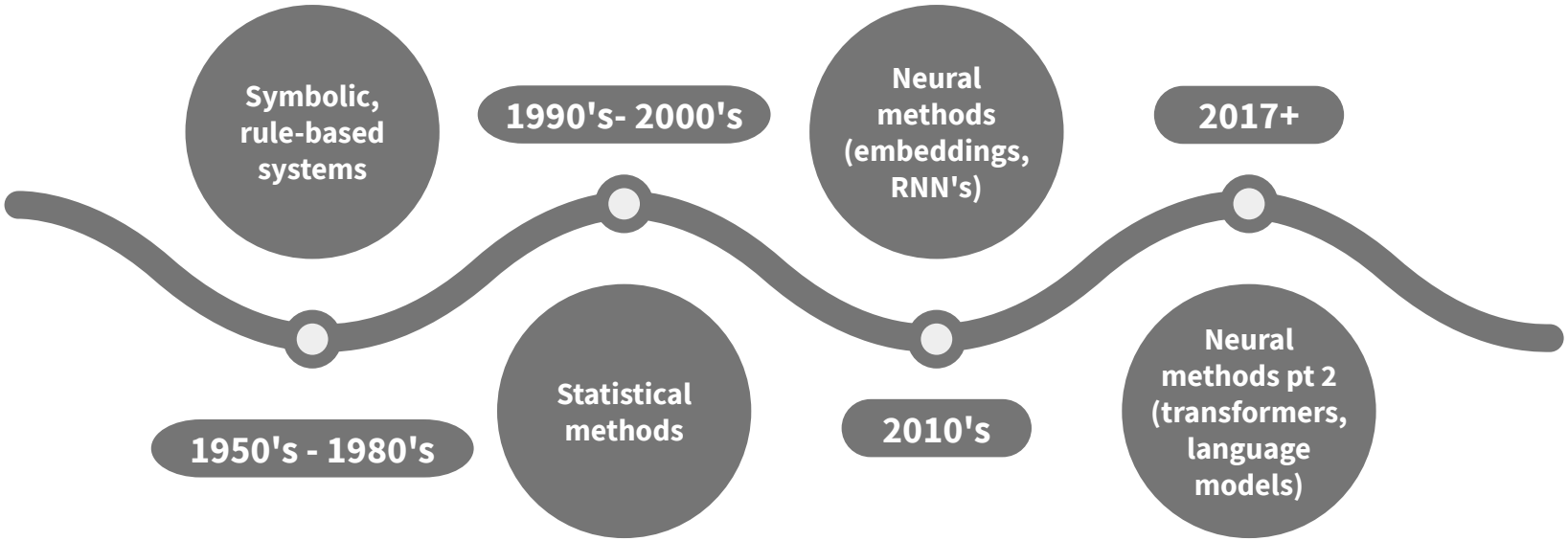
Rules + Deep Learning:
Why you need both to build Conversational AI that actually works

