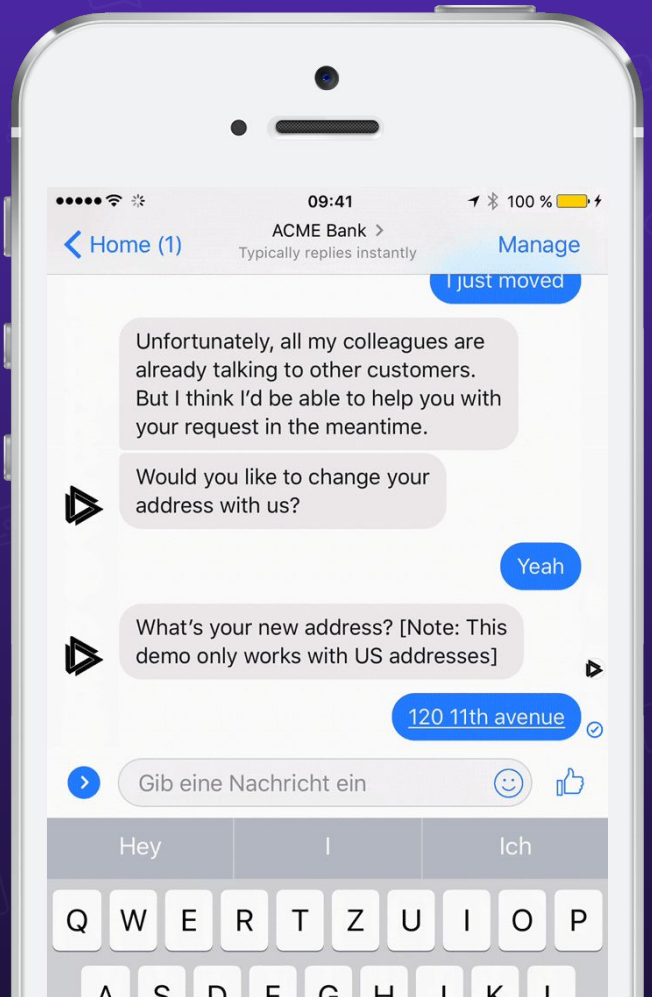
Deprecating the state machine: building conversational AI with Rasa stack

Justina Petraityte, Developer Advocate

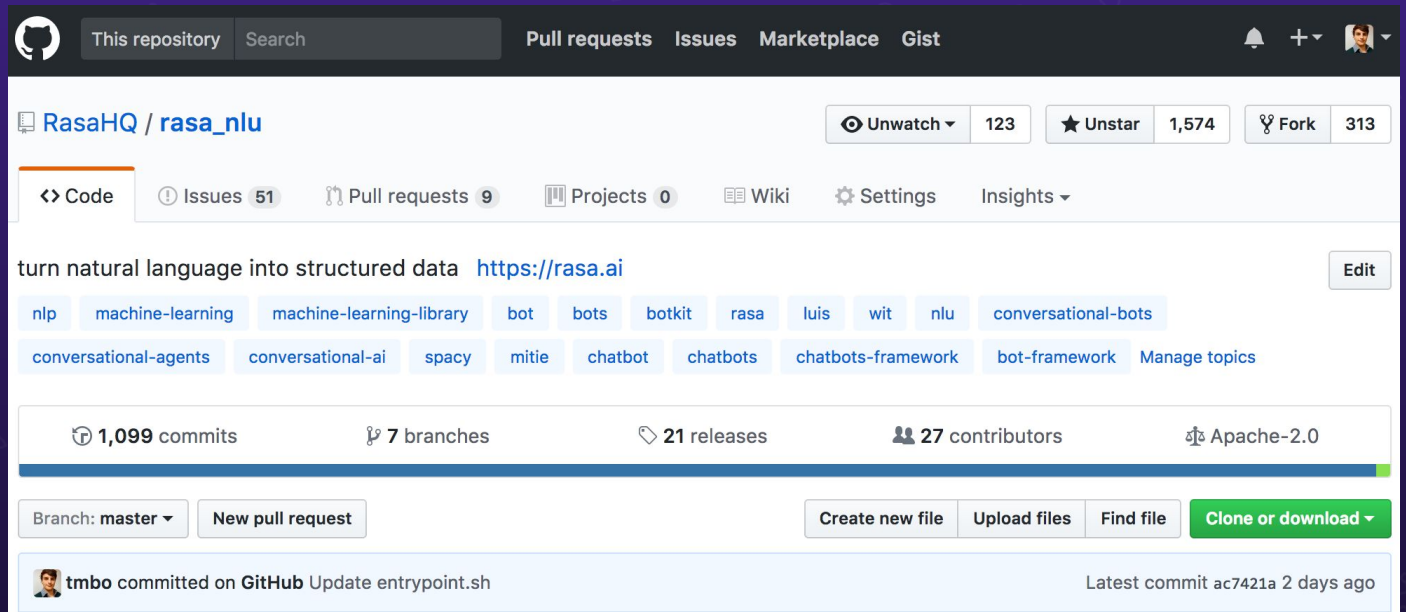
Conversational AI will dramatically change how your customers interact with you.



