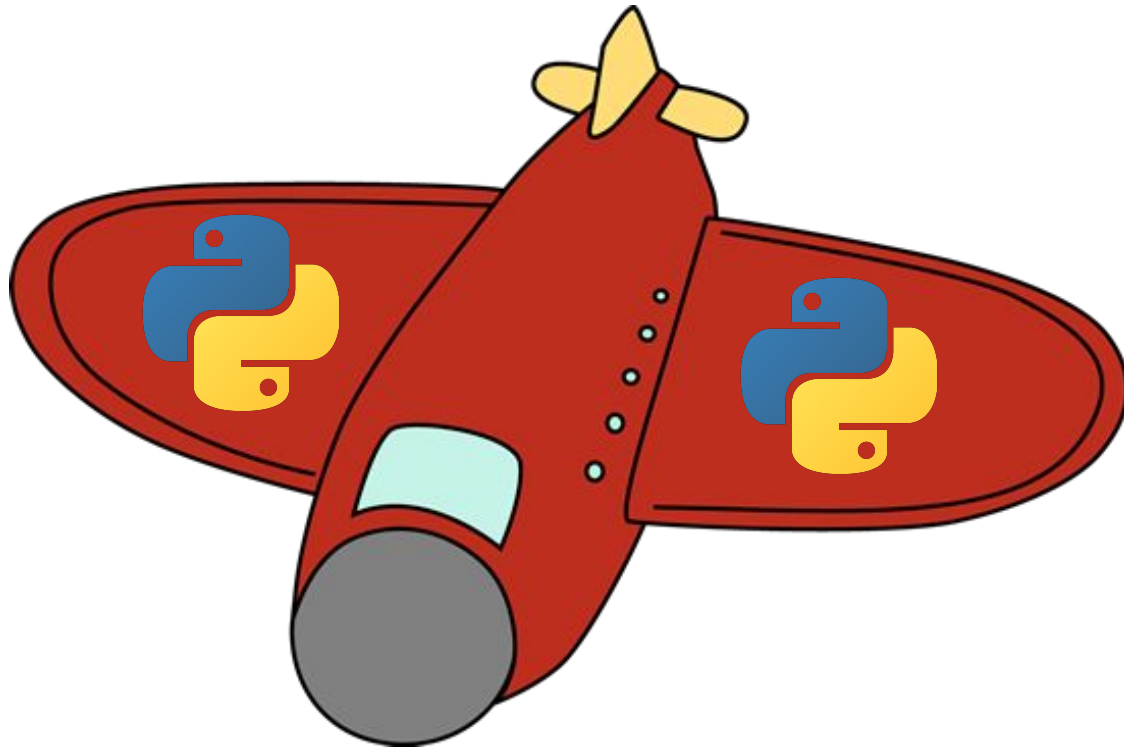




Representation is King: **The Journey to Quality Dialog Embeddings**

Adam Zika from Salted CX
EuroPython 24, Prague

Use-case: **PythonAir**



Use-case: **PythonAir**



George

Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

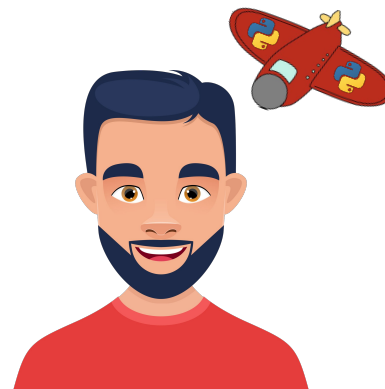
Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

Ok, please proceed.

Ticket is booked.

Thank you for your assistance.

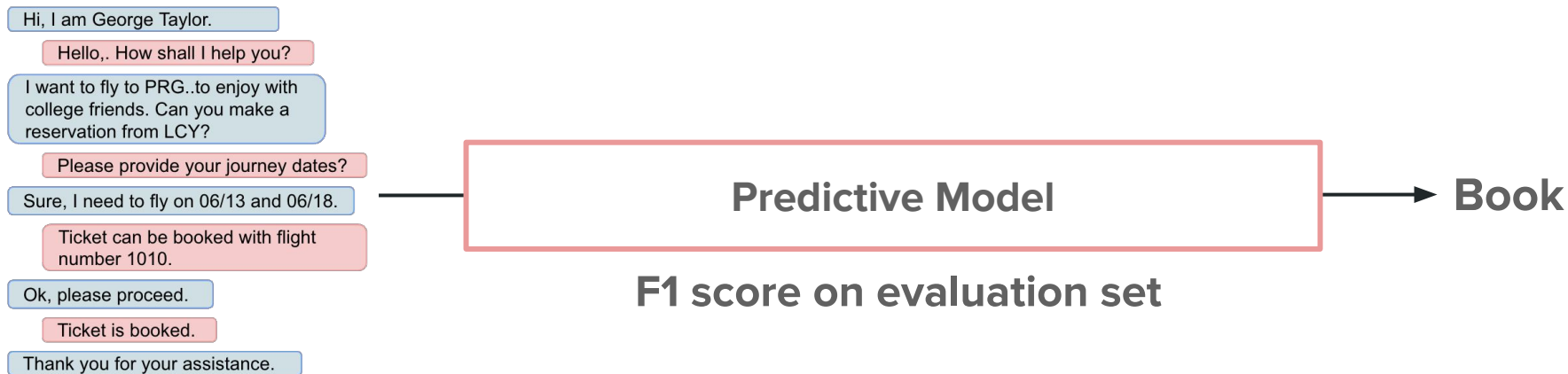


Ben

Book

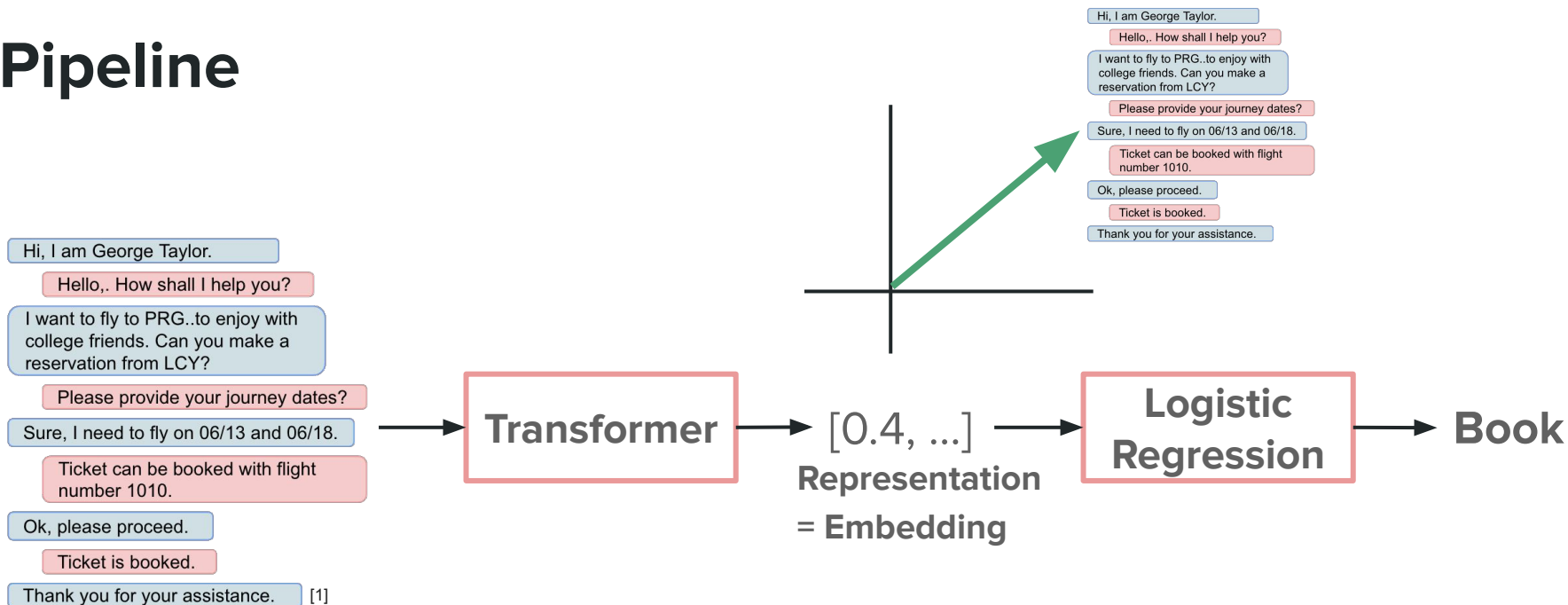
Cancel, Change, No flight, No reservation

Problem



- **Labelled:** 10 dialogs / action
- **Not labelled:** 15K dialogs

Pipeline



Transformer

Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

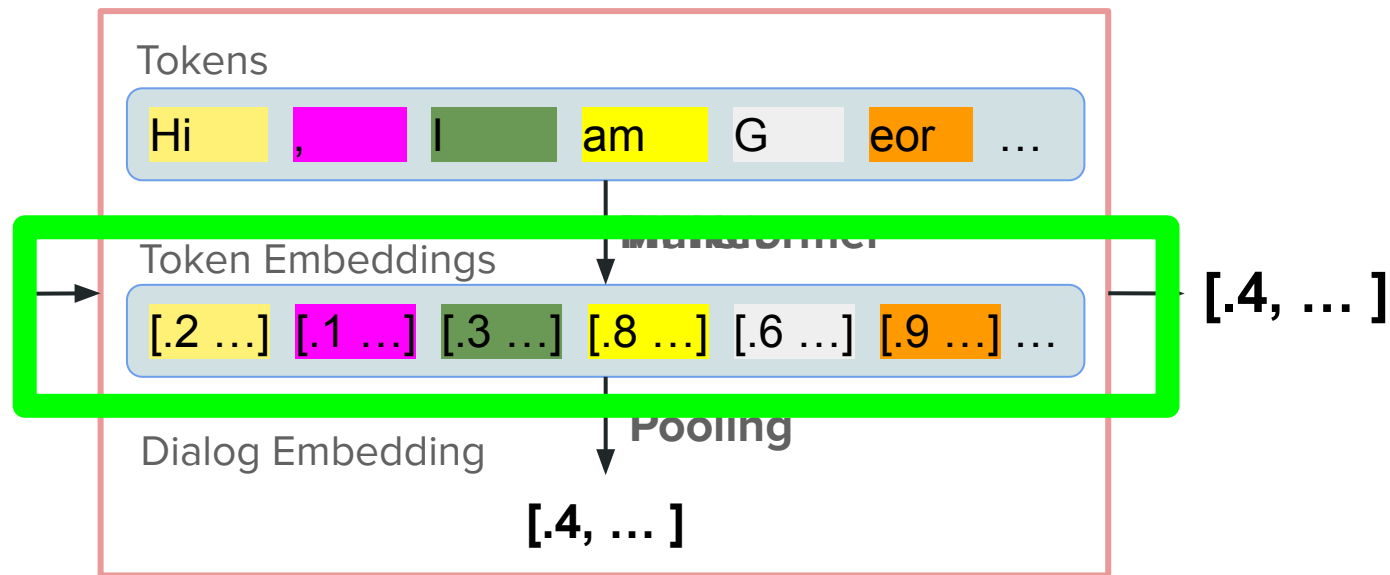
Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

Ok, please proceed.