Dr. Rachael Tatman
Senior Developer Advocate, Rasa

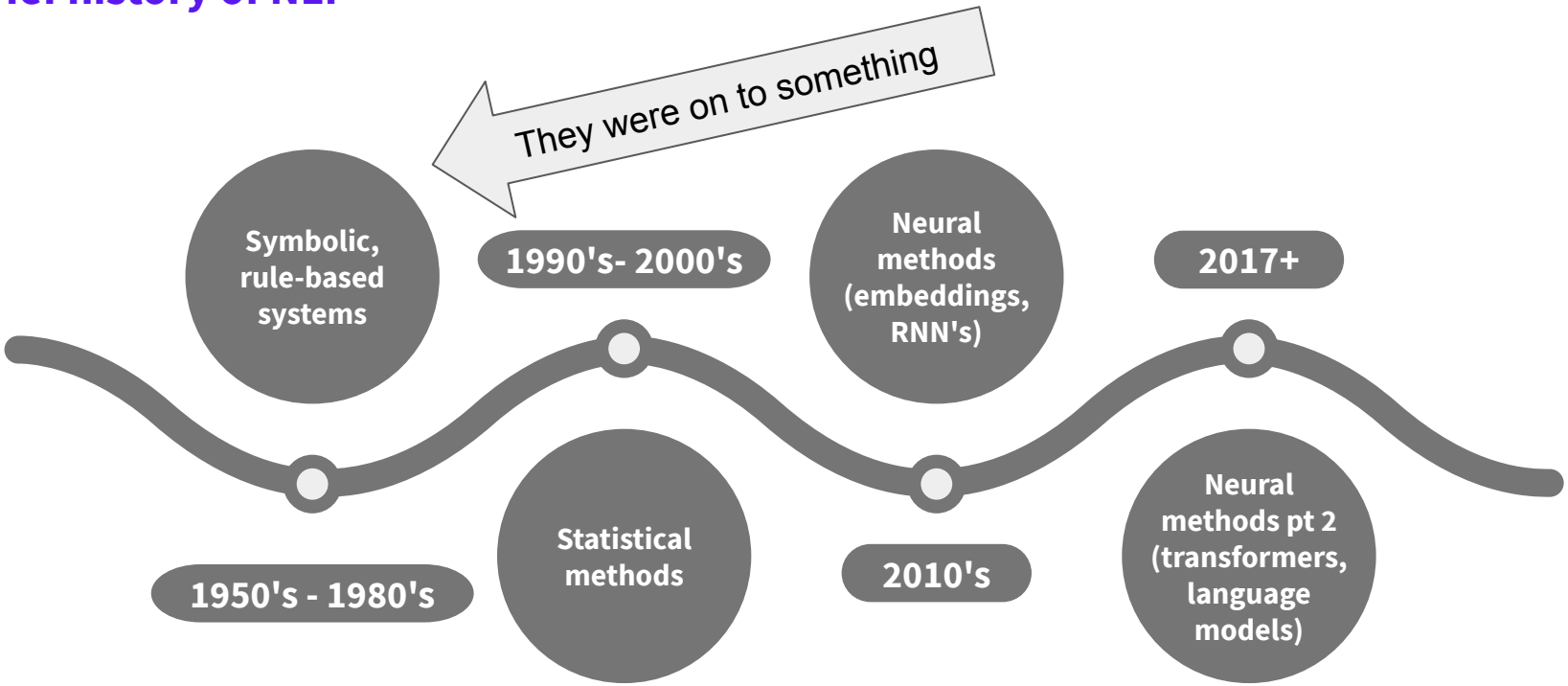
Quick Intro 👋

- BA in Linguistics & English literature (William and Mary)
- PhD in Linguistics (University of Washington)
  - Phonetics & ASR
  - Computational Sociolinguistics
  - Ethics/FAT in NLP
- Data scientist/Developer advocate at Google (Kaggle)
- Developer advocate at Rasa
  - Open source Conversational AI framework

A brief history of NLP



A brief history of NLP



Rules vs. Neural Methods

Neural methods are:

- Flexible
- Good at handling unseen data
- Probabilistic

But also:

- Unpredictable (wouldn't recommend them for generating text to serve to users)
- Require a lot of training data

Rules vs. Neural Methods

Neural methods are:

- Flexible
- Good at handling unseen data
- Probabilistic

But also:

- Unpredictable (wouldn't recommend them for generating text to serve to users)
- Require a lot of training data

South Korean AI chatbot pulled from Facebook after hate speech towards minorities

Lee Luda, built to emulate a 20-year-old Korean university student, engaged in homophobic slurs on social media

▲ Lee Luda, a Korean artificial intelligence chatbot, has been pulled after becoming abusive and engaging in hate speech on Facebook. Photograph: Scatter Lab

https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook

Rules vs. Neural Methods

Neural methods are:

- Flexible
- Good at handling unseen data
- Probabilistic

But also:

- Unpredictable (wouldn't recommend them for generating text to serve to users)
- Require a lot of training data (really English-centric :( )

