A Grading Rubric for AI Applications

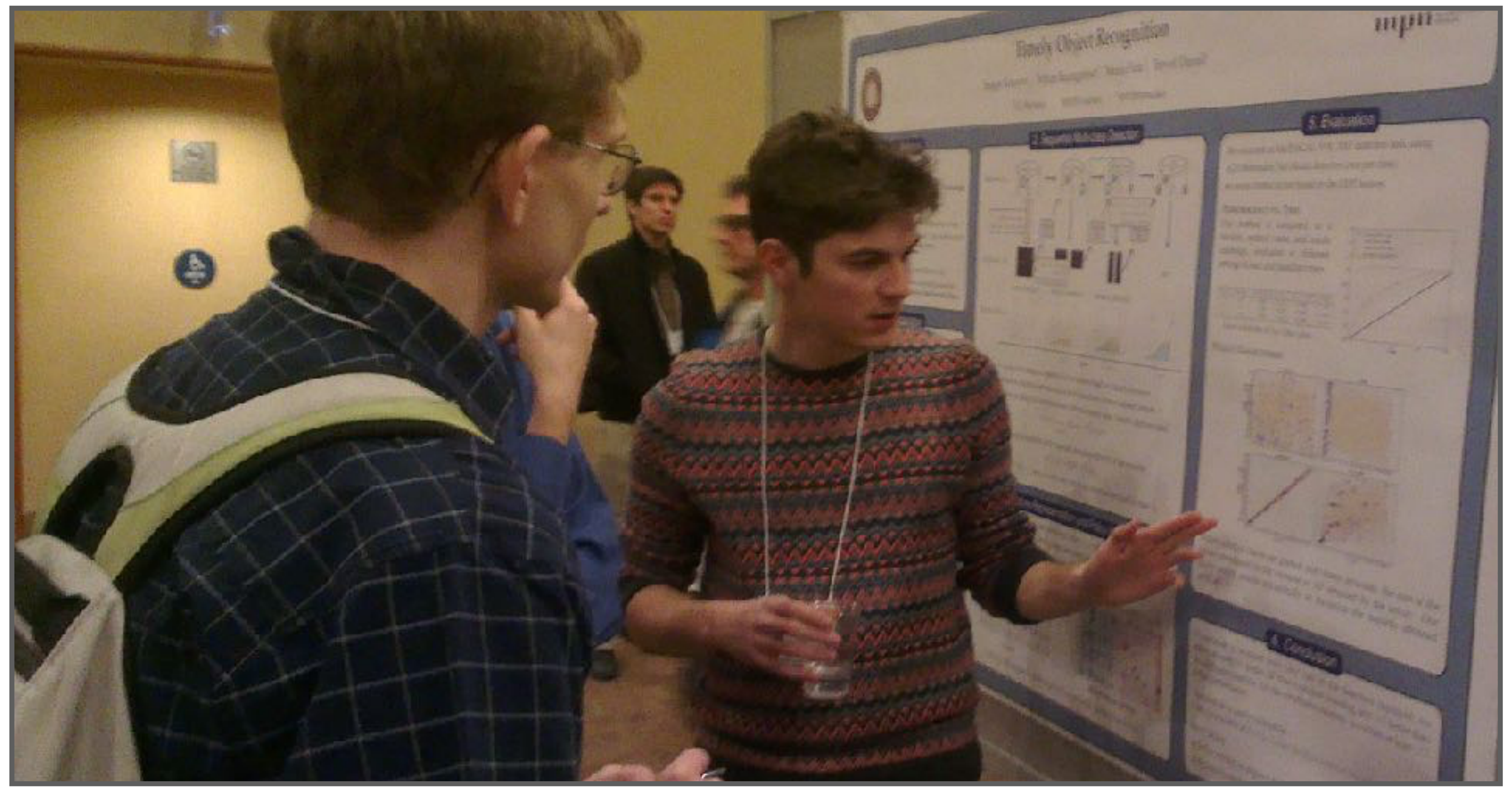
Sergey Karayev
Head of AI for STEM
Turnitin, Inc.

turnitin

2019 May 14

About Me

- Head of AI for STEM at **Turnitin**
  - 35M students served at 150+ countries.
- Co-founder of **Gradescope**
- Co-organizer of **Full Stack Deep Learning** program
- PhD Computer Science at **UC Berkeley**

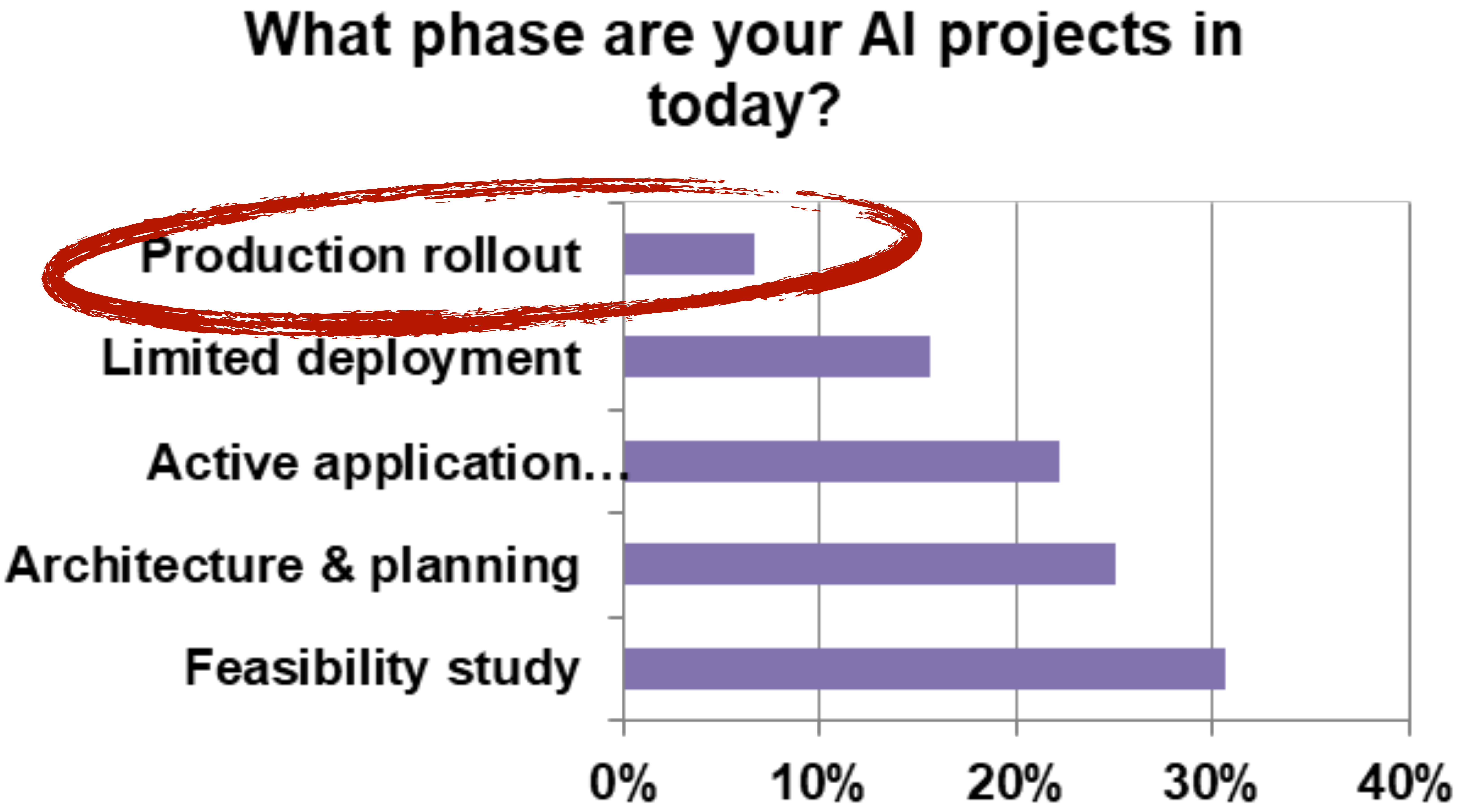


My Goal

- Criteria for evaluating AI projects and applications
  - ...for funding (by governments, investors)
  - ...for deployment (by school systems, companies)
  - **...for development (by startups)**

AI Products: It's Early Days!