Ticket is booked.

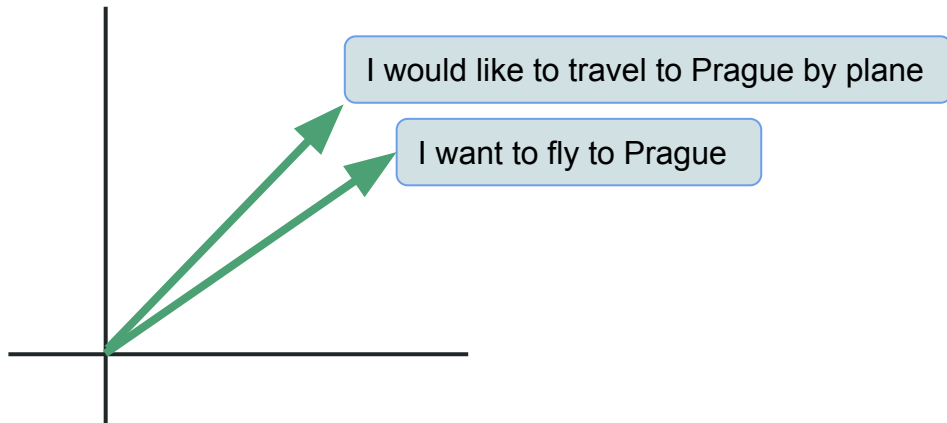
Thank you for your assistance.



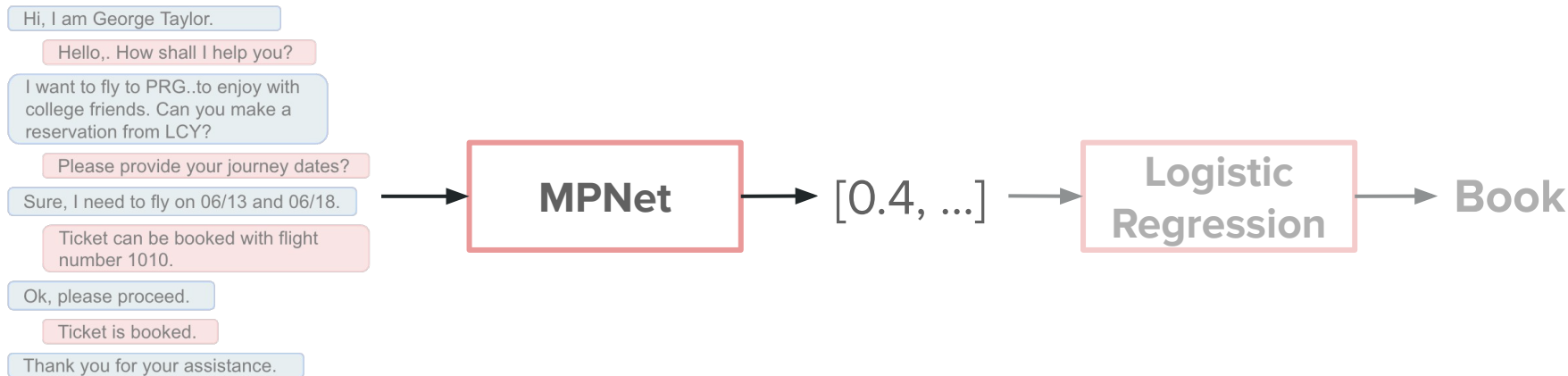
Masked and Permuted Language Modelling

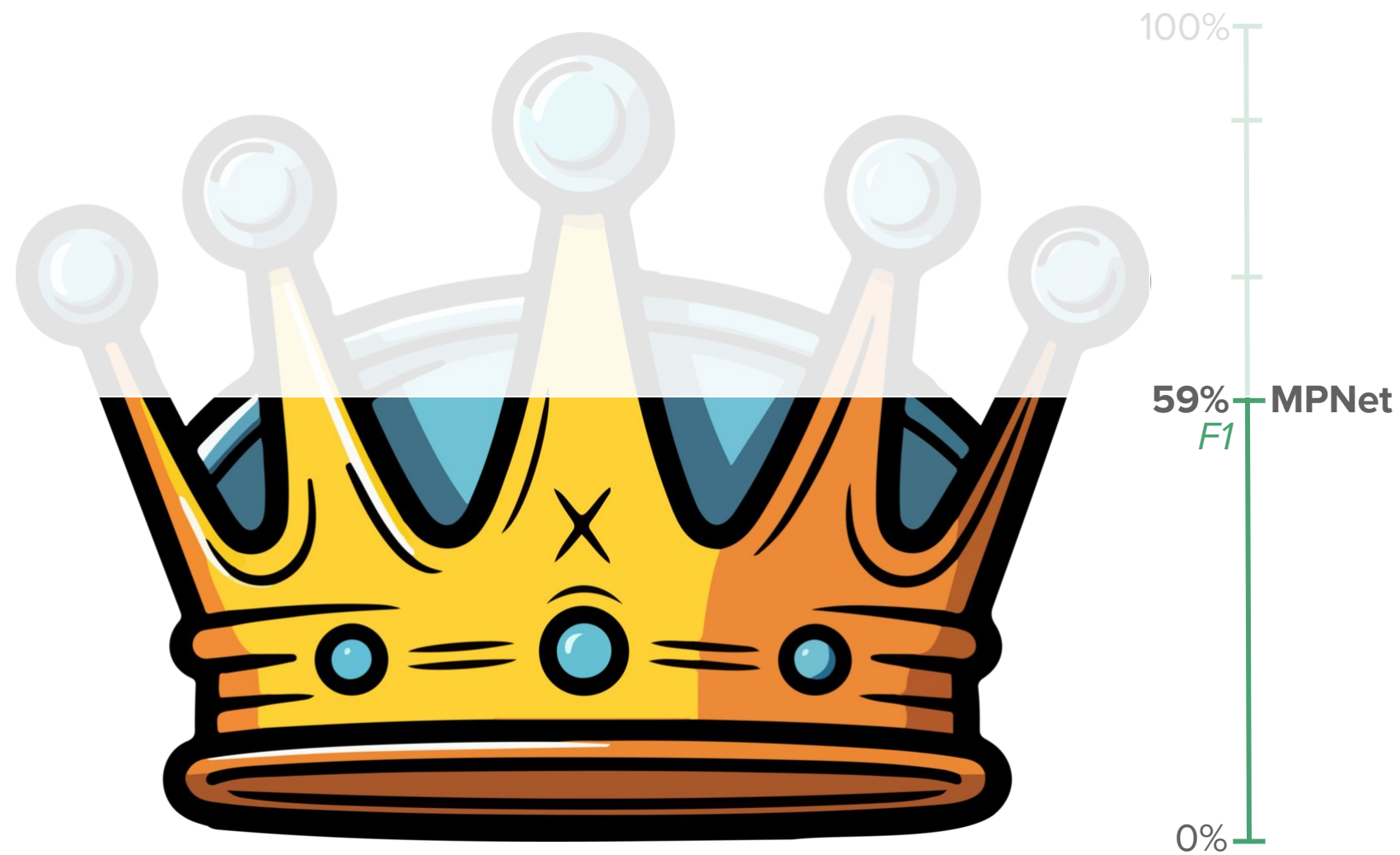
Code

```
1 from sentence_transformers import ...
2
3 model = SentenceTransformer(
4     modules=[
5         Transformer("microsoft/mpnet-base"),
6         Pooling(
7             word_embedding_dimension=768,
8             pooling_mode_mean_tokens=True,
9         )
10    ]
11 )
12 model.encode("I want to fly to Prague")
```



Pipeline





Transformer

Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

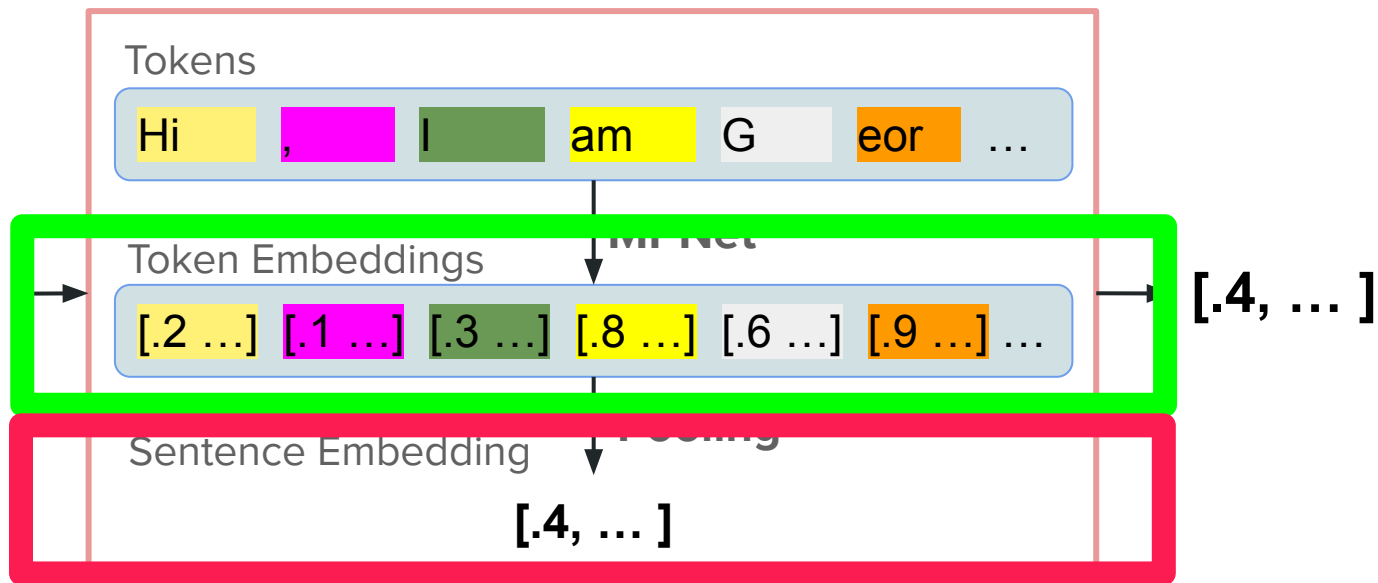
Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

Ok, please proceed.

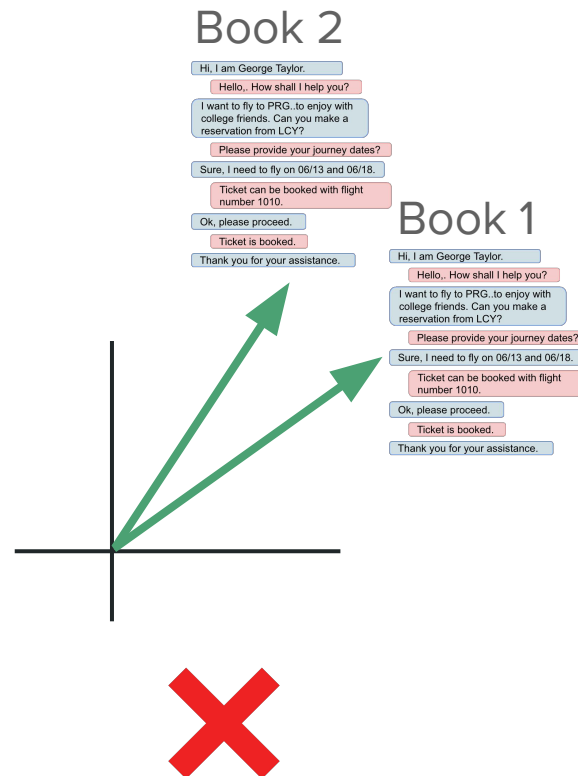
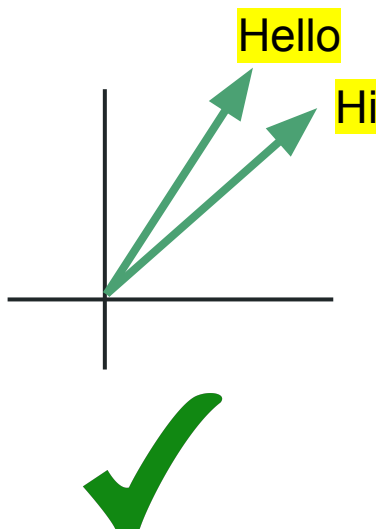
Ticket is booked.