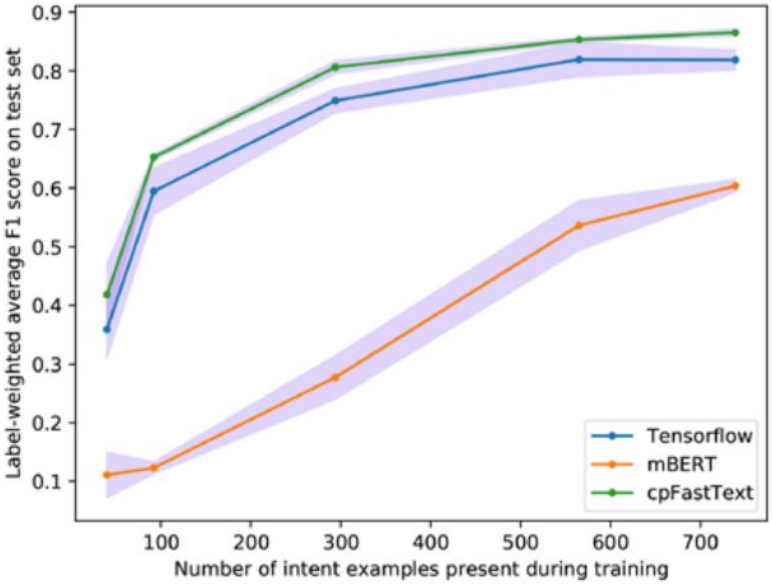
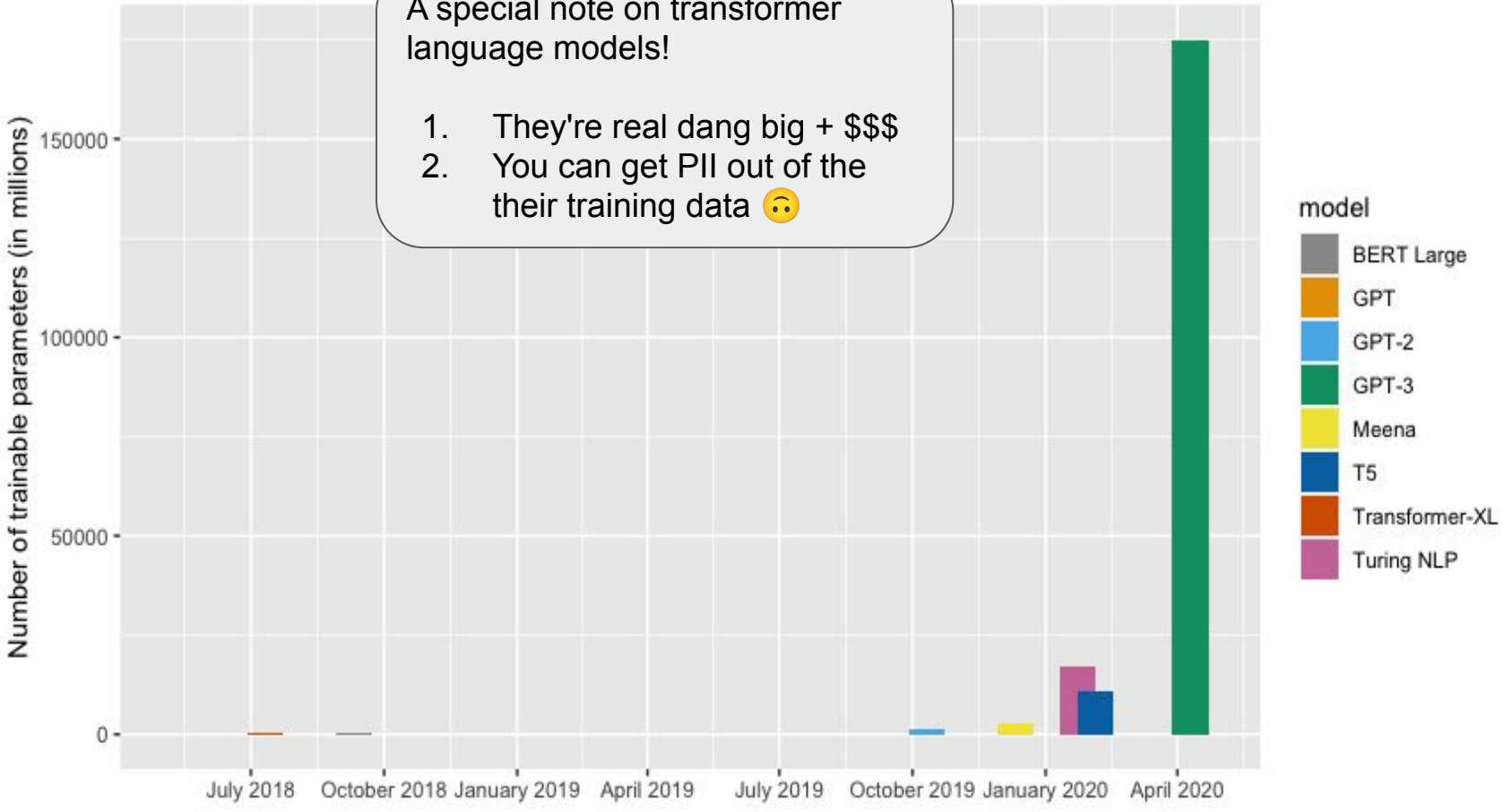


