

# IT'S TRICKY: CHATBOTS & QA



Lina Zubyte  
QA Consultant at ThoughtWorks®  
 @buggylina

A young girl with dark hair, wearing a pink floral hoodie and a matching backpack, reaches out to touch the hand of a white humanoid robot. The robot's hand is adorned with a branch of pink cherry blossoms. The background is a blurred indoor setting with warm, colorful lights.

# PERCEPTION



chatbots are



chatbots are **they really useful**

chatbots are **dead**

chatbots are **also known as**

chatbots are **the future**

chatbots are **the new apps**

chatbots are **useless**

chat bots are **stupid**

chatbots are **capable to understand the**

chatbots are **not ai**

chatbots are **not the future**

A **chatbot** is a computer program or an artificial intelligence which conducts a conversation via auditory or textual methods.

## Popular



Robbie Williams

⚡ Automated messaging

Chat with Robbie Williams!



beziehungsweise

⚡ Automated messaging

Dein News-Update rund um die Liebe



Job Bot

⚡ Automated messaging

Let the bot find a new job for you.



Jason Derulo

⚡ Automated messaging

Get close and personal with Jason Derulo



ShoeDazzle

⚡ Automated messaging

ShoeDazzle is the first monthly fashion ser...



ZAFUL

⚡ Automated messaging

The Official Zaful Facebook: hot fashion....

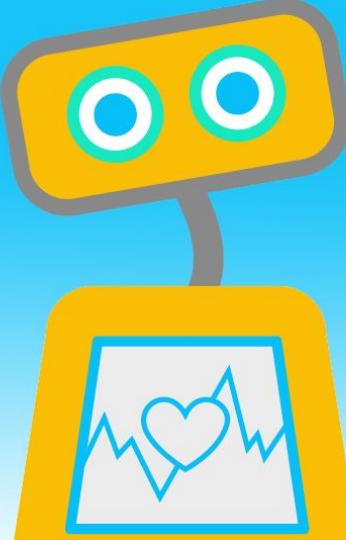
See More

# Chatbot lawyer overturns 160,000 parking tickets in London and New York



News > Health > Apple

## The free app that can help you conquer depression and anxiety



Hi, I'm Woebot.

I'm here for you, 24/7. No couches, no meds, no childhood stuff. Just strategies to improve your mood. Plus the occasional dorky joke. 😊



Home



People



Games



Discover

# **SOME TYPES OF CHATBOTS**

**Menu/Button-Based**

**Keyword Recognition-Based**

**Contextual using Machine Learning**

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.

This implementation by Norbert Landsteiner 2005.

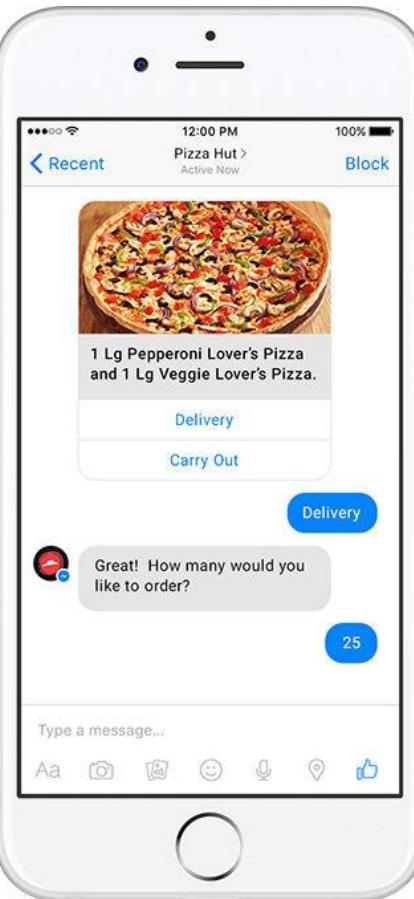
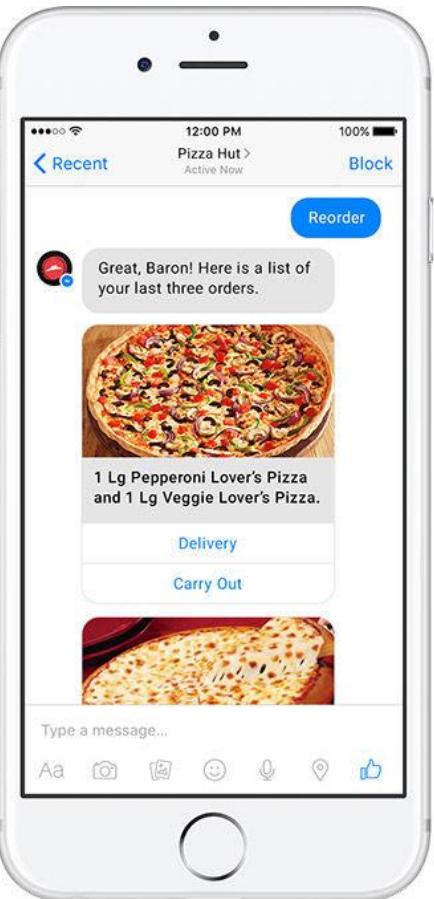
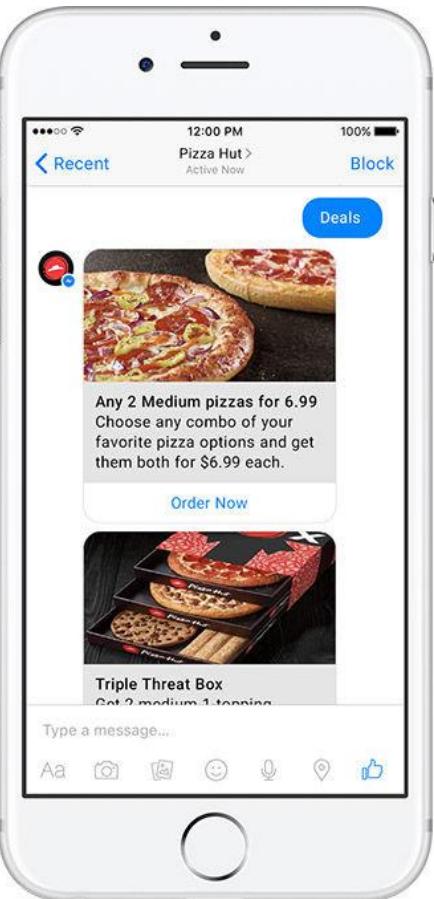
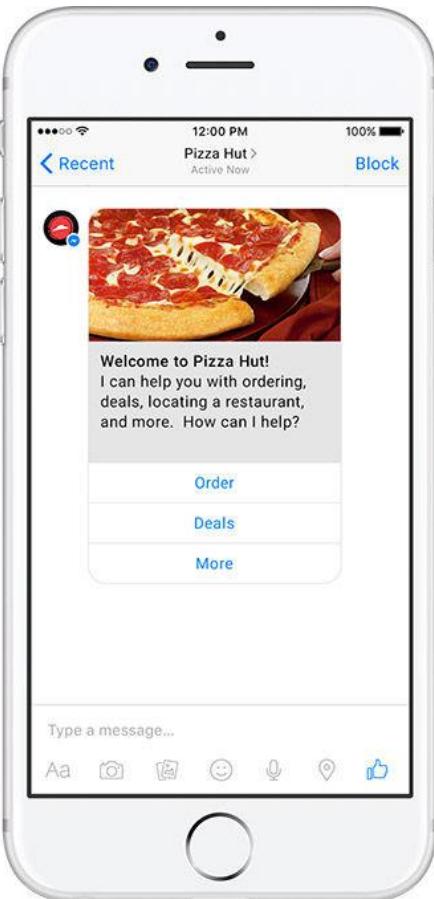
ELIZA: Is something troubling you ?