Thank you for your assistance.



Masked and Permuted Language Modelling

Token vs Sentence Representations



MPNet

	Cosine Similarity
I want to fly to Prague.	
I would like to travel to Prague by plane.	0.984
Sparta Prague is the best.	0.895
I don't want to fly to Prague.	★ 0.985



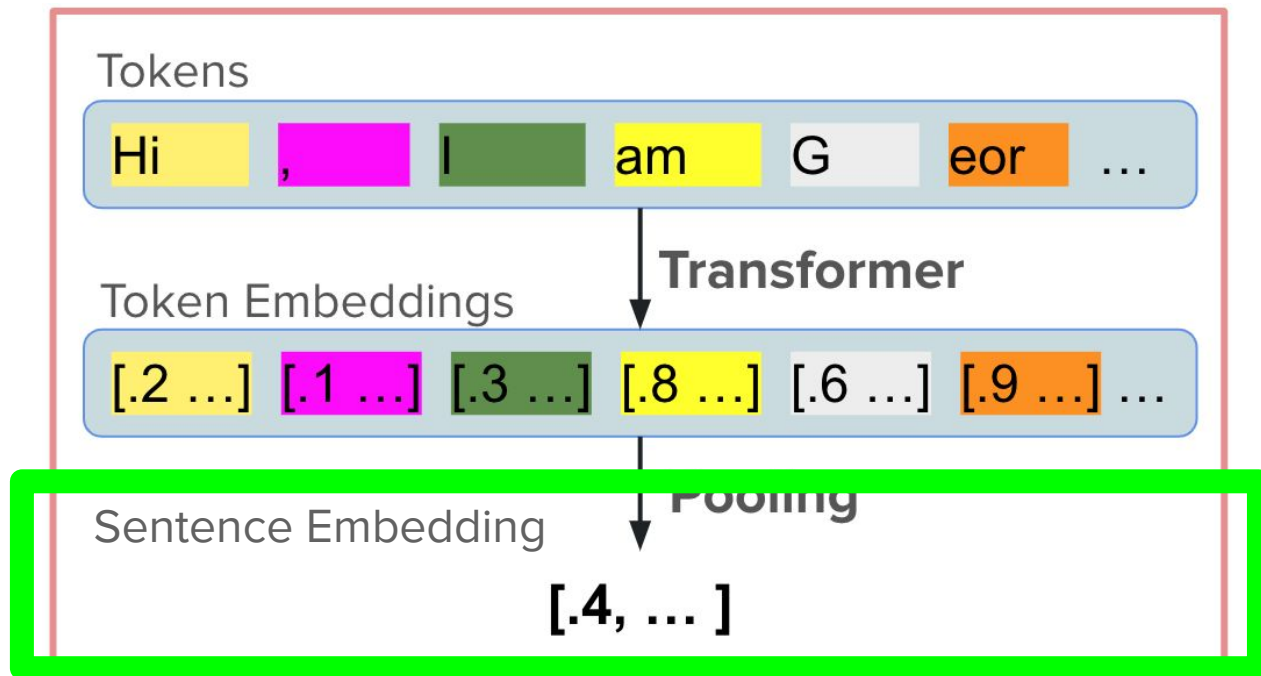
Not representing sentences very well

Suitable for **sentences**?