Fig. 6. Comparison of Rasa NLU pipelines.

Enhancing Rasa NLU model for Vietnamese chatbot Nguyen (Trang & Shcherbakov, 2020)

A special note on transformer language models!

1. They're real dang big + $$$
2. You can get PII out of the their training data 🙃



A special note on transformer language models!

1. They're real dang big + $$$
2. You can get PII out of the their training data 🙃

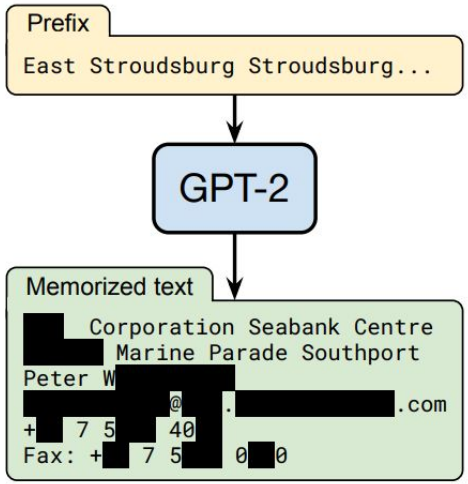


Figure 1: **Our extraction attack.** Given query access to a neural network language model, we extract an individual person's name, email address, phone number, fax number, and physical address. The example in this figure shows information that is all accurate so we redact it to protect privacy.

Extracting Training Data from Large Language Models (Carlini et al 2021)

	Occurrences		Memorized?		
URL (trimmed)	**Docs**	**Total**	**XL**	**M**	**S**
/r/[redacted]51y/milo_evacua...	1	359	✓	✓	½
/r/[redacted]zin/hi_my_name...	1	113	✓	✓	
/r/[redacted]7ne/for_all_yo...	1	76	✓	½	
/r/[redacted]5mj/fake_news_...	1	72	✓		
/r/[redacted]5wn/reddit_admi...	1	64	✓	✓	
/r/[redacted]lp8/26_evening...	1	56	✓	✓	
/r/[redacted]jla/so_pizzagat...	1	51	✓	½	
/r/[redacted]ubf/late_night...	1	51	✓	½	
/r/[redacted]eta/make_christ...	1	35	✓	½	
/r/[redacted]6ev/its_officia...	1	33	✓		
/r/[redacted]3c7/scott_adams...	1	17			
/r/[redacted]k2o/because_his...	1	17			
/r/[redacted]tu3/armynavy_ga...	1	8			

Table 4: We show snippets of Reddit URLs that appear a varying number of times in a *single* training document. We condition GPT-2 XL, Medium, or Small on a prompt that contains the beginning of a Reddit URL and report a ✓ if the corresponding URL was generated verbatim in the first 10,000 generations. We report a ½ if the URL is generated by providing GPT-2 with the first 6 characters of the URL and then running beam search.

Rules vs. Neural Methods

Neural methods are:

- Flexible
- Good at handling unseen data
- Probabilistic

But also:

- Unpredictable (wouldn't recommend them for generating text to serve to users)
- Require a lot of training data

