

DragonCrawl: Generative AI for High-Quality Mobile Testing

Juan Marciano
Uber

Who we are



Juan

ML, 5 years @Uber
GenAI, NLP, CV, RL

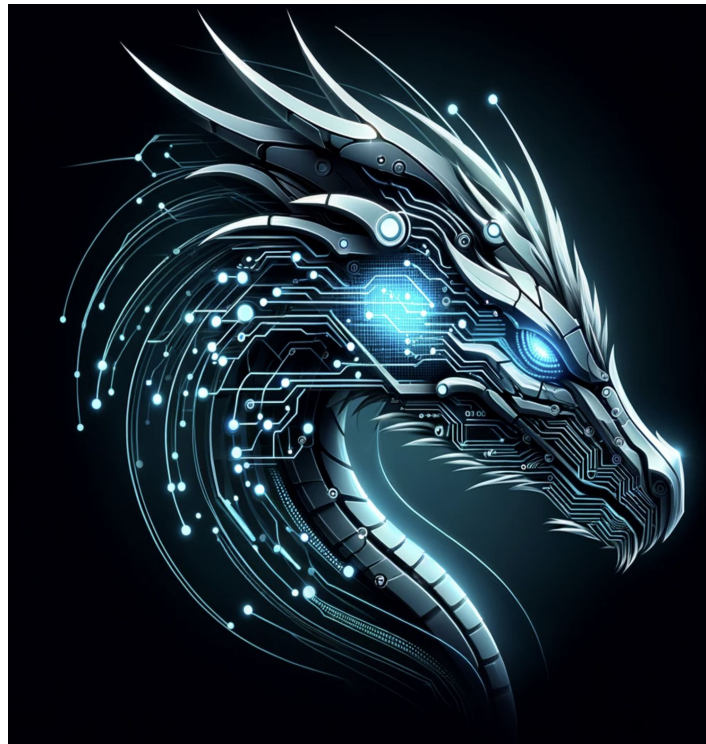
Background

Pain points of traditional E2E testing:

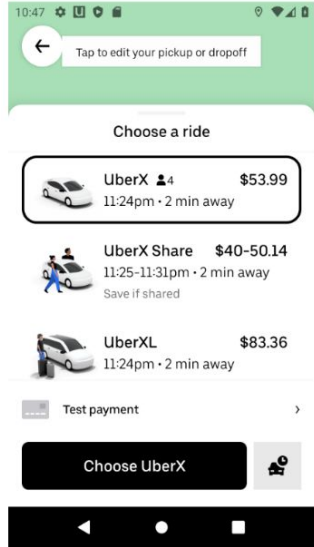
- SDETs spends **30-40%** time maintaining automated tests
- **Impossible** to shift left
- Extremely **expensive** to test all combinations

DragonCrawl: LLMs that

- High pass rate, high SNR
- Trained once, run in many cities. **Code Free**
- Scales: **1000x** smaller than ChatGPT