2018 survey of 300+ ML-involved developers from many industries



https://www.nextplatform.com/2018/04/24/lagging-in-ai-dont-worry-its-still-early/

A Grading Rubric for AI Applications

Category	✗	✓-	✓+
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

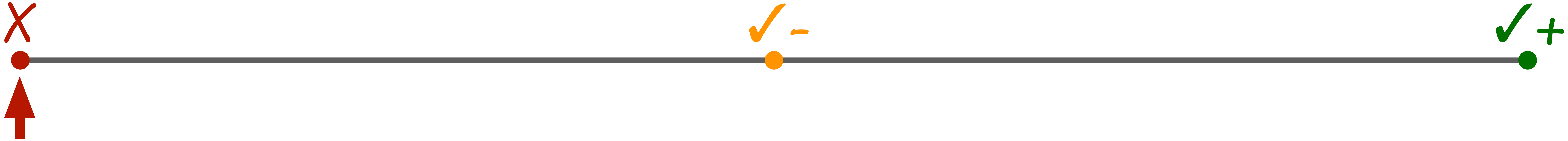
A Grading Rubric for AI Applications

Category	✗	✓-	✓+
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

Task Formulation

- Is the problem well defined, with clear metrics?

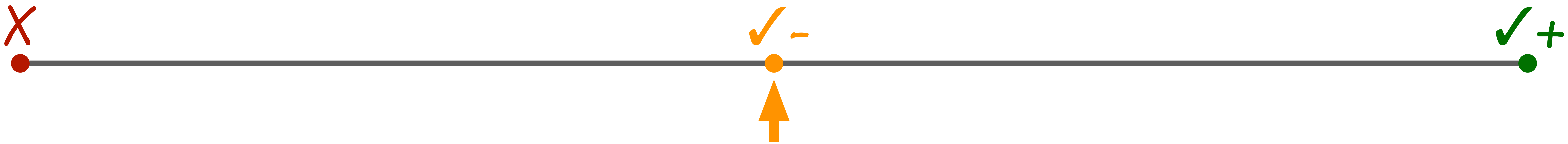
Task Formulation



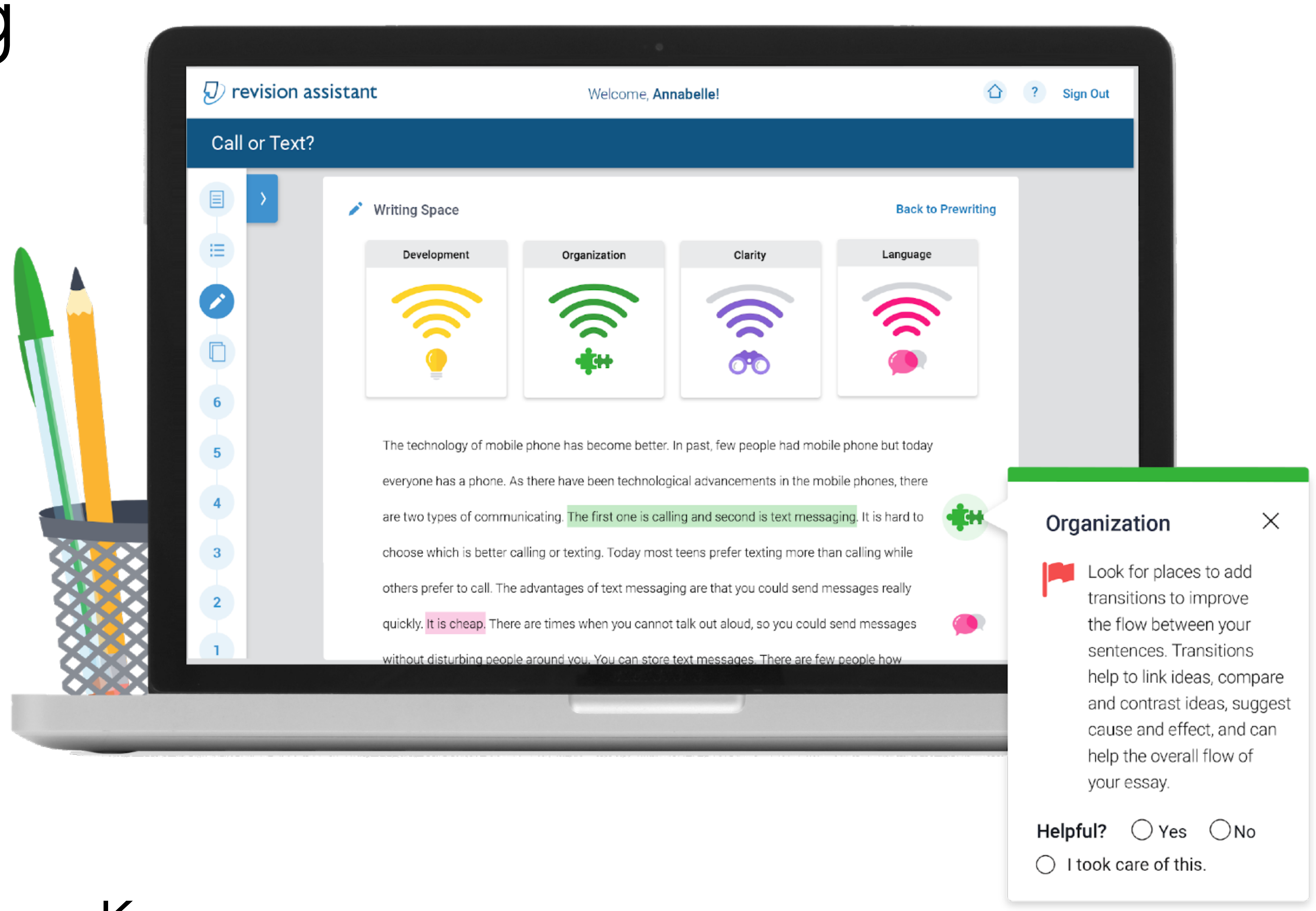
- No, AI is a magical solution to an unsolved, unmeasurable problem.

