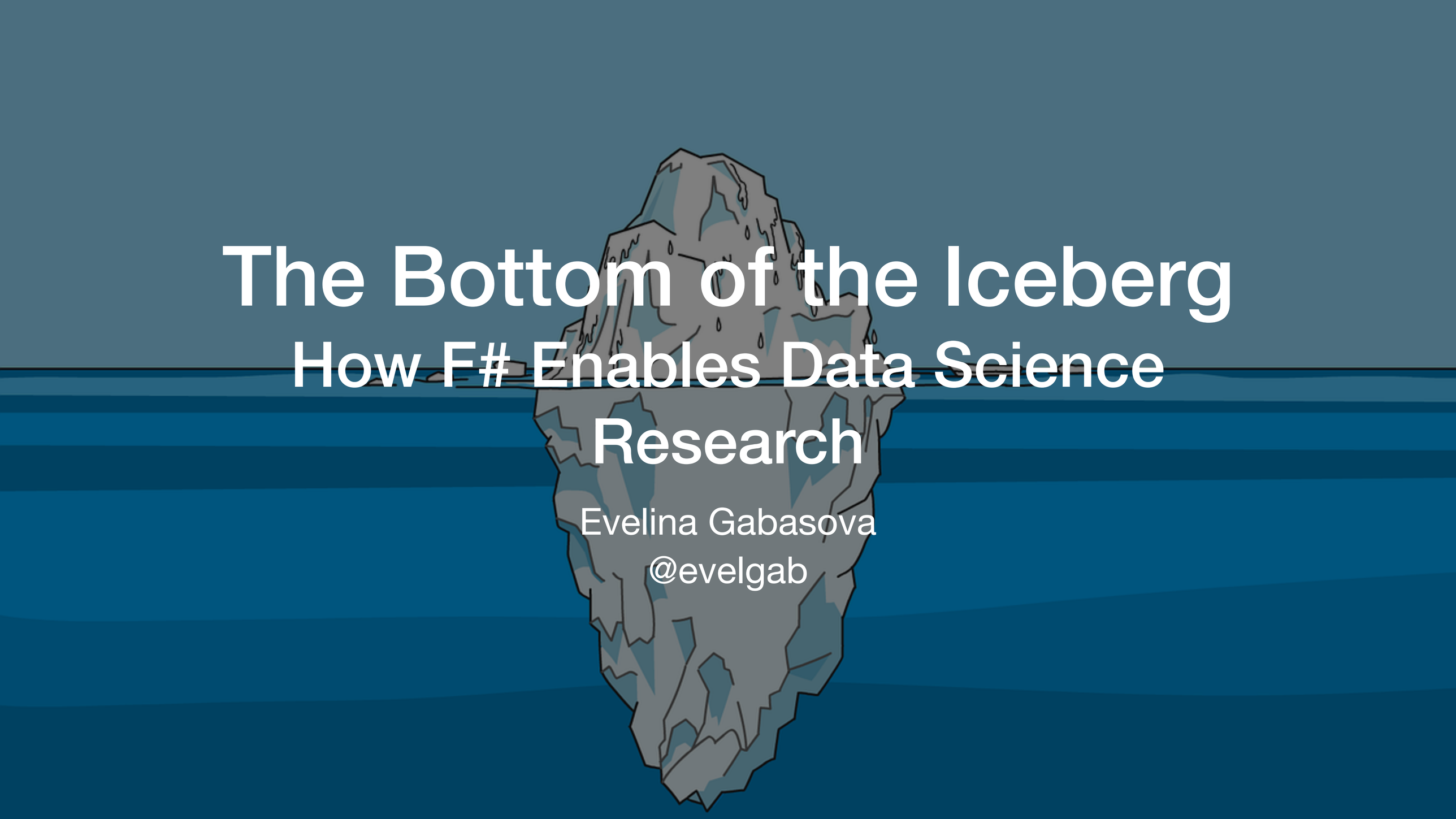


The Bottom of the Iceberg

How F# Enables Data Science Research

Evelina Gabasova
@evelgab

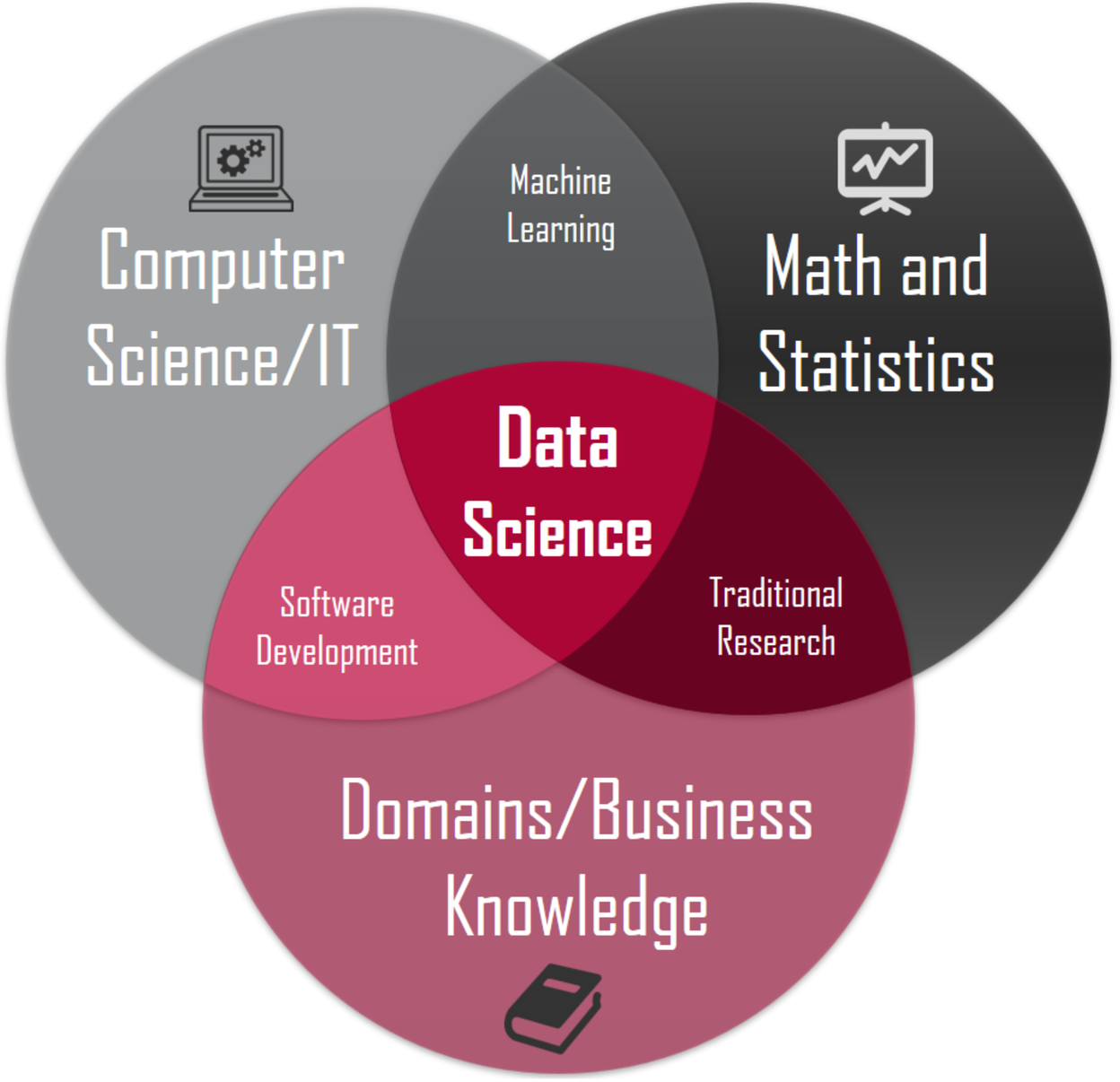
The Alan Turing Institute

Dr Evelina Gabašová

Principal Research Data Scientist

Research Engineering group

@turinghut23


Computer Science/IT
Machine Learning
Math and Statistics
Data Science
Software Development
Traditional Research
Domains/Business Knowledge

I HAVE NO
IDEA WHAT
I'M DOING

Research software engineering

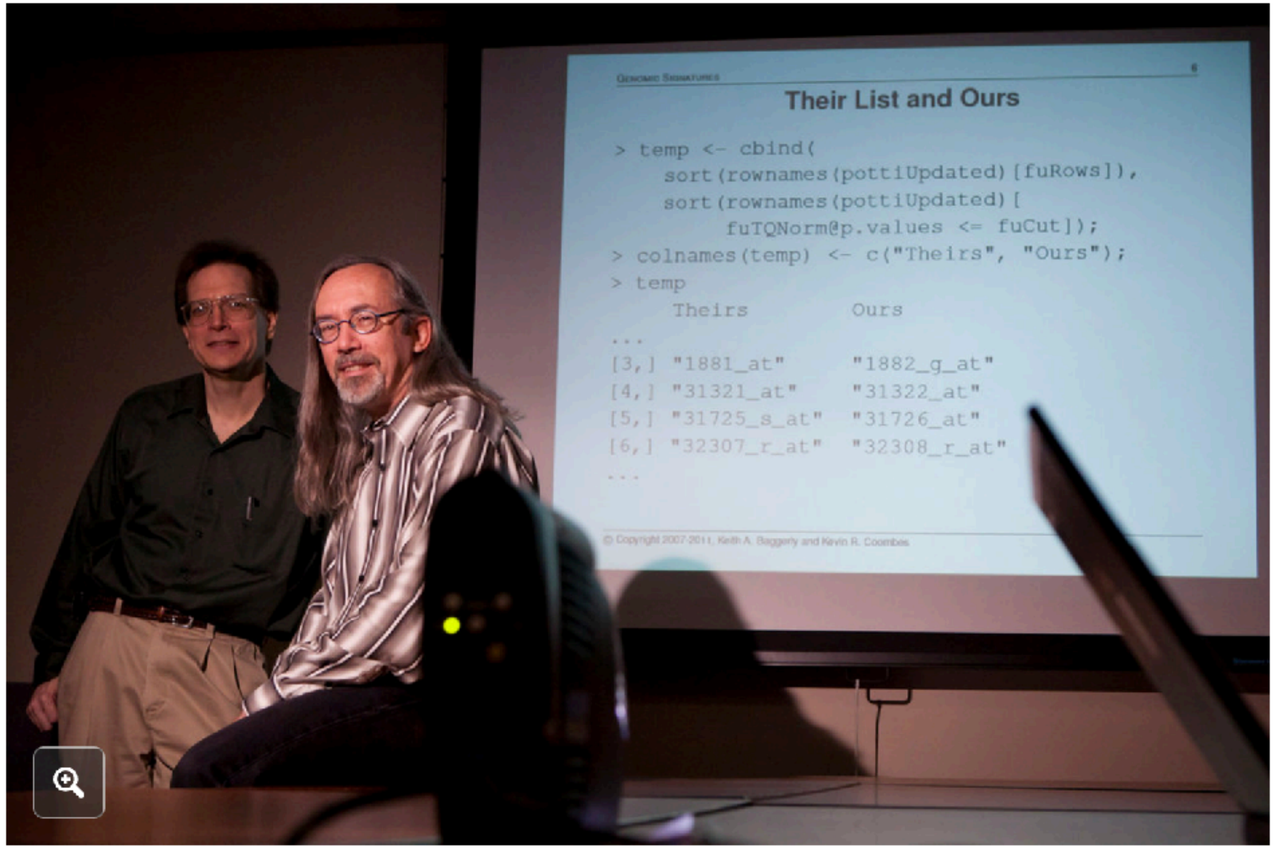
Research software engineering

Reproducibility

RESEARCH

How Bright Promise in Cancer Testing Fell Apart

By GINA KOLATA JULY 7, 2011



Keith Baggerly, left, and Kevin Coombes, statisticians at M. D. Anderson Cancer Center, found flaws in research on tumors. Michael Stravato for The New York Times

