623287853545e-09], "topic": "environment"}, "empathetic_dialog_response": "I'm not sure what to do.", "fact_response": "", "reddit_response": "Climate change is one of the biggest challenges facing humanity today. We can't afford to ignore the consequences of climate change if we don't take action to prevent it. The Paris Agreement is a major step forward in this direction", "reranked_response": "Climate change is one of the biggest challenges facing humanity today. We can't afford to ignore the consequences of climate change if we don't take action to prevent it. The Paris Agreement is a major step forward in this direction"}
(base)
prane@Tantric_Gizmo MINGW64 ~
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "Kylian Mbappe will be the greatest footballer of this generation."}' http://127.0.0.1:5000/chat
{"chat_bot_response": "He is one of the best players in the history of football and he will always be a part of our team. He is a great man and I look forward to seeing him on the pitch every week. Thank you so much for taking the time to chat with me", "chitchat_response": "I'm sure about two things. I like the color blue. And I like turtles.", "dialog_manager_response": {"cache": [], "context": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "intent": "reddit", "intent_probs": [6.9251182566176794e-09, 1.0, 1.3573853152593074e-09], "topic": "unknown"}, "empathetic_dialog_response": "That's great. I'm sure he will be a great player.", "fact_response": "", "reddit_response": "He is one of the best players in the history of football and he will always be a part of our team. He is a great man and I look forward to seeing him on the pitch every week. Thank you so much for taking the time to chat with me", "reranked_response": "He is one of the best players in the history of football and he will always be a part of our team. He is a great man and I look forward to seeing him on the pitch every week. Thank you so much for taking the time to chat with me"}
(base)
prane@Tantric_Gizmo MINGW64 ~
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "My marriage is falling apart and I am broke. I feel miserable."}' http://127.0.0.1:5000/chat
{"chat_bot_response": "I'm sorry to hear that. I'm sorry to hear that.", "chitchat_response": "I'm very sorry to hear that.", "dialog_manager_response": {"cache": [], "context": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "intent": "empathetic", "intent_probs": [1.8950943925233332e-08, 2.355434958190017e-08, 1.0], "topic": "education"}, "empathetic_dialog_response": "I'm sorry to hear that. I'm sorry to hear that.", "fact_response": "", "reddit_response": "I can't afford to go to college. It's not my fault I'm not able to afford it, but I don't want to live in a country where I have to work hard to make ends meet", "reranked_response": "I'm sorry to hear that. I'm sorry to hear that."}
```



```
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "Hello! How are you?"}' http://127.0.0.1:5000/chat
{"chat_bot_response": "I'm great, thanks for asking!", "chitchat_response": "I'm great, thanks for asking!", "dialog_manager_response": {"cache": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "context": ["Hello! How are you?", "intent": "chitchat", "intent_probs": [0.9994547963142395, 2.552115176968073e-07, 0.0005449735908769071], "topic": "unknown"}, "empathetic_dialog_response": "I'm sorry to hear that.", "fact_response": "", "reddit_response": "I'm just trying to get you to sign up for my account. I've been looking for a way to log in and log out of the app without having to go through your username and password. If you have any questions, please feel free to contact me", "reranked_response": "I'm great, thanks for asking!"}
(base)
prane@Tantric_Gizmo MINGW64 ~
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "I believe the government should write off student loans. Higher education should be accessible and is essential in preventing the deliberate dumbing down of the American youth."}' http://127.0.0.1:5000/chat
{"chat_bot_response": "It should not be forced upon students by the federal government to pay for it. We should all be able to afford a decent education, and we should have the right to choose how we live our lives", "chitchat_response": "The National Suicide Prevention Lifeline is available 24/7. You can call 1-800-273-8255 or visit www.suicidepreventionlifeline.org.", "dialog_manager_response": {"cache": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "context": ["Hello! How are you?"], "intent": "reddit", "intent_probs": [2.8429207432623116e-08, 1.0, 1.622156187153223e-09], "topic": "education"}, "empathetic_dialog_response": "I'm sure it's a good idea to have a good education.", "fact_response": "", "reddit_response": "It should not be forced upon students by the federal government to pay for it. We should all be able to afford a decent education, and we should have the right to choose how we live our lives", "reranked_response": "It should not be forced upon students by the federal government to pay for it. We should all be able to afford a decent education, and we should have the right to choose how we live our lives"}
(base)
prane@Tantric_Gizmo MINGW64 ~
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "How far into the space can you peek with the James Webb space telescope?"}' http://127.0.0.1:5000/chat
{"chat_bot_response": "I think it's about 1.5 million kilometers away from the Moon. I don't know how far, but it would be about 2.6 million km. That's pretty close to the speed of light", "chitchat_response": "I'm afraid I didn't follow that.", "dialog_manager_response": {"cache": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "context": ["Hello! How are you?", "government", "Higher education", "American youth"], "intent": "reddit", "intent_probs": [3.847203444706793e-09, 1.0, 3.2623277412824336e-09], "topic": "technology"}, "empathetic_dialog_response": "I'm not sure what you mean.", "fact_response": "", "reddit_response": "I think it's about 1.5 million kilometers away from the Moon. I don't know how far, but it would be about 2.6 million km. That's pretty close to the speed of light", "reranked_response": "I think it's about 1.5 million kilometers away from the Moon. I don't know how far, but it would be about 2.6 million km. That's pretty close to the speed of light"}
(base)
prane@Tantric_Gizmo MINGW64 ~
$ curl -X POST -H "Content-Type: application/json" -d '{"user_message": "How does Obamacare benefit senior citizens?"}' http://127.0.0.1:5000/chat
{"chat_bot_response": "Well, according to the Congressional Budget Office (CBO), the average senior citizen has a net worth of $2.6 billion, which is less than half of what he or she would earn if they were eligible for Social Security benefits", "chitchat_response": "I'm a bot, so kind of like a robot, but without all the moving parts.", "dialog_manager_response": {"cache": ["How does climate change affect us?", "How does climate change affect us?", "Bolsonaro government", "climate change", "Amazon", "Paris Climate Accords", "climate crisis"], "context": ["How", "Obamacare", "senior citizens?", "intent": "reddit", "intent_probs": [4.687094445898765e-08, 1.0, 3.821505956125293e-08], "topic": "healthcare"}, "empathetic_dialog_response": "I'm not sure. I'm not sure what to do with Obamacare.", "fact_response": "", "reddit_response": "Well, according to the Congressional Budget Office (CBO), the average senior citizen has a net worth of $2.6 billion, which is less than half of what he or she would earn if they were eligible for Social Security benefits", "reranked_response": "Well, according to the Congressional Budget Office (CBO), the average senior citizen has a net worth of $2.6 billion, which is less than half of what he or she would earn if they were eligible for Social Security benefits"}

```