YOU: I would like to introduce you to the AADays audience - how should I do that?

ELIZA: Do you would like to introduce anyone else ?

YOU: No, only you.

ELIZA: You are being a bit negative.



# BALU

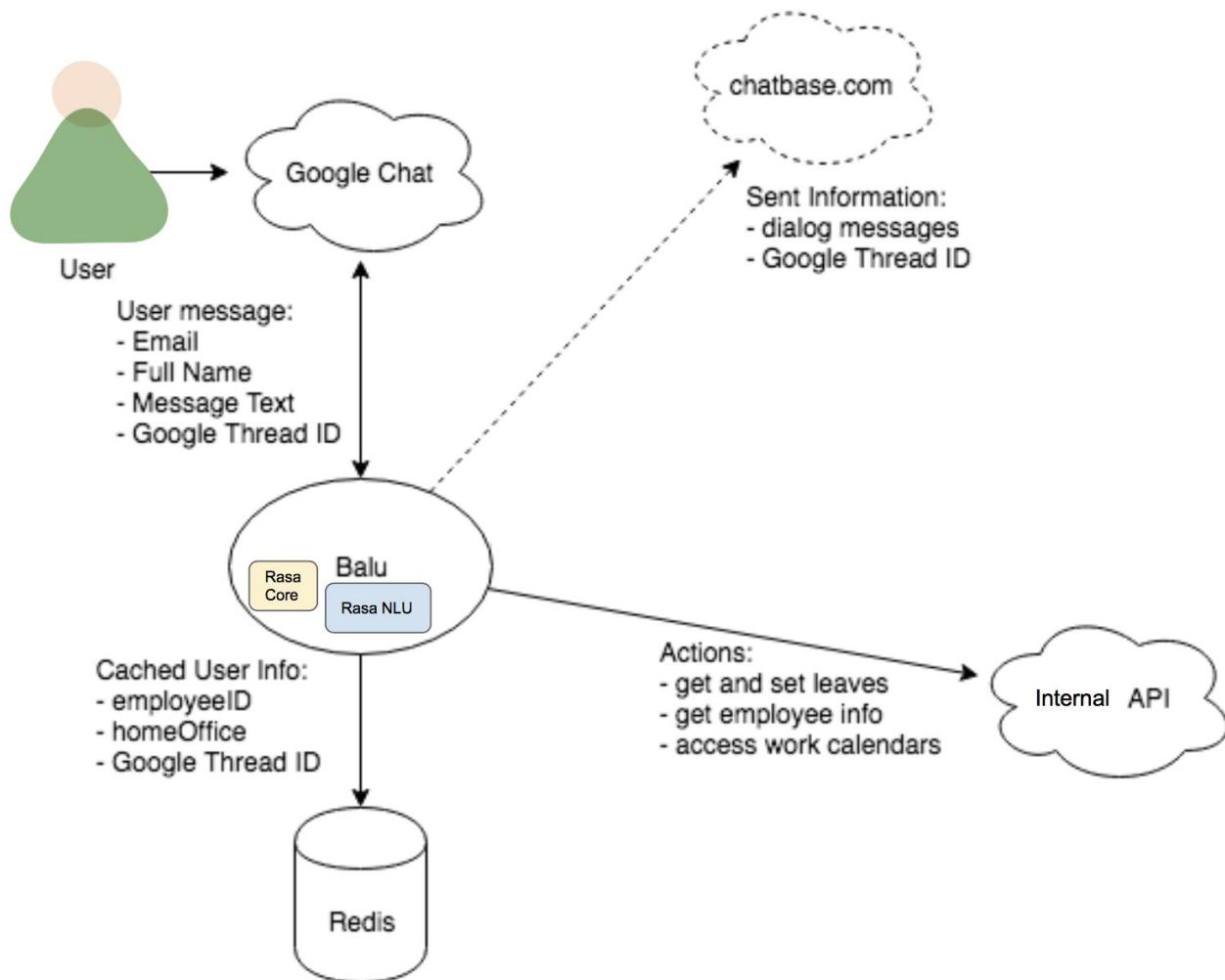


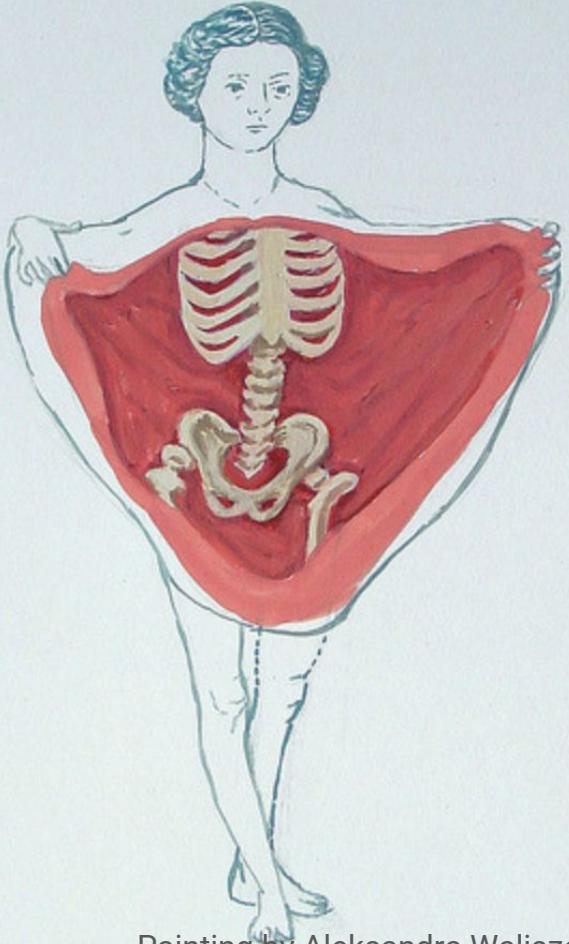
Balu [staging] BOT Jun 11, 11:20 AM

Hello! I'm Balu, your TWDE chatbot 😊

You can ask me about your total leave days for this year, what leaves you have planned, how many leaves you have already taken, what the next public holidays are and you are also able to book/cancel your upcoming vacation.

## Internal vacation handling chatbot using machine learning





Painting by Aleksandra Waliszewska

# Intents

# Entities

**data/nlu.md +  
data/domain.yaml**

