Walking the Talk: Random Exploration of a Chatbot API

James Thomas
Association for Software Testing webinar 2022-10-14
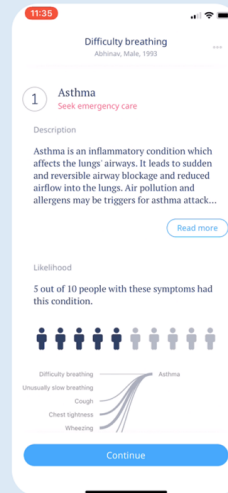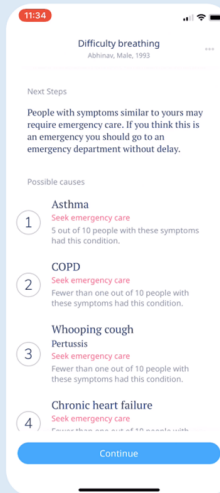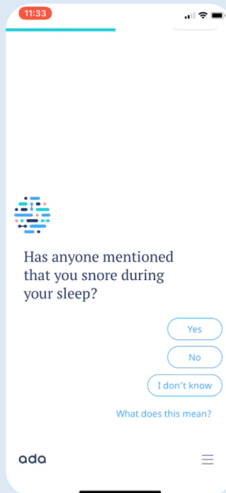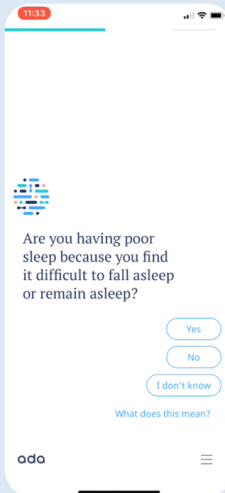
@qahiccupps

www.ada.com

@qahiccupps

11:31

Thank you, Abhinav. I have enough info for now and can help you find out what's going on. Just start a symptom assessment.

Start symptom assessment

ada

11:31

ok. Just start a symptom assessment.

Start symptom assessment

Okay, Abhinav. Who is the assessment for?

Myself

Someone else

ada

11:31

you last answer by tapping on it.

Okay

Let's start with the symptom that's troubling you the most.

Enter a symptom, e.g. headache

ada

11:31

shortness   cancel

shortness of breath
shortness of breath at rest
shortness of breath with exertion
shortness of breath during exertion
shortness of breath when lying flat
shortness of breath during gentle exercise
shortness of breath during physical activity
shortness of breath during physical exertion

q w e r t y u i o p
a s d f g h j k l
z x c v b n m

123   space   Search

11:33

Are you having poor sleep because you find it difficult to fall asleep or remain asleep?

Yes

No

I don't know

What does this mean?

ada

11:33

Has anyone mentioned that you snore during your sleep?

Yes

No

I don't know

What does this mean?

ada

11:34

Difficulty breathing
Abhinav, Male, 1993

Next Steps

People with symptoms similar to yours may require emergency care. If you think this is an emergency you should go to an emergency department without delay.

Possible causes

1 Asthma
   Seek emergency care
   5 out of 10 people with these symptoms had this condition.

2 COPD
   Seek emergency care
   Fewer than one out of 10 people with these symptoms had this condition.

3 Whooping cough
   Pertussis
   Seek emergency care
   Fewer than one out of 10 people with these symptoms had this condition.

4 Chronic heart failure
   Seek emergency care

Continue

11:35

Difficulty breathing
Abhinav, Male, 1993

1 Asthma
   Seek emergency care

Description

Asthma is an inflammatory condition which affects the lungs' airways. It leads to sudden and reversible airway blockage and reduced airflow into the lungs. Air pollution and allergens may be triggers for asthma attack...

Read more

Likelihood

5 out of 10 people with these symptoms had this condition.