- **Example**: AI for achieving happiness

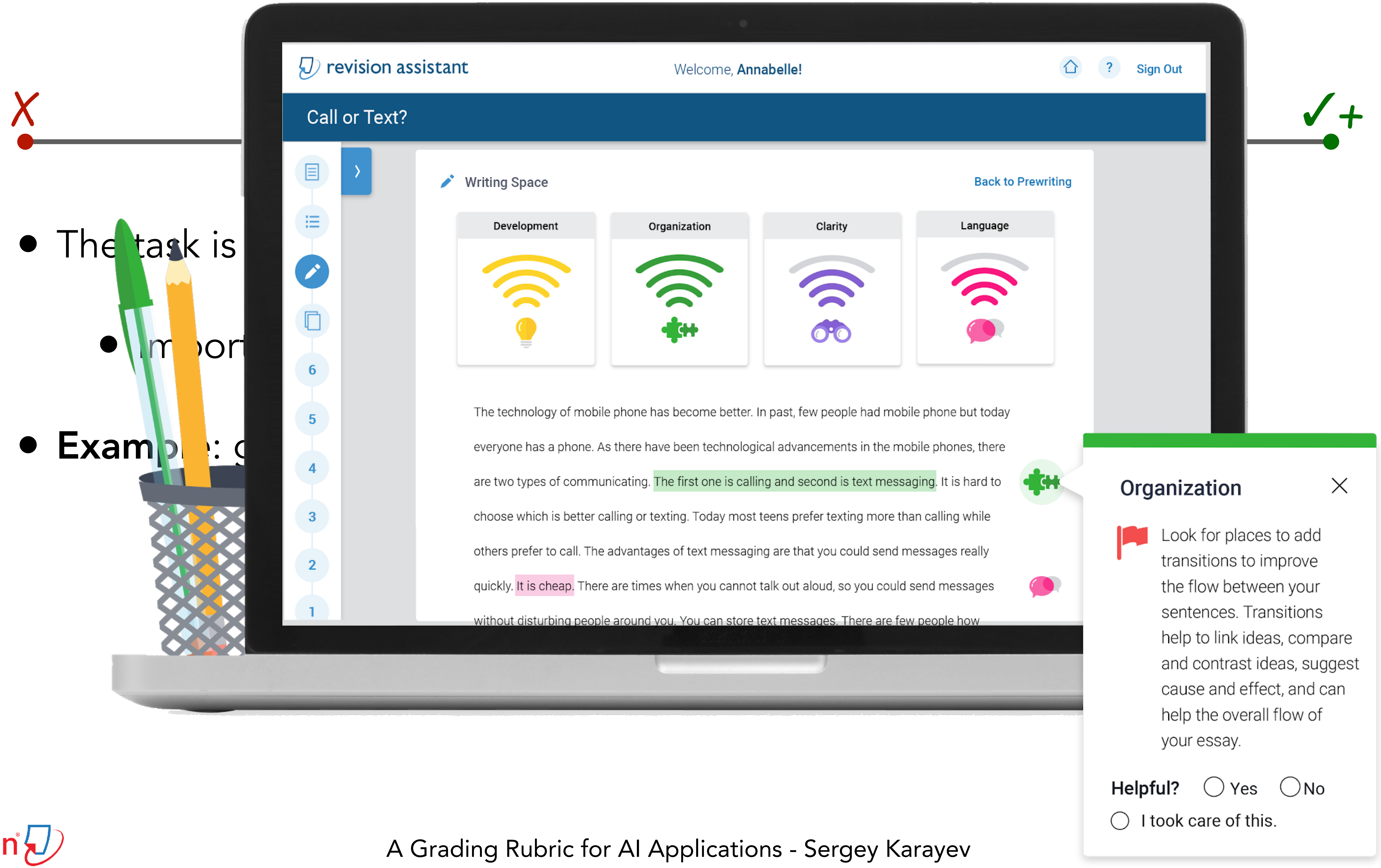
Task Formulation



- The task is complicated, but possible to measure performance.
    - Important to have experts involved in designing the product!
- **Example**: giving students feedback on writing



- The task is
  - Import
- Example: g



Task Formulation

✗ ✓- ✓+

- The task is well-defined and there is a good baseline solution.
- **Example**: Facebook face recognition

https://techcrunch.com/2017/12/19/facebook-facial-recognition-photos/

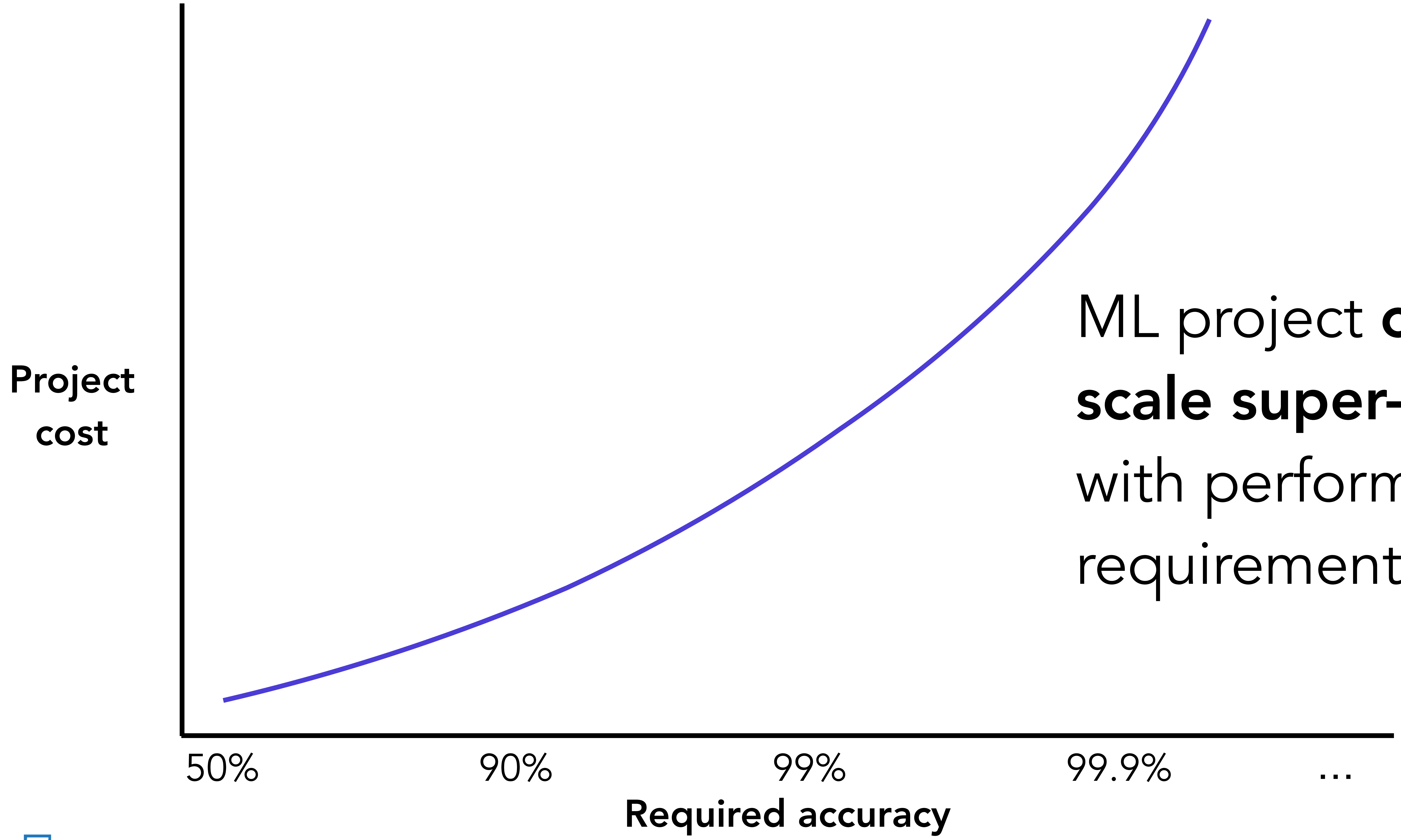
A Grading Rubric for AI Applications

Category	✗	✓-	✓+
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

User Interface and Performance Requirements

- Is the AI autonomous or assisting the user?
- What % of time does AI need to make a prediction?
- How bad is it to make a mistake?