# displaCy Named Entity Visualizer

the topic more. In this session, I will share my learnings on chatbots and discuss these questions: What should we think about when we build chatbots? How can we as QAs help to ensure the quality of a chatbot? What in the end is a good quality chatbot? Join me in my talk and get ready for the future of chatbots.

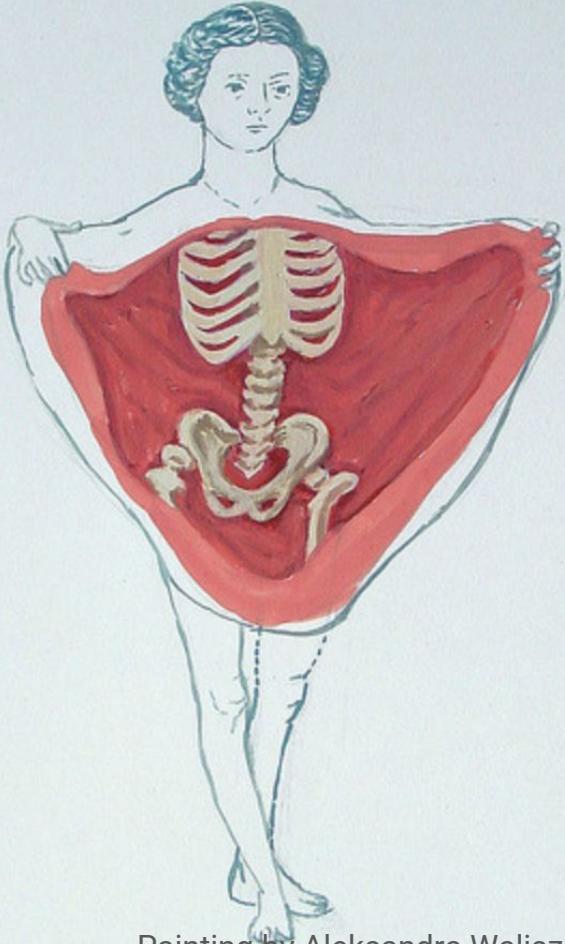


Model ?

English - en\_core\_web\_sm (v2.0.0)



In 2016 DATE, as part of a survey run by Oracle PERSON regarding tech trends, over 80% PERCENT of businesses answered that they already have or will implement a chatbot by 2020 DATE. A few years DATE passed, the craze for chatbots is not as sky-rocketing, but the trend is still there: more and more businesses are experimenting with chatbots, especially, when it comes to customer-facing services. However, with chatbots comes a huge amount of uncertainty.