Difficulty breathing   Asthma
Unusually slow breathing
Cough
Chest tightness
Wheezing

Continue

@qahiccupps

@qahiccupps

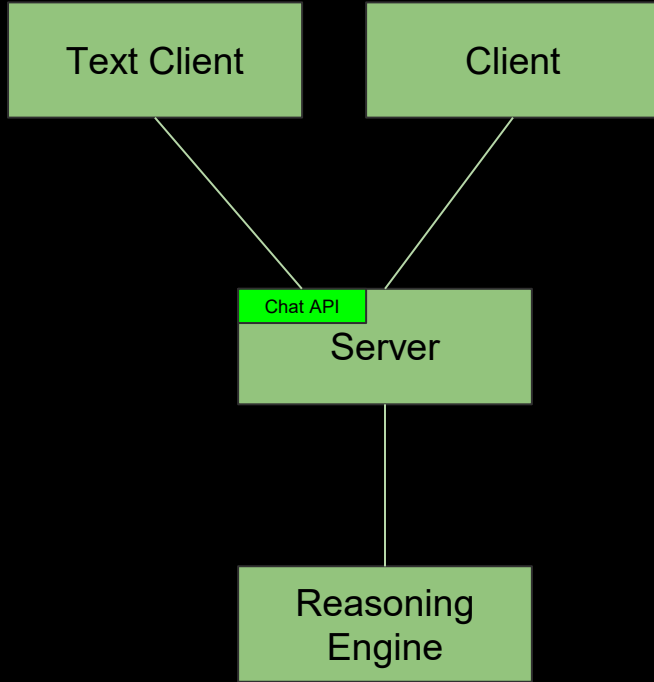# Project: Extract a Basic Chatbot API

@qahiccupps

The Task

Presentation-agnostic API.

Integration with text-based clients.

Turn-based medical assessment.

Short deadline.

@qahiccupps
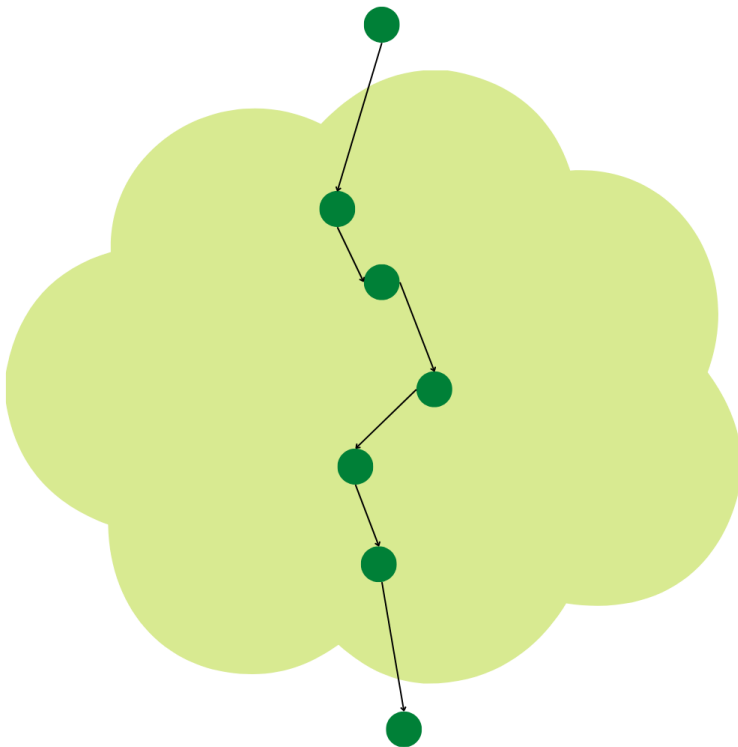
Risks, Questions, Test Ideas

Covering the Assessment Space

Unintended consequences.

Discover edges and corners.