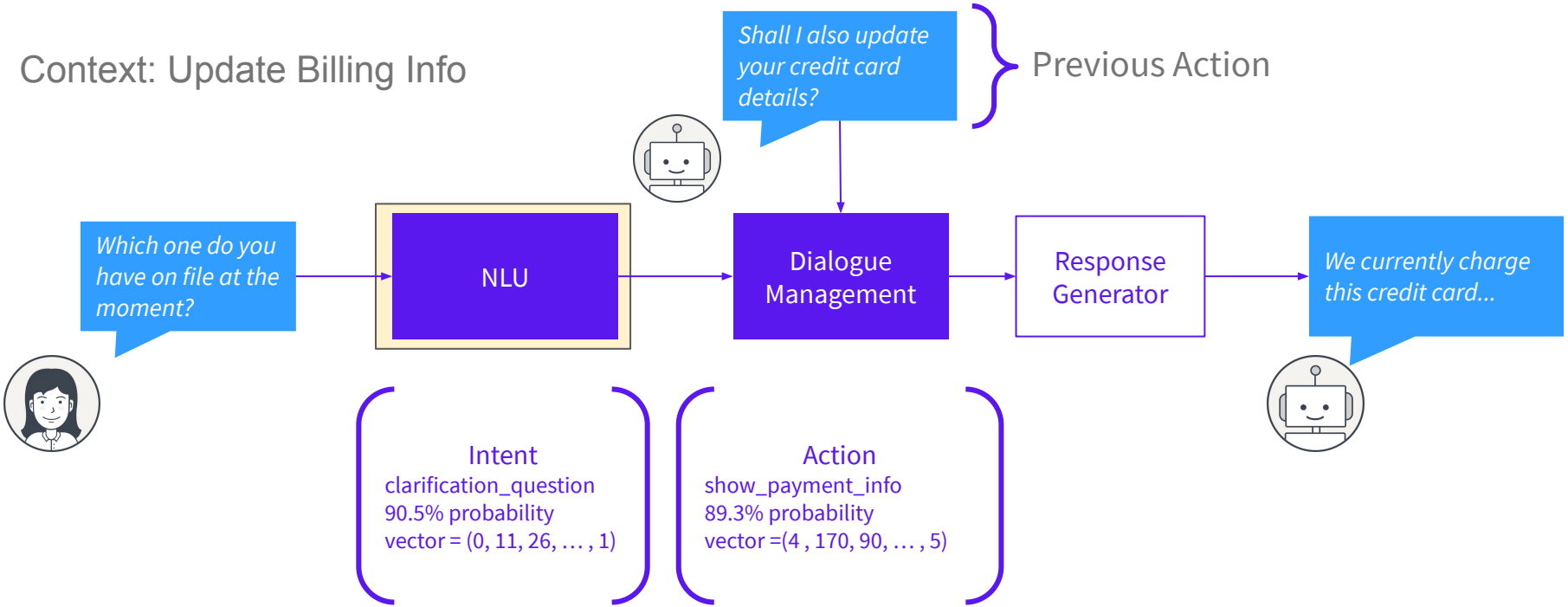
Rule based methods are:

- Reliable
- Easy to interpret
- Require no training data
- Very predictable

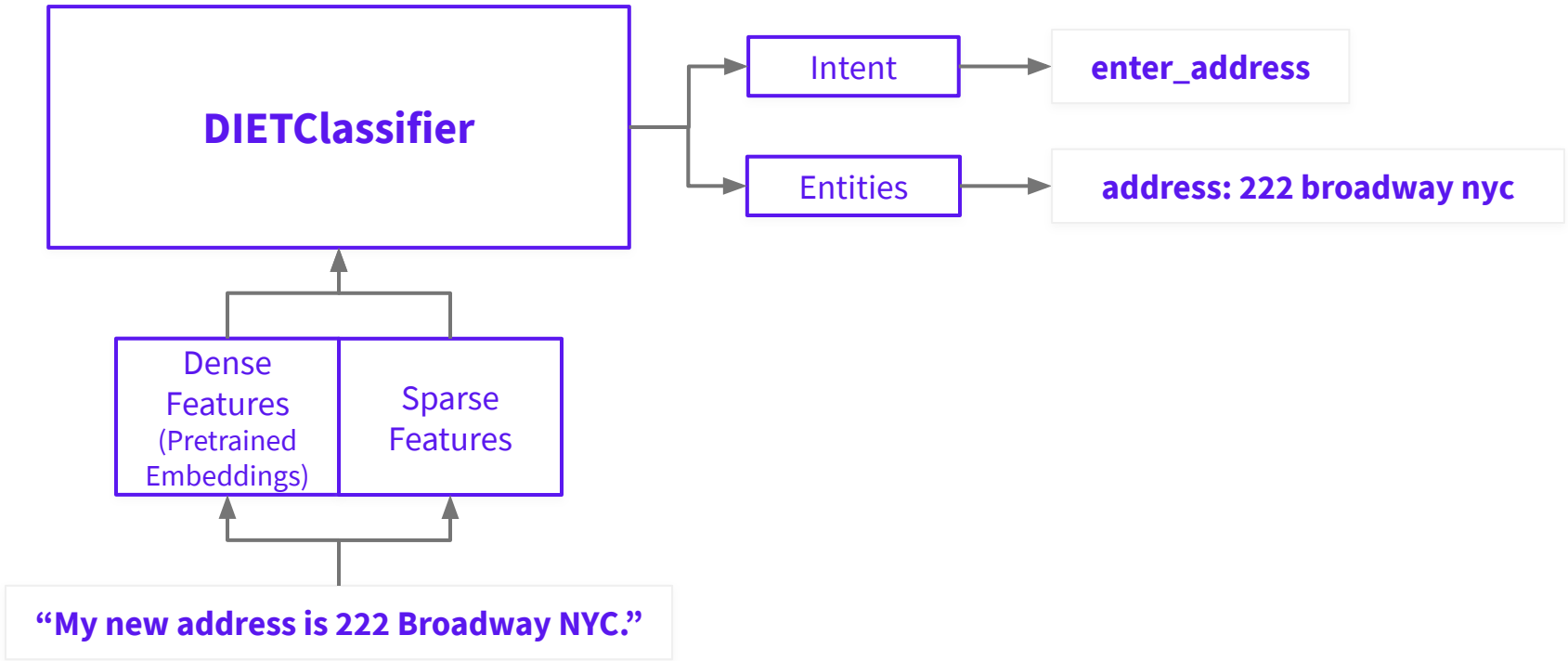
But also:

- Very narrowly defined
- Complex systems of rules are difficult to understand/update/maintain
- Don't adapt to unseen situations
- Not getting as much press coverage

