

NLP LAB 2021 - VDA

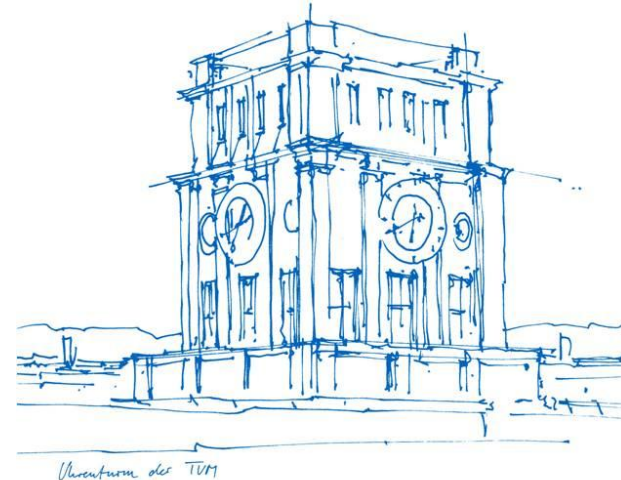
Ananta Bhattarai, Sophie Schoen, Viviane Rehor

Technische Universität München

Department of Informatics

Research Group Social Computing

Munich, May 25 2021



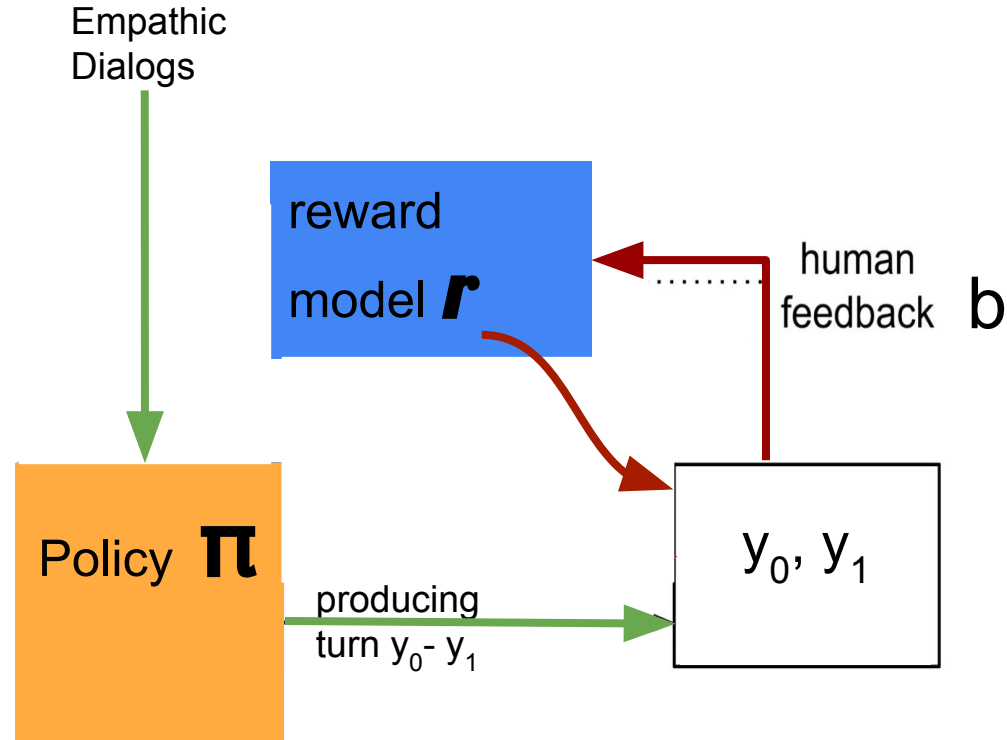
Our Vision - Part 1

1. Creating a Baseline: Fine-Tuning GPT-2 with Empathic Dialogues datasets

- a. Data Preprocessing (Encoding, Decoding & Importing Data) ✓
 - b. Fine-Tuning to get π ✓