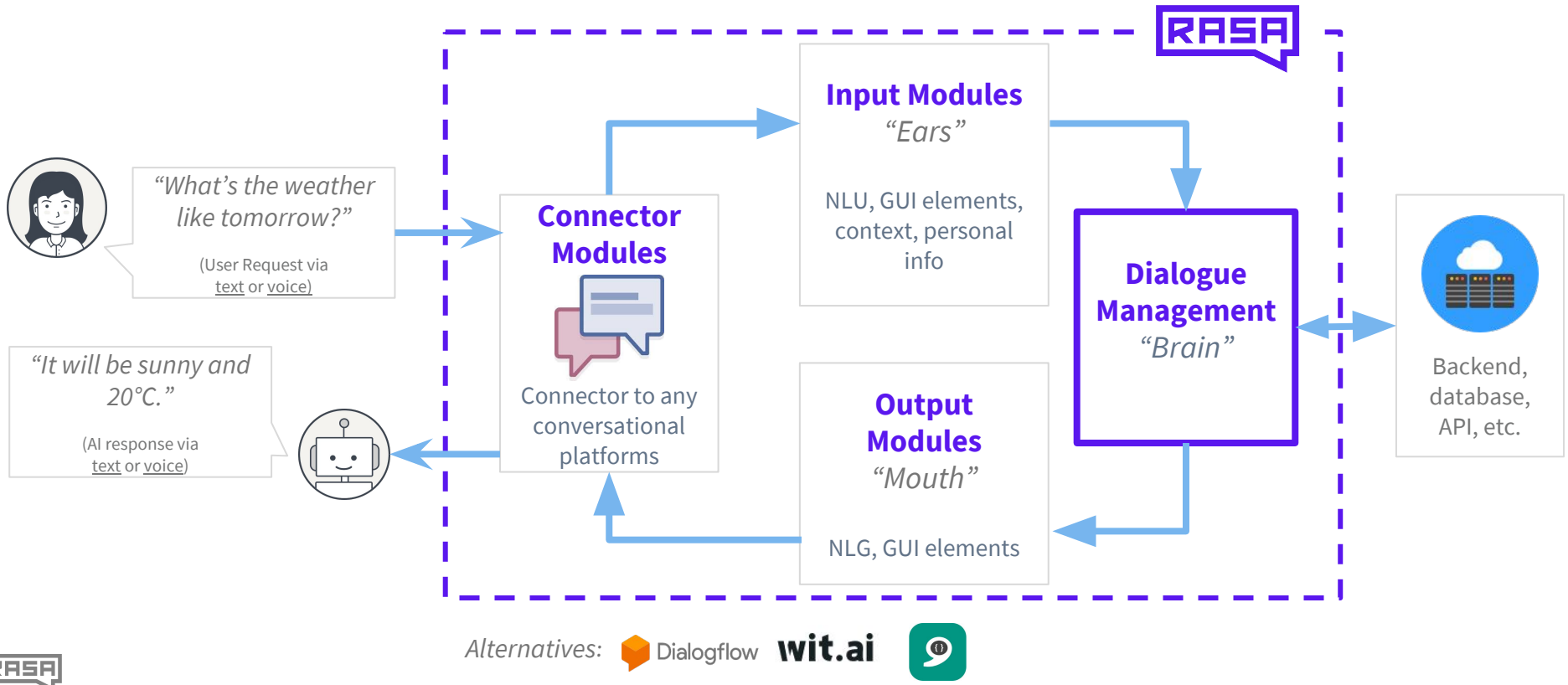
Example of a live Skill:
A customer can change her address via Facebook Messenger

RASA

An open source, highly scalable ML framework to build conversational software



Rasa the OSS to build conversational software with ML



Why Rasa?

Runs Locally

- No Network Overhead
- Control QoS
- Deploy anywhere

Own Your Data

- Don’t hand data over to big tech co’s
- Avoid vendor lock-in

Hackable

- Tune models for your use case

What we are focusing on today

Goal:

build & understand a bot based on machine learning

Roadmap:

1. Natural Language Understanding
    i. Theory
    ii. Let’s Code
2. Dialogue Handling
    i. Theory
    ii. Let’s Code
3. Research
4. Questions