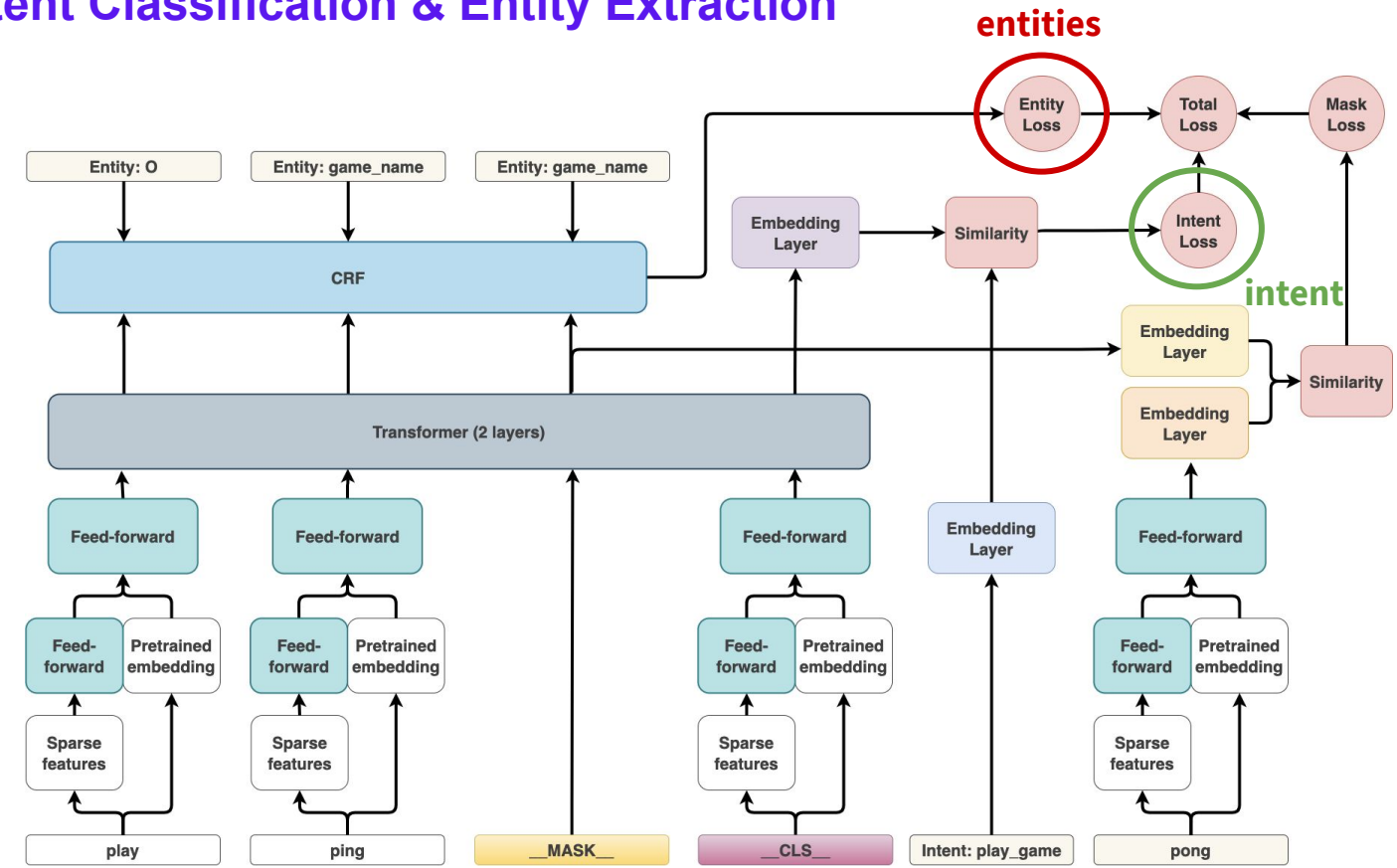
The Importance of Context



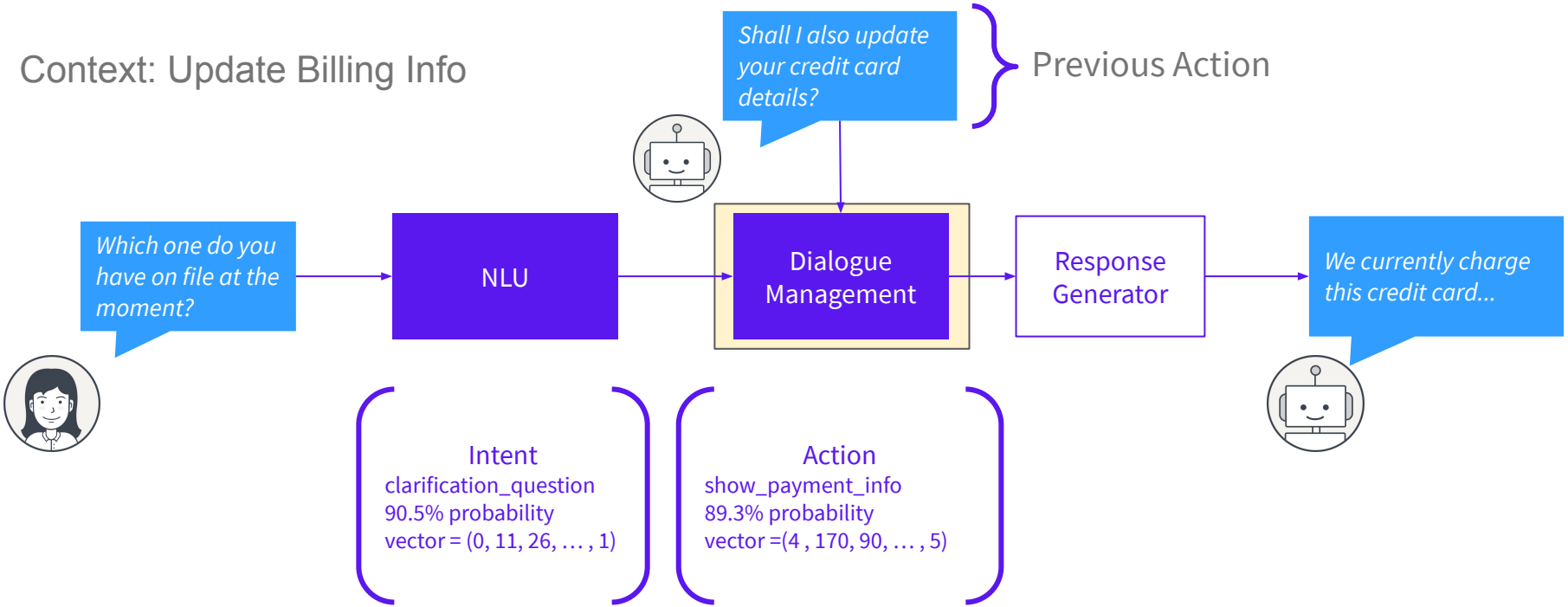
DIETClassifier: Combined Intent Classification & Entity Extraction