Painting by Aleksandra Waliszewska

# Intents

# Entities

# SLOTS

# TEMPLATES

# ACTIONS

**data/nlu.md +  
data/domain.yaml**

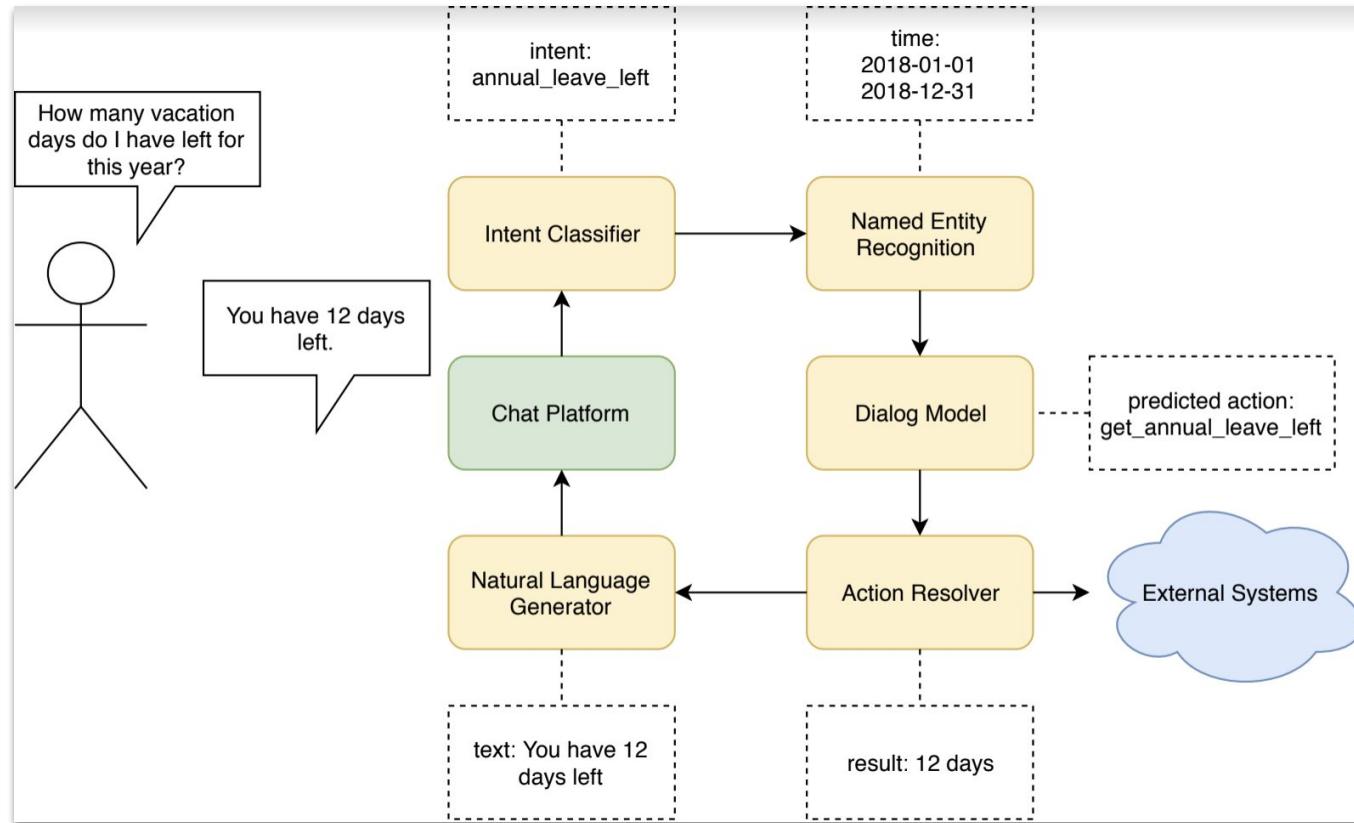
## Book annual leave with from and to dates provided

---

- greet
  - utter\_greet
- leave\_annual\_take{"time": {"to": "2018-07-01T00:00:00.000Z", "from": "2017-06-02T00:00:00.000Z"}}
  - slot{"time": {"to": "2018-07-01T00:00:00.000Z", "from": "2017-06-02T00:00:00.000Z"}}
  - action\_get\_booking\_interval
  - slot{"start\_date": "2017-06-02T00:00:00.000Z"}
  - slot{"end\_date": "2018-07-01T00:00:00.000Z"}
  - utter\_leave\_annual\_take\_confirm
- confirm
  - action\_book\_leave
  - slot{"start\_date": null}
  - slot{"end\_date": null}
  - slot{"time": null}
- thanks
  - utter\_thanks
  - action\_restart