RECENT COMMENTS

texas ta July 9, 2011
This story makes me so sad, since I work in this field as well. As a graduate student, I was witness to many flawed analyses with either...

tulipsinyard July 9, 2011
A mildly different perspective: after years teaching mathematics and statistics at universities in North America, I remain shocked by how...

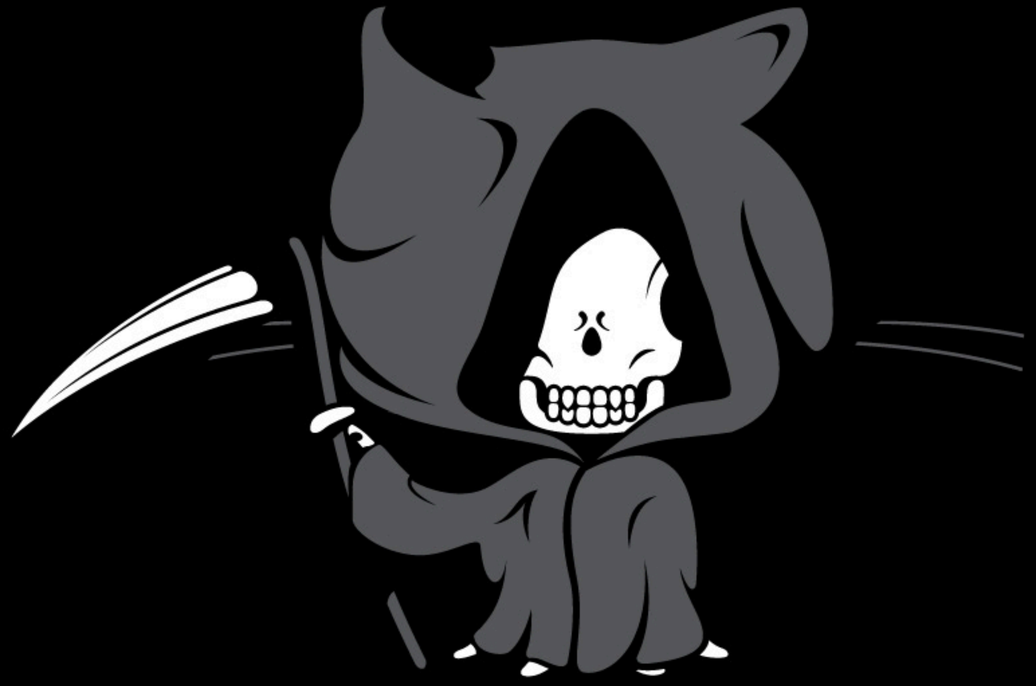
Peter Melzer July 9, 2011
The widely-publicized promises of human genomics research for novel cures exert enormous pressure on the research community, paired with...

SEE ALL COMMENTS

Research software engineering

Reusability and sustainability

Academic GitHub



Research software engineering

Software development practices

CRAN Repository Policy

- Submitting updates should be done responsibly and with respect for the volunteers' time. Once a package is established (which may take several rounds), "no more than every 1–2 months" seems appropriate.

Research software engineering

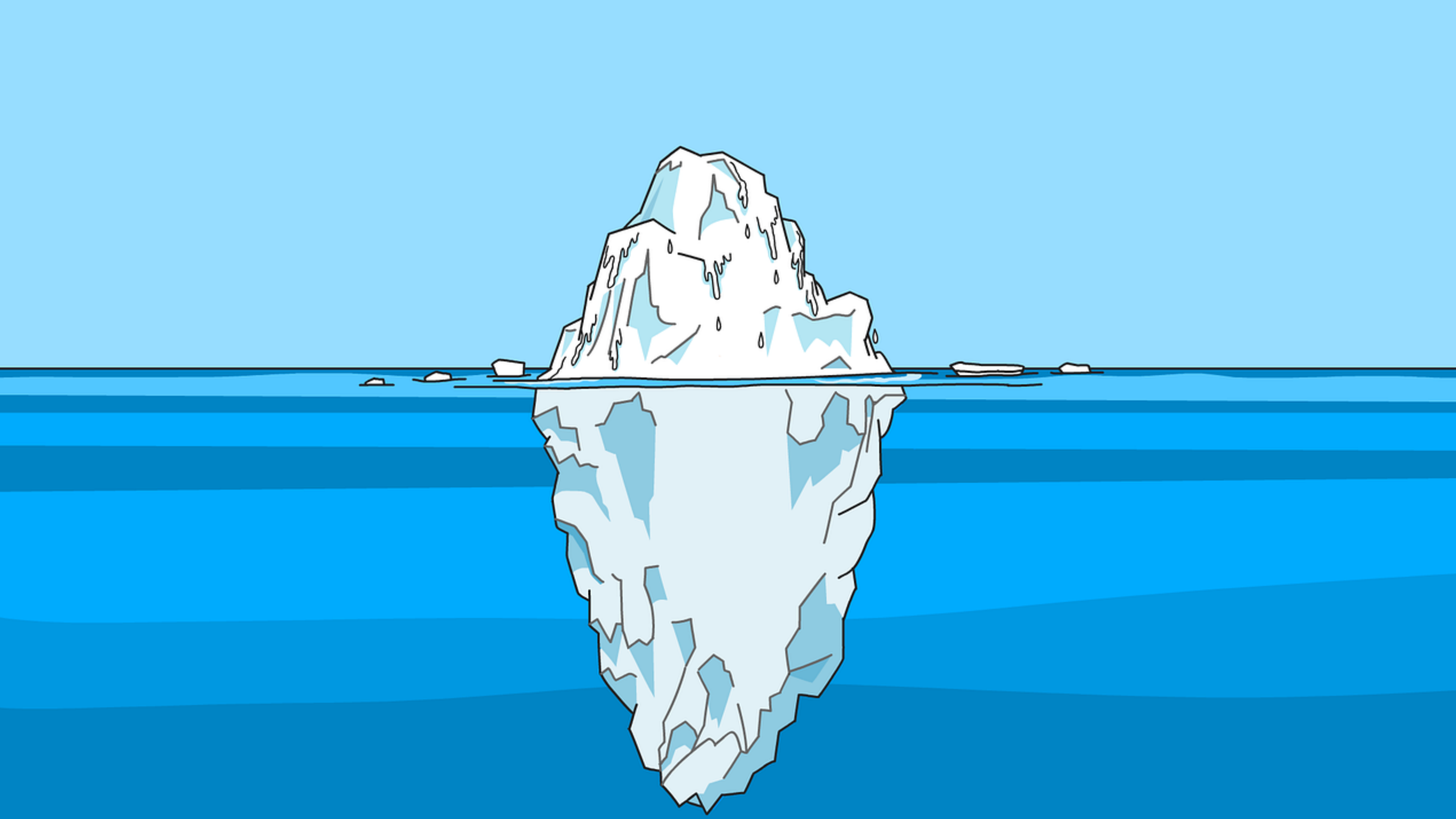
Data science

Research software engineering

Research

+

Real-world applications



Big Data Borat
@BigDataBorat

Following

In Data Science, 80% of time spent prepare data, 20% of time spent complain about need for prepare data.

2:47 AM - 27 Feb 2013

538 Retweets **343** Likes

12 538 343

F# and research engineering

The (mis)informed citizen

Computational social science

Fake news

"Fake news"

- Actual fabricated news
- Conspiracy theories
- Extreme bias
- Hate groups
- ...