Setup

1. Jupyter notebook in python 3.6 (2.7 should work as well)

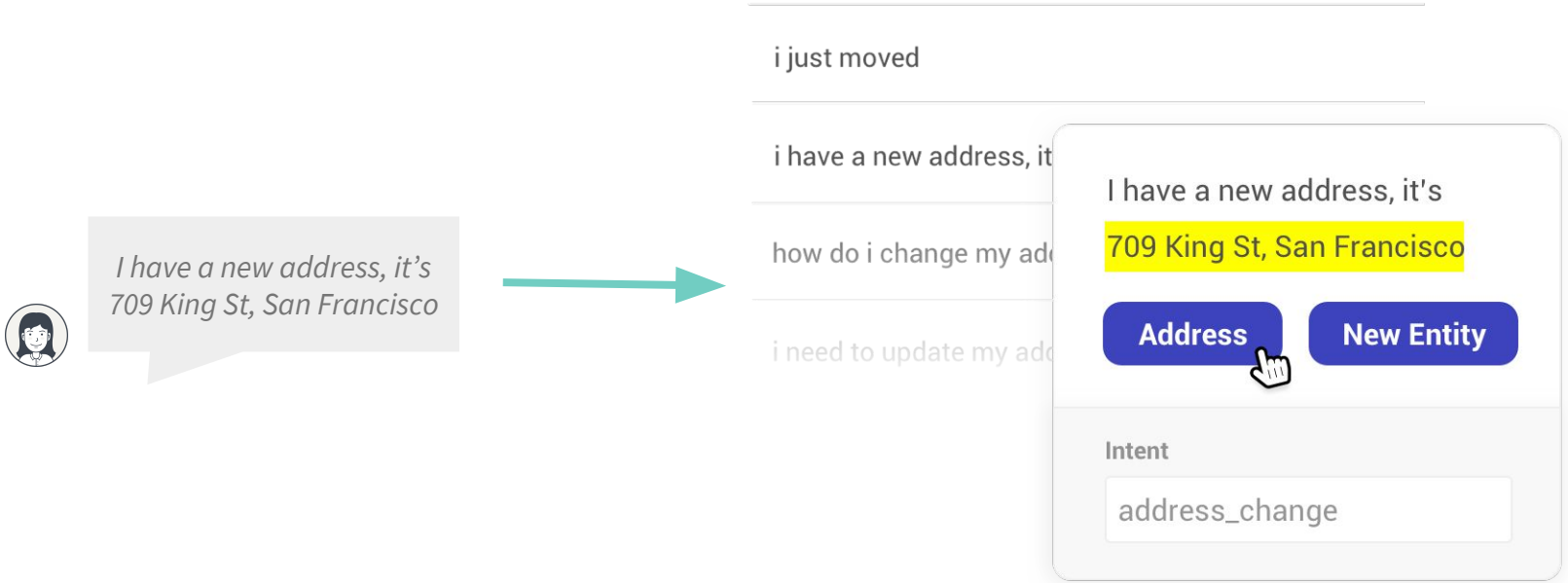
2. Download:
   Repository: https://github.com/RasaHQ/rasa-workshop-pydata-berlin