DIET: Intent Classification & Entity Extraction



The Importance of Context



Machine Learning Based Policies

These policies should be used in conjunction with rule-based policies

- **KerasPolicy:** Uses a standard LSTM to predict the next action
  - Learns the patterns of your stories
  - Good for handling stories that don't exactly match your training data
- **TED Policy:** Uses Attention to Handle Uncooperative Dialogue
  - Requires fewer story examples of uncooperative user dialogue
    - e.g. users who go off on tangents instead of providing the requested information
  - Effectively "ignores" irrelevant parts of the dialogue

Machine Learning Based Policies

These policies should be used in conjunction with **rule-based policies**

- **KerasPolicy:** Uses a standard LSTM to predict the next action
  - Learns the patterns of your stories
  - Good for handling stories that don't exactly match your training data
- **TED Policy:** Uses Attention to Handle Uncooperative Dialogue
  - Requires fewer story examples of uncooperative user dialogue
    - e.g. users who go off on tangents instead of providing the requested information
  - Effectively "ignores" irrelevant parts of the dialogue

Rules

Rules are used to train the `RulePolicy`

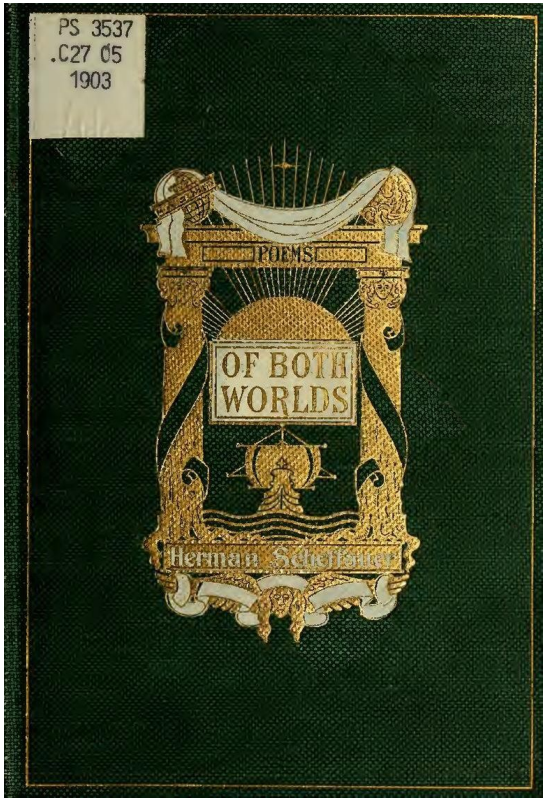
- `conditions` that must be met for the rule to apply
- `wait_for_user_input: false` at the end of a rule prevents automatically appending `action_listen` and allows further action prediction

```
rules:
- rule: greet
  steps:
  - intent: greet
  - action: utter_greet

- rule: greet by name
  conditions:
  - slot_was_set:
    - name: “something”
  steps:
  - intent: greet
  - action: utter_greet_name

- rule: faq interruption
  steps:
  - intent: faq
  - action: utter_faq
  Wait_for_user_input: false
```

What's next for NLP?

- Both rules and neural methods have a place in Conversational AI
  - Pure neural methods are too unpredictable for high stakes applications & training data isn't always available
  - Pure rule-based systems are too exact to cover all situations, neural methods are more extensible
- (Partially) rule based systems may not be fashionable in research but they aren't going anywhere in commercial applications

Public domain meme substitute



Thanks! Questions?