Supervised learning

Serious article
Fake news

Supervised learning

94 % accuracy

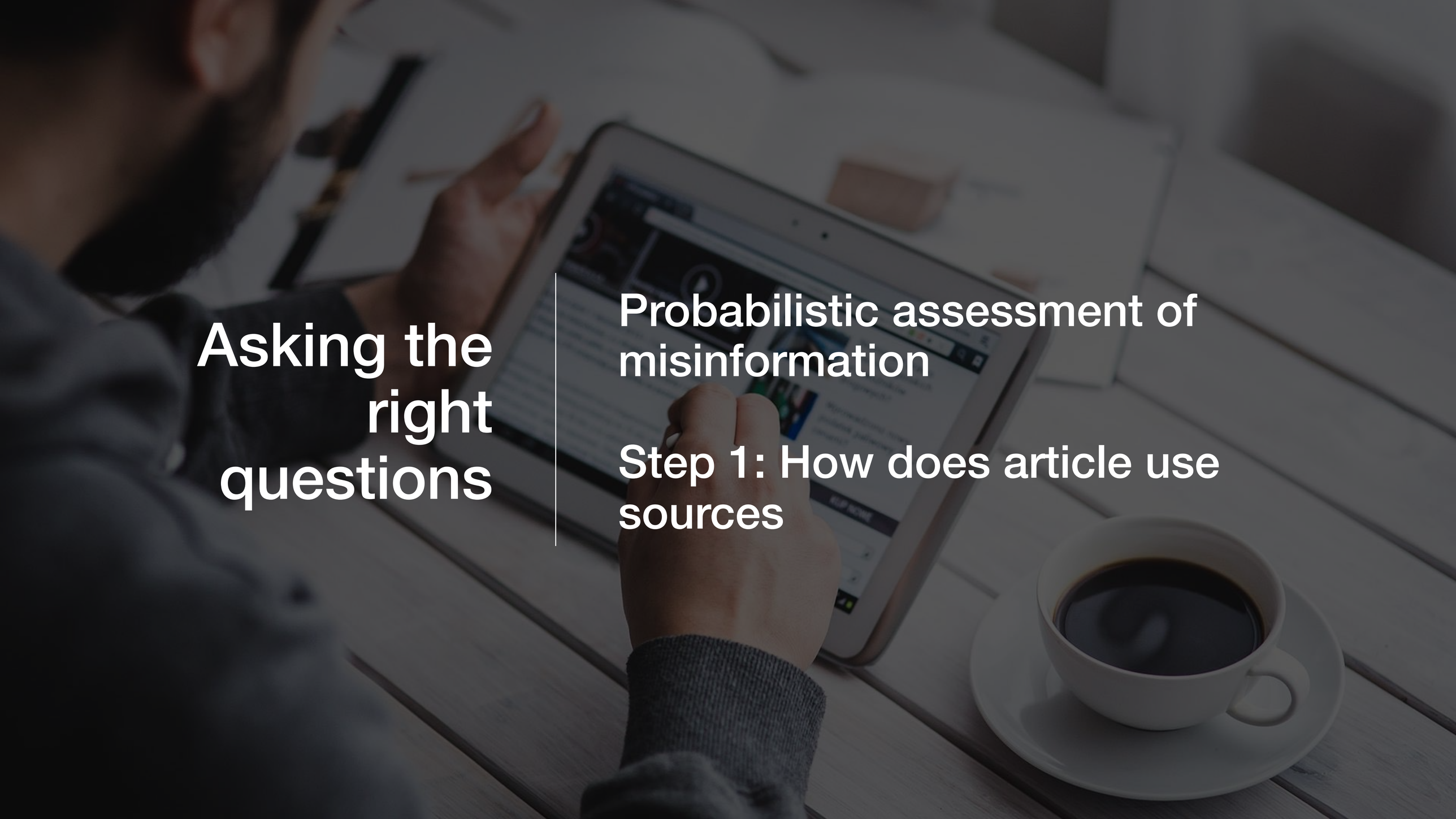
Serious article
Fake news

Supervised learning

Input data → Machine learning → Labels

Serious article
Fake news

Machine learning is only as good as your data



Asking the right questions

Probabilistic assessment of misinformation

Step 1: How does article use sources

Getting the right data

325 websites

Mix of "quality" websites and "fake news" websites

Crawler + extracting article texts

196
210
225
239
254
197
211
226
240
255
198
212
256
199
213
257
200
214
229
243
258
safe

Misinformation project

Annotation tool

The Alan Turing Institute

Deserunt sint tempor ipsum aliqua

Ut aliqua duis elit eu cupidatat ea. Sint ipsum enim exercitation id pariatur cupidatat sunt laborum non eu ex eu. Fugiat quis labore deserunt amet velit ex ad. Sunt ipsum incididunt veniam eu do eu eu. Ipsum magna ipsum deserunt duis adipisicing qui adipisicing cupidatat adipisicing tempor. Id dolore quis proident nostrud consequat irure. Pariatur proident cillum do ea sunt ullamco cillum.

Id ex tempor amet tempor id ipsum voluptate. Do ad consectetur deserunt amet. Et mollit eiusmod do sit dolore dolore.

Aliquip aliquip adipisicing laborum non elit qui. Nostrud in est nostrud laboris eiusmod commodo officia pariatur culpa. Id magna commodo sunt reprehenderit elit id. Ullamco mollit Lorem esse est sunt anim nisi consequat voluptate culpa tempor est pariatur. Officia incididunt consequat dolore do do adipisicing ullamco sint consequat. Qui aute ea veniam excepteur tempor ea ea cillum officia aliquip veniam est.

Demo

on a key campaign promise to turn the court more conservative for generations.

Trump spoke briefly from the White House lawn before the Senate vote session started, saying that "in the end maybe the process, it was really unattractive, but the extra week was something that I think was really good. I thought it was really good. I think a lot of very positive things happened in the last week. It didn't look that way but in the end, that's what happened."

As for Kavanaugh, Trump said "he's going in looking really good."

White House press secretary Sarah Sanders said that Trump called Kavanaugh to congratulate him while onboard Air Force One as he headed to a campaign rally in Kansas.

Trump told reporters that his speech in Mississippi on Tuesday, days before the vote, had an effect on Kavanaugh's nomination. During that speech, Trump appeared to mock Christine Blasey Ford's testimony about her alleged assault at the hands of Kavanaugh. The judge has denied the accusations.

"I think the Mississippi speech had a great impact, yes. I think it was a very important thing," Trump said on the plane Saturday.

"He was chosen for the reason of his temperament, his incredible past, his outstanding years on the court. He's had an outstanding record, a brilliant scholar, a totally brilliant scholar -- top -- and we're very honored that he was able to withstand this horrible horrible attack by the Democrats," Trump said.

The drama of the vote tally

Kavanaugh's confirmation essentially became a done deal on Friday afternoon, when Maine Republican Sen. Susan Collins, a key undecided vote, announced on the Senate floor that, despite allegations of sexual assault and misconduct against him, "I will vote to confirm Judge Kavanaugh."

Just hours earlier, with Kavanaugh's fate still uncertain, the full Senate had

Does the article mention any sources?

Yes | No | Mark as not relevant

Source number 1

Notes

× Delete source

1. Highlight the portion of the text that refers to this source.
   There may be multiple sections in the text that refer to the same source - please highlight all of them. Click Start to start highlighting, and Finish to complete highlighting.

   Continue | Finish | Clear highlights

2. Is the source named or anonymous?

   Not anonymous | Anonymous

Source number 2

Notes

× Delete source

1. Highlight the portion of the text that refers to this source.
   There may be multiple sections in the text that refer to the same source - please highlight all of them. Click Start to start highlighting, and Finish to complete highlighting.