Exercise the API extensively.
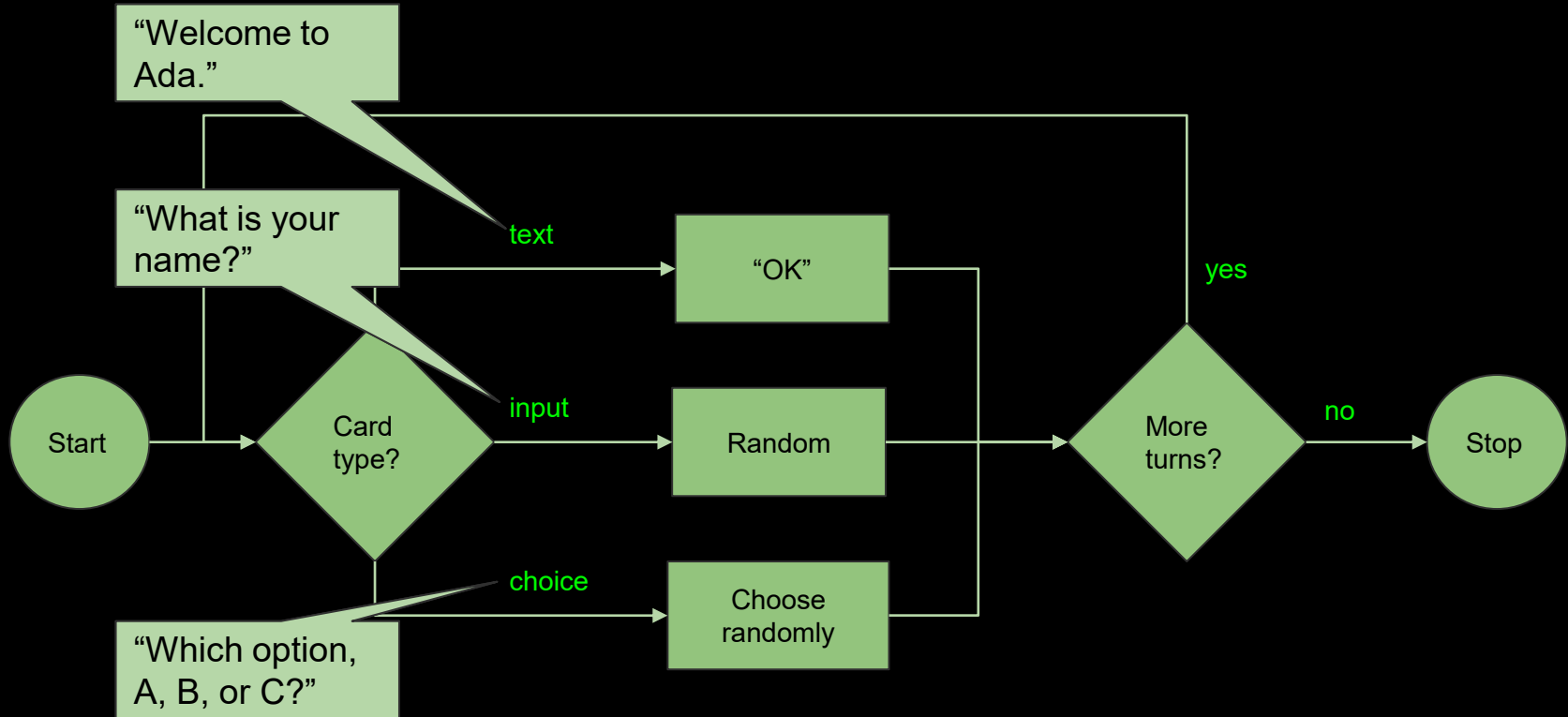
@qahiccupps

A Walker

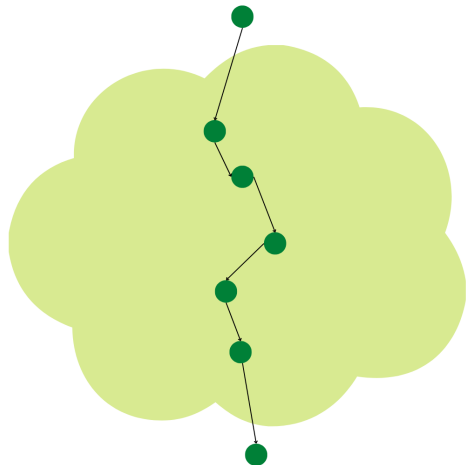# The idea

Run unattended medical assessments.