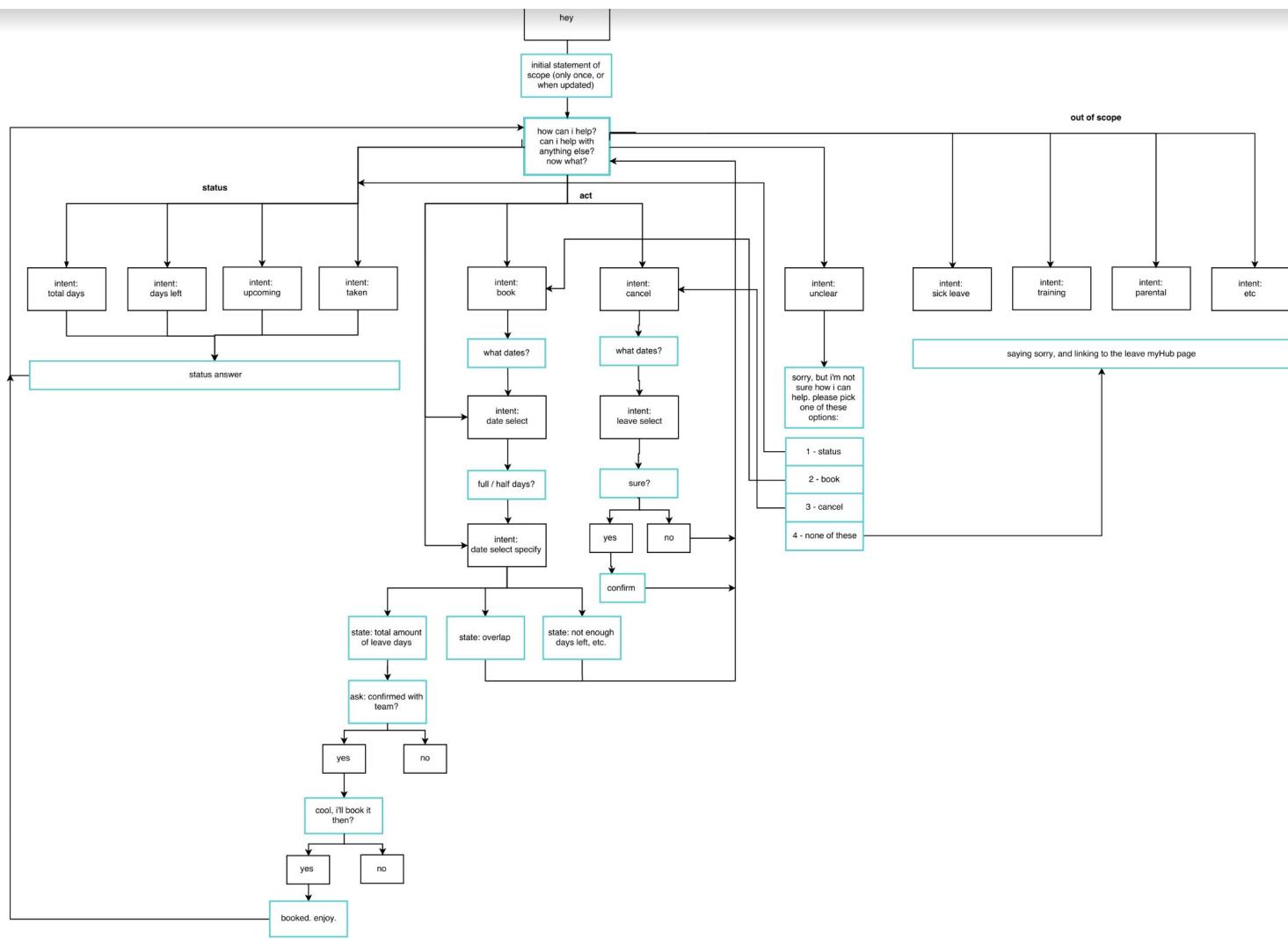
**data/stories.md**

# BALU EXAMPLE

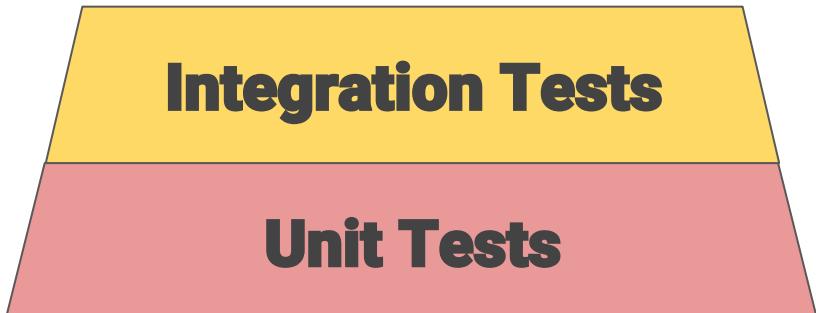




**HOW DO YOU TEST IT?**



# BALU QUALITY FEEDBACK



**Integration Tests**

**Unit Tests**

# INTEGRATION TESTS



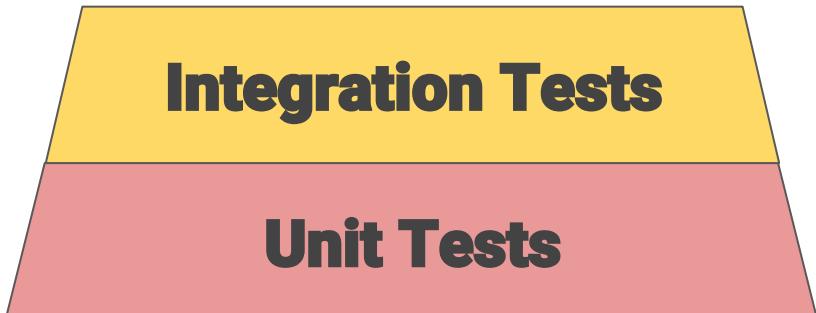
```
cancel_leaves.json
1  {
2    "I'd like to cancel some vacation?": {
3      "templates": [
4        "utter_choose_leave_annual_to_be_cancelled"
5      ],
6      "messages": [
7      ]
8    },
9    "16344": {
10      "templates": [
11        "utter_confirm_leave_annual_to_be_cancelled"
12      ]
13    },
14    "yes": {
15      "templates": [
16        "utter_cancel_leave_done"
17      ]
18    }
19  }
```

# INTEGRATION TESTS

```
@freeze_time("2018-04-24")
def test_cancel_leaves(mocker):
    mock_external_systems(mocker)
    helpers.conversation_tester(
        client, 'cancel_leaves.json',
        template_filler={
            'leave': '16 Feb 2018 - 17 Feb 2018',
        },
    )
```