User Interface and Performance Requirements



ML project **costs scale super-linearly** with performance requirements

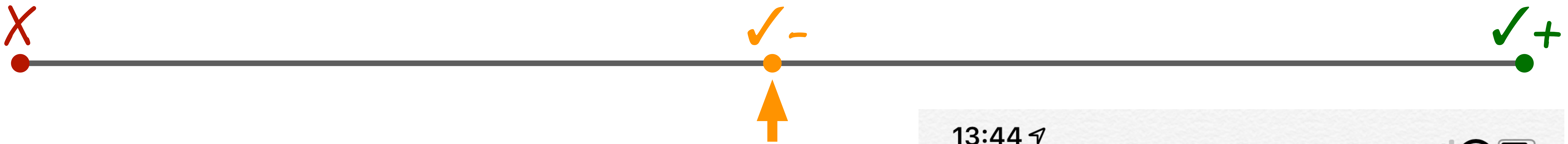
Diagram courtesy of Josh Tobin

User Interface and Performance Requirements

✗ ✓- ✓+

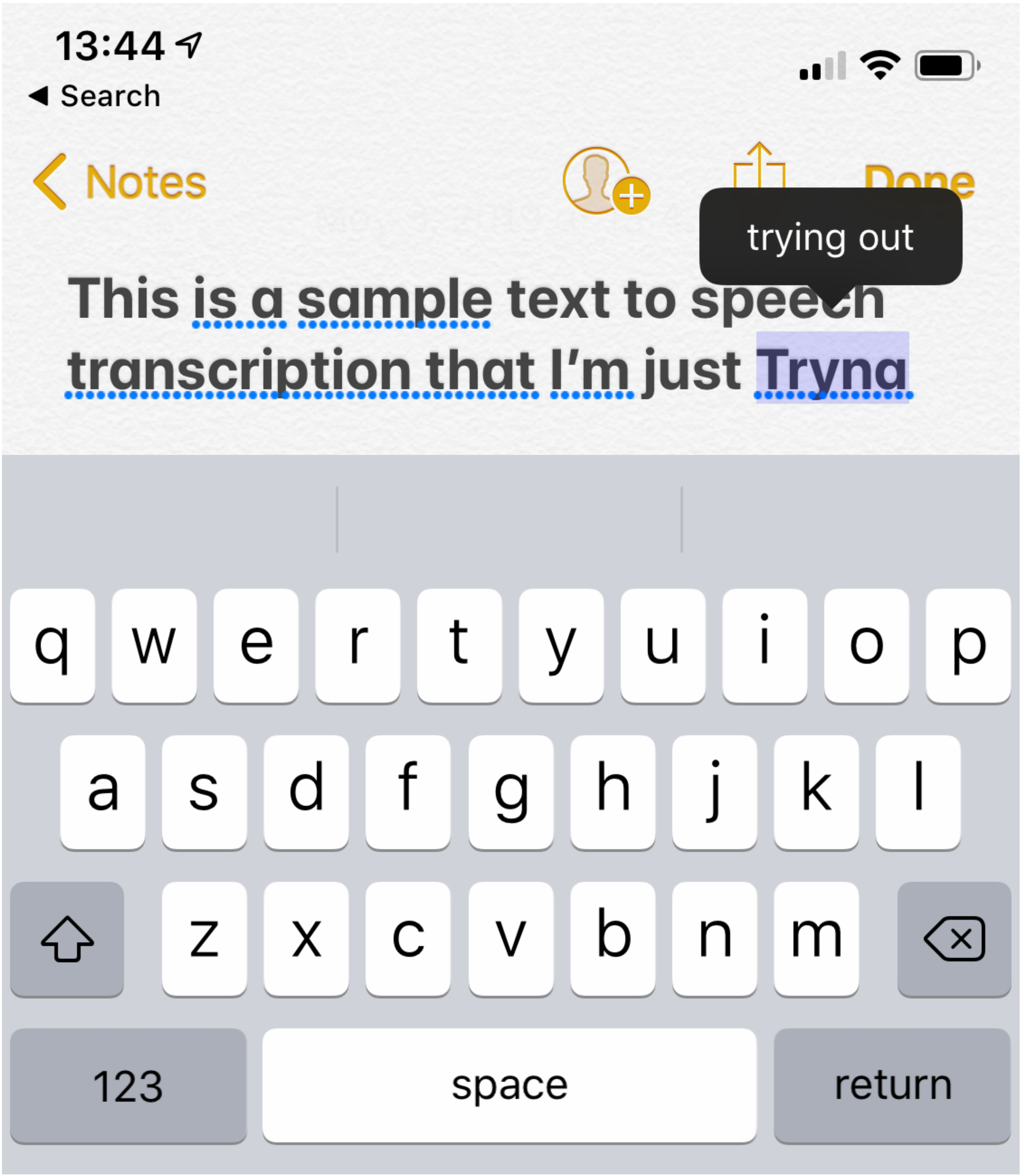
- AI must be 100% autonomous, predicting all the time, and mistakes are fatal.
- **Example**: self-driving cars

User Interface and Performance Requirements

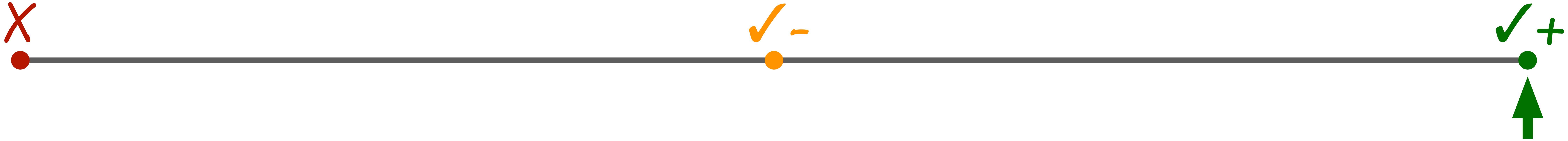


- AI always has to predict, and user resents correcting mistakes.

- **Example**: speech-to-text transcription

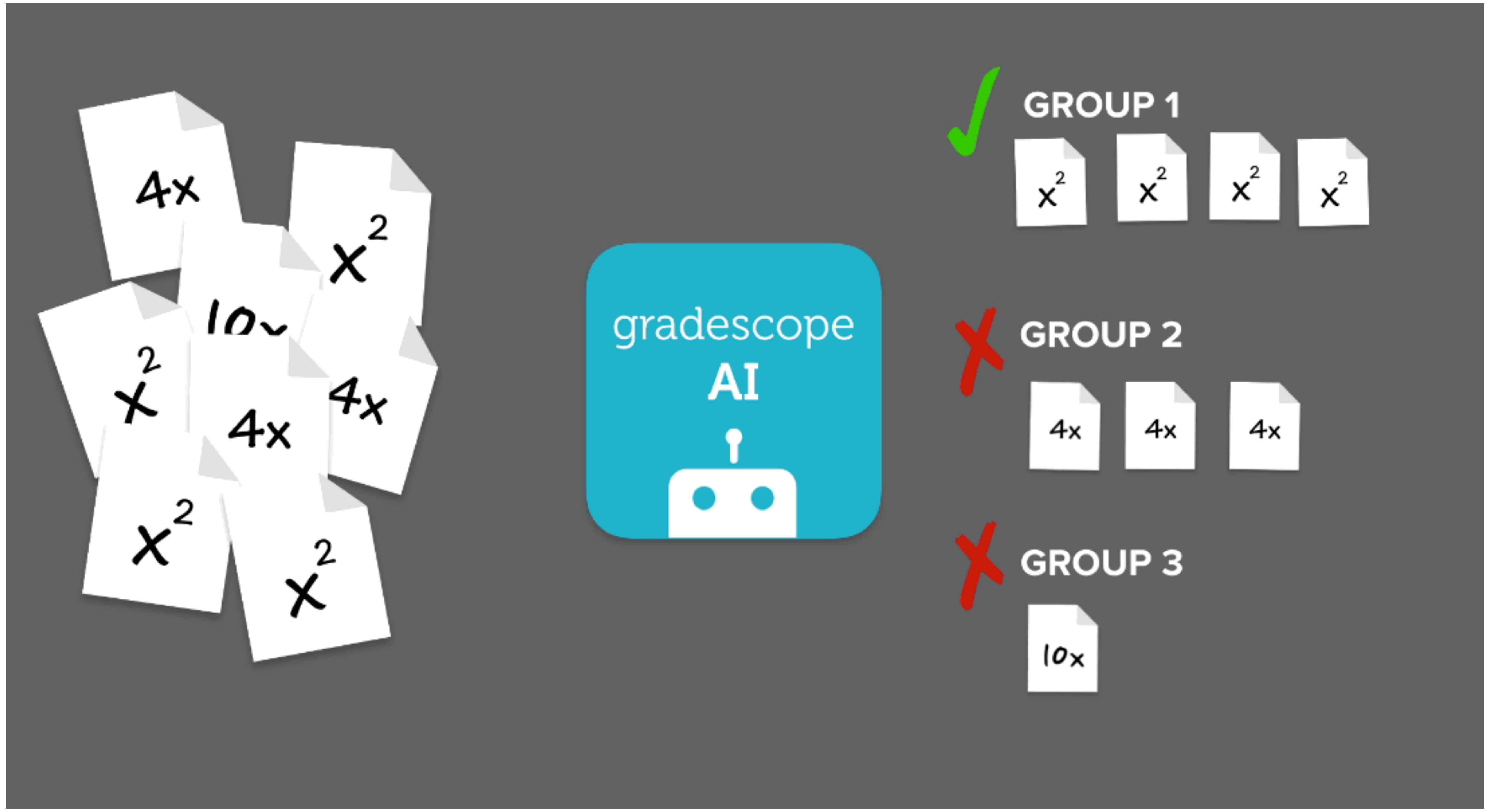


User Interface and Performance Requirements



- AI doesn't have to predict every time, and user feels fine about correcting mistakes.