Under the hood
Natural Language Understanding

Under the Hood

Rasa NLU: Natural Language Understanding

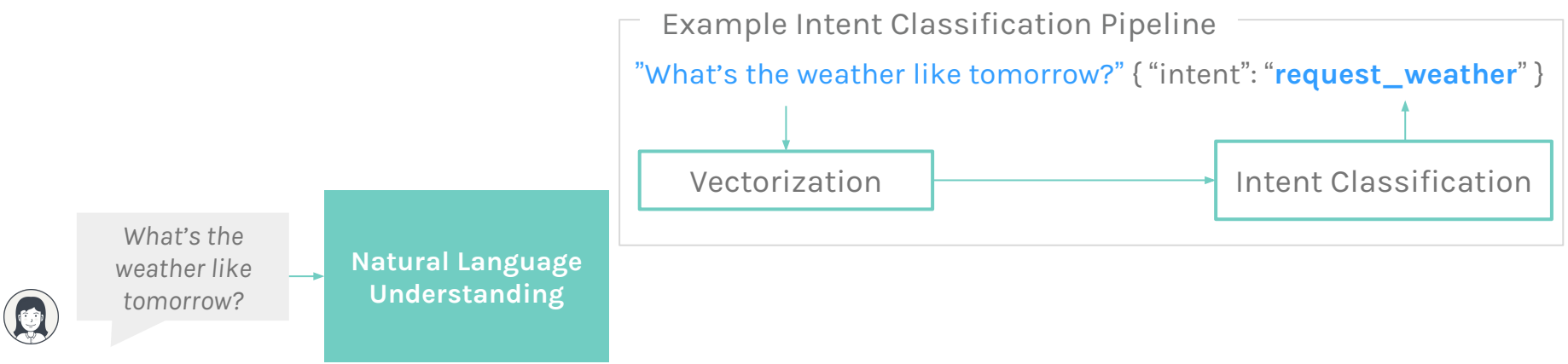
Goal: create structured data



https://github.com/RasaHQ/rasa-workshop-pydata-berlin

Under The Hood

Natural Language Understanding

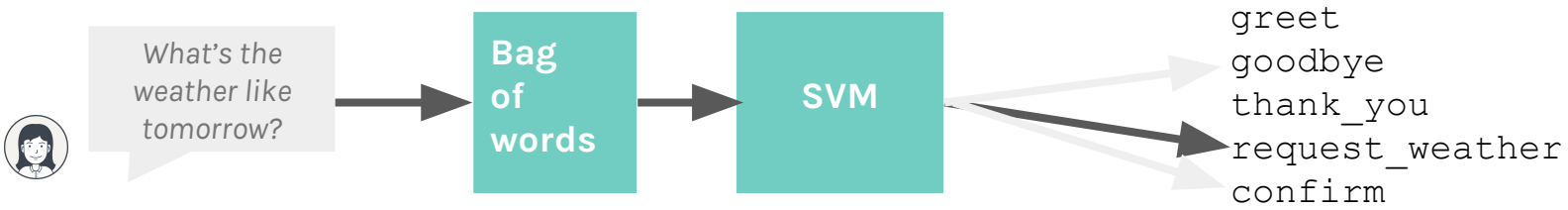


https://github.com/RasaHQ/rasa-workshop-pydata-berlin

Rasa NLU: Natural Language Understanding

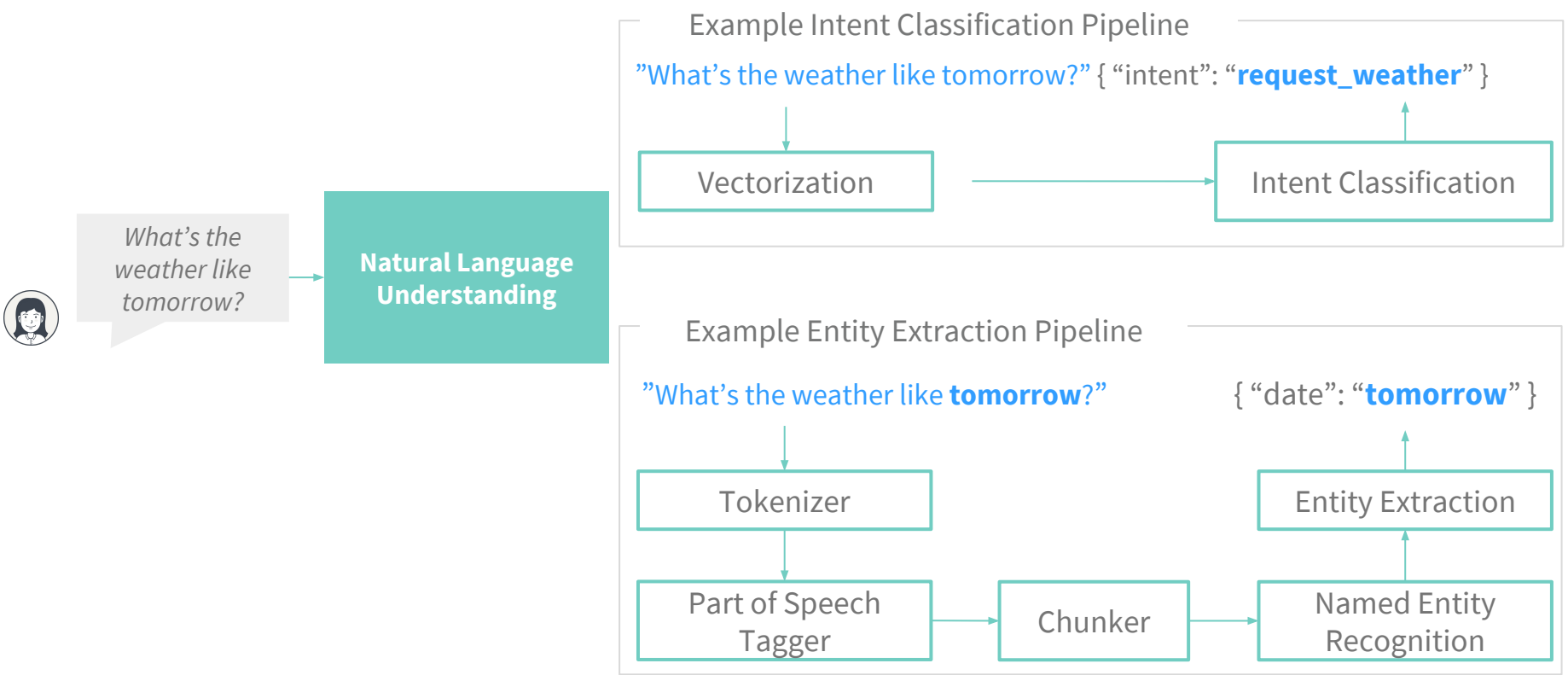
Bags are your friend

$$\{v_1, \ldots, v_s\} \rightarrow \frac{1}{s} \sum_i v_i$$



Under the Hood

Rasa NLU: Natural Language Understanding



https://github.com/RasaHQ/rasa-workshop-pydata-berlin

Rasa NLU: Entity Extraction

Where can I get a burrito in the 2nd arrondissement ?

cuisine

location

averaged perceptron

$$\hat{y} = \text{sign}\left(\sum_{k=1}^{K} c^{(k)}\left(w^{(k)} \cdot \hat{x} + b^{(k)}\right)\right)$$

1. Binary classifier `is_entity` & then `entity_classifier`
2. Direct structured prediction

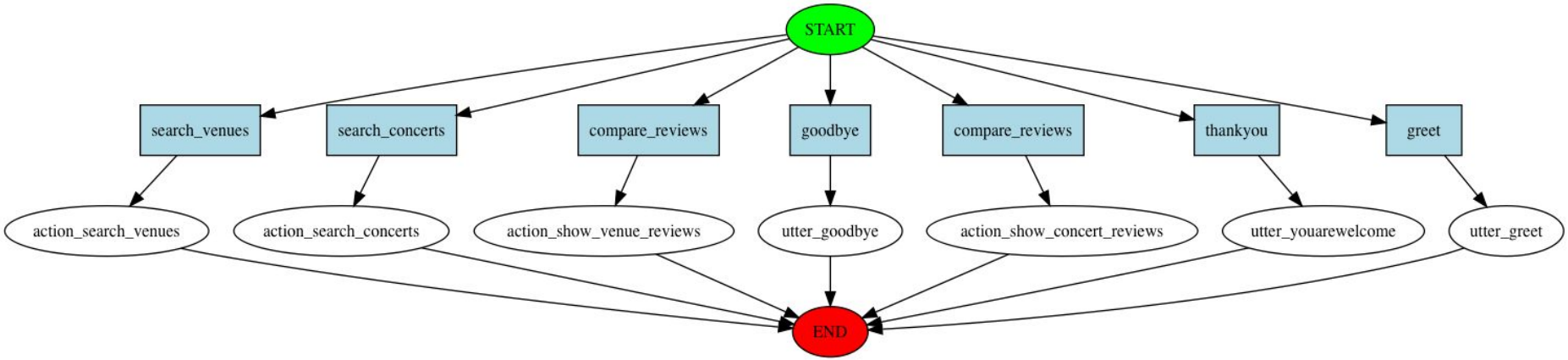
Let’s code!