How: LLMs



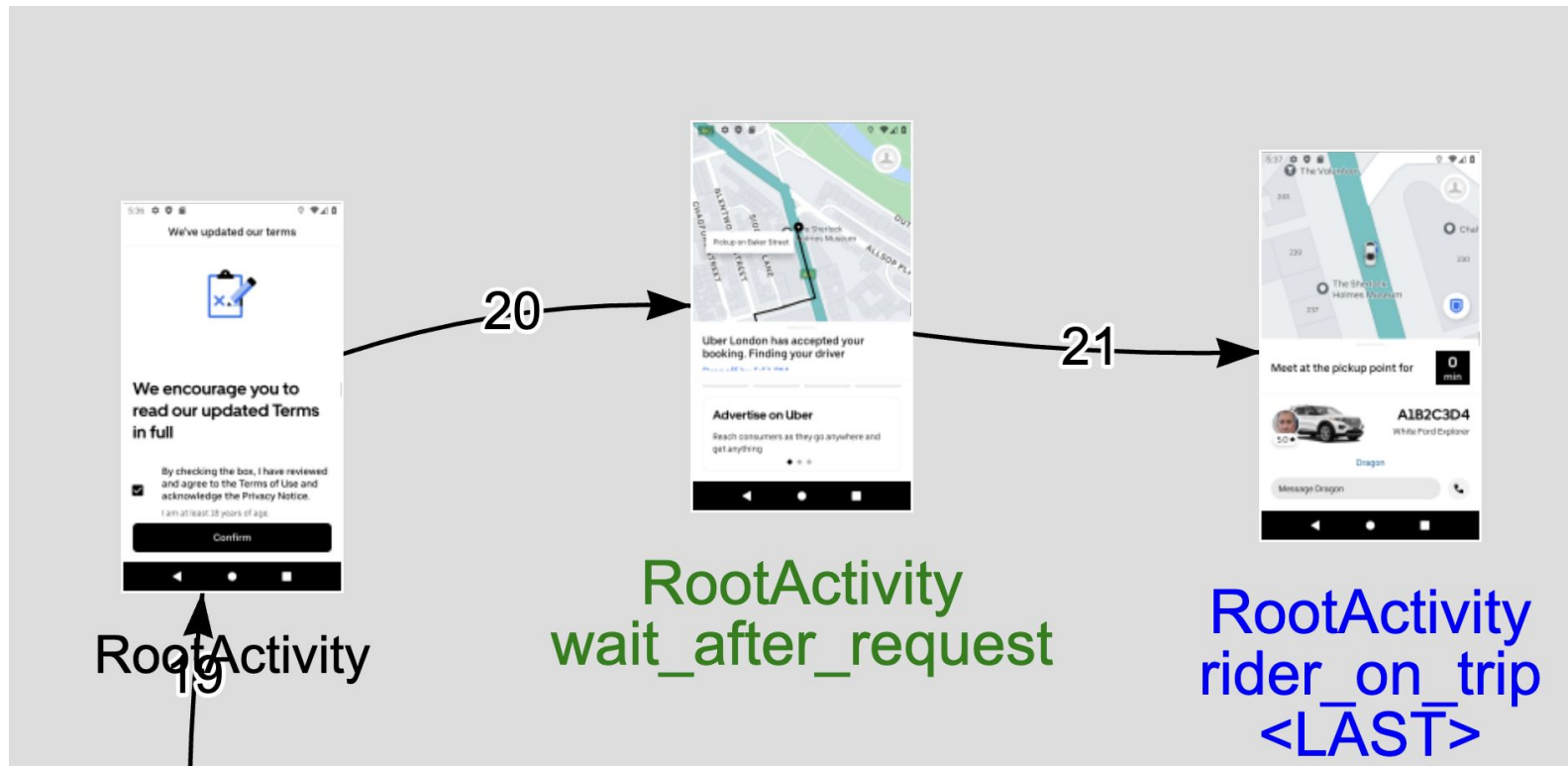
Goal:
"Request a
trip"

UI: "Button1:
Choose
Uberx,
Button2: ..."

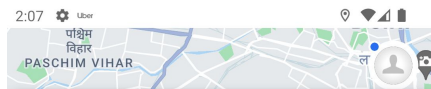


Touch "Choose
UberX"

How: Flow execution separate from assertions



Future of Assertions



राइड के लिए रिवेस्ट की गई

आपके आस-पास ड्राइवर पार्टनर को ढूँढा जा रहा है

Ride details

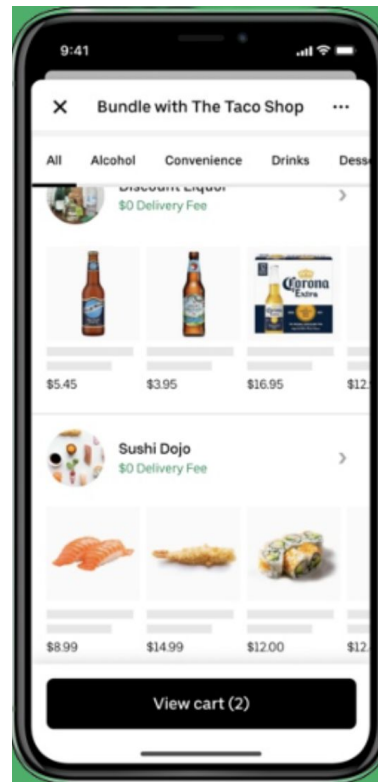
Rajiv Chowk , Gate 2 पर मिलें

2:37 pm बजे तक ड्रॉप ऑफ़



```
dragon_assertions_model.evaluate_assertion("is there an ad?", image)
```

(True,
'Yes, there is an ad for the movie "Baaton Mein Ullu Jhakaas" in the image.')



```
dragon_assertions_model.evaluate_assertion("are we showing alcohol drinks?", image)
```

(True, 'Yes, we are showing alcohol drinks in the image.')