   Continue | Finish | Clear highlights

My SAFE stack experience

The (mis)informed citizen

Mix of "quality" websites and "fake news" websites

Dataset paper submitted

2,490 articles from 47 outlets

6,000 annotations created

Hate speech

Putting AI in air traffic control

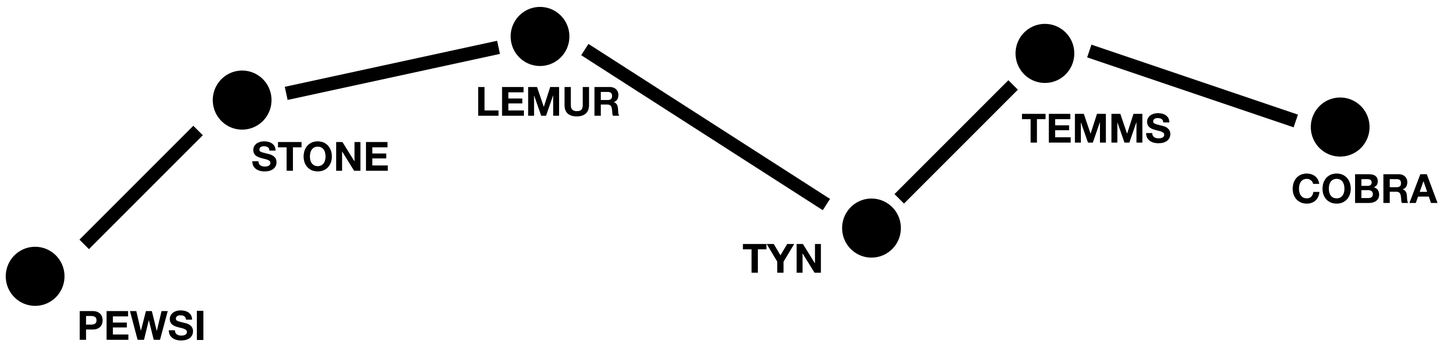
Automated game playing agents

RUSH HOUR

Air traffic control 101

How does air traffic control work

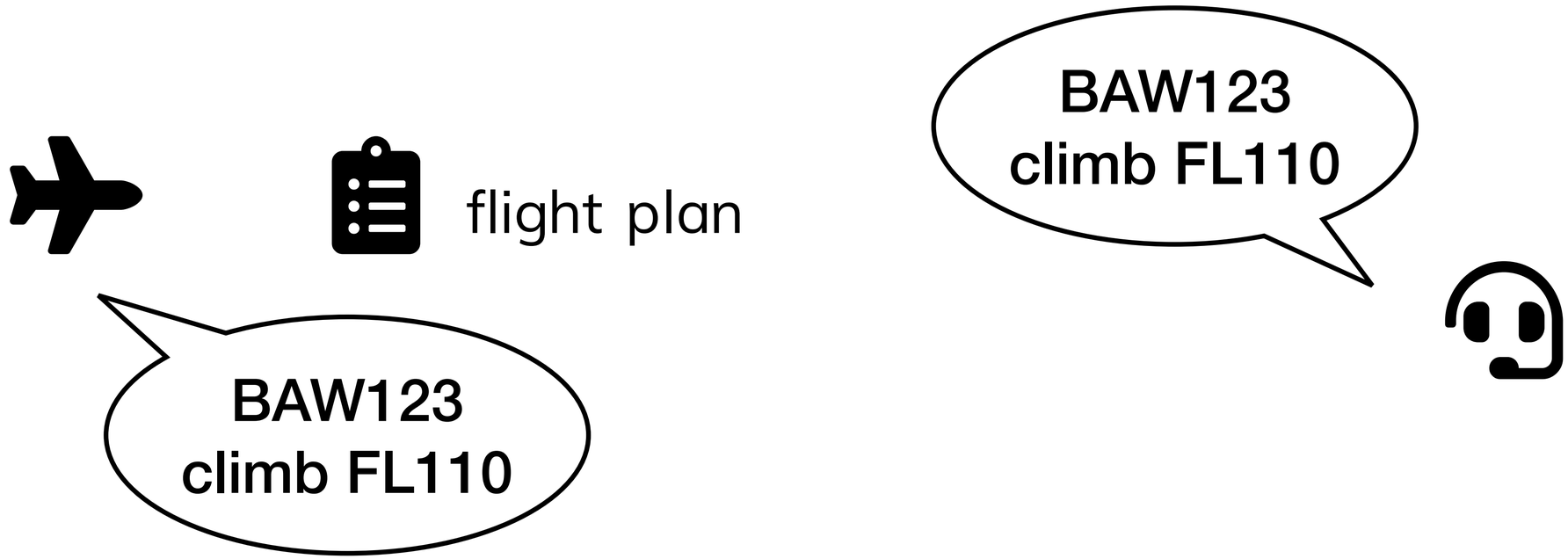
flight plan



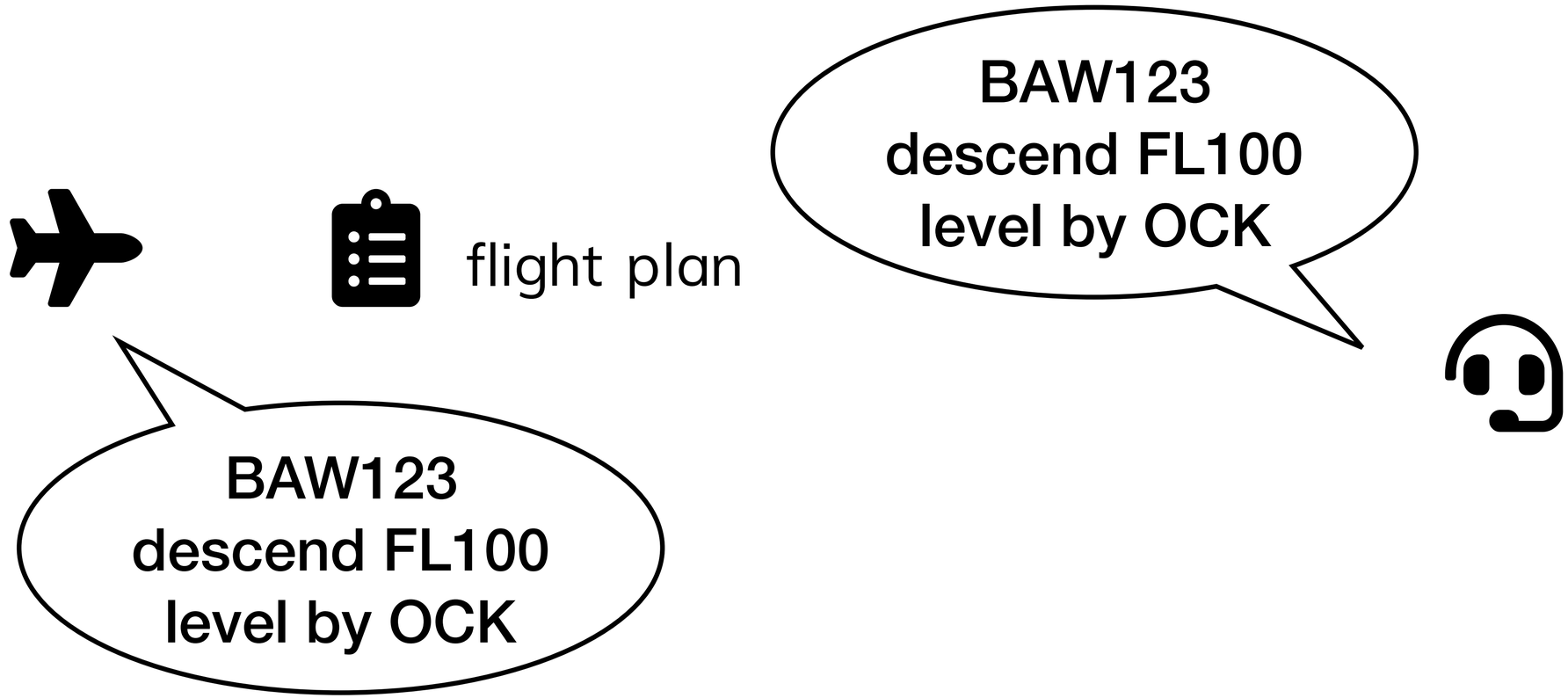