Under the hood
Dialogue Management

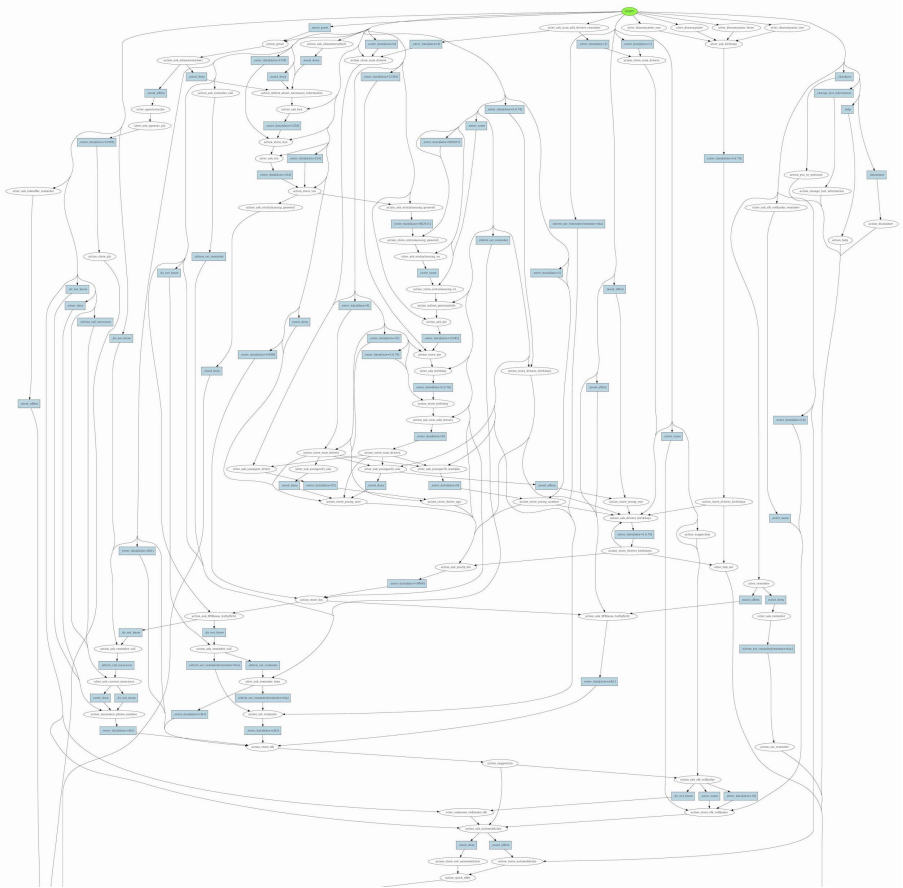
Why Dialogue Handling with Rasa Core?

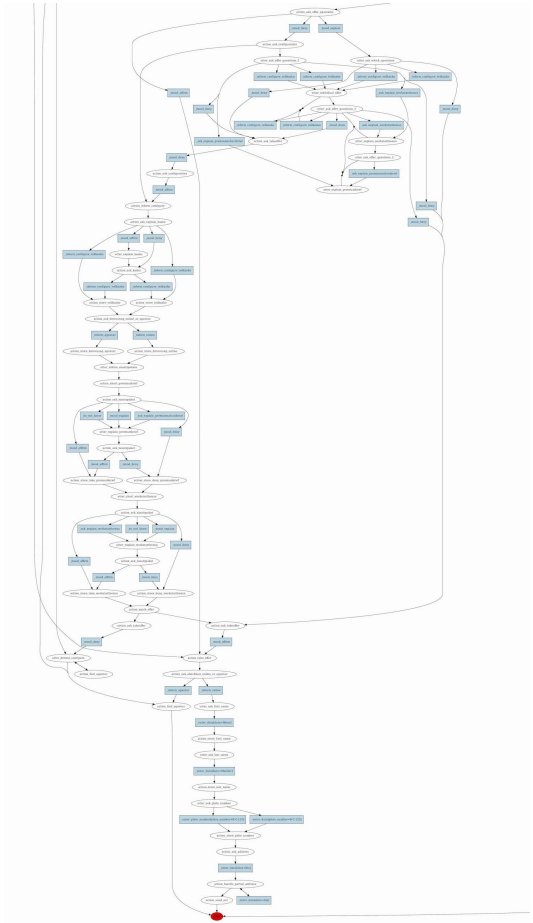
- No more state machines!
- Reinforcement Learning: too much data, reward functions...
- Need a simple solution for everyone

Why Machine Learning?