Note: image taken from [here](#)

Quality Wins

Bugs: 11 P0 mobile bugs

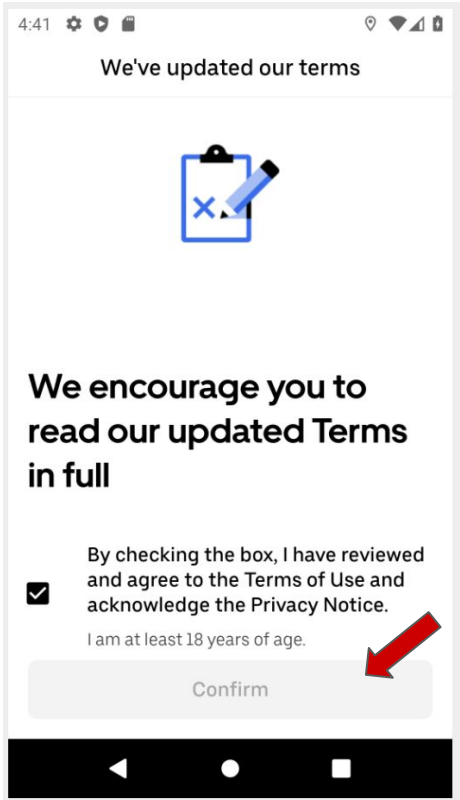
Localization: 100+

Breaking into new markets: Catalan, Bengali, Pashto, Dahri

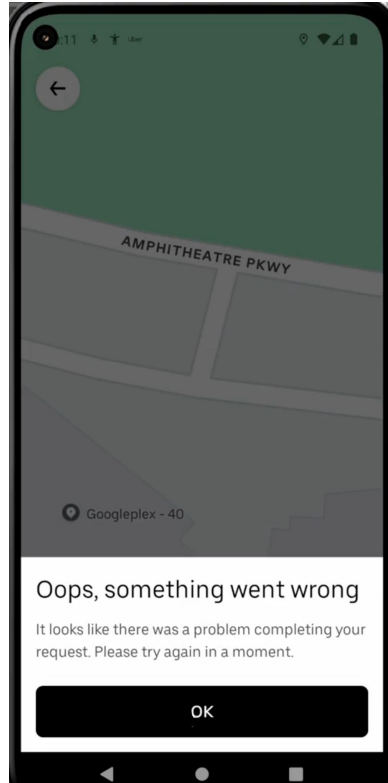
Dragon + Chaos testing: Finding backend bugs (19 P0 bugs already)