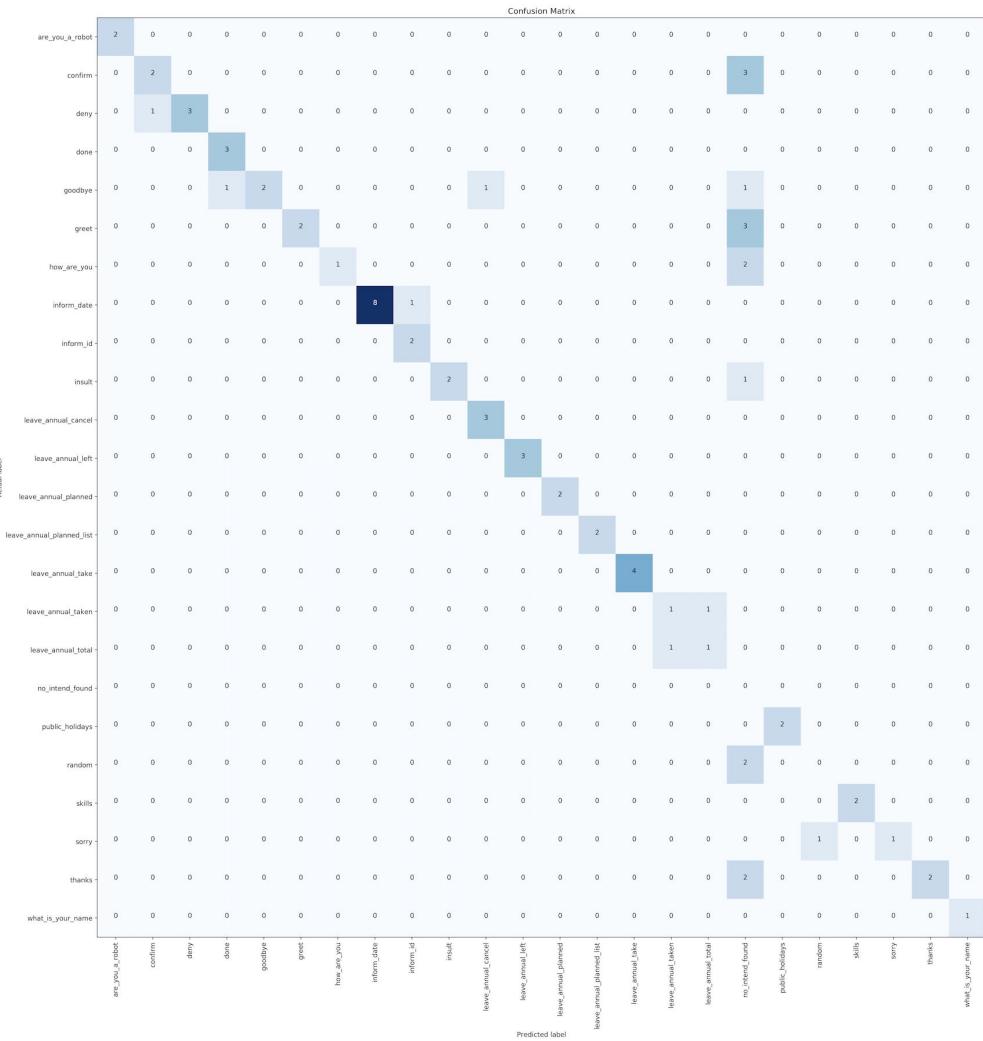
# BALU QUALITY FEEDBACK

?

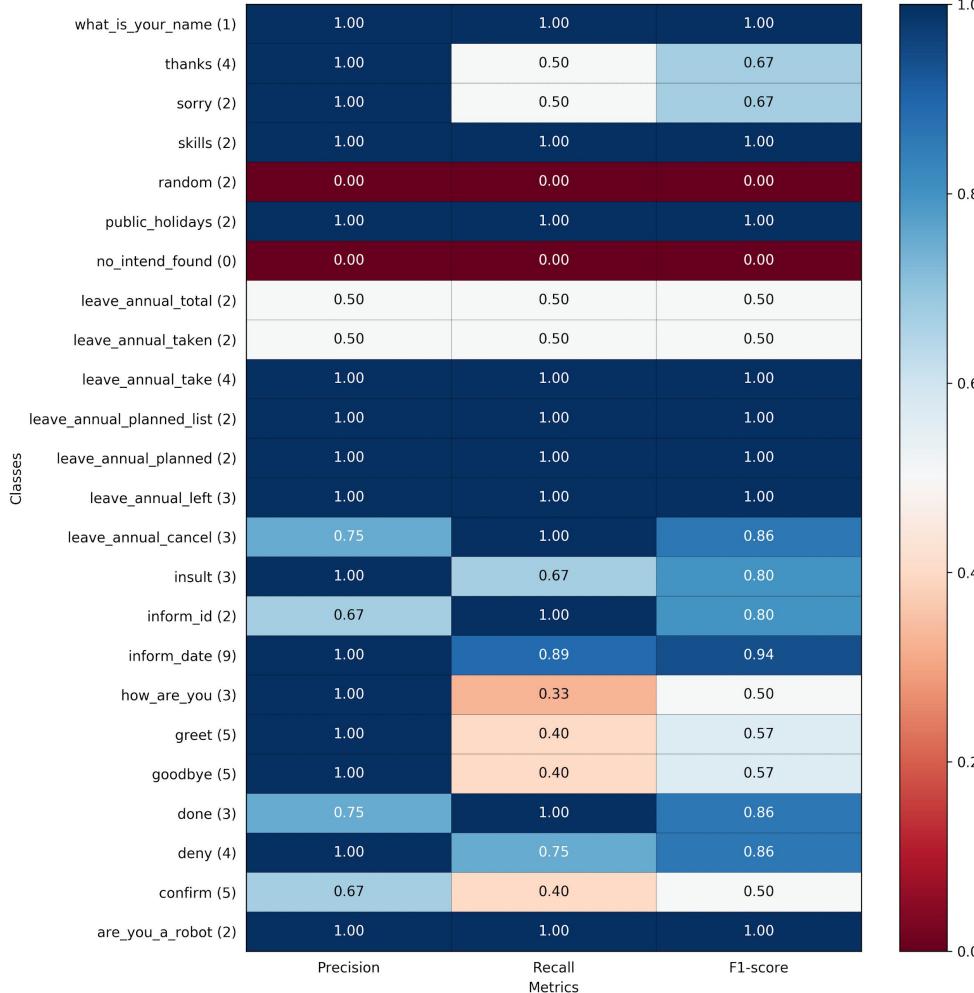


Integration Tests

Unit Tests



Intent Classification Report



# WHAT'S TRICKY?



Lina Zubyte Now

hi



Balu [staging] BOT Now

Hi! What can I do for you?



Lina Zubyte Now

I want a vacation



Balu [staging] BOT Now

Am I? Kidding. Yes.

# WHAT'S TRICKY?

```
'intent': {'confidence': 0.949816107749939, 'name': 'inform_date'},  
'intent_ranking': [{"confidence": 0.949816107749939, "name": "inform_date"},  
 {"confidence": 0.18554848432540894, "name": "greet"},  
 {"confidence": 0.17427736520767212, "name": "inform_id"},  
 {"confidence": 0.17015646398067474, "name": "random"},  
 {"confidence": 0.08837316930294037,  
  "name": "what_is_your_name"},  
 {"confidence": 0.07831726968288422,  
  "name": "leave_annual_taken"},  
 {"confidence": 0.0738229900598526,  
  "name": "leave_annual_planned"},  
 {"confidence": 0.05114135146141052, "name": "goodbye"},  
 {"confidence": 0.03761787712574005, "name": "skills"},  
 {"confidence": 0.03345265984535217, "name": "deny"}],  
'text': '1 july'}
```

# CHALLENGES

- Lack of dialog flow change metrics
- Flaky integration tests
- There are no clear automation solutions to make sure quality of dialogs is good

# EXPLORATORY TESTING



Balu [staging] BOT Now NEW

Did you check with your team?



Lina Zubyte Now NEW

yep



Balu [staging] BOT Now NEW

I'm sorry, I don't understand what you mean.



**Balu [staging]** BOT 6 mins •

Great, I booked the following leave days for you: 26 May 2018 - 01 Aug 2018, that counts for 48.0 day(s).



**Lina Zubyte** Now **NEW**

how many days have I planned



**Balu [staging]** BOT Now **NEW**

You have 48.0 days planned. Would you like to see the list of your upcoming leaves?