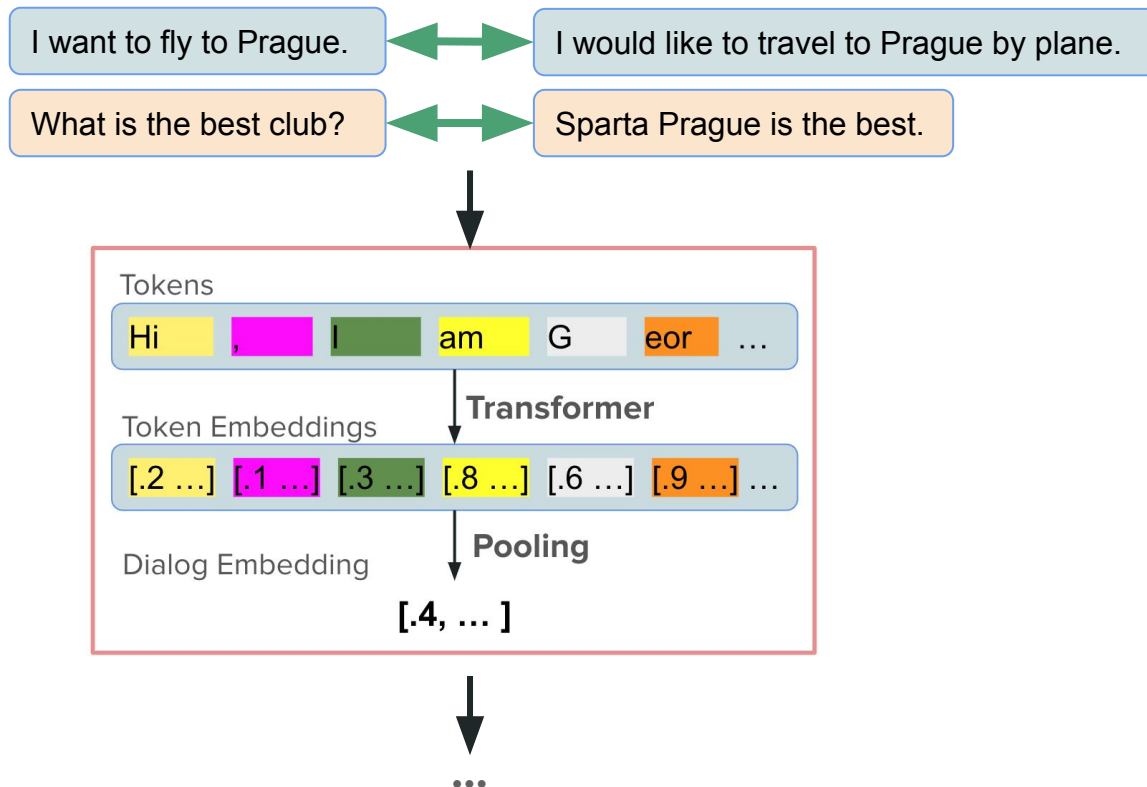


Sentence Transformers

Sentence Transformers

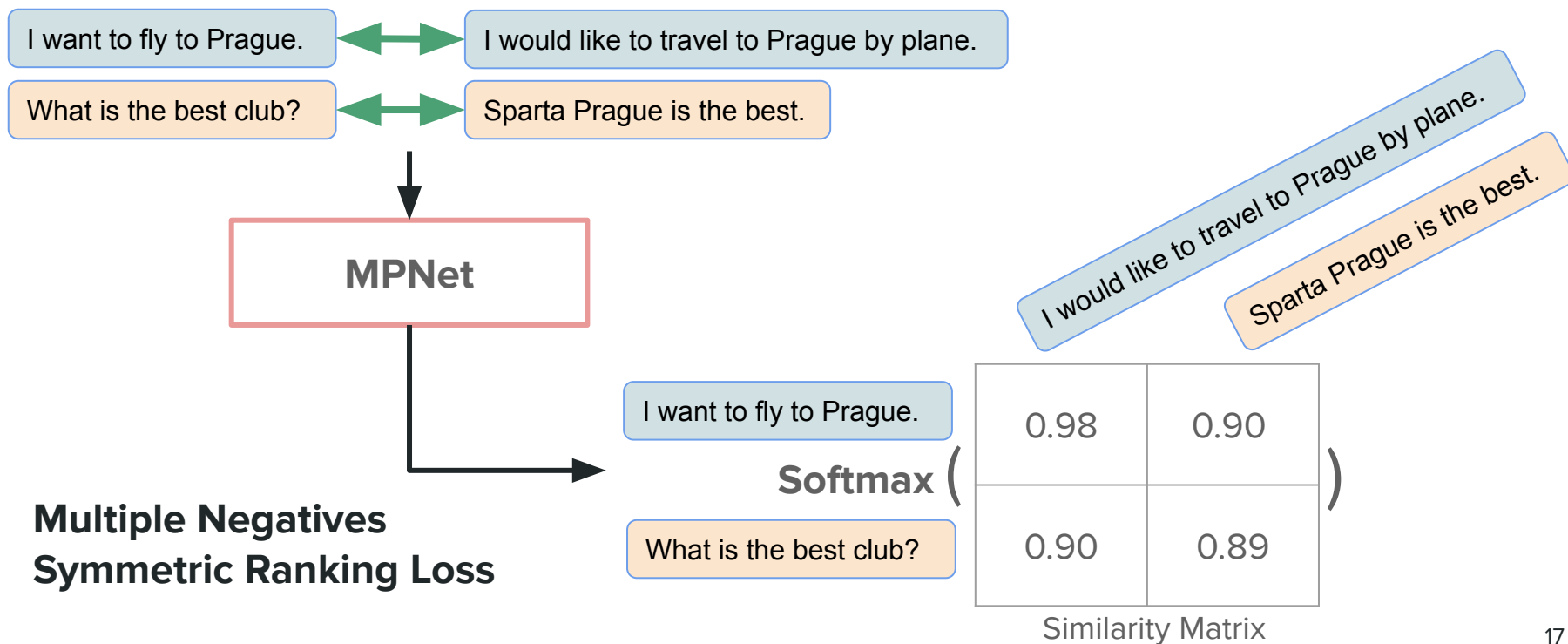


Sentence Transformers



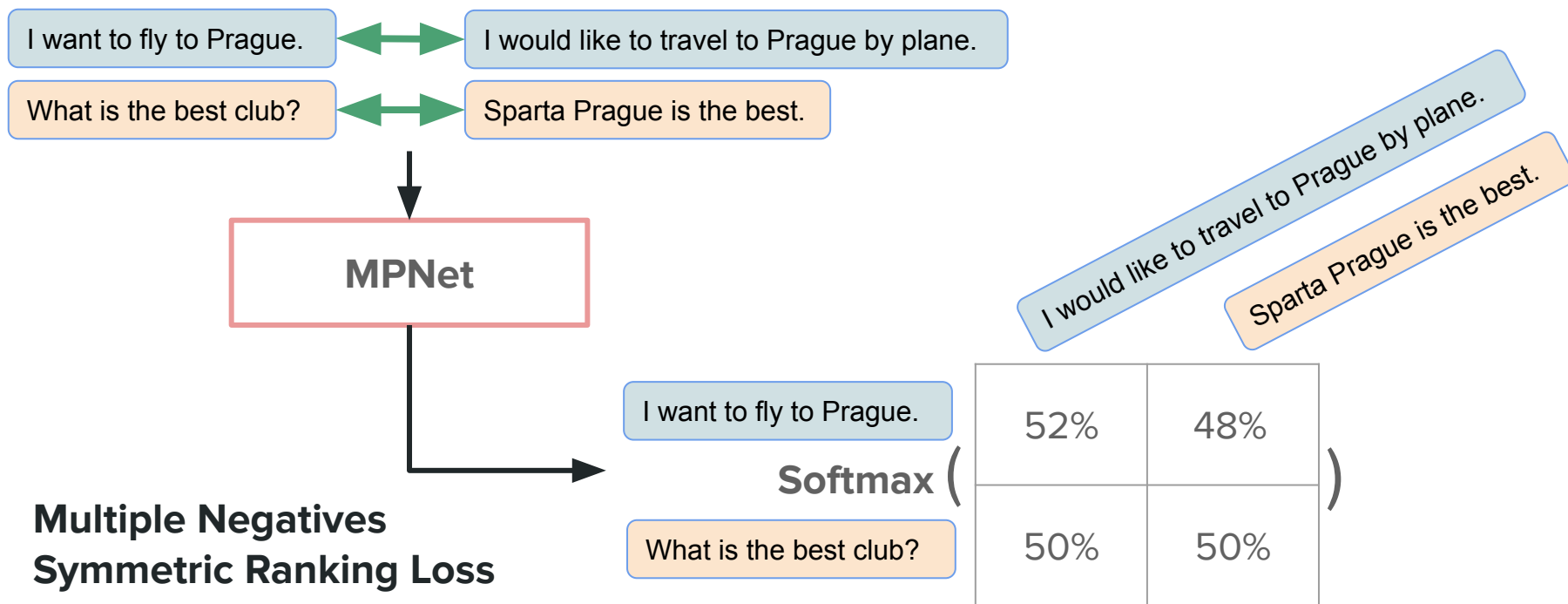
Sentence-MPNet

all-mpnet-base-v2



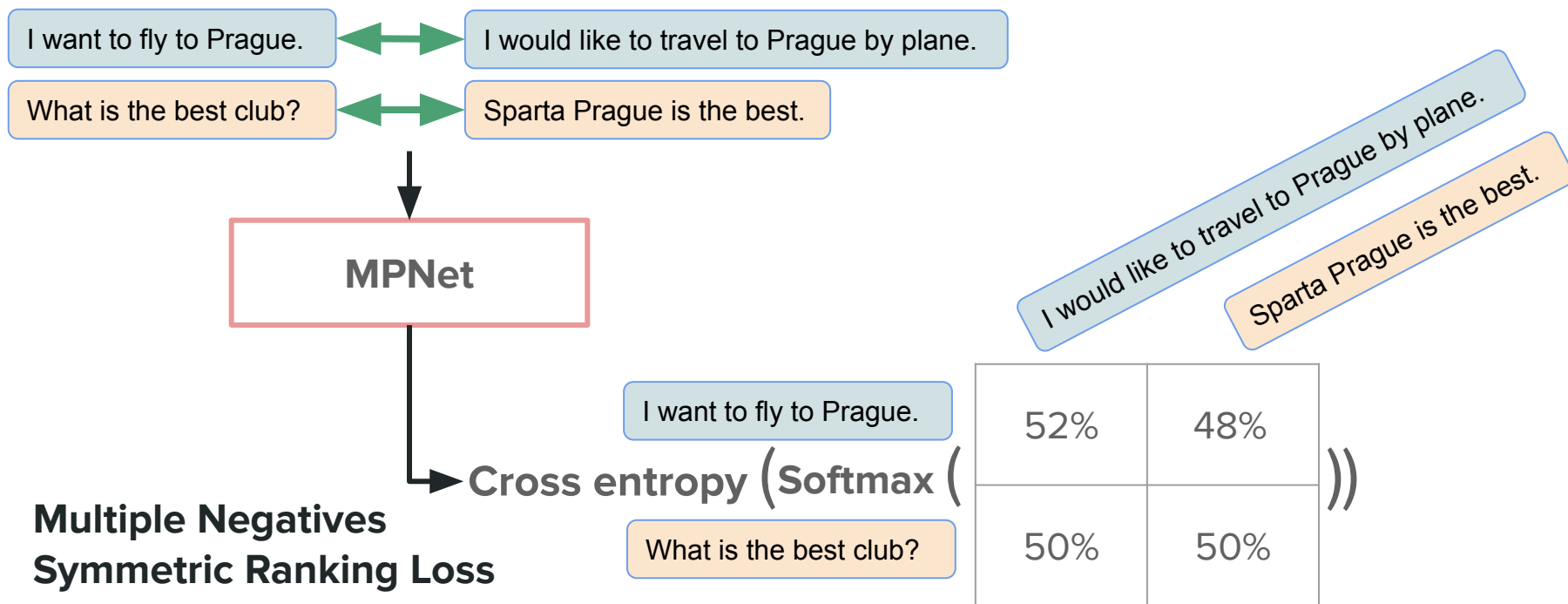
Sentence-MPNet

all-mpnet-base-v2



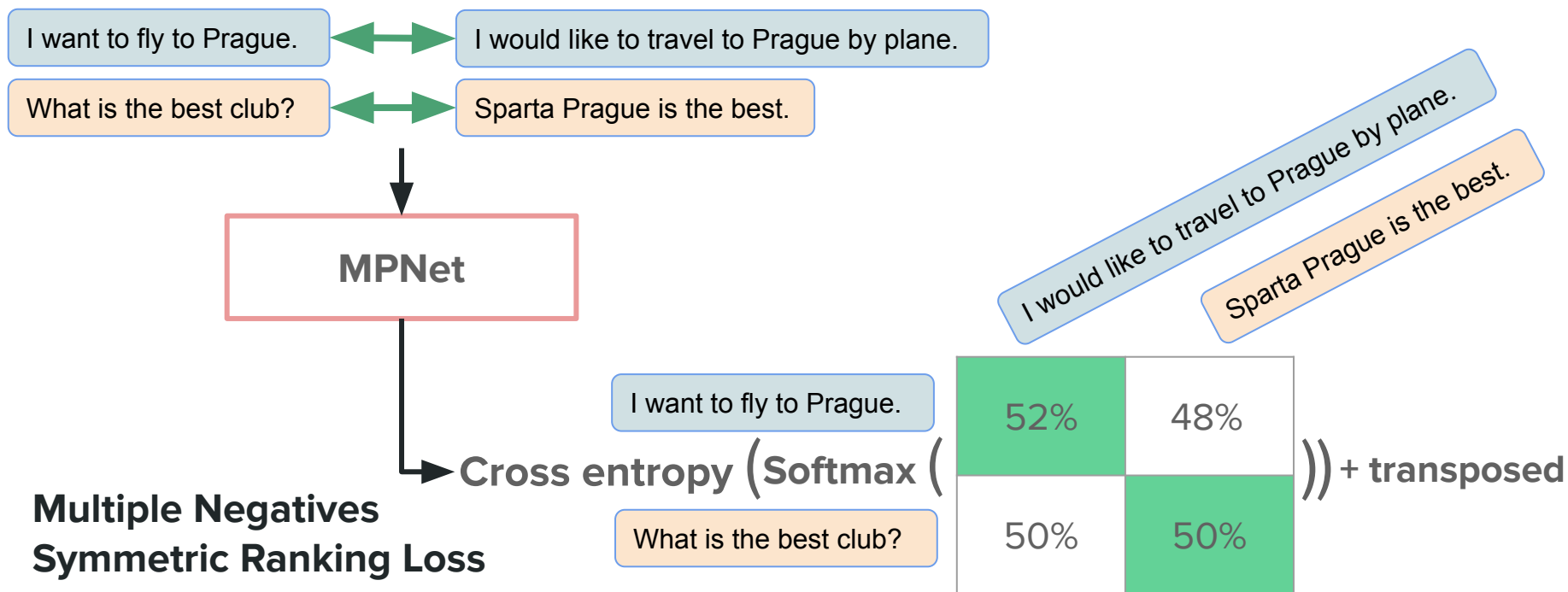
Sentence-MPNet

all-mpnet-base-v2



Sentence-MPNet

all-mpnet-base-v2



Code

```
1 from sentence_transformers import ...
2
3 model = SentenceTransformer(
4     modules=[
5         Transformer("sentence-transformers/all-mpnet-base-v2"),
6         Pooling(
7             word_embedding_dimension=768,
8             pooling_mode_mean_tokens=True,
9         )
10     ]
11 )
12 model.encode("I want to fly to Prague")
```

Sentence-MPNet

all-mpnet-base-v2

I want to fly to Prague.

Cosine Similarity

I would like to travel to Prague by plane.

★ 0.86

Sparta Prague is the best.

0.46

I **don't** want to fly to Prague.