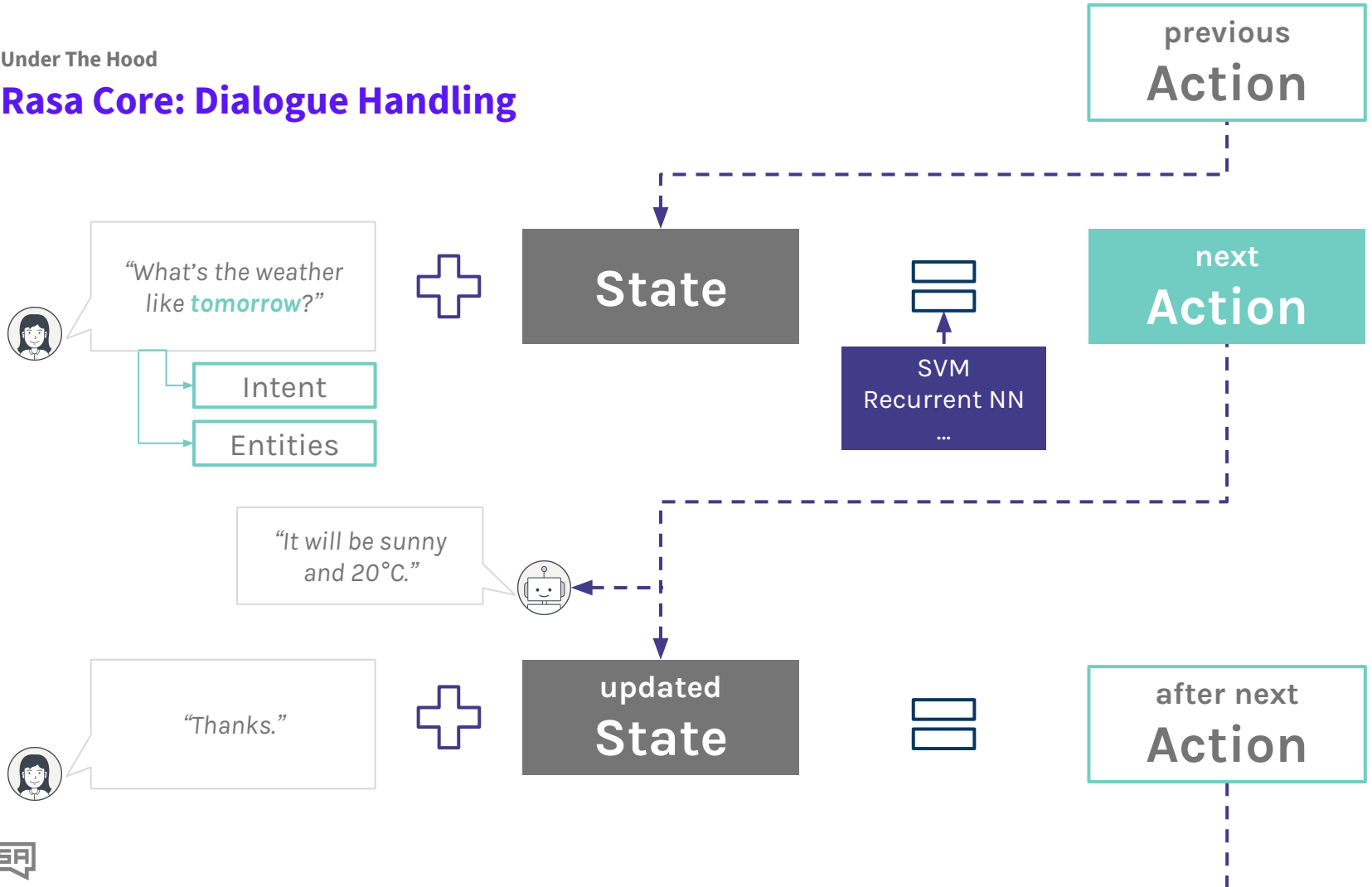
State Machines are infeasible





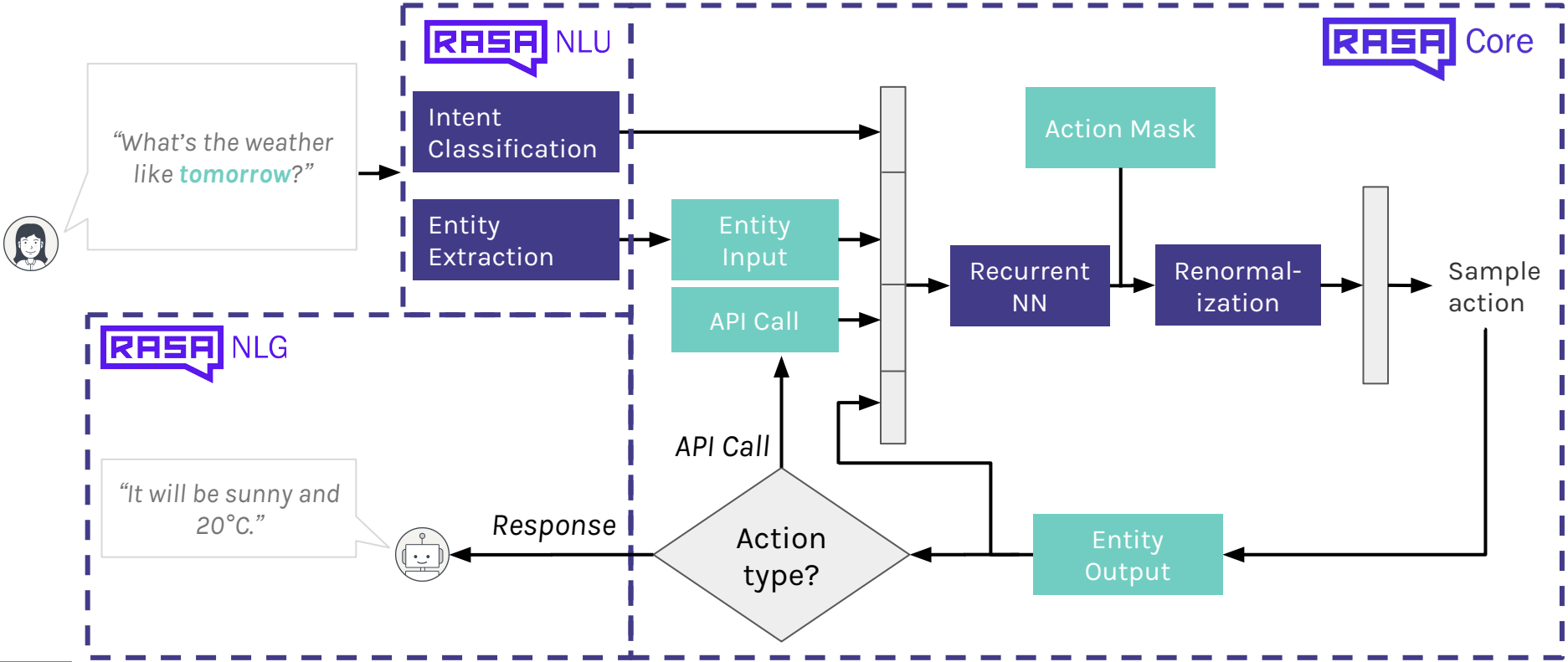
Under The Hood

Rasa Core: Dialogue Handling



Rasa Core: Dialogue Handling

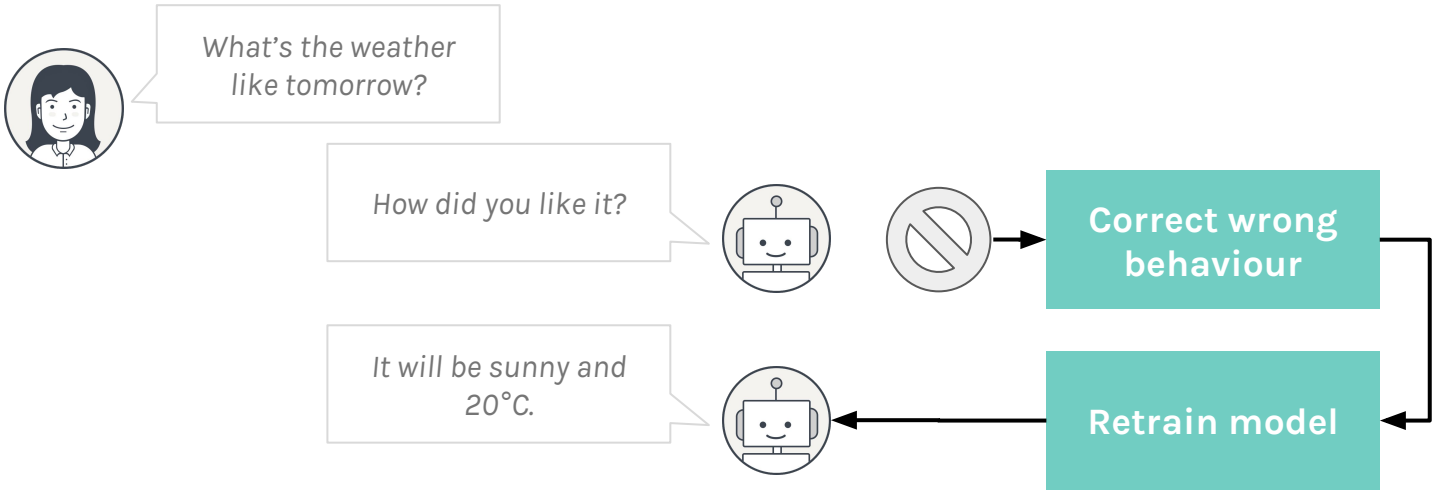
Similar to LSTM-dialogue prediction paper: https://arxiv.org/abs/1606.01269



Let’s code!

Rasa Core: Dialogue Training

Issue: How to get started? → Online Learning



Let’s Code

Interactive Learning

Research

Research

Training NLU models without initial word vectors

Goal: Learn an **embedding** for the intent labels based on the user messages

- Learns joined embeddings for intents & words at the same time
- Allows multi-intent labels
- Knows about similarity between intent labels
- Based on Starspace Paper

https://medium.com/rasa-blog/supervised-word-vectors-from-scratch-in-rasa-nlu-6daf794efcd8

https://medium.com/rasa-blog/how-to-handle-multiple-intents-per-input-using-rasa-nlu-tensorflow-pipeline-75698b49c383

Research

Training NLU models without initial word vectors

Goal: Learn an **embedding** for the intent labels based on the user messages

Multi-Intent:

Text	Intent
Hey how are you? i don't really care	`greet+dontcare`
ok something else then? thanks a bunch	`deny+thankyou`
cool! Who is the mayor or New York City?	`state_happy+random`

Evaluation:

Pipeline	train F1-score	test F1-score
spacy (small)	0.684 (0.020)	0.325 (0.018)
tensorflow_embedding	0.984 (0.001)	**0.898** (0.017)

Research

Generalisation *across* dialogue tasks

Why do we need this complex architecture? For generalisation between domains!