- **Example**: Gradescope
  AI-assisted answer grouping



?

x^2

$\frac{1}{2}x^2 + C$

$\frac{1}{2}x^2 + C$

$\frac{1}{2}x^2 + C$

x^2

$\frac{1}{2}x^2 + C$

x^2

User Interface and Performance Requirements for Education Applications

Aiding instructor

Aiding student

Replacing instructor

Lower → Higher

Performance Requirements

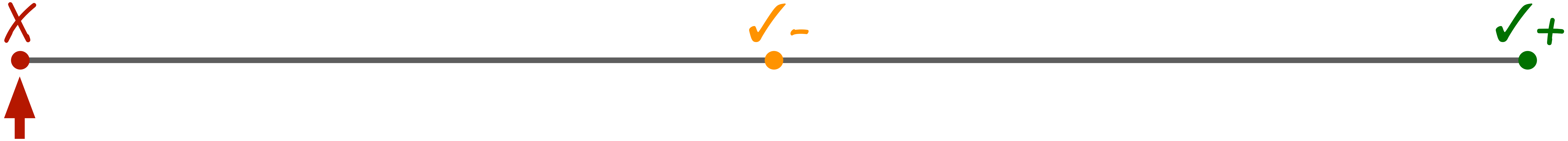
A Grading Rubric for AI Applications

Category	✗	✓-	✓+
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

Technical Difficulty

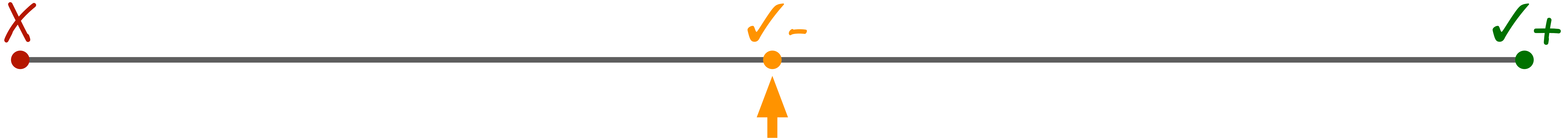
- What are the technical resources needed to solve the problem?
  - (From analogs, academic literature, and expert opinions).

Technical Difficulty

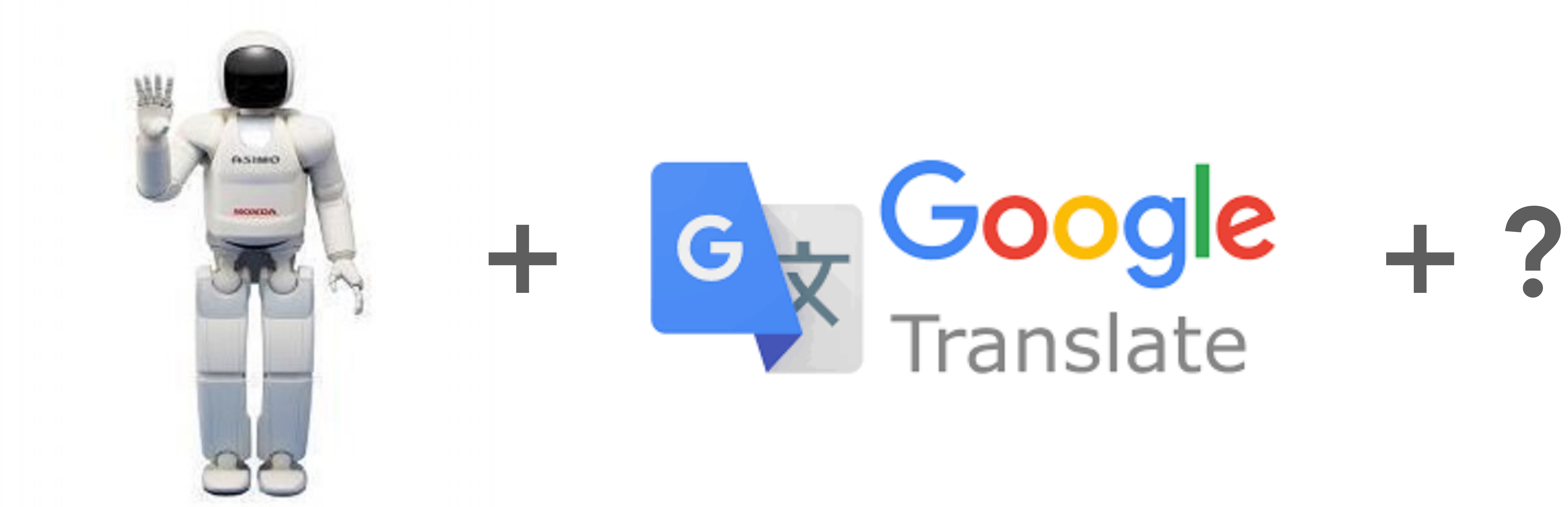


- The problem is several steps removed from today's technology.
- Requires world-class research team and unknown number of years.
- **Example**: intelligent tutor in all subjects and grade levels

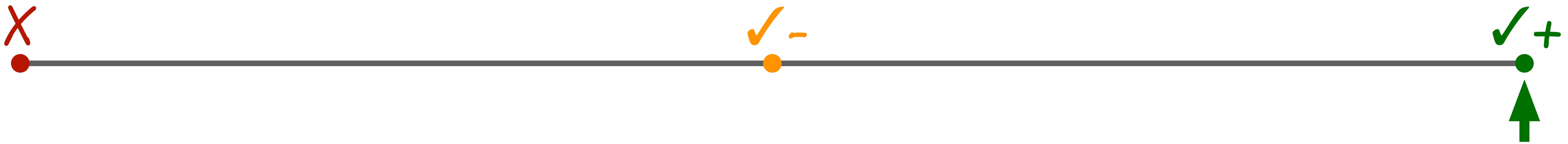
Technical Difficulty



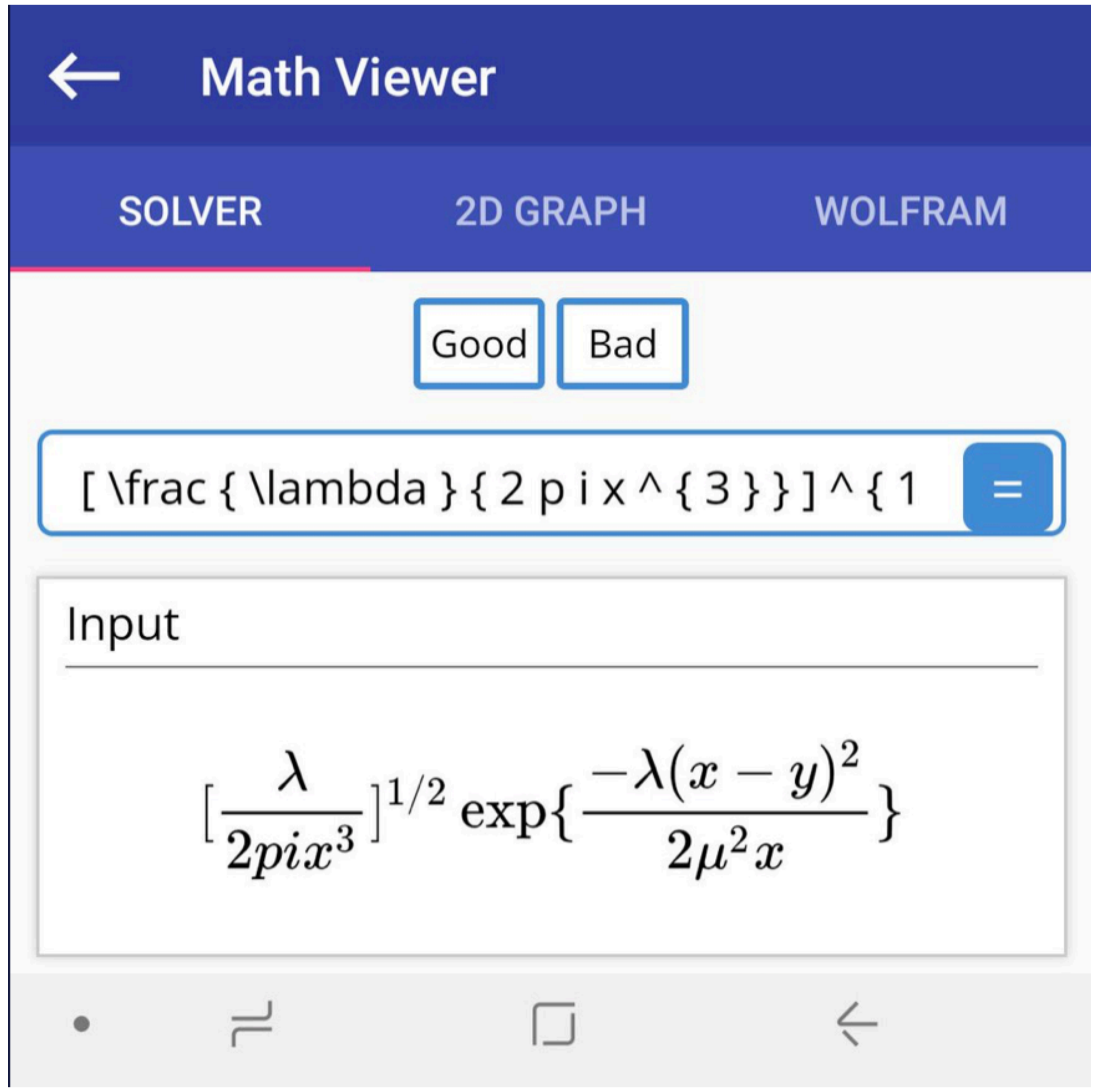
- Reason to believe problem is solvable, but no working analog.
- Requires a research team and possibility of failure.
- **Example**: robot that teaches infants a second language.