Sample Bugs

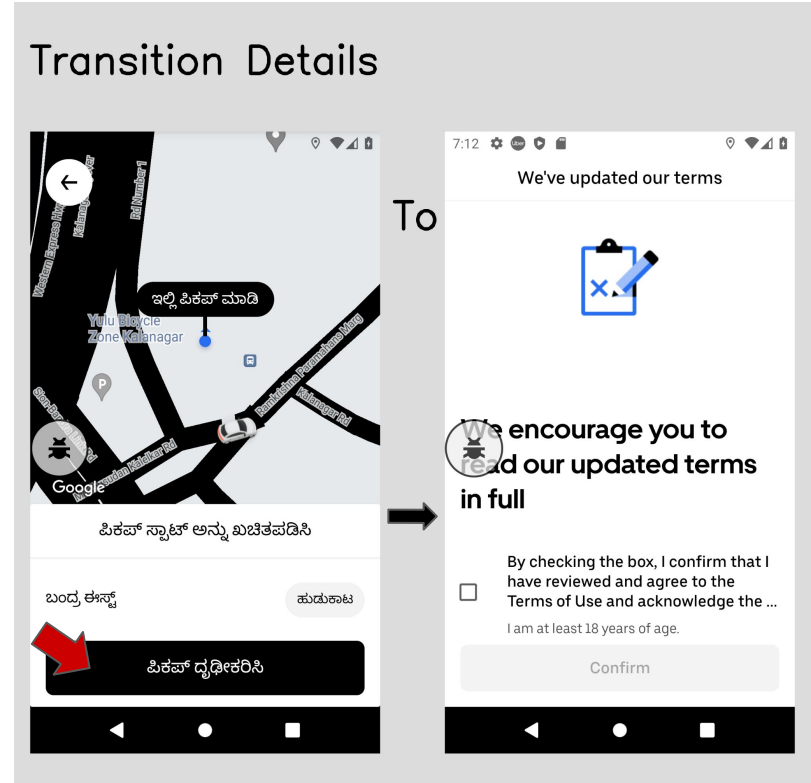
Mobile bug



Backend bug



100+ localization bugs



Productivity Wins

Pass Rate: **99%+** since Dec 2023. **No Maintenance**

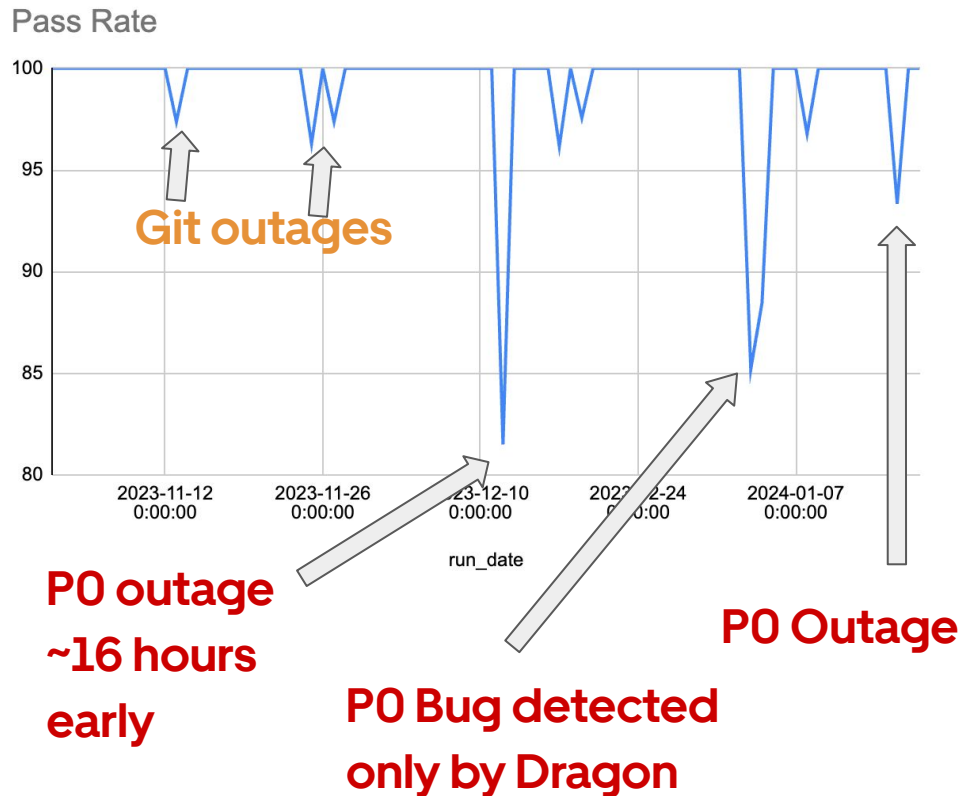
Bugs SNR: ~90% YTD, 100% in Q2

Triaging: **0 hours**

Coverage:

- 5 cities, 2 langs ~90 builds/day
- 184 cities, 55 langs weekly

~30K dev hours saved in test authorship



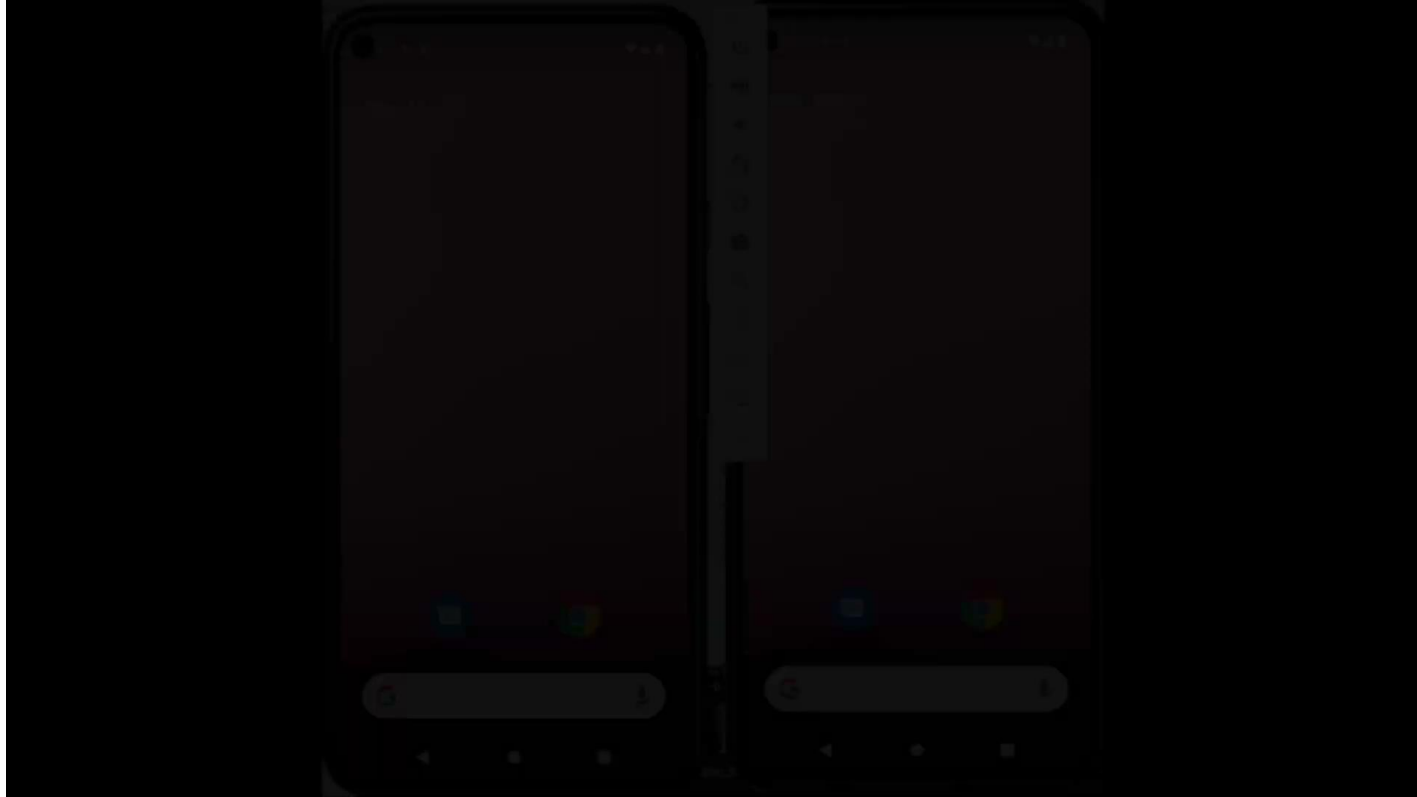
Demo

Scenario:

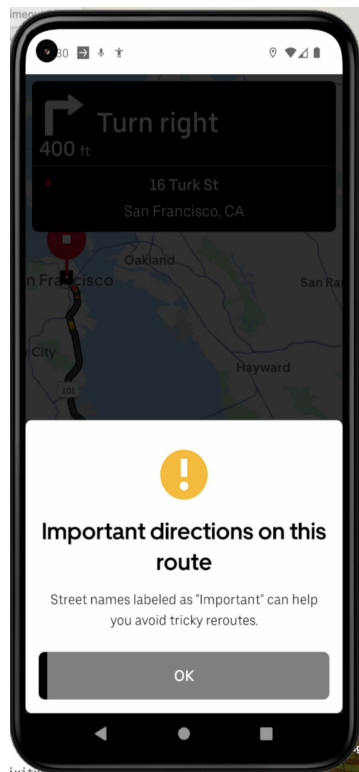
1 Dragon Driver

1 Dragon Rider

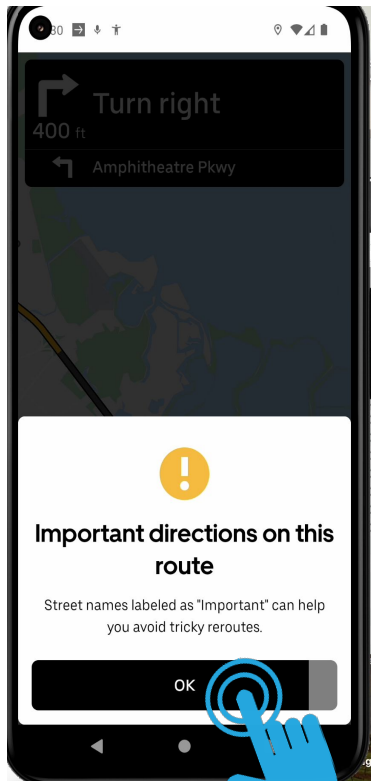
Goal: Core trip flow



Demo Recap



Wait



Resilient to UI flow changes!

99% pass rate!