**Lina Zubyte** Now **NEW**

how many days I have left



**Balu [staging]** BOT Now **NEW**

You have -20.0 leave day(s) left for 2018



Lina Zubyte Now •

hi



Balu [staging] BOT Now •



Lina Zubyte Now •

damn



Balu [staging] BOT Now •



Lina Zubyte Now •

Balu



Balu [staging] BOT Now •



Lina Zubyte Now •

no good



Balu [staging] BOT Now NEW

Welcome back! What can I do for you?

**“YOU DON’T TEST THE  
CHATBOT -  
THE CHATBOT TESTS  
YOU”**

# QUALITY PRINCIPLES

# FUNCTIONALITY IS THE PRIORITY



Balu [staging] BOT 1 min •

Hey there! How can I help you?



Lina Zubyte 1 min •

I'd like to book a vacation



Balu [staging] BOT 1 min •

Starting when?



Lina Zubyte 1 min •

august 7



Balu [staging] BOT 1 min •

When do you want your vacation to end?



Lina Zubyte 1 min •

august 17



Balu [staging] BOT 1 min •

To?



Lina Zubyte 1 min •

august 17



Balu [staging] BOT 1 min •

To?



Lina Zubyte 1 min •

august 17

# USER TESTING + EXPLORATORY TESTING



Lina Zubyte Now NEW

hi I'd like to book a vacation



Balu [staging] BOT Now NEW

From when?



Lina Zubyte Now NEW

end of September



Balu [staging] BOT Now NEW

To?



Lina Zubyte Now NEW

end of November



Balu [staging] BOT Now NEW

Did you check with your team?



Lina Zubyte Now NEW

yes



Balu [staging] BOT Now NEW

Great, I booked the following leave days for you: 01 Sep 2018 - 01 Nov 2018, that counts for 43.0 day(s).

# FOCUS ON CHARACTER & COMPREHENSION



Lina Zubyte Now •  
bye



Balu [staging] BOT  
A large yellow smiley face emoji with a wide open mouth showing white teeth and black outlines for the eyes.



Lina Zubyte Now •



Balu [staging] BOT Now •

Let me know what I can do for you!

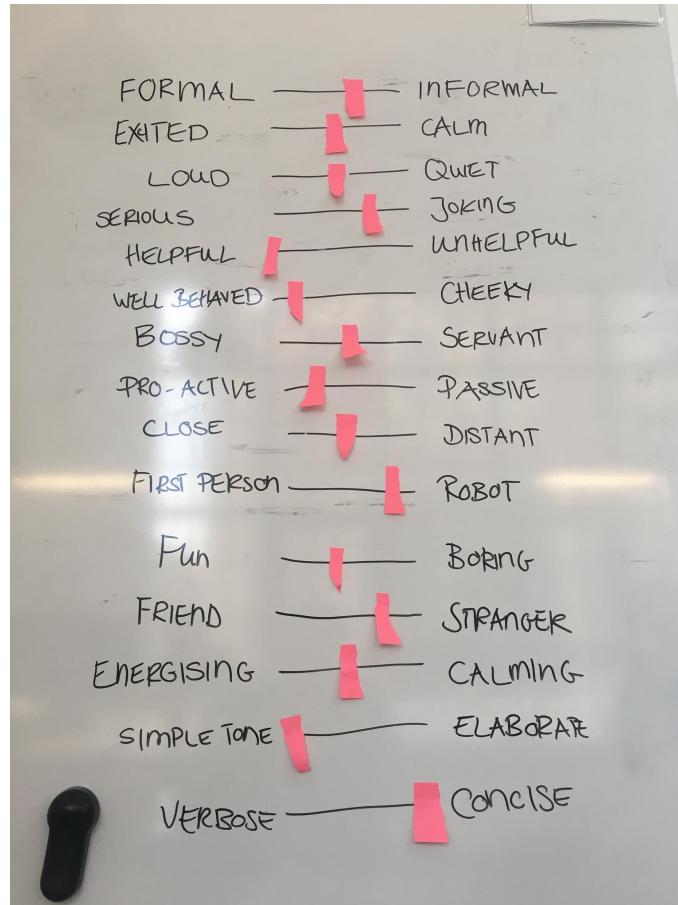


Lina Zubyte Now •  
sad



Balu [staging] BOT Now •  
Lovely, thanks.

# DEFINE THE CHARACTER & ADJUST UTTERS



# DIALOG MATTERS



Balu [staging] BOT Now NEW

Alright then. Anything else I can help you with?