How does air traffic control work

flight plan

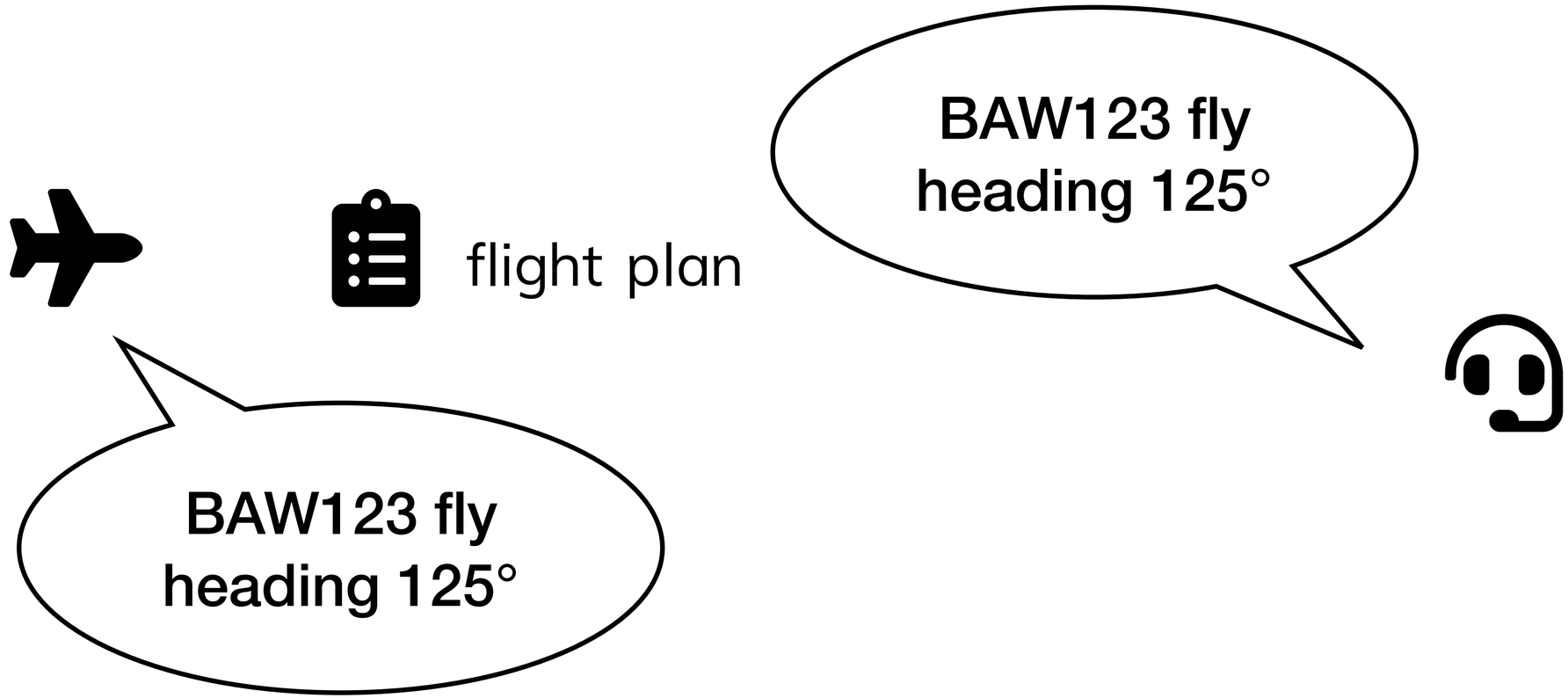
How does air traffic control work



How does air traffic control work



How does air traffic control work



Loss of separation

5 nautical miles laterally

1000 feet vertically

Loss of separation

5.76 miles

5 nautical miles laterally

304.8 meters

1000 feet vertically

Loss of separation

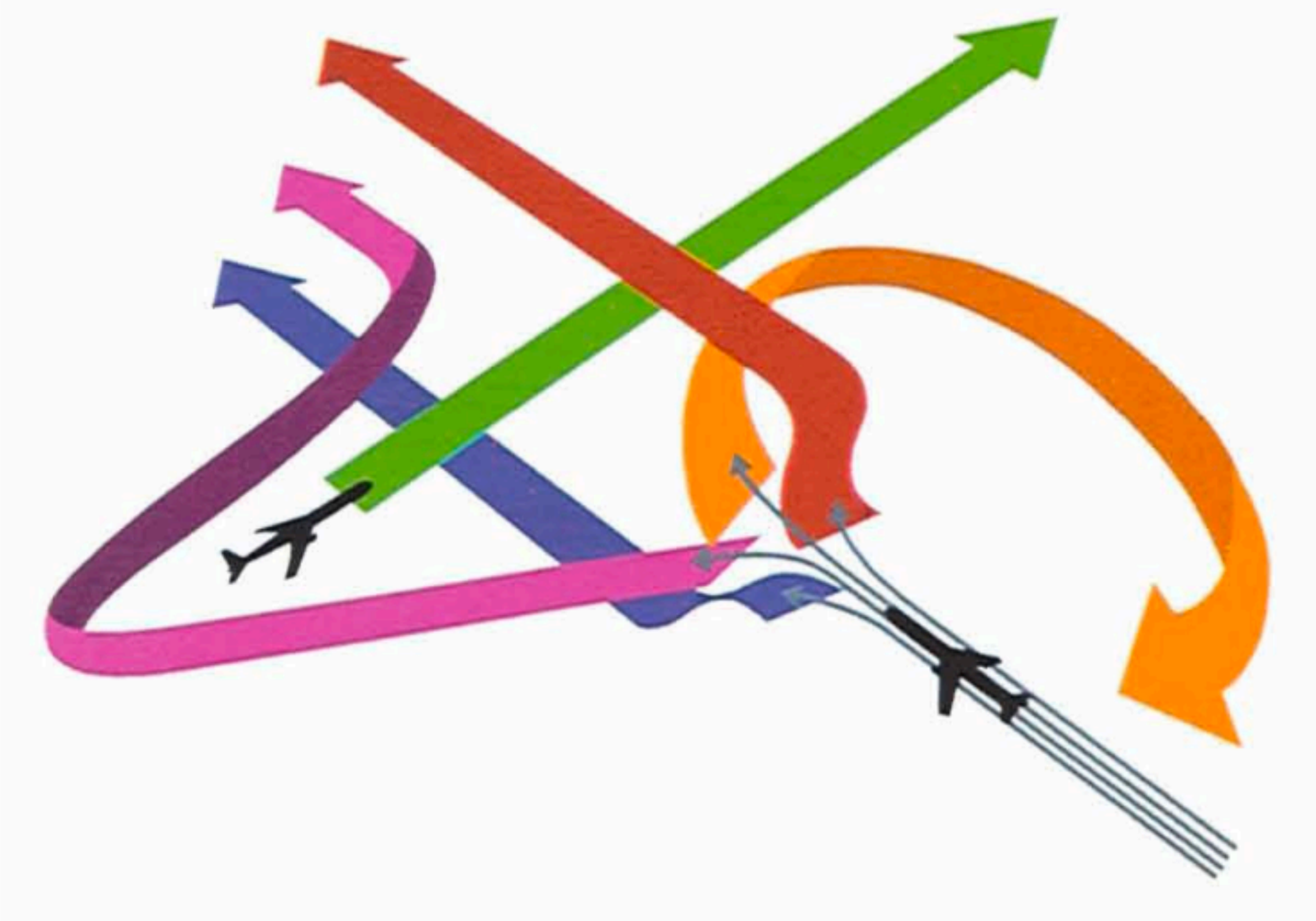
5.76 miles, 9.26 km

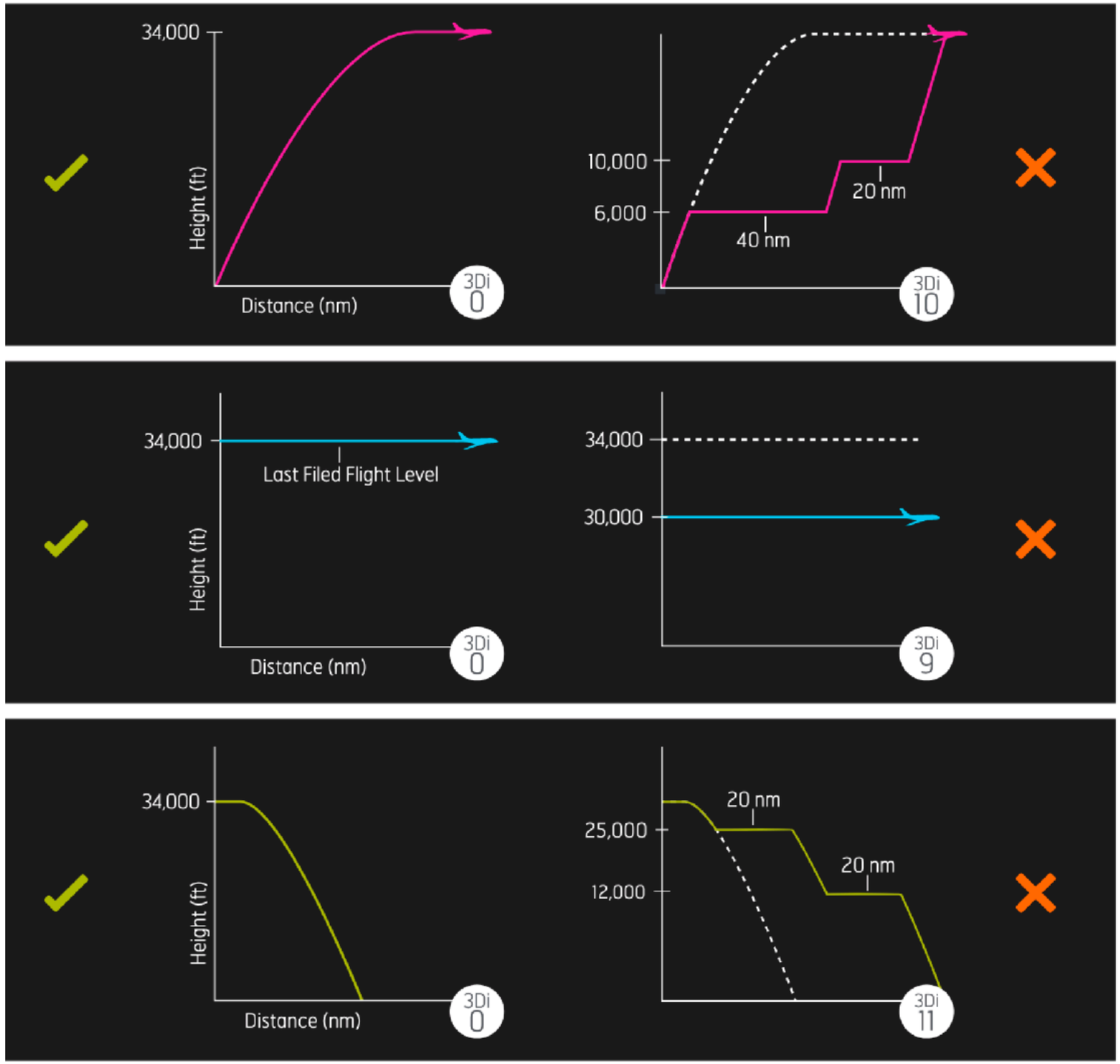
5 nautical miles laterally

304.8 meters

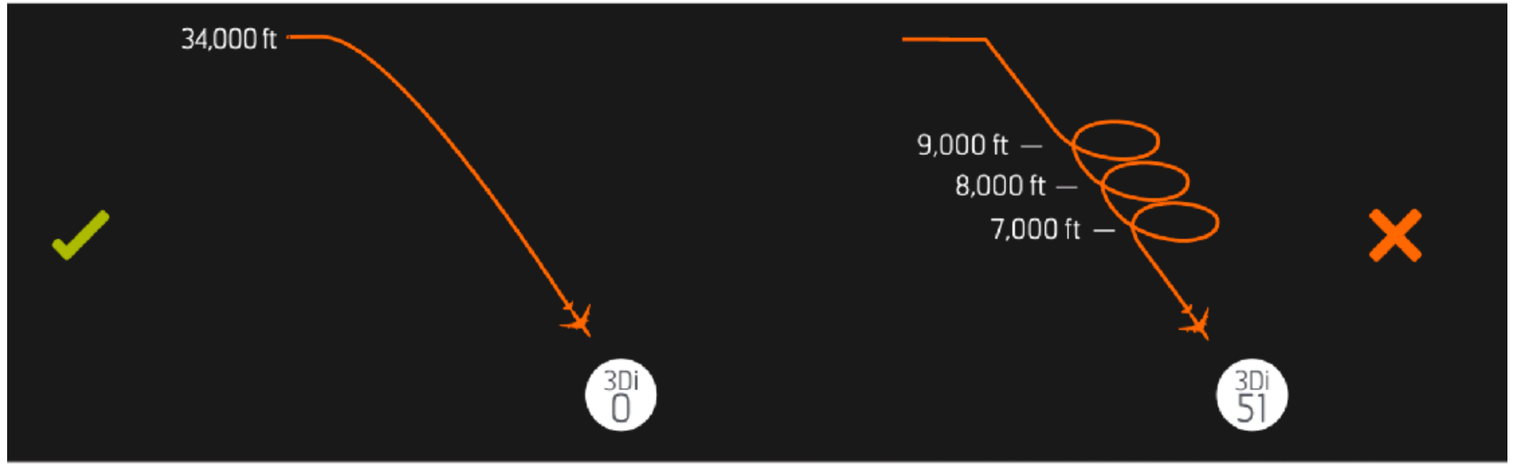