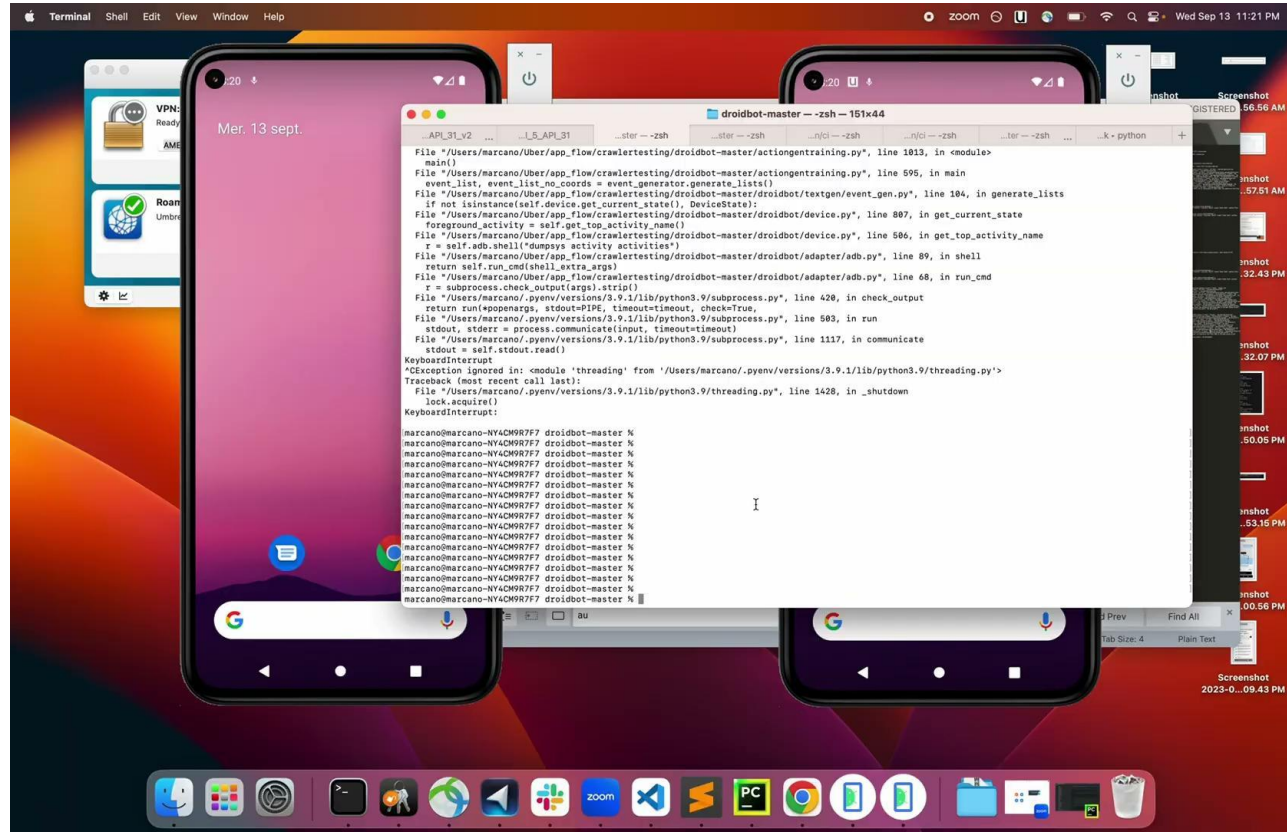
Demo 2

Scenario:

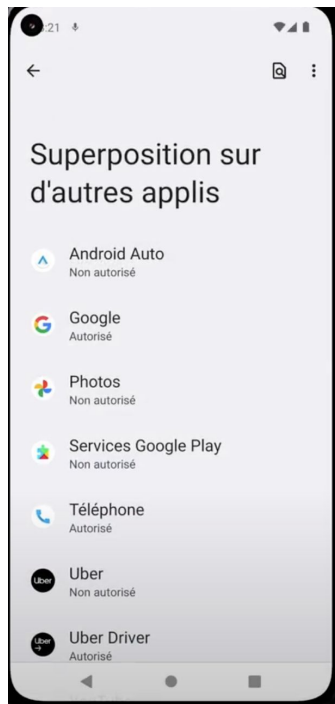
1 Dragon Driver

1 Dragon Rider

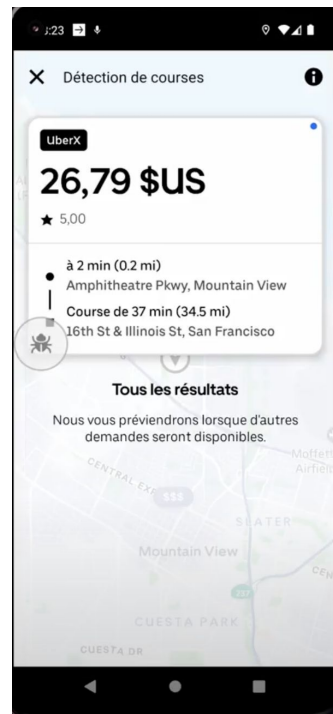
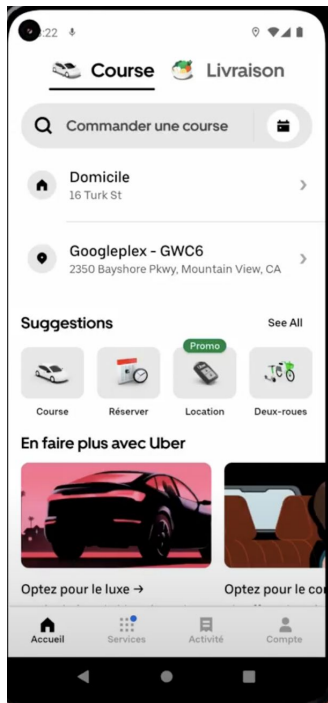
Goal: French Core trip flow



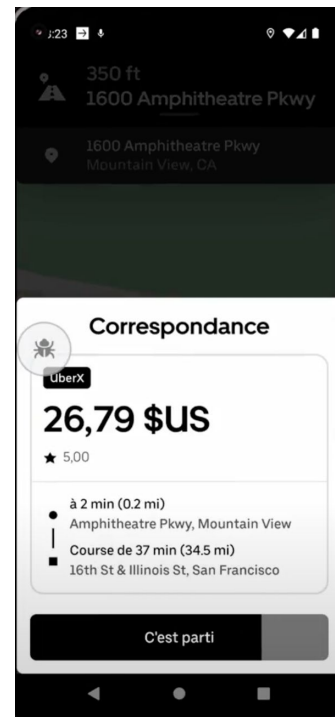
Demo Recap



Multiple actions



Accept the trip twice



What's next!

- Generalized Assertions Model
- DragonCrawl Cross-platform
- Increase coverage
- DragonCrawl for bug reproduction

Keep in touch!



marcano@uber.com

[linkedin](#)