Lina Zubyte Now NEW

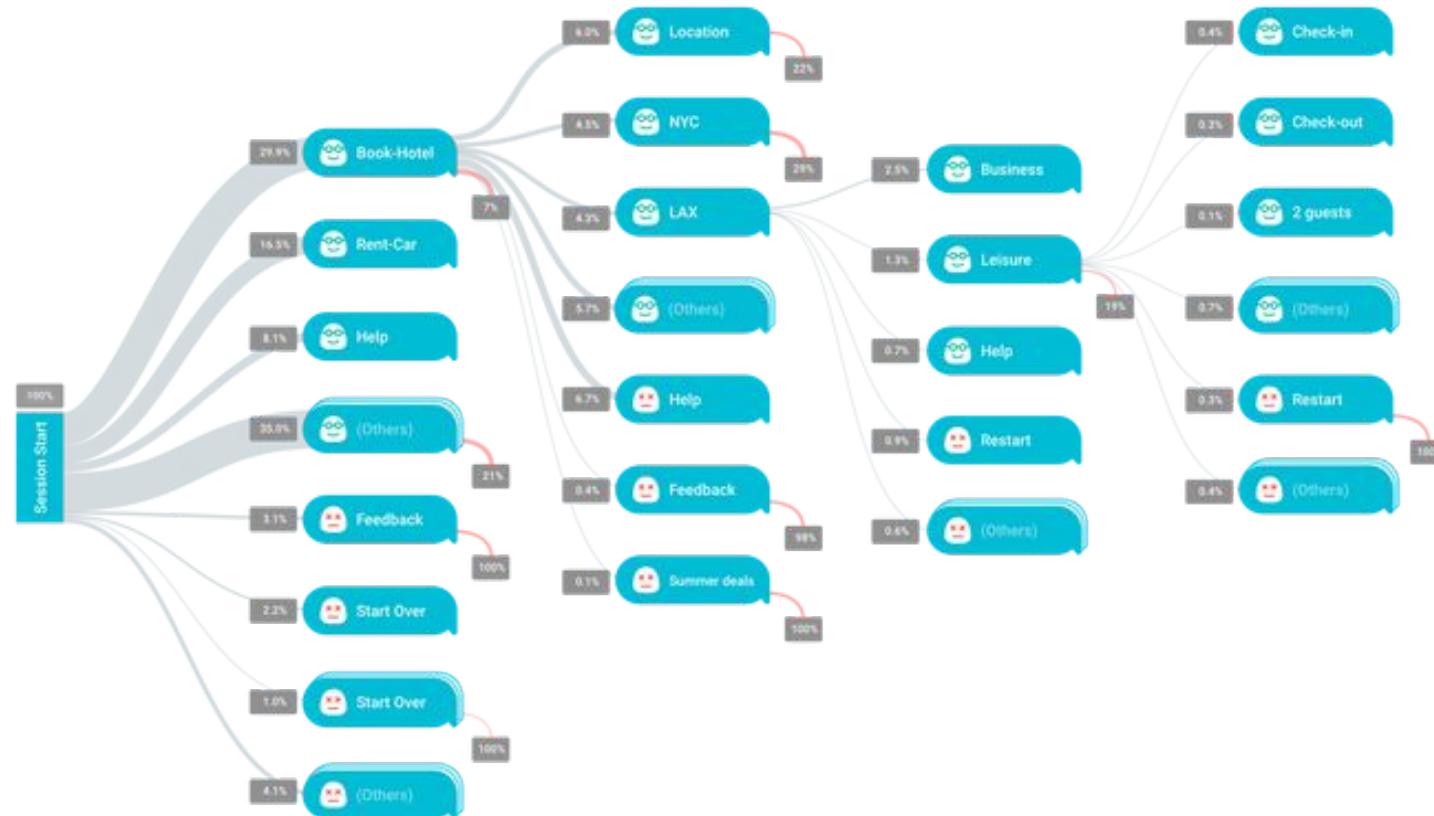
yes



Balu [staging] BOT Now NEW

You don't have any vacation planned currently.

# DIALOG METRICS & STRATEGIC ANALYSIS



# QUALITY TABLE

**Monitoring & Strategic Analysis**

**Exploratory Testing**

**Machine Learning Metrics**

**Integration Tests**

**Unit Tests**

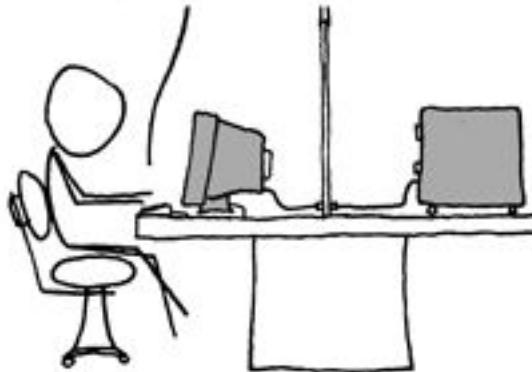
A wide-angle photograph of a landscape featuring numerous rolling green hills. The hills are covered in lush vegetation, with sunlight casting soft shadows that emphasize their rounded forms. The perspective is from a low angle, looking across the terrain towards a distant horizon where the hills meet a clear, pale blue sky.

**THE QUALITY OF MACHINE  
LEARNING CHATBOTS  
NEEDS A BETTER  
UNDERSTANDING OF HUMAN  
COMMUNICATION FIRST.**

TURING TEST EXTRA CREDIT:  
CONVINCE THE EXAMINER  
THAT HE'S A COMPUTER.

YOU KNOW, YOU MAKE  
SOME REALLY GOOD POINTS.

I  
I'M ... NOT EVEN SURE  
WHO I AM ANYMORE.



# RESOURCES & READINGS

- Photos (except for screenshots are taken from Unsplash): <https://unsplash.com>
- Illustration by Saatchi Art: <https://www.saatchiart.com/art/Painting-v/147928/1127926/view>
- Data sets: [https://en.wikipedia.org/wiki/Training,\\_test,\\_and\\_validation\\_sets](https://en.wikipedia.org/wiki/Training,_test,_and_validation_sets)
- About the Turing test: <http://isturingtestpassed.github.io>
- Rasa framework: <https://rasa.com/>
- Entity visualizer: <https://explosion.ai/demos/displacy-ent>
- Some stats on chatbot expectations: <https://ideal.com/5-stats-chatbots/>
- Chatbots were the next big thing: what happened? <https://blog.growthbot.org/chatbots-were-the-next-big-thing-what-happened>
- 7 Important Testing Chatbots areas: <https://blog.appliedai.com/chatbot-testing-frameworks/>
- 3 Types of Chatbots:  
<https://chatbotsmagazine.com/the-3-types-of-chatbots-how-to-determine-the-right-one-for-your-needs-a4df8c69ec4c>
- Pizza hut chatbot image: <http://www.digitaljournal.com/img/3/7/9/5/3/4/i/3/0/5/o/pizza-hut-bot.jpg>
- Eliza: <https://en.wikipedia.org/wiki/ELIZA>
- [Paper on evaluating chatbot quality \(includes research on other papers, too\)](#)
- [General overview of ChatBot testing with main areas to concentrate on](#)
- [Fundamental tips of testing before/pre/during/after launch](#)
- [7 Tests \(mainly for conversational logic\) Your ChatBot Should Pass](#)
- Chatbot which replies to spam: <https://www.theverge.com/2017/11/10/16632724/scam-chatbot-ai-email-rescam-netsafe>
- Do not pay chatbot:  
<https://www.theguardian.com/technology/2016/jun/28/chatbot-ai-lawyer-donotpay-parking-tickets-london-new-york>