Can be iterated and customised quickly.
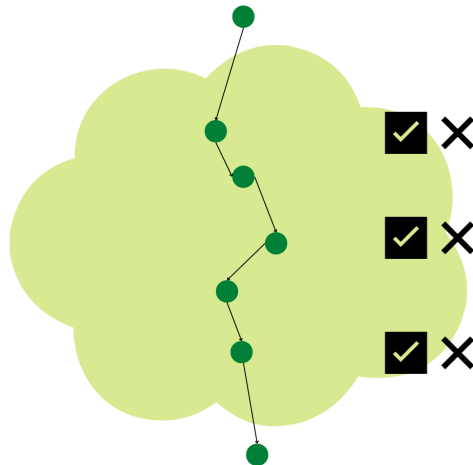
Identify places for deeper inspection.
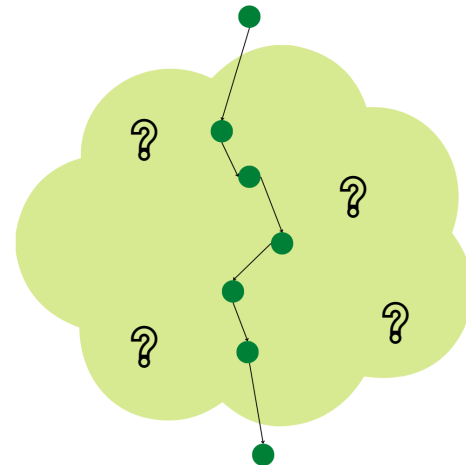
@qahiccupps

The Simplest Thing That Could Possibly Work

@qahiccupps

Navigate   Survey   Explore

@qahiccupps

Demonstration

@qahiccupps

Implementation

Configuration

Error checks
~1/200 turns

Another
symptom
~1/5 turns

```json
"previous_step": {
    "doc": "weighting to choose the previous endpoint when available",
    "probability": 0.01
},

"abort": {
    "doc": "weighting to choose the abort endpoint when available",
    "probability": 0.0005
},

"validate": {
    "doc": "weighting to check validation on INPUT cards. Do this sparingly it slows assessments down",
    "probability": 0.0005
},

"self": {
    "doc": "weighting to choose the self endpoint when available",
    "probability": 0.01
},

"more_symptoms__no": {
    "doc": "weighting to answer 'no' when asked if there are more symptoms. Set higher for faster assessments",
    "probability": 0.8
},

"symptom_list__none_of_these": {
    "doc": "weighting to say 'none of these' when give a list of standard symptoms. Set higher to search more often",
    "probability": 0.8
},

"symptom_search_result__none_of_these": {
    "doc": "weighting to say no search results match. Set higher to search again more often",
    "probability": 0.02
},

"specific_presenting_complaint_search": {
    "doc": "presenting complaints to search for that, if the only PC, are very likely to have the given advice levels. This is a very weak
medical sanity check.",
    "probability": 0.05,
    "value": [
        { "head lice": ["selfCare","selfCarePharma"] }
    ]
},
```

@qahiccupps

Detailed, Parsable Logs

Archive config

```
{
  "max_dialog_turn": 100,
  "locale": "en-GB",
  "standard_headers": {
    "Accept-Language": "en-GB",
    "X-Ada-UserId": "user1"
  },
  ...
  "previous_step": {
    "doc": "weighting to choose the previous endpoint when available",
    "probability": 0.01
  },
  ...
}

...

-- 3
{"cardType": "INPUT", "step": 3, "title": {"en-GB": "Patient Information"}, ...}

<-- card:  3 INPUT What is the person's name?
--> reply: DLFLプQsc上8万J
--> response: next {'step': 3, 'value': 'DLFLプQsc上8万J'}

-- 4
{"cardType": "CHOICE", "step": 4, "title": {"en-GB": "Patient Information"}, ... "options": [...], ...}

<-- card:  4 CHOICE What is DLFLプQsc上8万J's sex? Biological sex is a risk factor for some conditions. Your answer is
necessary to give you accurate recommendations.
--> reply: 0 Female
--> response: next {'step': 4, 'optionId': 0}

...

-- 10
{"cardType": "TEXT", "step": 10, ...}

<-- card:  10 TEXT Sorry, we haven't found any matches for the symptom you typed in. Try using fewer words or use diffe
rent words to describe your symptom. Only type in one symptom at a time.
--> reply:
--> response: next {'step': 10}

...

 ------ state -----
{'seen_years_question': True, 'seen_months_question': True, 'state_reset_counter': 0, 'presenting_complaint': ['Floppy
baby'], 'assessment_id': '13ce7d7a-e08f-4ba6-9b2b-71fe7b891366', 'primary_secondary': 'Someone else', 'age_value': '1',
 'found_pdf_report': True, 'start_time': '2022-09-10T15:21:08.878123', 'end_time': '2022-09-10T15:21:49.864989'}
```