2. Producing turns $y_0 - y_1$ to give human feedback on
 - a. With test/validation set of Empathic Dialogs dataset

3. Train reward model r with those turns and human feedback b

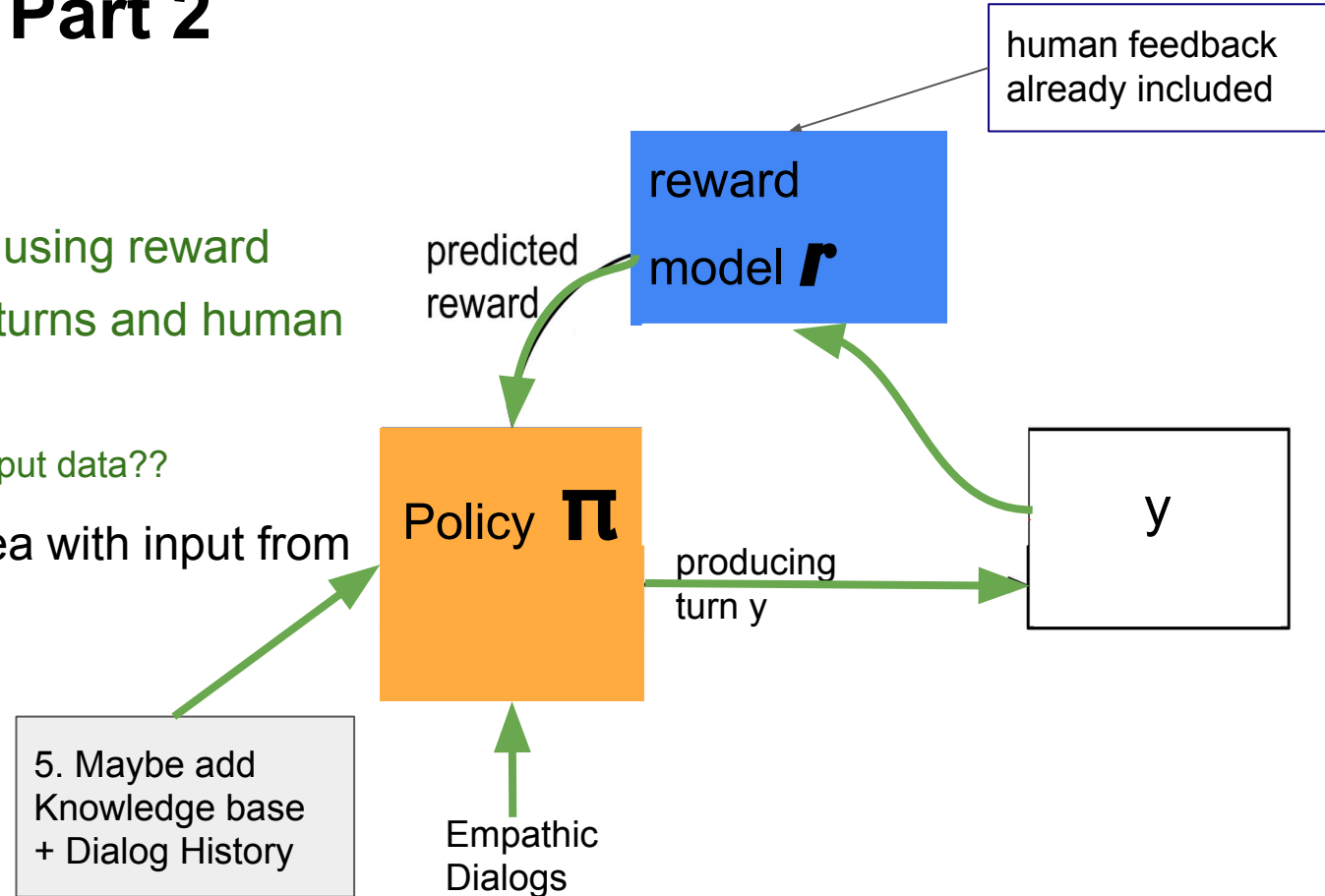


Our Vision - Part 2

4. Train policy π using reward model r with those turns and human feedback b

a. With which input data??