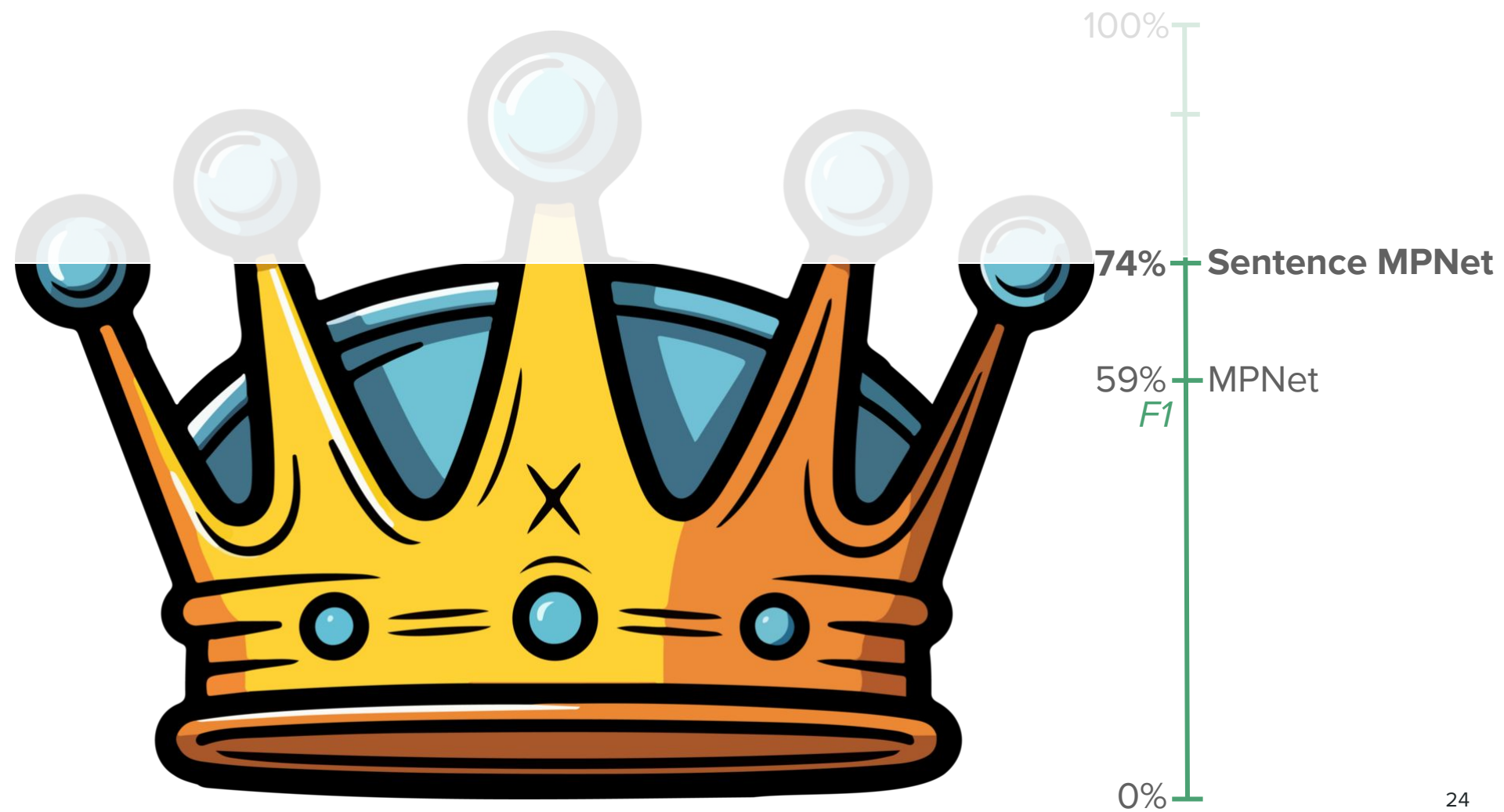
0.74

Suitable for sentences



Pipeline



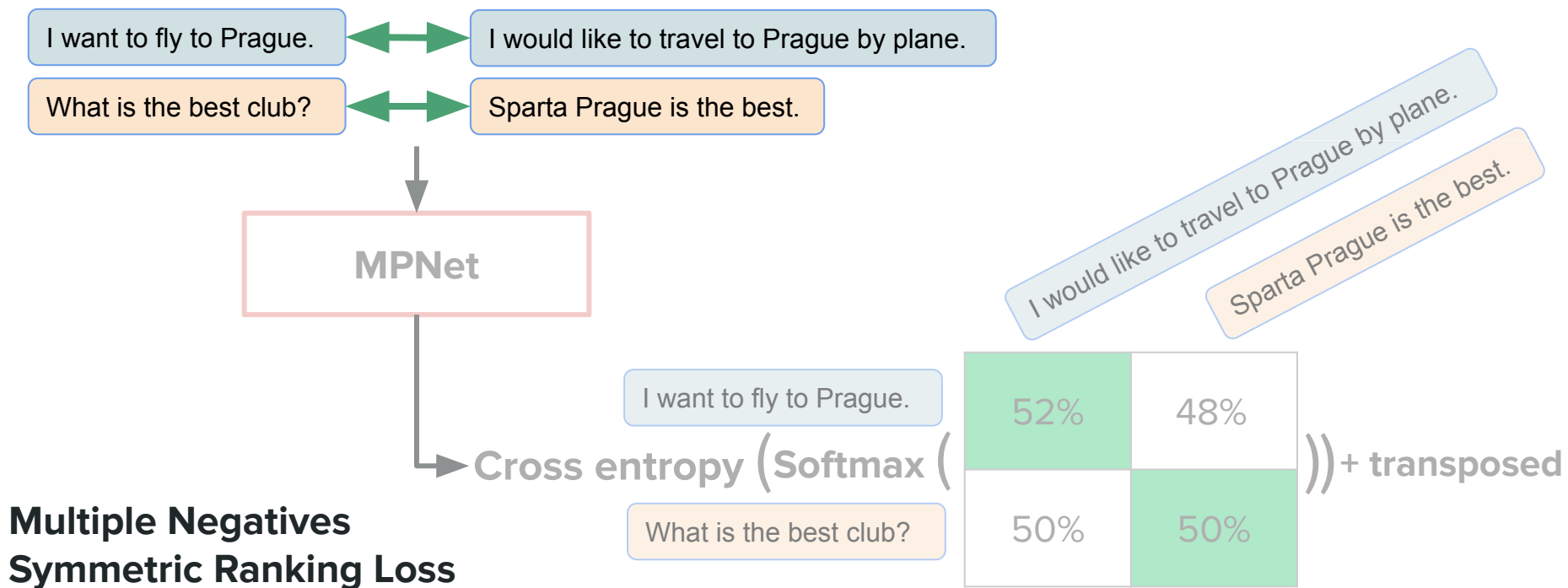




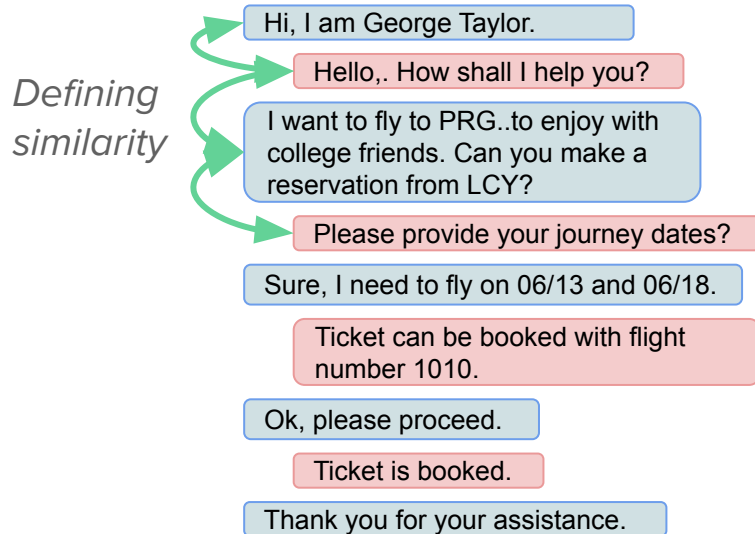
**Sentence
MPNet**

Sentence-MPNet

all-mpnet-base-v2



Train Sentence MPNet

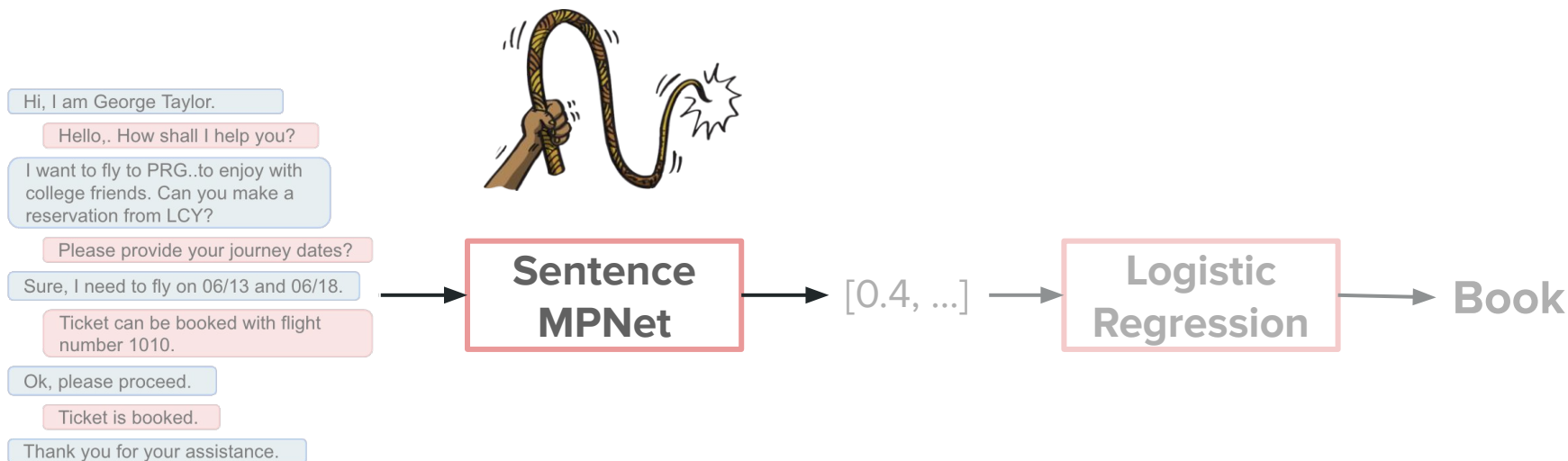


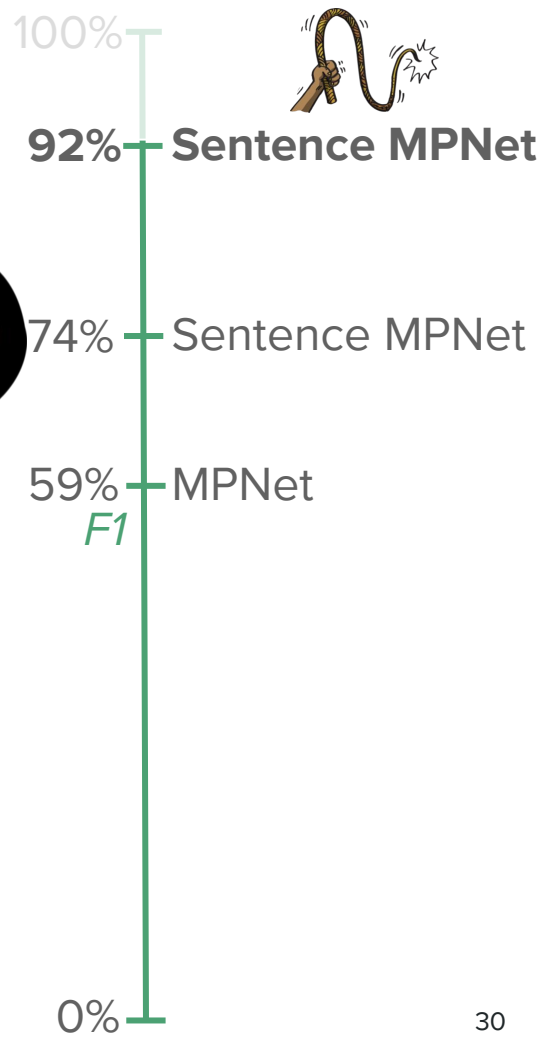
**Sentence
MPNet**

Code

```
1 from sentence_transformers import ...
2
3 model.fit(
4     train_objectives=[
5         (
6             DataLoader(train_pairs, batch_size=64),
7             MultipleNegativesSymmetricRankingLoss(model)
8         )
9     ],
10 )
```

Pipeline





Yes, but...

I want to fly to Prague.

Cosine Similarity

I would like to travel to Prague by plane.

0.922

Sparta Prague is the best.

0.905

I **don't** want to fly to Prague.