1000 feet vertically

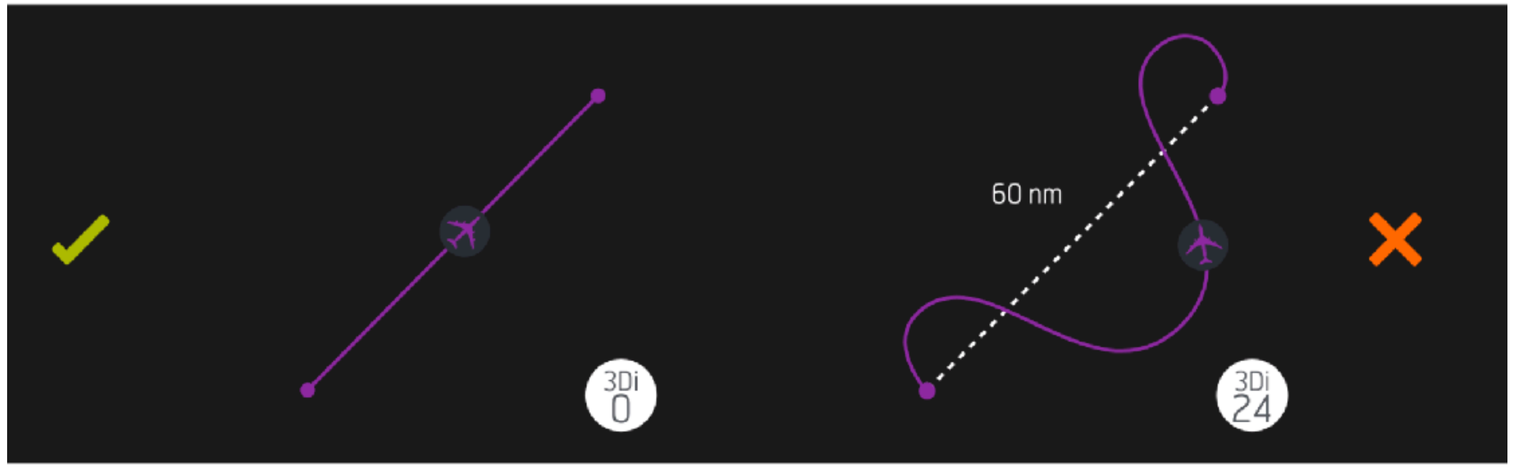
Avoiding loss of separation

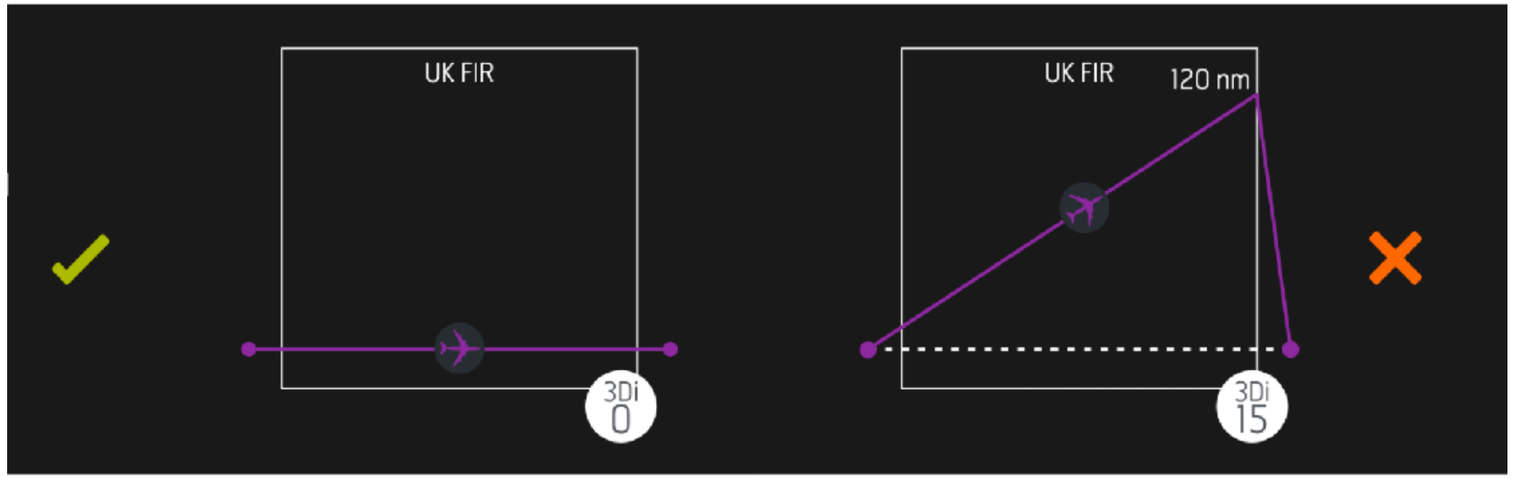




Additional objectives







Additional objectives

Free route airspace

Main sources of uncertainty

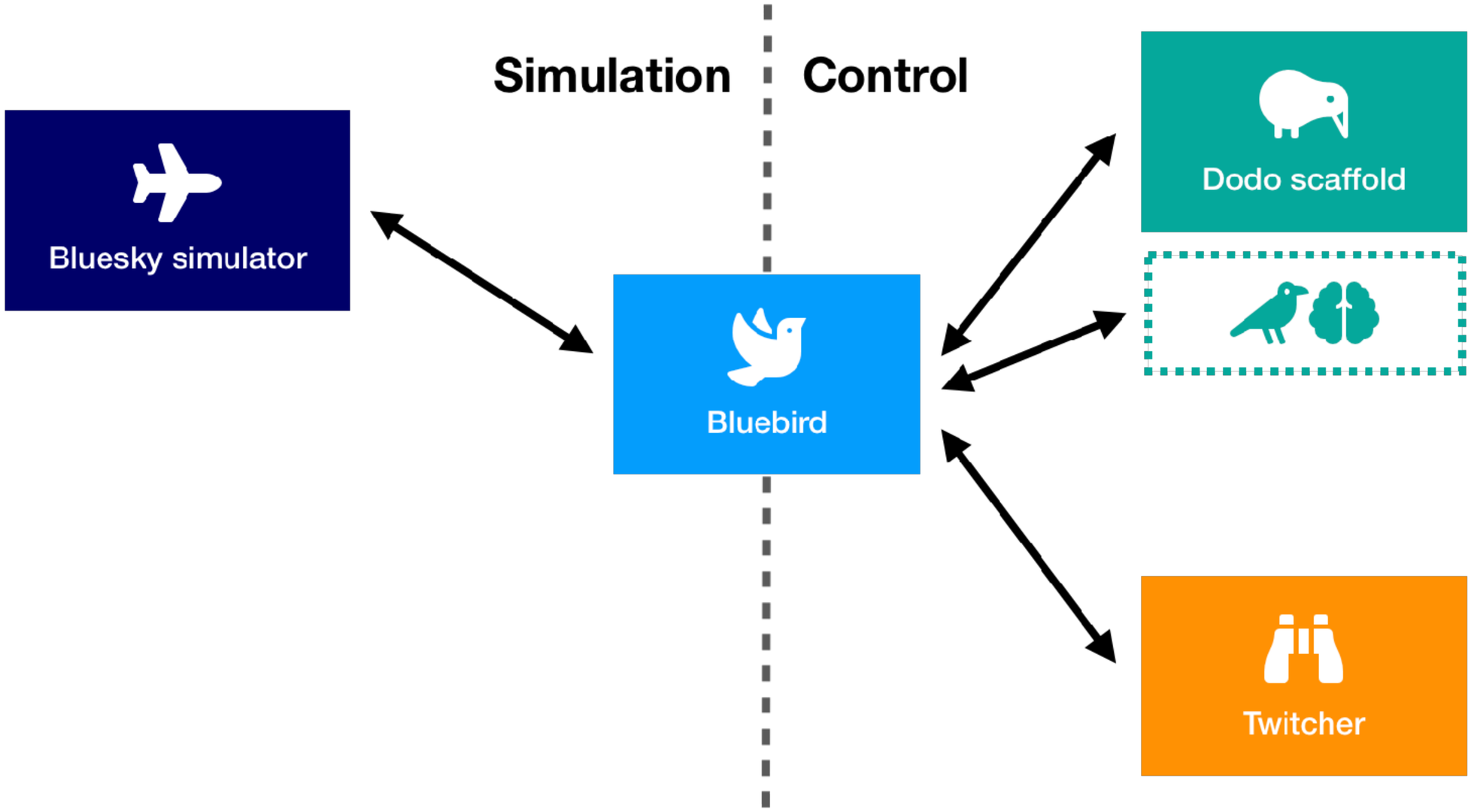
Pilot behaviour

Aircraft load

Weather

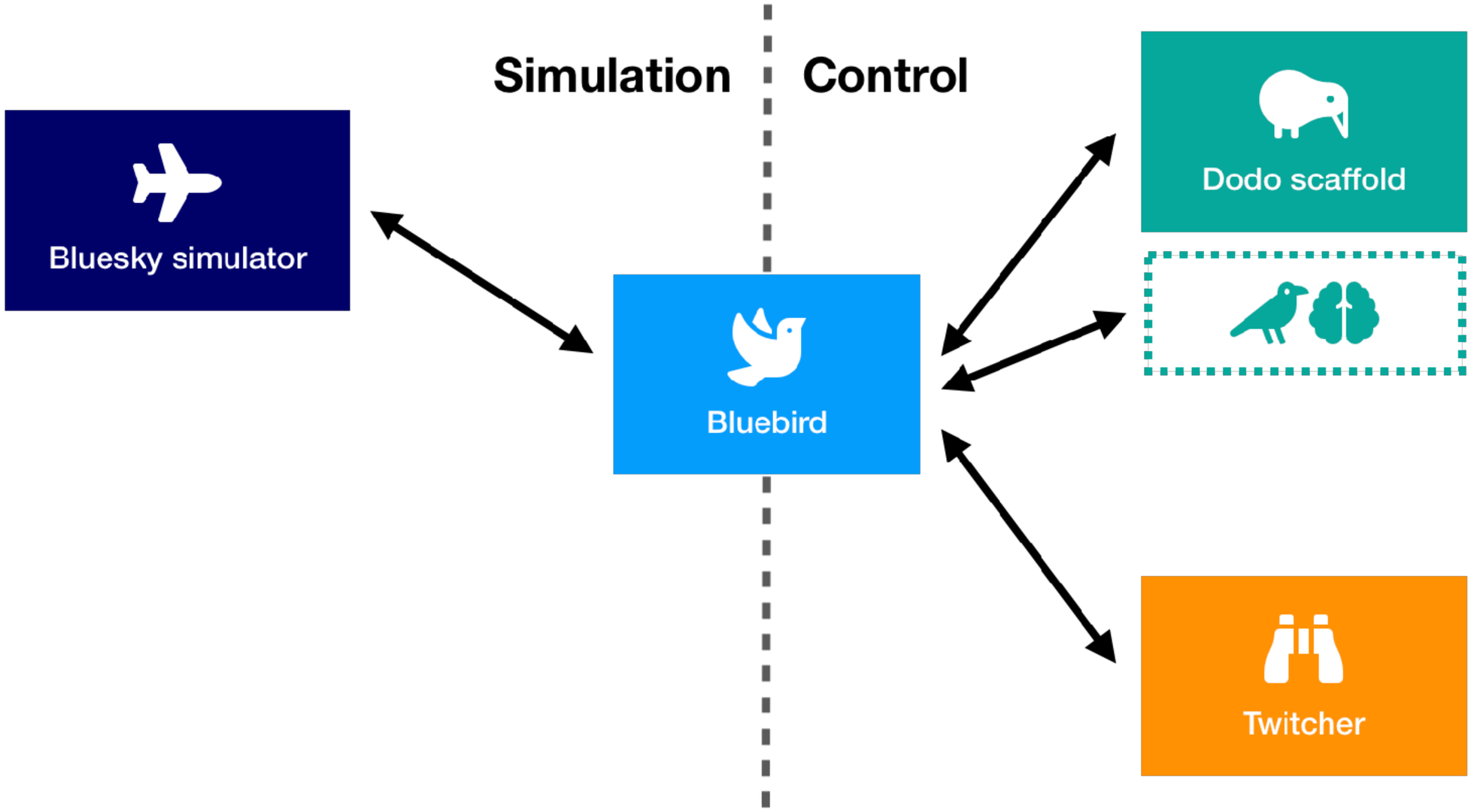
t: 00:00:50, **Δt:** 0.05, **Speed:** 1.0x, **UTC:** 2019-04-04 00:00:50, **Mode:** Operate, **Aircraft:** 6, **Conflicts:** 0/0, **LoS:** 0/0 Node 1:1