5. Improve the idea with input from other papers



Sampling Responses

- Use validation set of the Empathetic dialogues dataset to create a context.

A: Hello

B: Hi there, How are you?

A: I am having a bad day.

B: What happened?

A: I was on diet but I ate pizza today

B: Don't worry. Think of it as your cheating day.



1. Hello </s> (Context 1)

2. Hello <SOC> Hi there, How are you? </s> (Context 2)

3. Hello <SOC> Hi there, How are you? <SOC> I am having a bad day </s> (Context 3)

...



Sample 2 responses to each context

Sampling Responses

A: I am really looking forward to this season of Lakers, with LeBron coming in

B: Oh wow, I didn't even know about the transfer. I bet lakers fan are delighted.

C: They are on cloud 9! With such a legend on their team, who wouldn't be, right?

GPT-2 355M and Temperature = 0.7

sample0: I think so. The fact that people are excited is great. I hear a lot of stories about the move.

sample1: Wow, it's going to be a blast! I am sure you will be back!

GPT-2 355M and Temperature = 0.5

sample0: I bet many of them.

sample1: That is a great sentiment. I am sure that many of them are.

GPT-2 124M and Temperature = 0.7

sample0: I am very confident of the team, and I am sure they will make a good team.

sample1: I am really looking forward to it. I believe they will be the best team in the league.

GPT-2 124M and Temperature = 0.5

sample0: I am sure they will be, I bet Lakers are going to be a great team.

sample1: I think they are the best. I think they are the best team in the league.

How do we label?

- Empathy
- Relevance (on topic?)
- Fluency (accurate grammatically and content-wise)
- Continuation (contributes to continue talking)
- What if the answer consists of two parts, one good and one bad one?
 - -> Remove the bad one!

Labelling Script

Input:

```
Annotator sophie: [Sample 305 of 2030]
```

```
A: I had a big fishing trip planned for this past weekend, but it got rained out.
```

```
B: Will you get to make it up?
```

```
Which next turn is better?
```

```
-----  
Dialog 1
```

```
Yeah I can't wait for the boat to go out on the lake.
```

```
Dialog 2
```

```
I am sure I will! I hope I won't. I am so glad I did!
```

Options for Interaction: (typing)

- 1 / 2 : Left or right sample is better
- 3 : Both are good, so random choice
- 4 : Both samples do not fit → kick them out
- 5 / 6 : Copy part of left or right sample to console

Do we agree?

Test run with 10 samples:

	Ananta	Vivi	Sophie
Ananta	0	5	6
Vivi	5	0	3
Sophie	6	3	0

Table shows the confusion matrix of agreement between 3 annotators. For example, (Ananta, Vivi) shows how many responses were labelled same by Ananta and Vivi out of 10 samples.

All agreement: 20%

All different: 0%

Ananta and Vivi: 50%

Ananta and Sophie: 60%

Vivi and Sophie: 30%

Do we agree?

Test run with 20 samples:

	Ananta	Vivi	Sophie
Ananta	0	11	12
Vivi	11	0	9
Sophie	12	9	0

Table shows the confusion matrix of agreement between 3 annotators. For example, (Ananta, Vivi) shows how many responses were labelled same by Ananta and Vivi out of 20 samples.

All agreement: 35%

All different: 10%

Ananta and Vivi: 55%

Ananta and Sophie: 60%

Vivi and Sophie: 45%

Do we agree?

Test run with 30 samples:

	Ananta	Vivi	Sophie
Ananta	0	17	18
Vivi	17	0	18
Sophie	18	18	0

Table shows the confusion matrix of agreement between 3 annotators. For example, (Ananta, Vivi) shows how many responses were labelled same by Ananta and Vivi out of 30 samples.