Technical Difficulty



- Clear analogs, but technology is not yet mainstream.
- Requires capable engineers and a fixed time frame.
- **Example**: recognize handwritten math



https://twitter.com/lvlzay_sci/status/1085501430767804416

A Grading Rubric for AI Applications

Category	✗	✓-	✓+
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

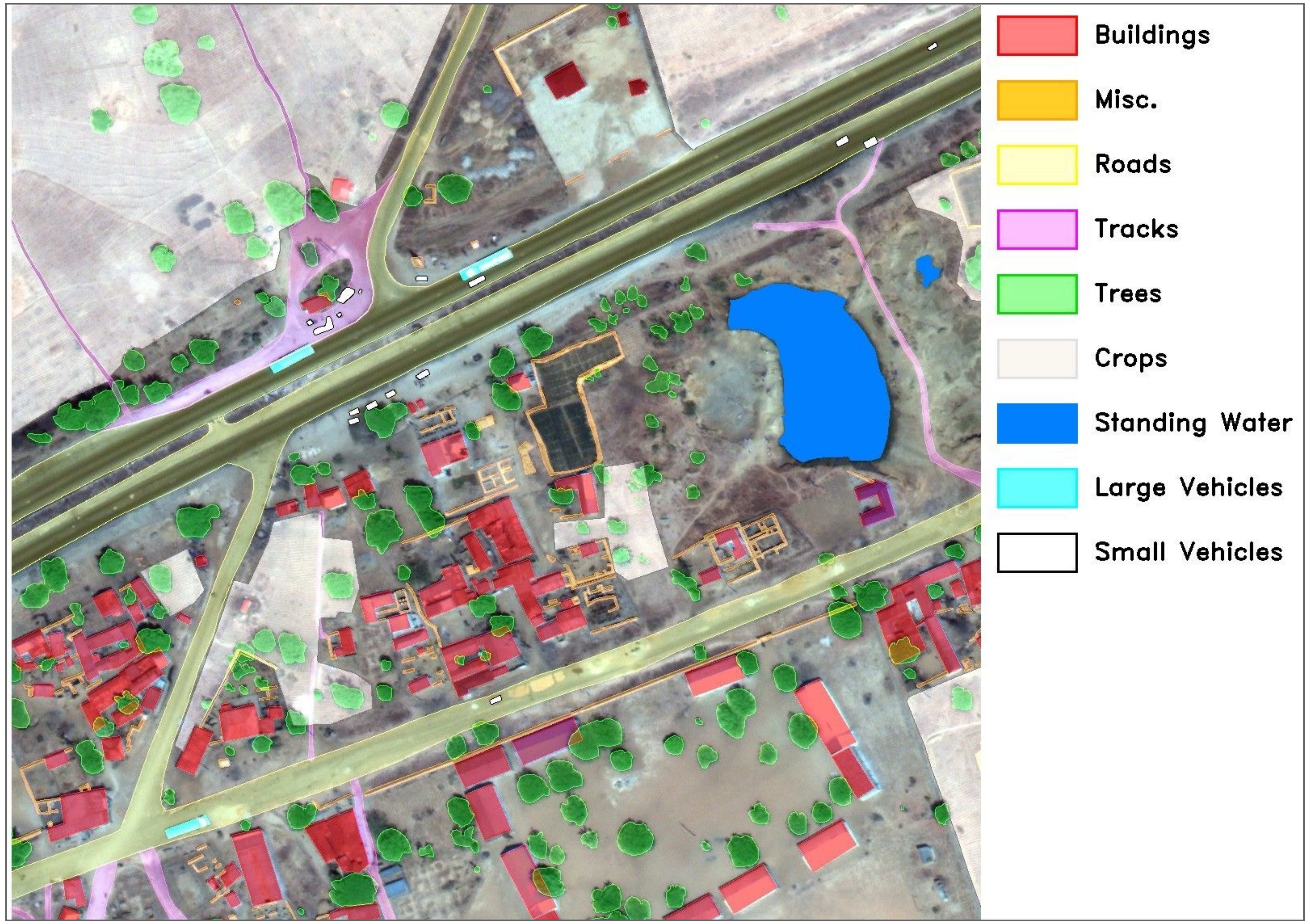
Initial Data Moat

- Defensible AI means proprietary data
    - (This is why Google/Facebook open-source their AI code)
- Is raw data available?
- How expensive it is to label?
- Can someone else get it?

Initial Data Moat

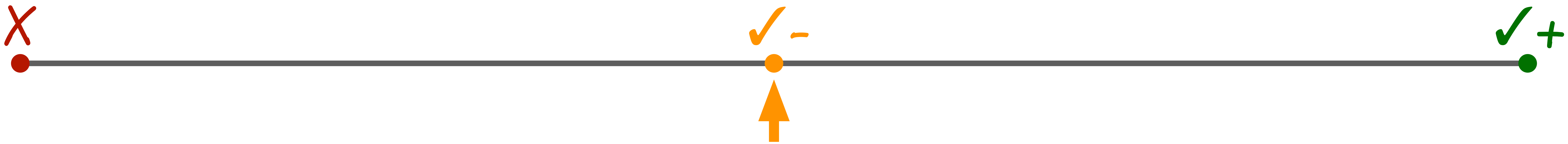
✗ ✓- ✓+

- Data is expensive to obtain and/or label, but there is no exclusivity.
- **Example**: satellite data



https://cdn-sv1.deepsense.ai/wp-content/uploads/2017/04/sample_image_from_the_training_set.jpg

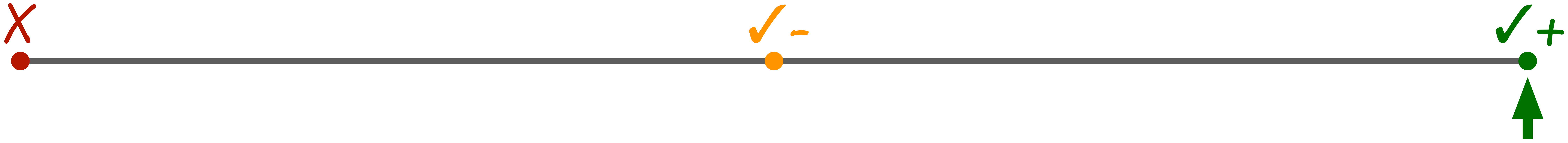
Initial Data Moat



- Data is expensive to obtain and/or label, but there is exclusivity.
- **Example**: radiology data obtained through exclusive partnership

NHS +

Initial Data Moat



- Data is exclusive and already labeled.
- **Example**: detailed educational records of every student in Singapore

A Grading Rubric for AI Applications

Category	✗	✓-	✓+
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

turnitin

Data Flywheel

- Are AI predictions monitored?
- Are wrong predictions corrected and system re-trained?
- Can users do this?