Op

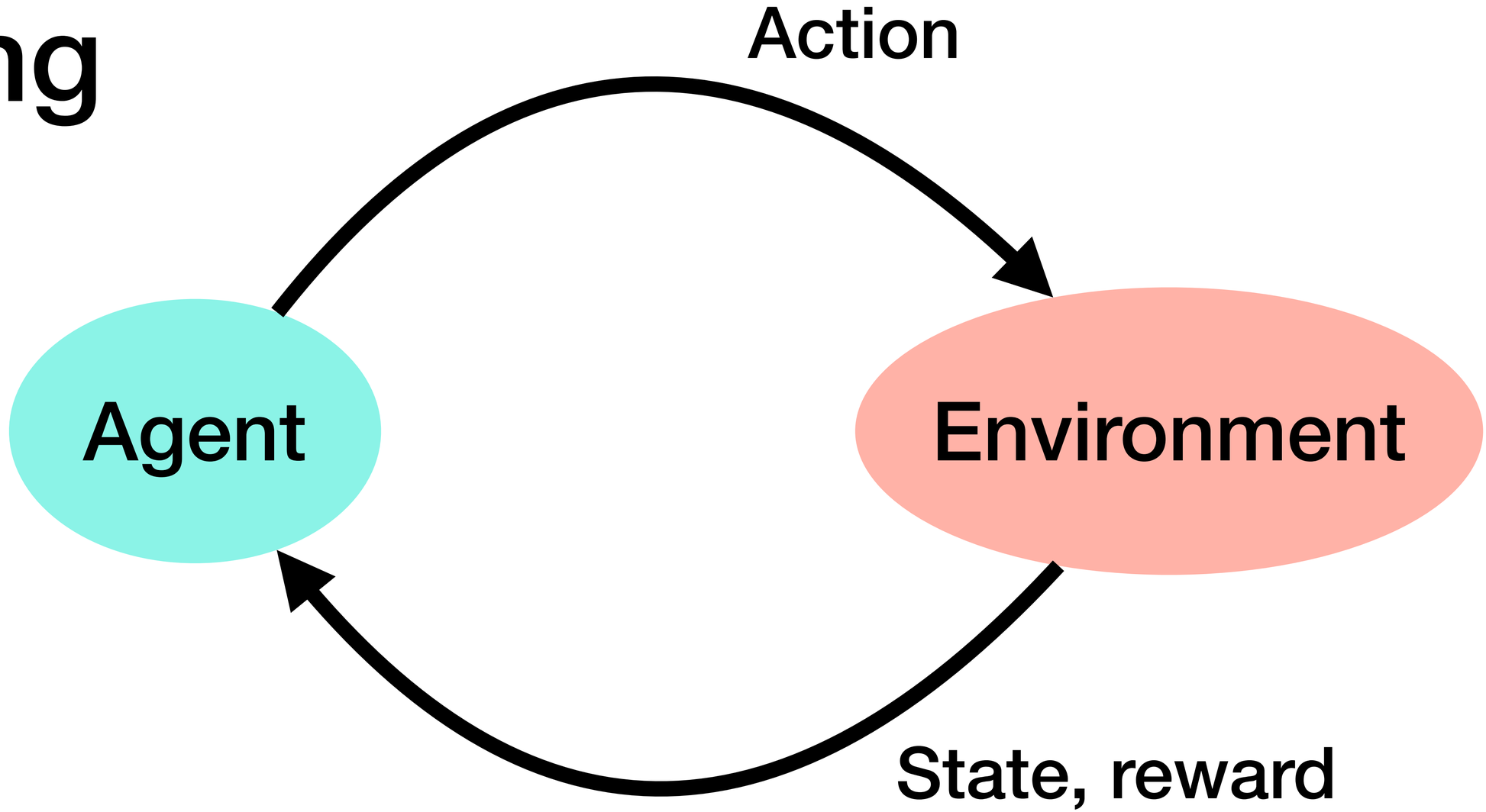

Simulation
Control
Bluesky simulator
Bluebird
Dodo scaffold
Twitcher

PLEASE DO NOT TOUCH THE SCREENS
PLEASE DO NOT TOUCH THE SCREENS

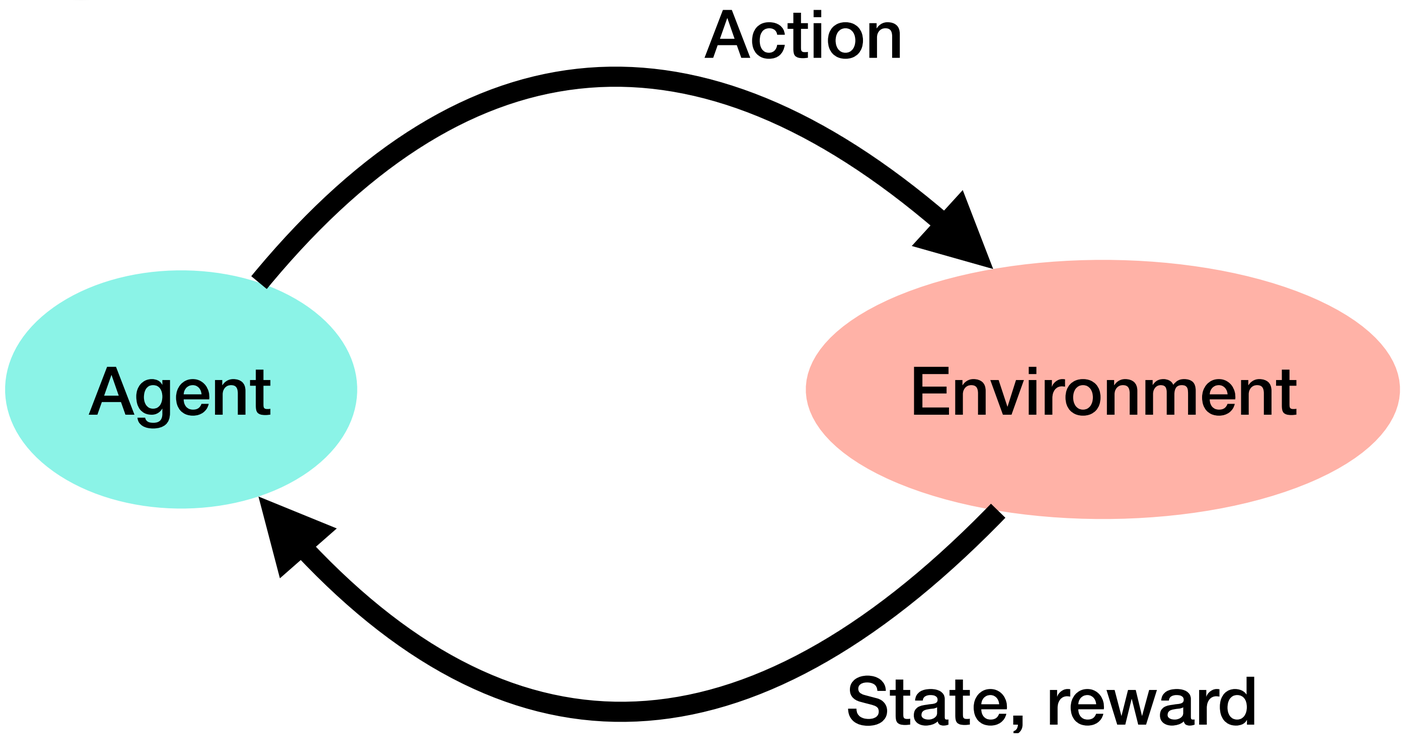
Demo


Simulation
Control
Bluesky simulator
Bluebird
Dodo scaffold
Twitcher

Reinforcement learning



Reinforcement learning



Step: State -> Action -> State', Reward

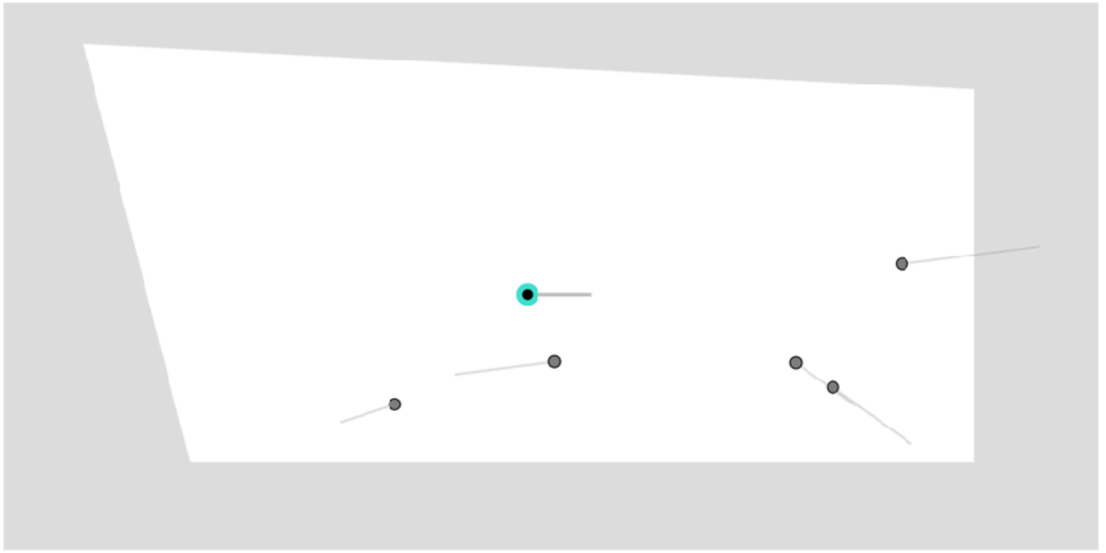
Open experimentation platform

Twitcher

localhost:8080

Twitcher

The Alan Turing Institute



GENERAL CONTROLS

- Run as observer
- Load test scenario
- Reset simulator

SIMULATION CONTROLS

- Play/Resume
- Pause
- Create aircraft

Aircraft ID	Latitude	Longitude	Altitude
ABC500	**51.100**	**-0.633**	**25000 ft**
BA1003	50.810	0.455	30000 ft
BA1004	50.890	-0.538	30000 ft
BA1006	51.195	0.702	35000 ft
EJ9009	50.887	0.324	20000 ft
MON100	50.757	-1.106	20000 ft