★ 0.925



George

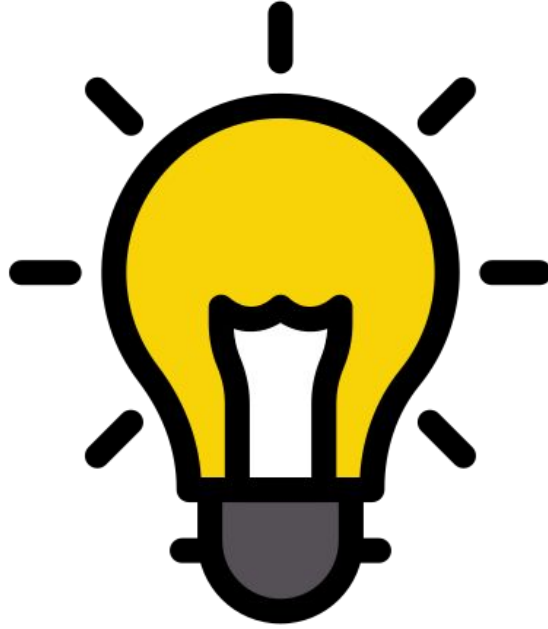
I want to fly to Prague.



I would like to travel to Prague by plane.

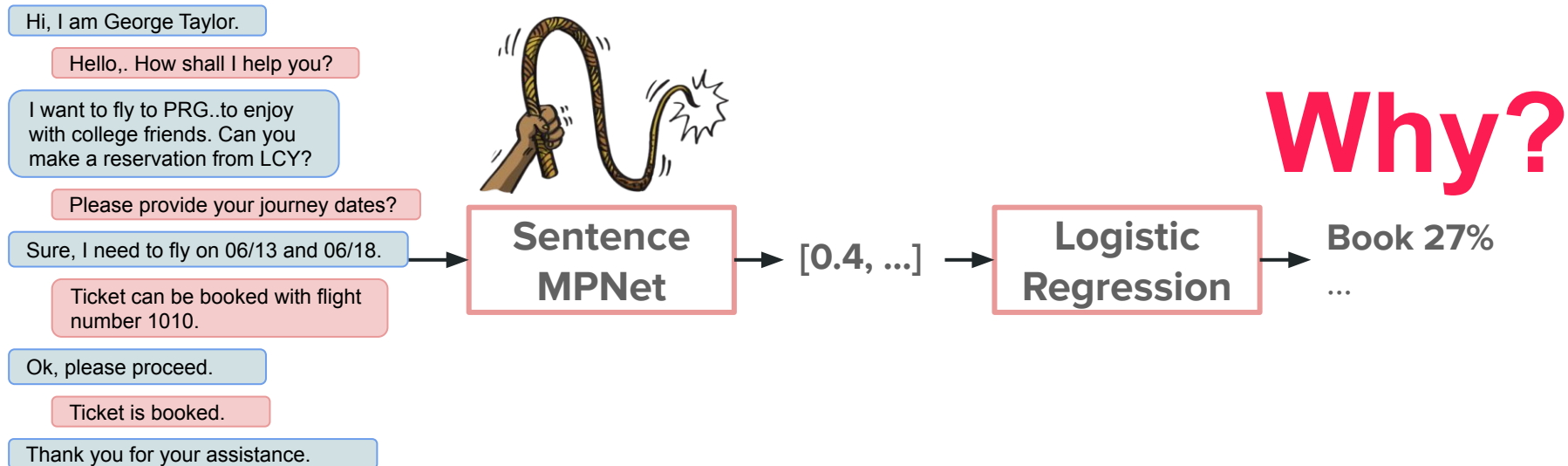


Ben

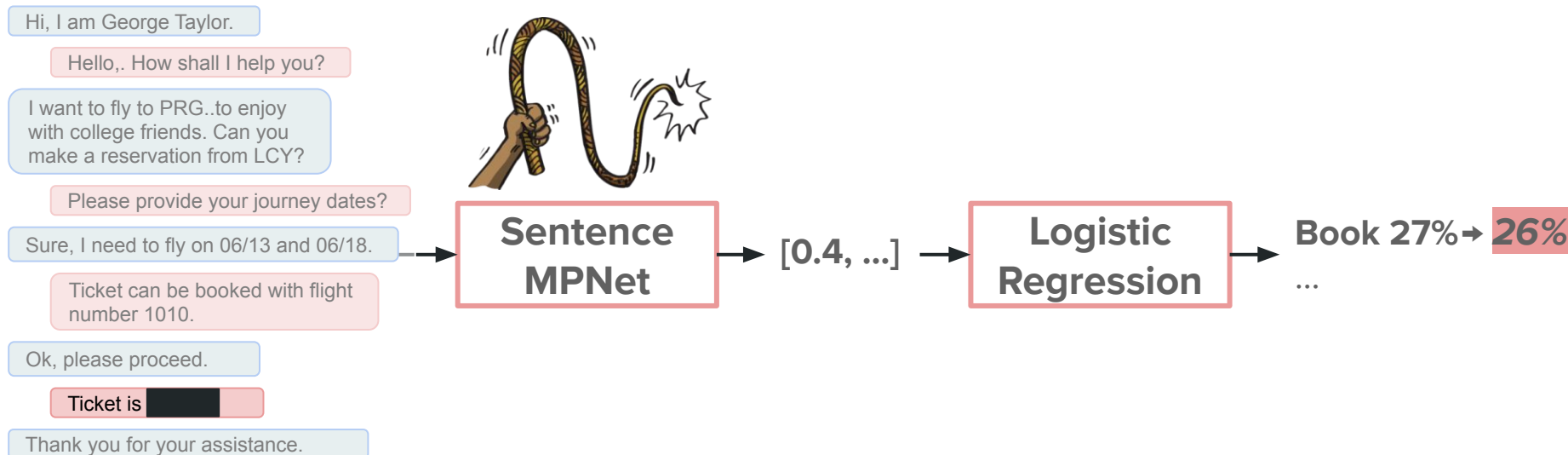


Intuition

Intuition



Intuition with SHAP



SHAP

Intuition with SHAP

Book 27%

Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

Ok, please proceed.

Ticket is booked.

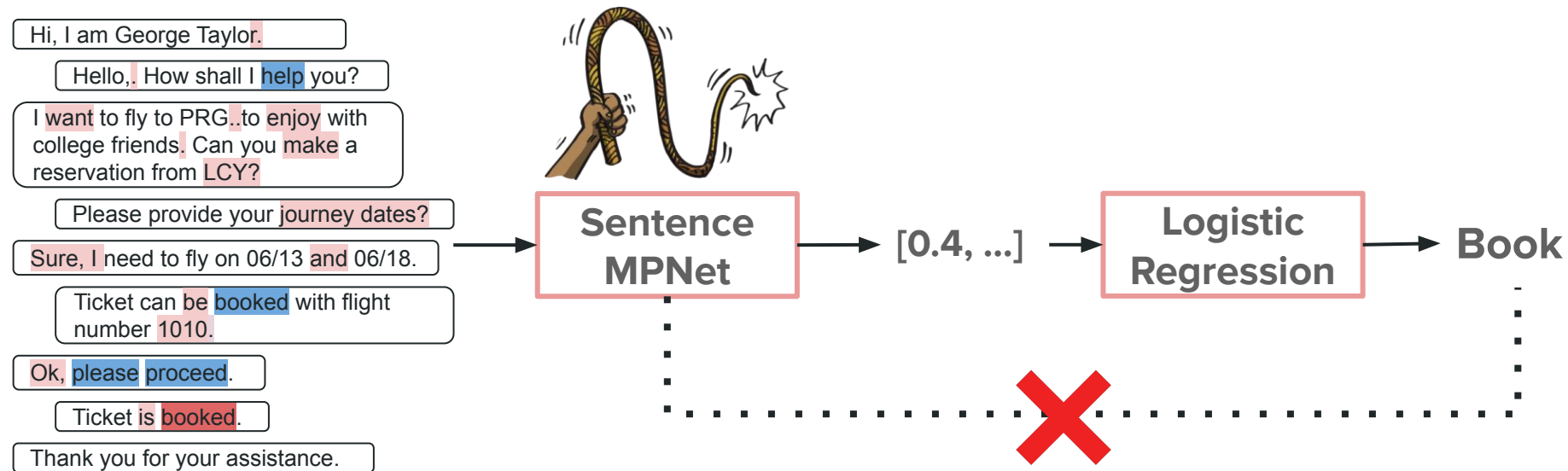
Thank you for your assistance.