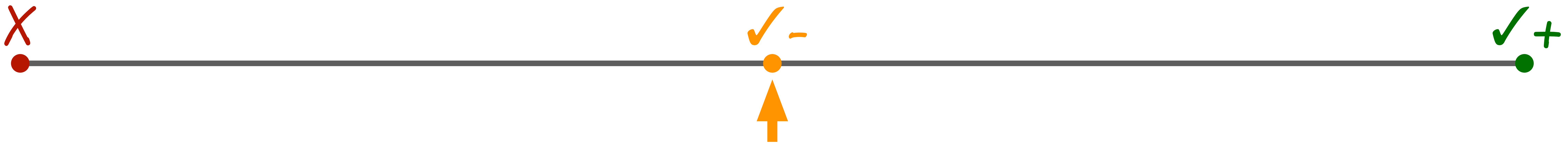
Data Flywheel

- Predictions are not monitored or corrected

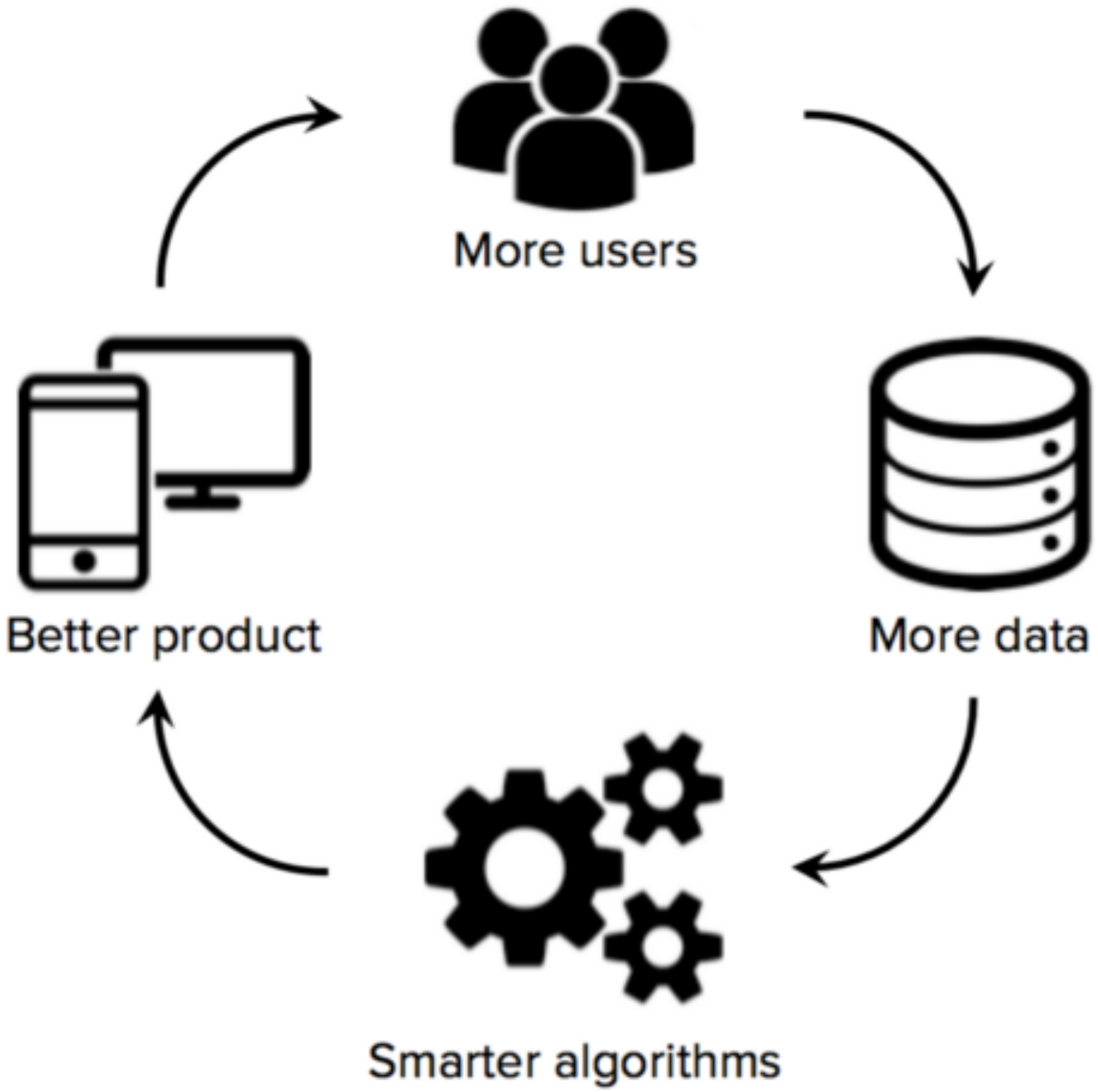
- **Example**: most AI products

https://rvpartners.com.au/avoid-flying-blind-what-you-can-measure-you-can-manage/