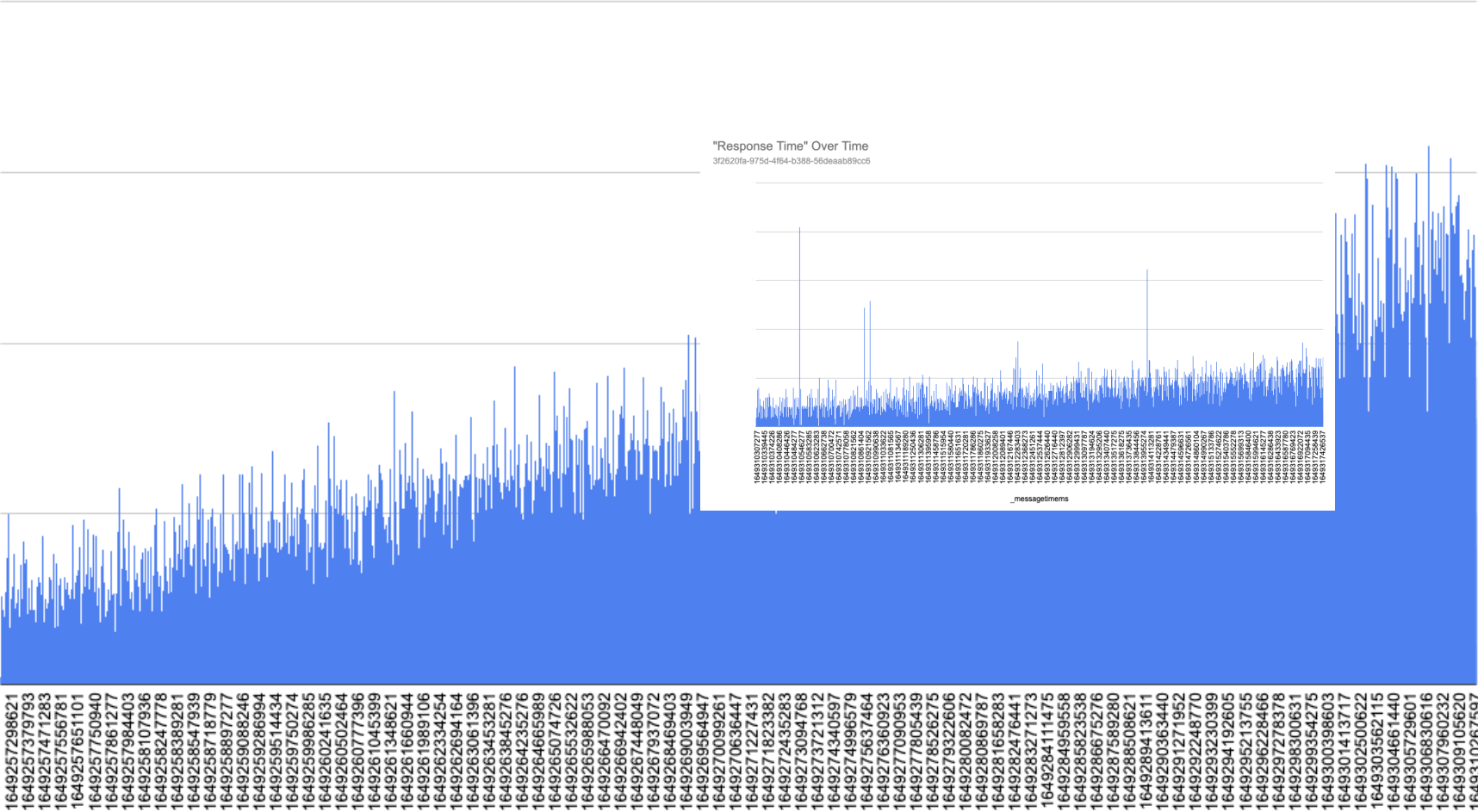