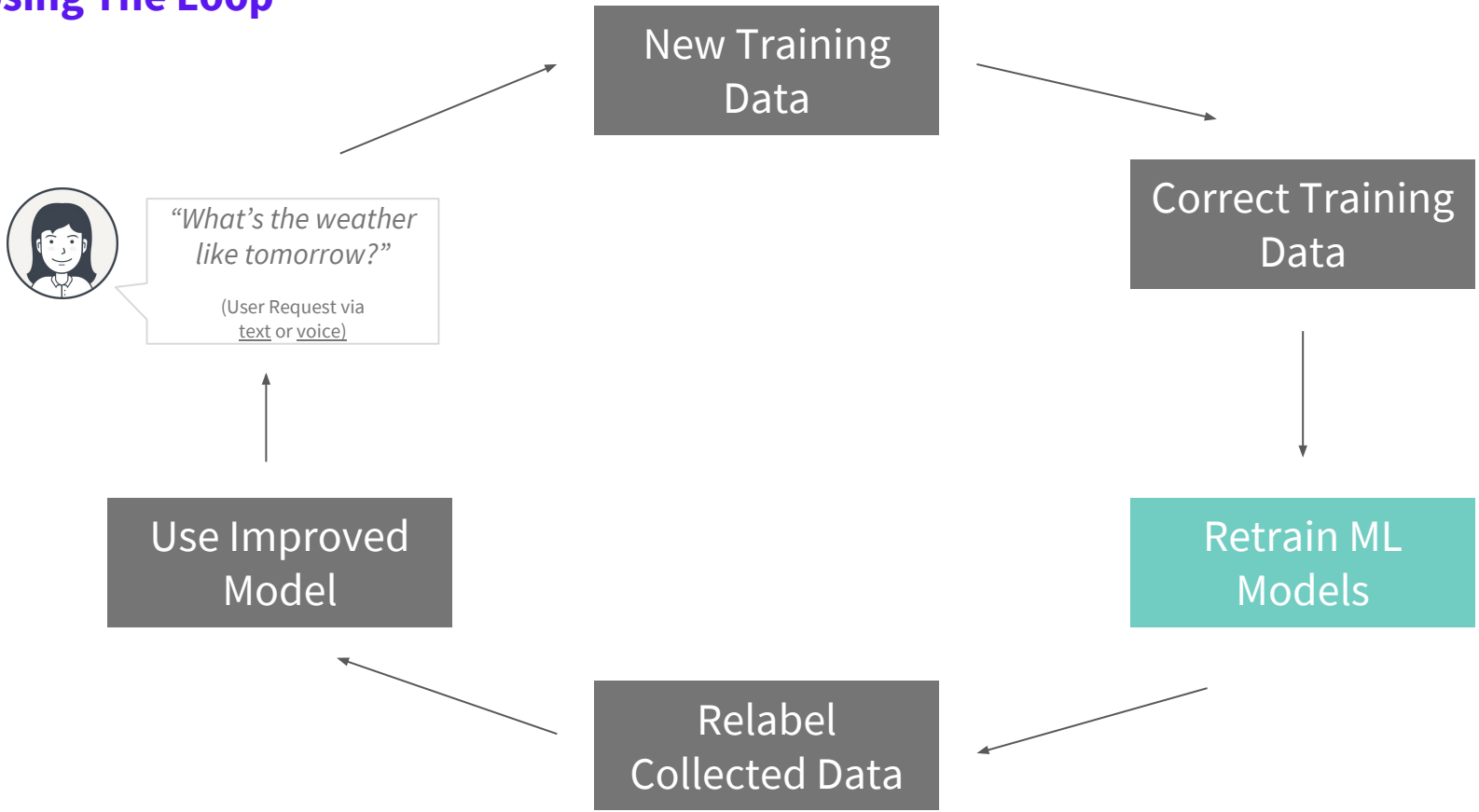
```
## hotel explain 1.3
* request_hotel
    - utter_ask_details
* inform{"location": "paris"}
    - utter_ask_people
* inform{"people": "4"}
    - utter_ask_price
* explain
    - utter_explain_price_hotel
    - utter_ask_price
```

```
## restaurant explain 1.3
* request_restaurant
    - utter_ask_details
* inform{"location": "paris"}
    - utter_ask_people
* inform{"people": "4"}
    - utter_ask_price
* explain
    - utter_explain_price_restaurant
    - utter_ask_price
```

Final Thoughts

Final Thoughts

Closing The Loop



Open challenges

For those that are curious:

- Handling OOV words
- Multi language entity recognition
- Combination of dialogue models

We’re constantly working on improving our models!

Current Research

Good reads for a rainy day:

- Last Words: Computational Linguistics and Deep Learning (blog)
  https://goo.gl/lGSRuj
- Starspace Embeddings (paper)
  https://arxiv.org/abs/1709.03856
- End-to-End dialogue system using RNN (paper)
  https://arxiv.org/pdf/1604.04562.pdf
- MemN2N in python (github)
  https://github.com/vinhkhuc/MemN2N-babi-python
- Sentence Embeddings (blog)
  https://medium.com/huggingface/universal-word-sentence-embeddings-ce48ddc8fc3a

Summary

4 take home thoughts:

- Techniques to handle small data sets are key to get started with conversational AI
- Deep ML techniques help advance state of the art NLU and conversational AI
- Combine ML with traditional Programming and Rules where appropriate
- Abandon flow charts

Get in touch!

Justina Petraityte
Developer Advocate

juste@rasa.ai
@juste_petr

We are hiring!

ML Product Success Engineer

Help the teams who are using Rasa Platform succeed.

ML Engineer

Help us push the limits of the conversational AI software.