All agreement: 43.33%

All different: 10%

Ananta and Vivi: 56.66%

Ananta and Sophie: 60%

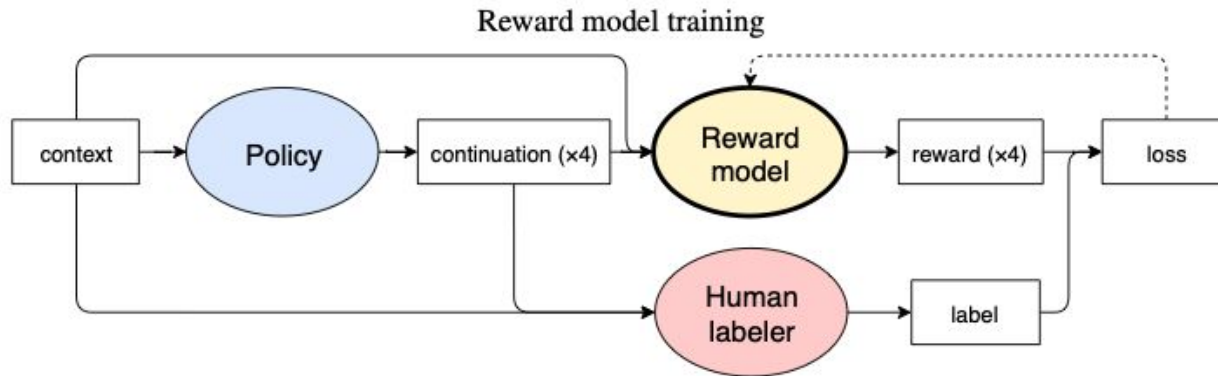
Vivi and Sophie: 60%

How much did we label so far?

- Ananta: 500 samples (~ 45 sec/sample)
- Sophie: 500 sample (~ 40 sec/sample)
- Vivi: 300 sample (~ 40 sec/sample)

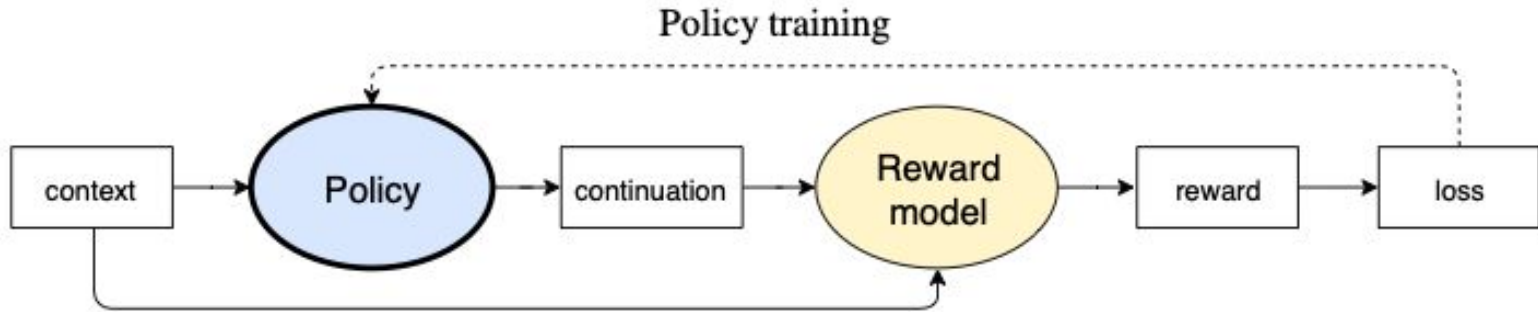
Next Steps

- Training the Reward Model from our human feedback
→ we teach the model to score the responses



Next Steps

- Fine Tuning reward model by optimizing policy → Minimization of KL divergence



Create an automatic evaluation method

- Start with the Empathetic Dialog Paper ideas