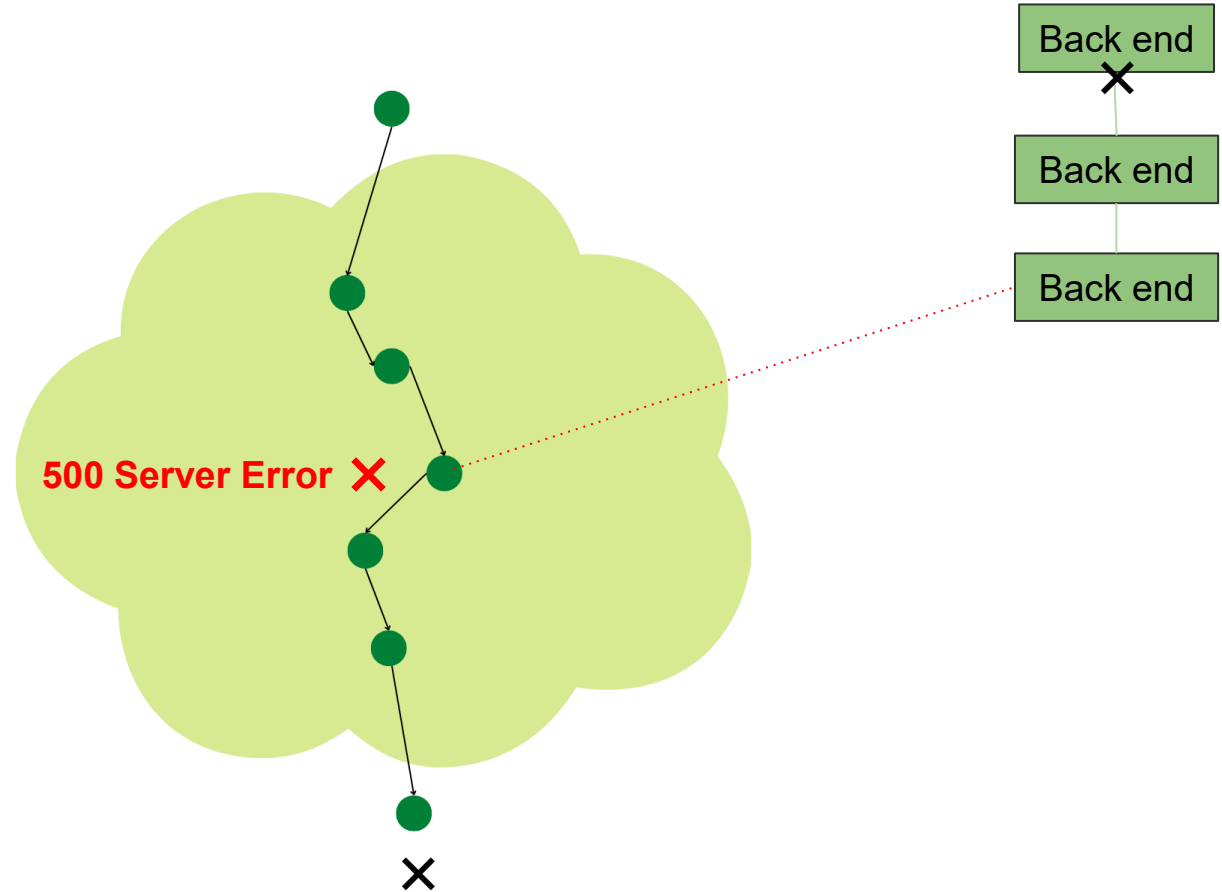Archive state