Why predicted as

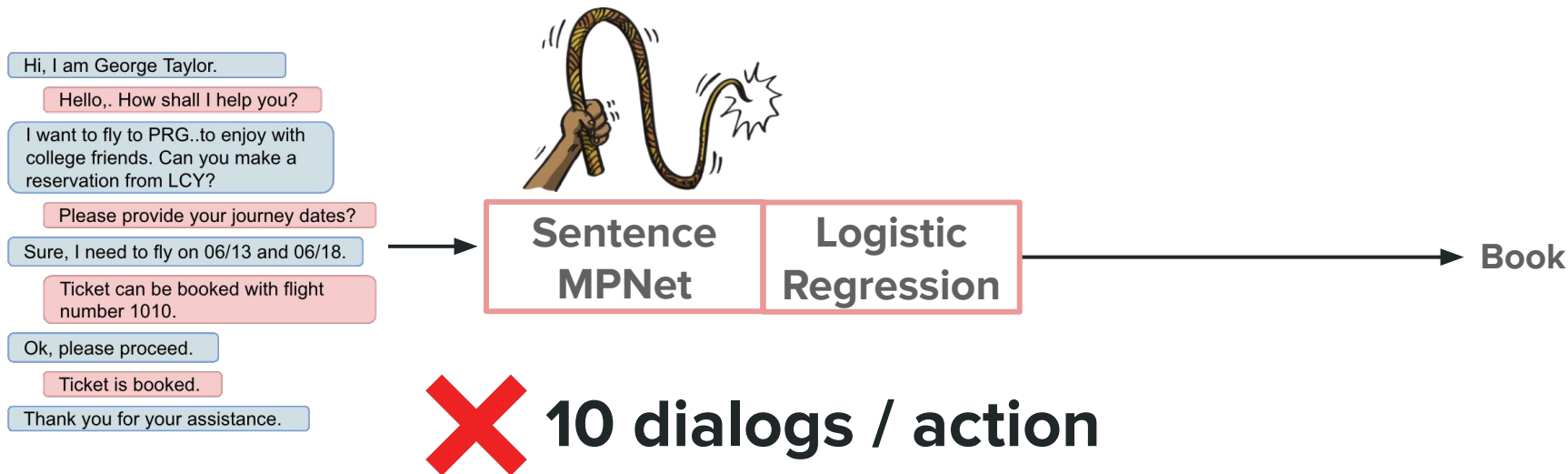
- book

- other

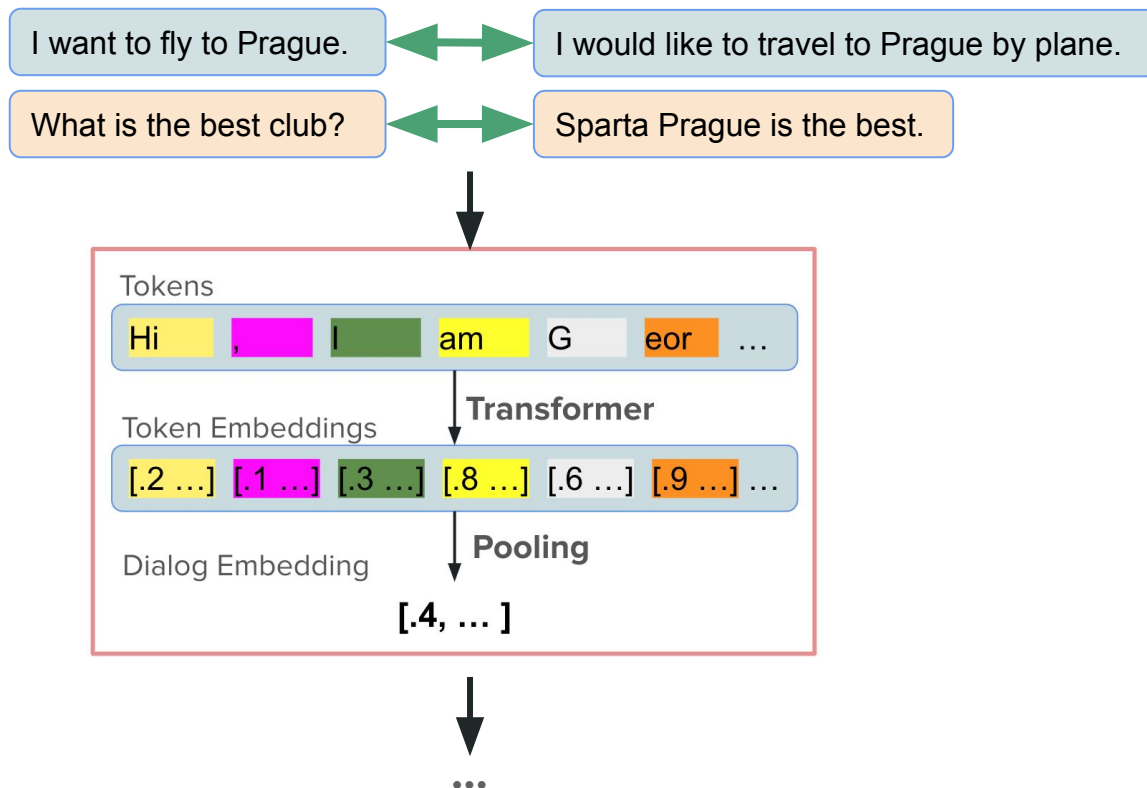
Why is this happening?



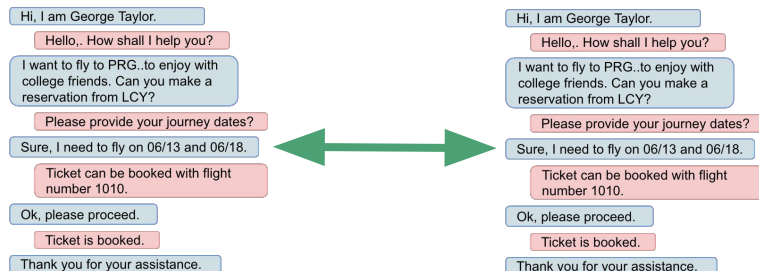
End-to-end training



Training Sentence Transformers

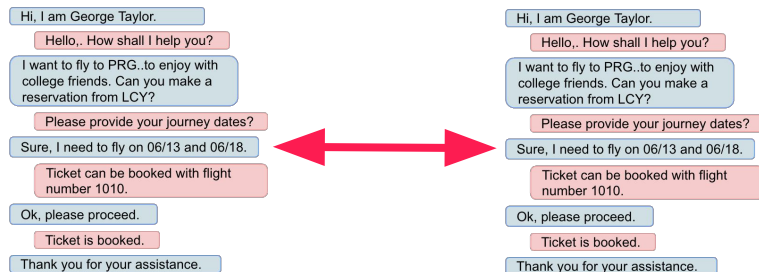


Training Sentence Transformers



Book 1

Book 2



Book 1

No flight

Similar pairs: 225
Dissimilar pairs: 1000

Contrastive Loss

Hi, I am George Taylor.
Hello,. How shall I help you?
I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?
Please provide your journey dates?
Sure, I need to fly on 06/13 and 06/18.
Ticket can be booked with flight number 1010.
Ok, please proceed.
Ticket is booked.
Thank you for your assistance.

Book 1



Hi, I am George Taylor.
Hello,. How shall I help you?
I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?
Please provide your journey dates?
Sure, I need to fly on 06/13 and 06/18.
Ticket can be booked with flight number 1010.
Ok, please proceed.
Ticket is booked.
Thank you for your assistance.

Book 2

Sentence MPNet

[.7, ...]

[.8, ...]

$$1 - \cos (\text{[.7, ...]} \text{[.8, ...]}) = 0.016^2$$

Code

```
1 from sentence_transformers import ...
2
3 model.fit(
4     train_objectives=[
5         (
6             DataLoader(train_pairs, batch_size=64),
7             ContrastiveLoss(model)
8         )
9     ],
10 )
```

Effects

Book 27%

Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

Ok, please proceed.

Ticket is booked.

Thank you for your assistance.

Sentence
MPNet



Book 77%

Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

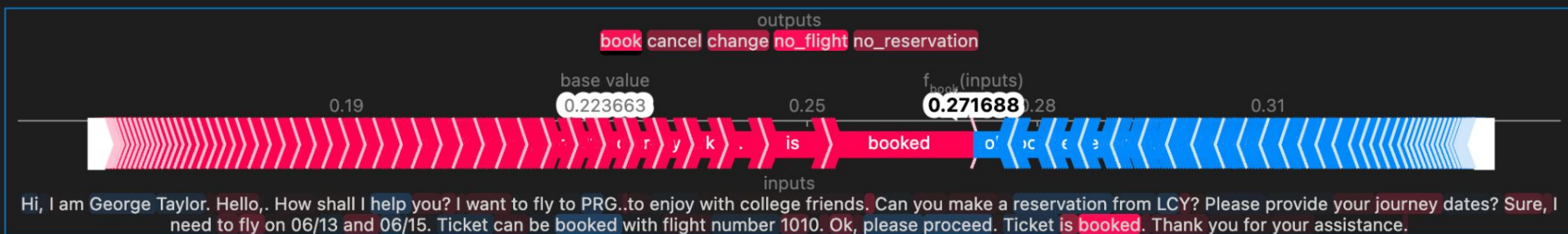
Ok, please proceed.

Ticket is booked.

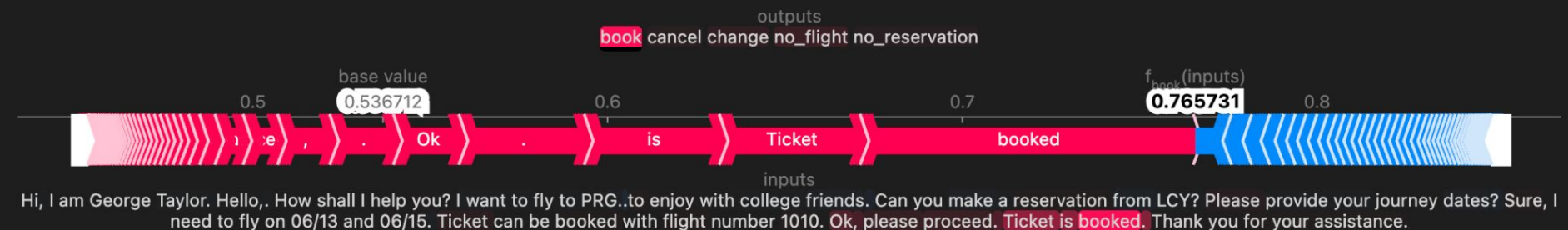
Thank you for your assistance.

Sentence
MPNet

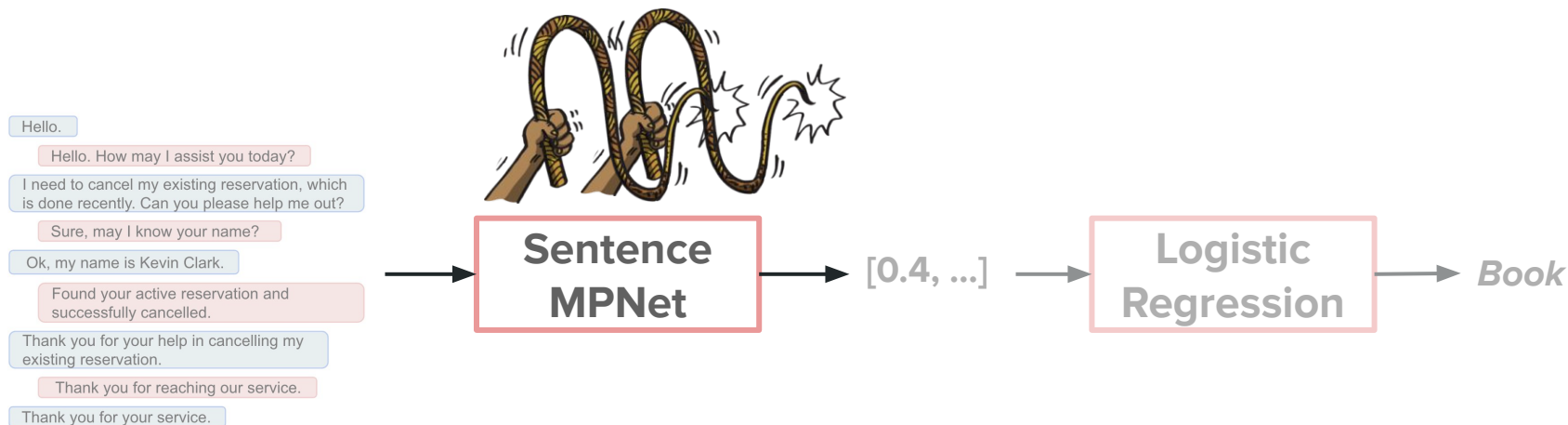


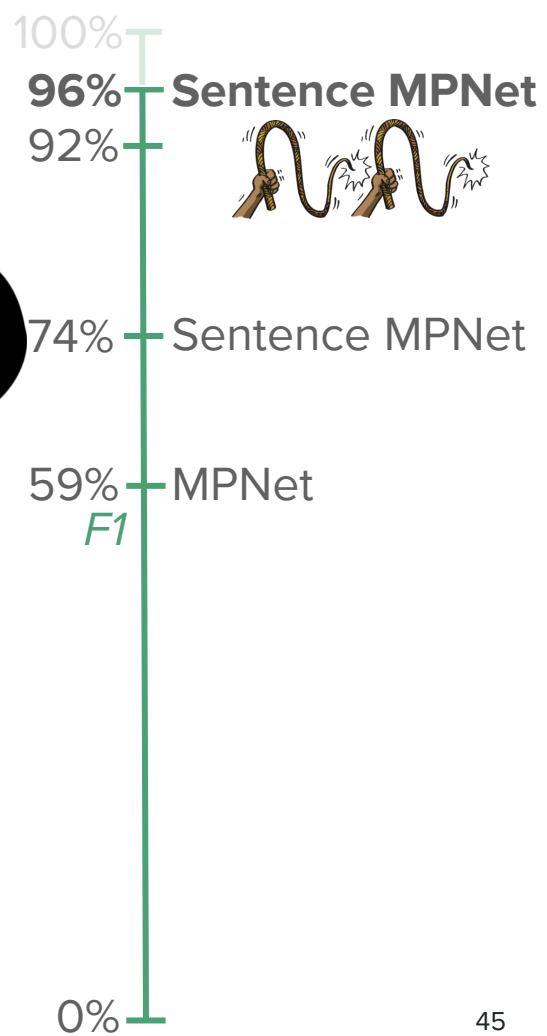


```
=== sent-mpnet-unlbl-lbl ===
True: book
Predicted: book
Probs: 76.6%, 4.0%, 7.1%, 9.3%, 3.0%
```