ABC500

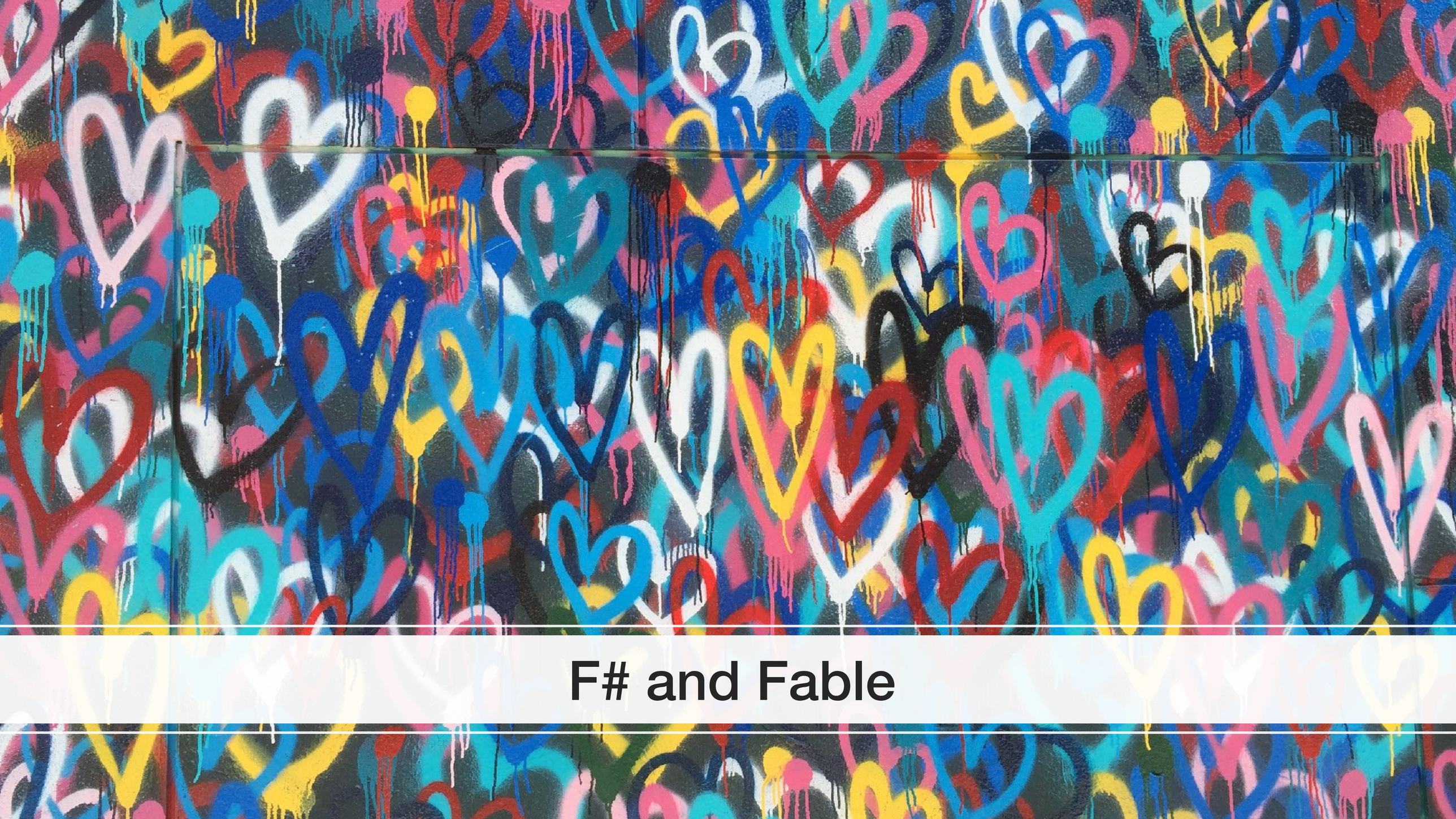
Longitude	-0.633	
Latitude	51.100	
Course	270.0°	Change heading
Altitude	25000 ft	Change
Ground speed	381 knots	Change calib. air speed
Vertical speed	0.0 ft/min	

github.com/alan-turing-institute/

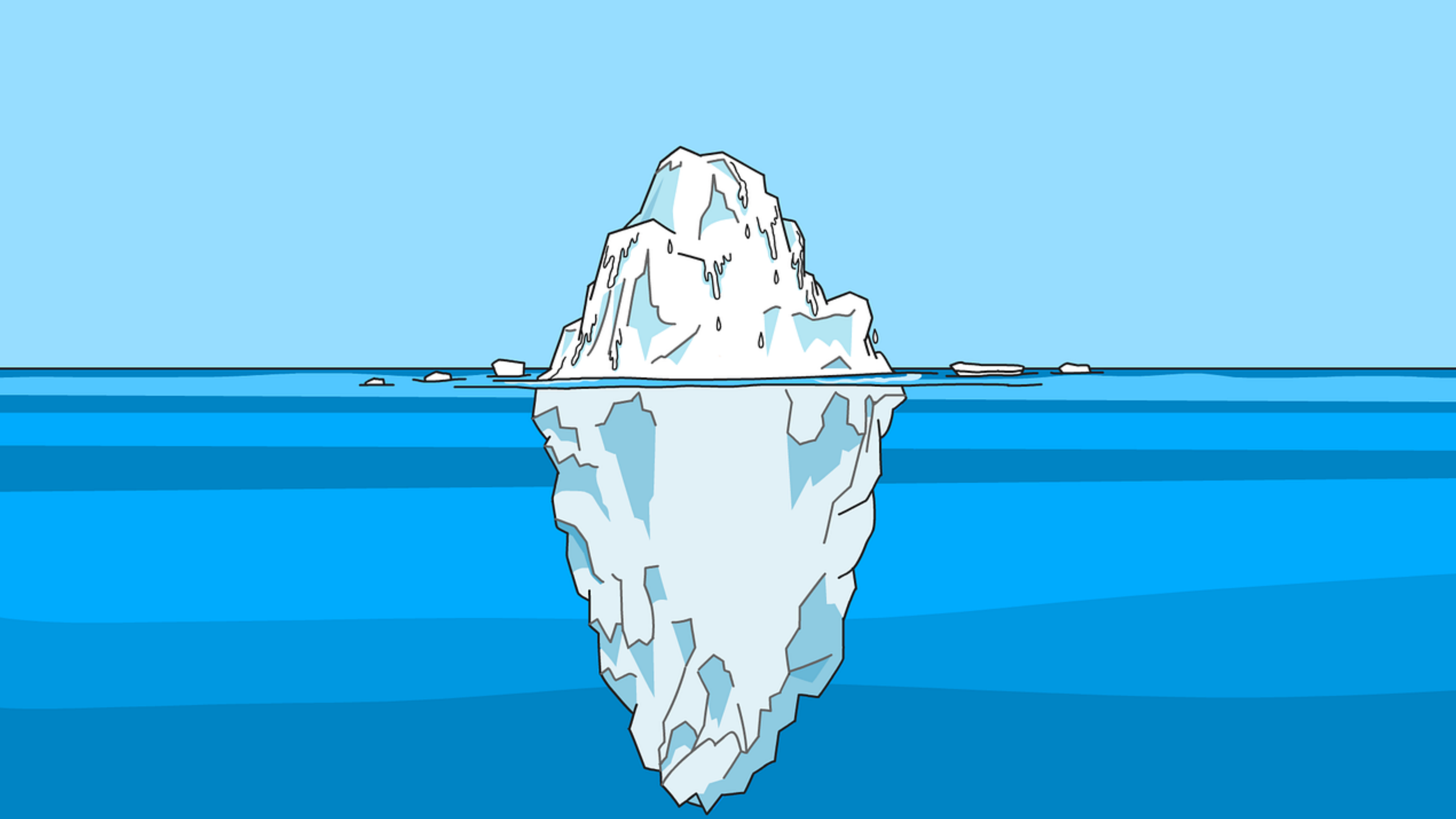
simurgh

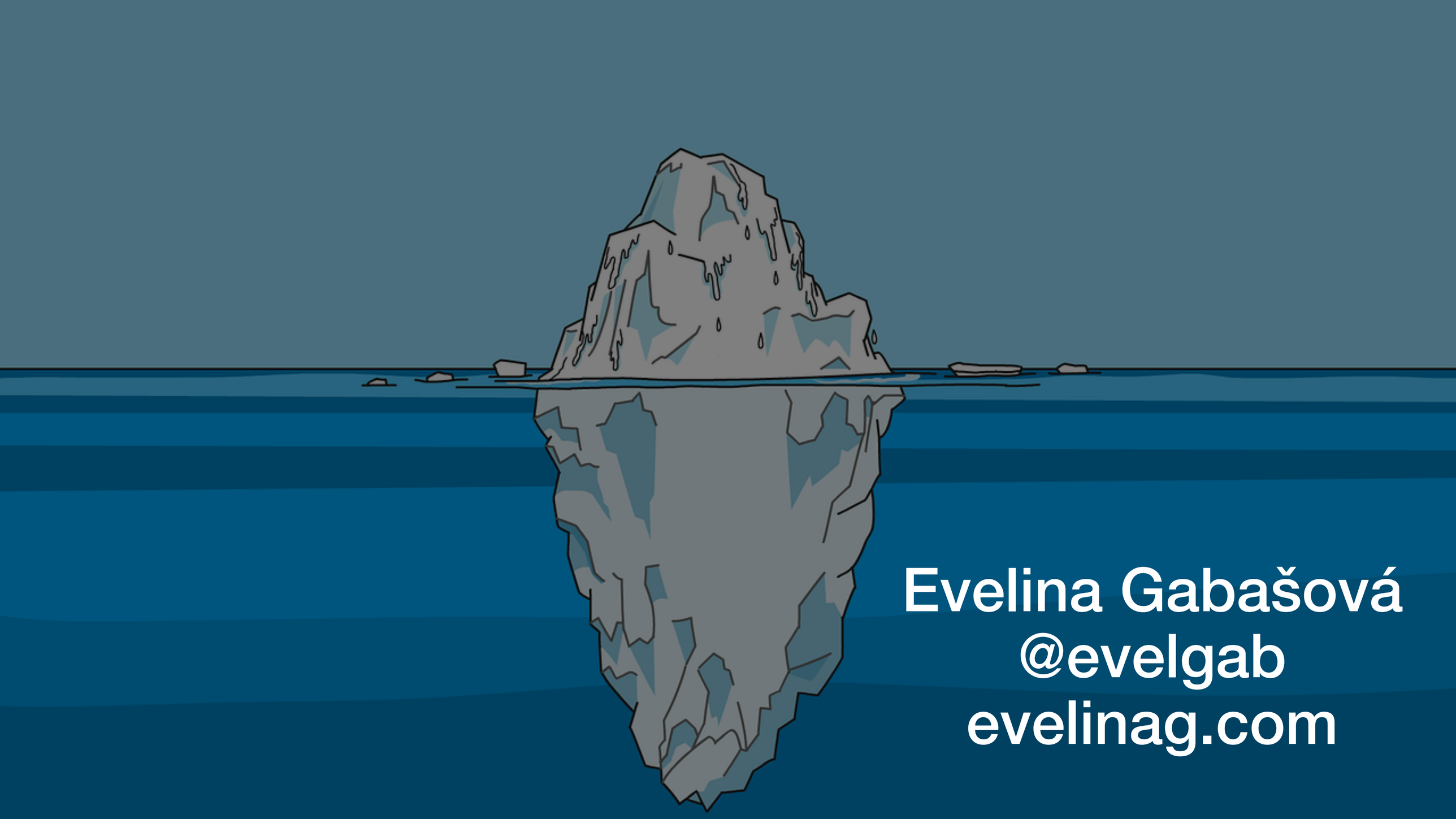
Example from a parallel universe

Shiny



F# and Fable





Evelina Gabašová
@evelgab
evelinag.com