@qahiccupps

Parallel

```
$ ./runner.sh demo
starting iteration 1 ...
starting iteration 2 ...
starting iteration 3 ...
starting iteration 4 ...
starting iteration 5 ...
$ ls demo
1.1           2.1           3.1           4.1           5.1           config.json
1.2           2.2           3.2           4.2           5.2
```

```bash
for i in $(seq 1 $NUM_ITERATIONS); do
    echo "starting iteration $i ..."
    printf "%s\0" $(seq 1 $NUM_PARALLEL) | xargs -0 -I @ -P $NUM_PARALLEL sh -c "python3 -u $DAPIC > $i.@"
done
```

@qahiccupps

"Response Time" Over Time

e256b083-1716-44a8-9e99-0d60b05ee219

"Response Time" Over Time

3f2620fa-975d-4f64-b388-56deaab89cc6

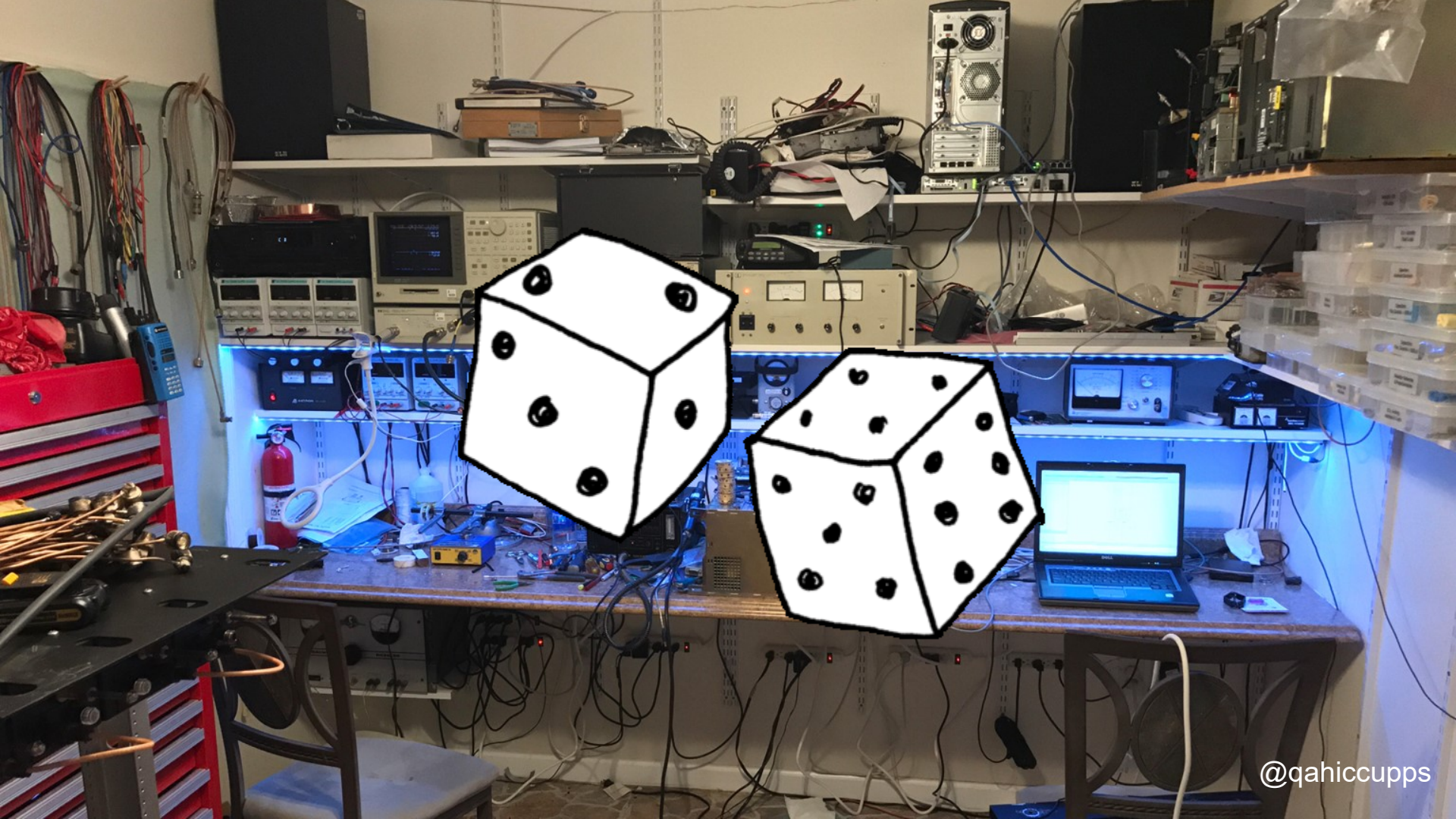Back end

Back end

Back end

500 Server Error ✗

✗

@qahiccupps

Reflection

What Worked Well … and What was Challenging

Verbose logs.

Asserting generally and on fixes.

Randomisation for unknowns.

Configuration for directed exploration.

Toolkit (replay, parallel, analysis, …).