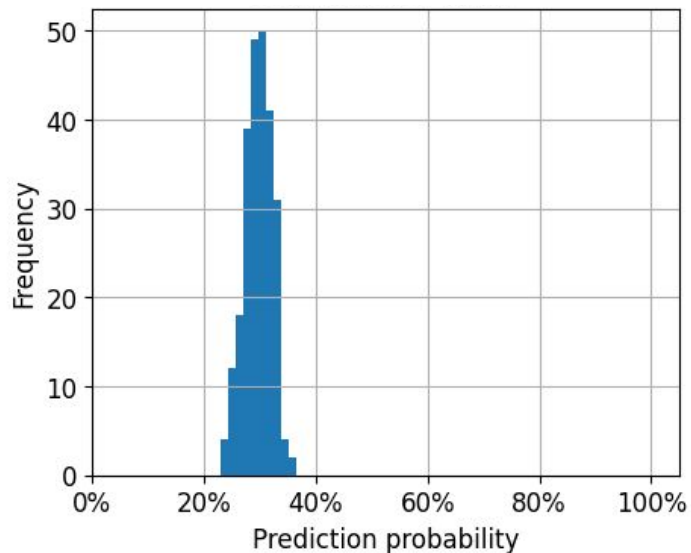


Pipeline

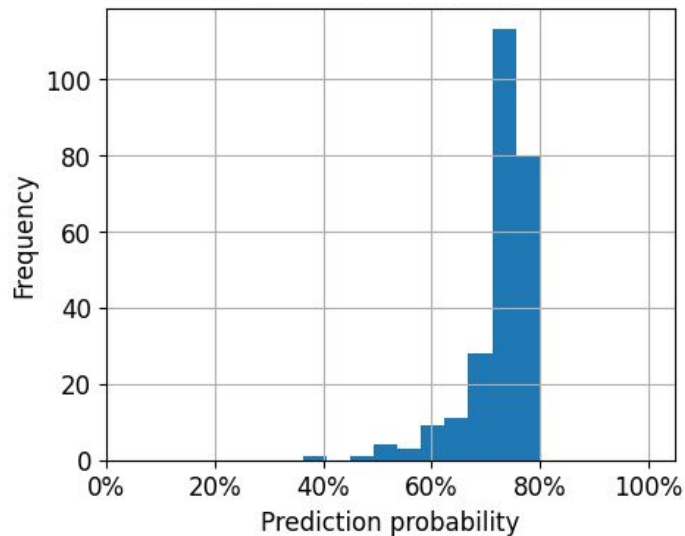




More Robustness



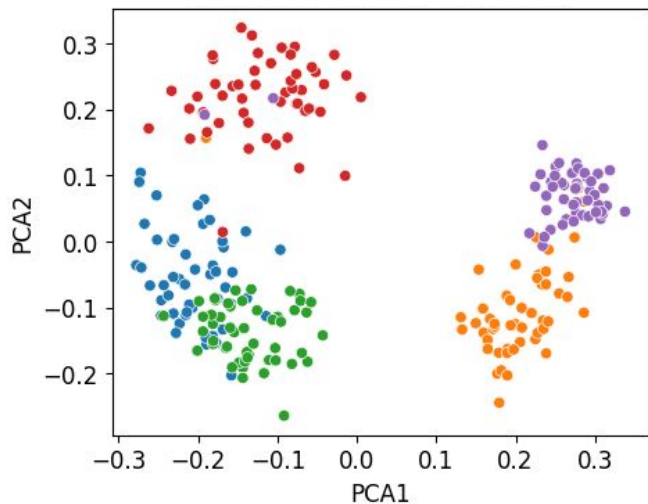
**Sentence
MPNet**



**Sentence
MPNet**



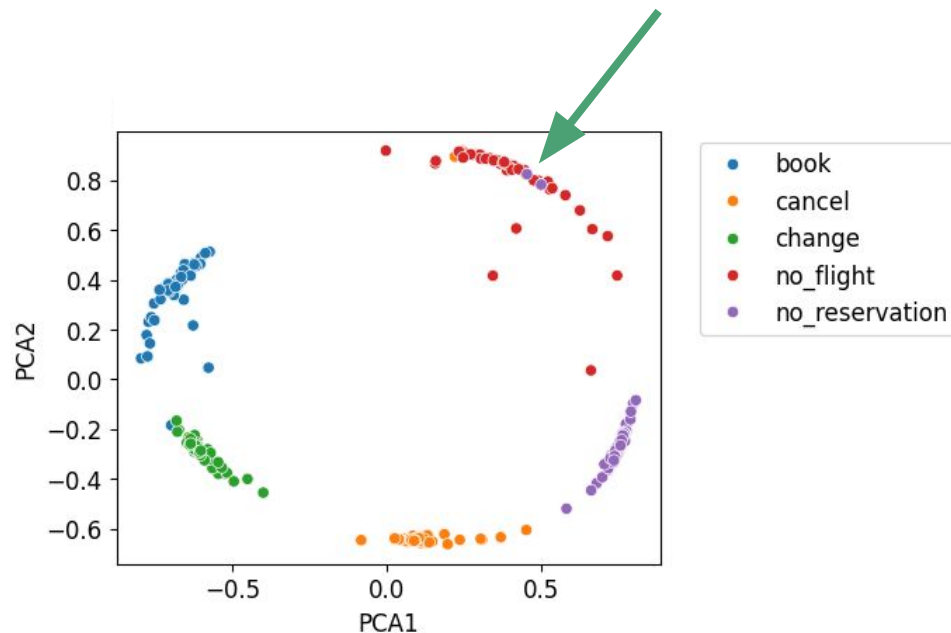
Better Separation



**Sentence
MPNet**

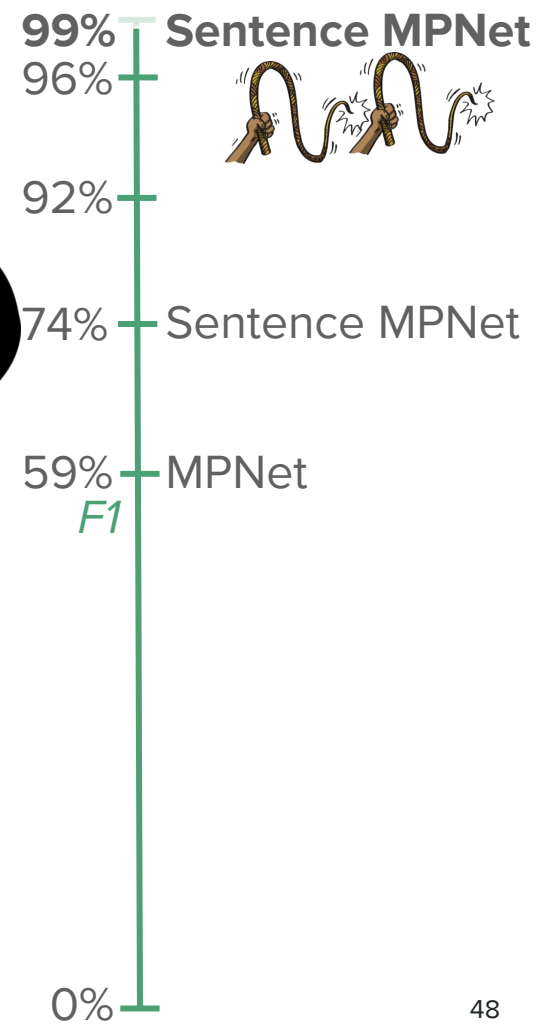


Incorrectly labelled

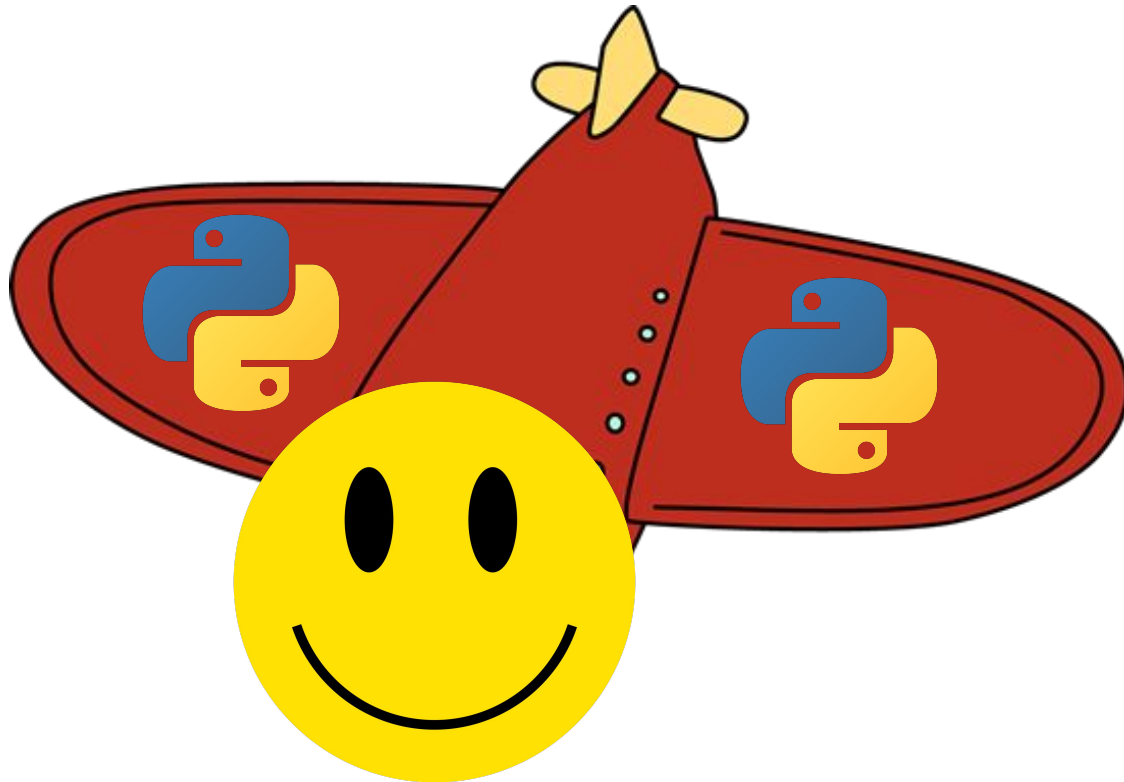


**Sentence
MPNet**





PythonAir



Takeaway



Hi, I am George Taylor.

Hello,. How shall I help you?

I want to fly to PRG..to enjoy with college friends. Can you make a reservation from LCY?

Please provide your journey dates?

Sure, I need to fly on 06/13 and 06/18.

Ticket can be booked with flight number 1010.

Ok, please proceed.

Ticket is booked.

Thank you for your assistance. [1]



Representation

Let's chat about embeddings



/in/zikaadam/



/azikoss/azikoss/europython24-dialog-embeddings

Salted^{CX}

Thanks to Salted, to my colleagues and friends for help, and to Sparta Praha for being the best ❤️

MPNet PCA