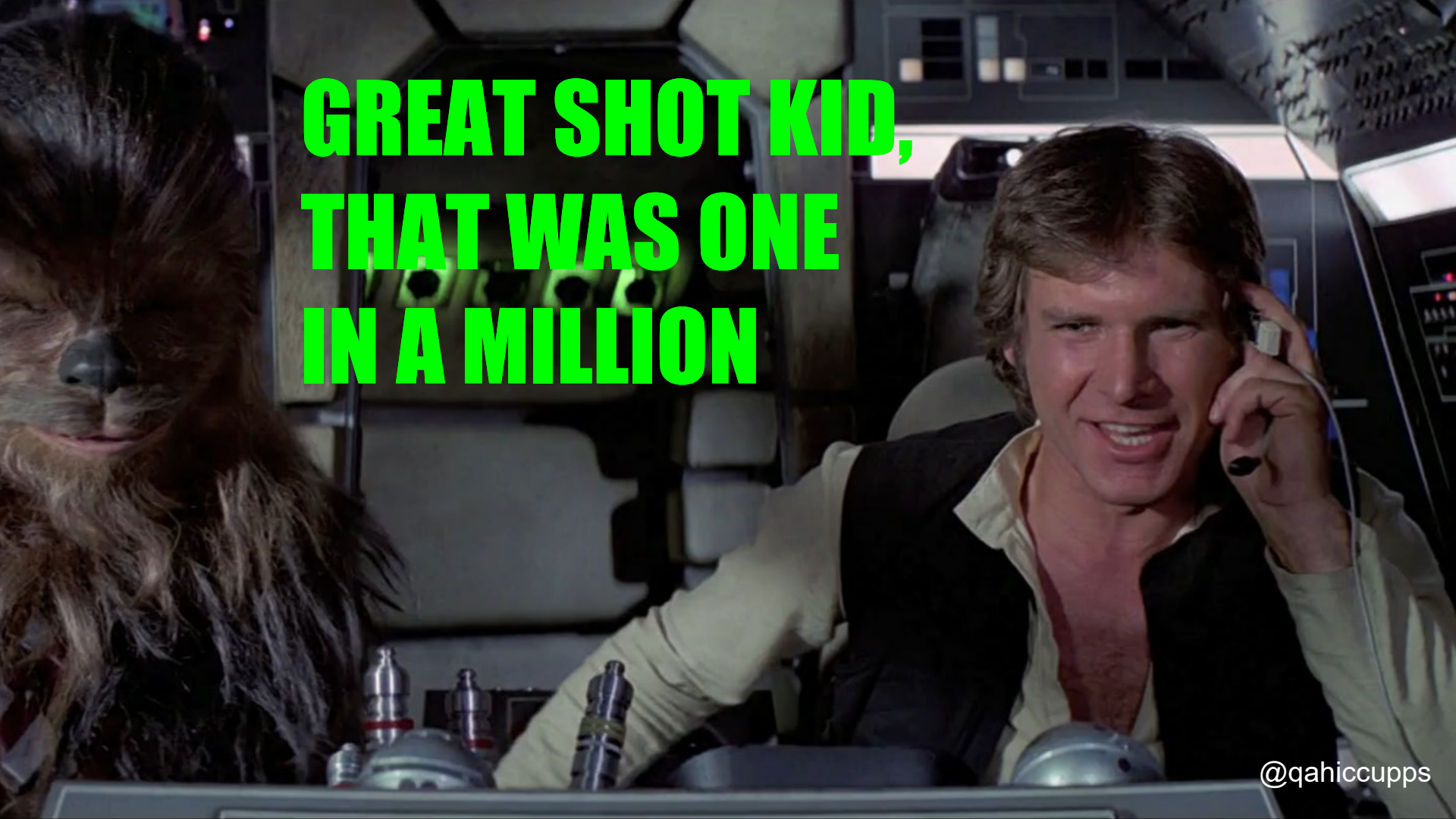Question-driven development.

Code changes and dependencies.

Card identifiers.

Medical testing.

State.

No explicit model.

@qahiccupps

GREAT SHOT KID, THAT WAS ONE IN A MILLION

@qahiccupps

Questions?

@qahiccupps

References

Ada screenshots: https://www.uisources.com/explainer/ada-diagnosing-via-chat-bot

Wipotec: https://www.wipotec-ocs.com/en/product-inspection/

Microscope: https://londonlaboquip.com/product/microscope-binocular-biological-sc302

Messy lab: https://imgur.com/gallery/bQiK6

Dice: https://www.richardhughesjones.com/luck-randomness/dice-gif/

Altwalker: https://altom.gitlab.io/altwalker/altwalker/

Star Wars: https://www.starwars.com/video/one-in-a-million-shot

@qahiccupps