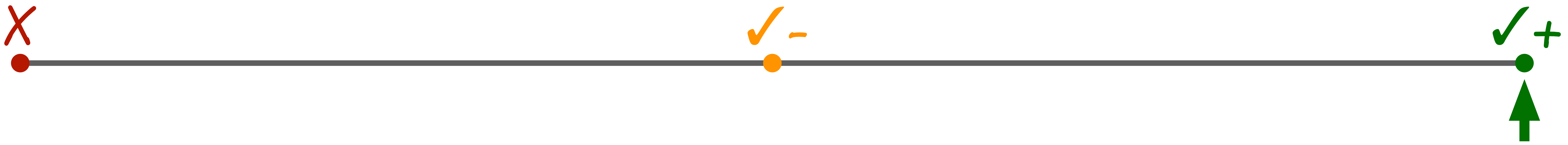
Data Flywheel



- Predictions are monitored, and low-confidence ones are QA'd

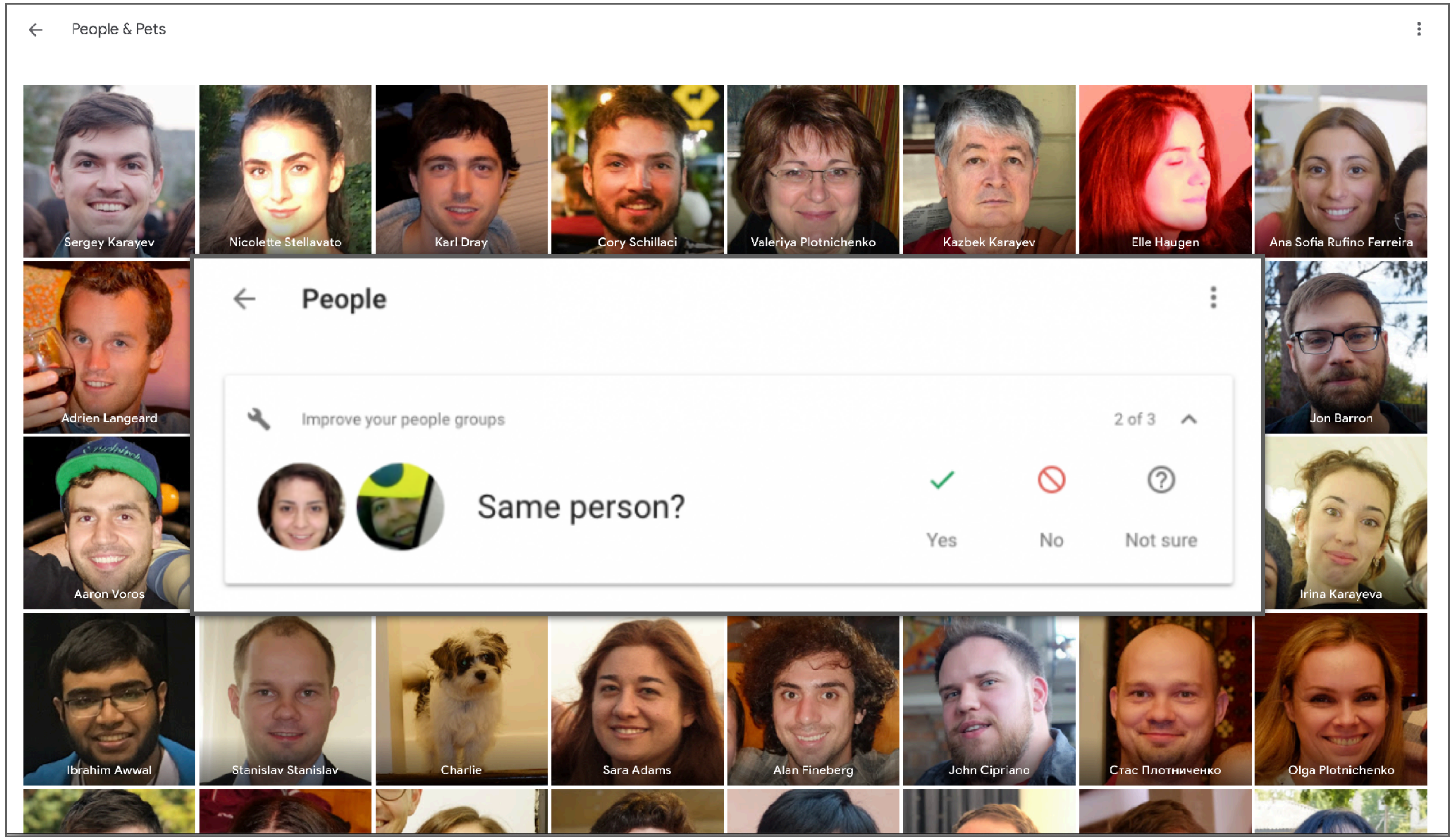


Data Flywheel



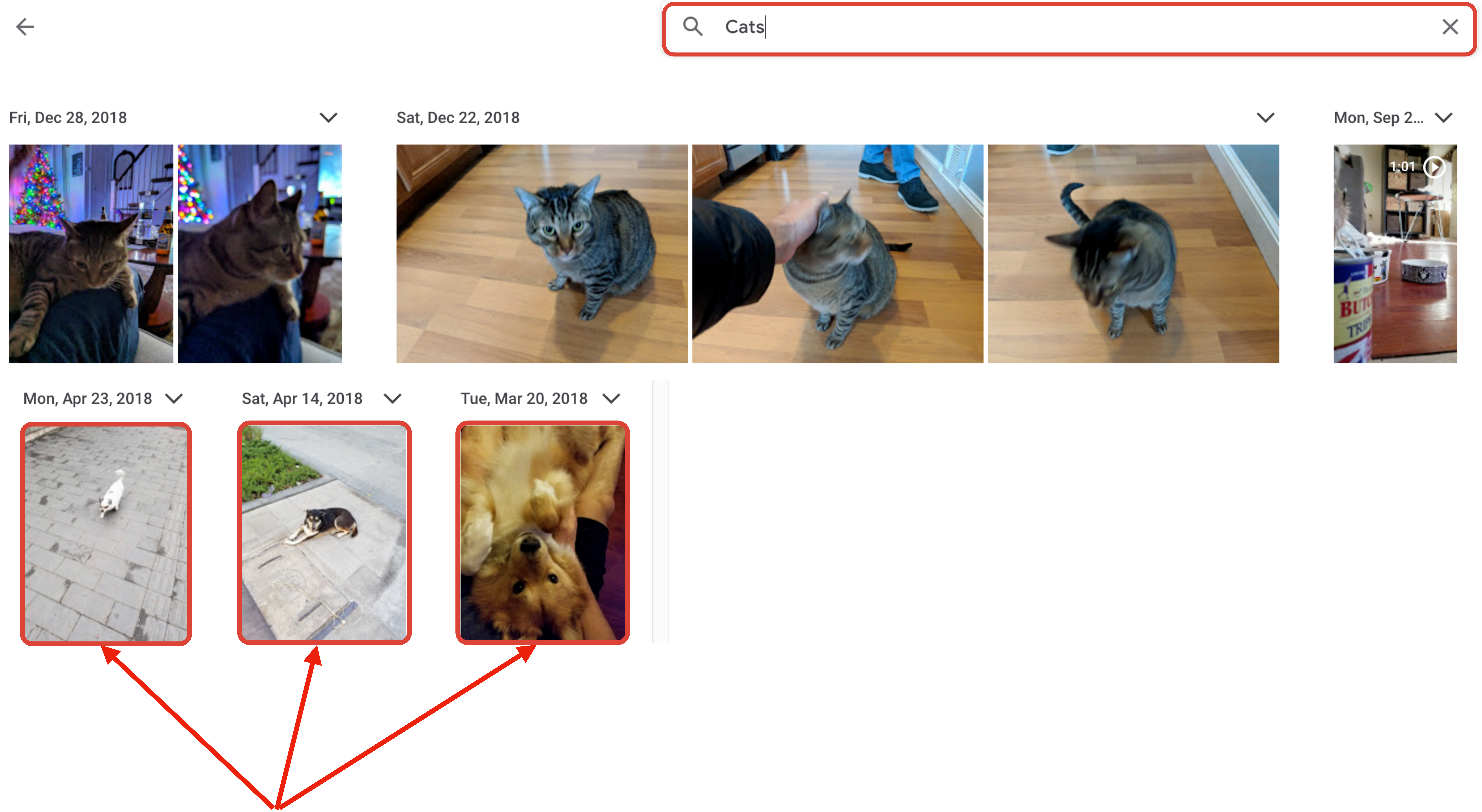
- Predictions are monitored, and users are able to improve them

- **Example**: Google Photos



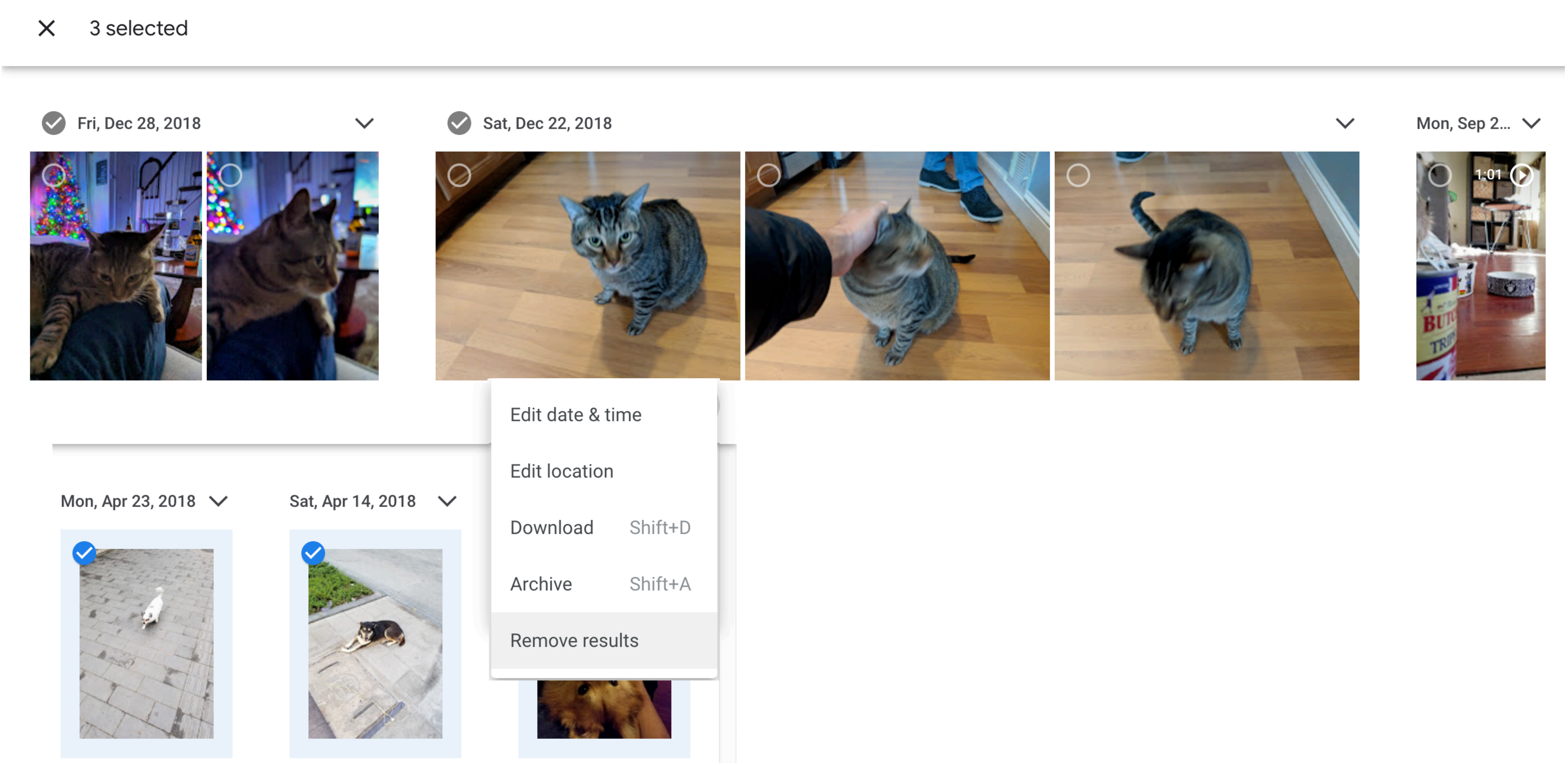
turnitin

Slide courtesy of Michael Yen



Dogs

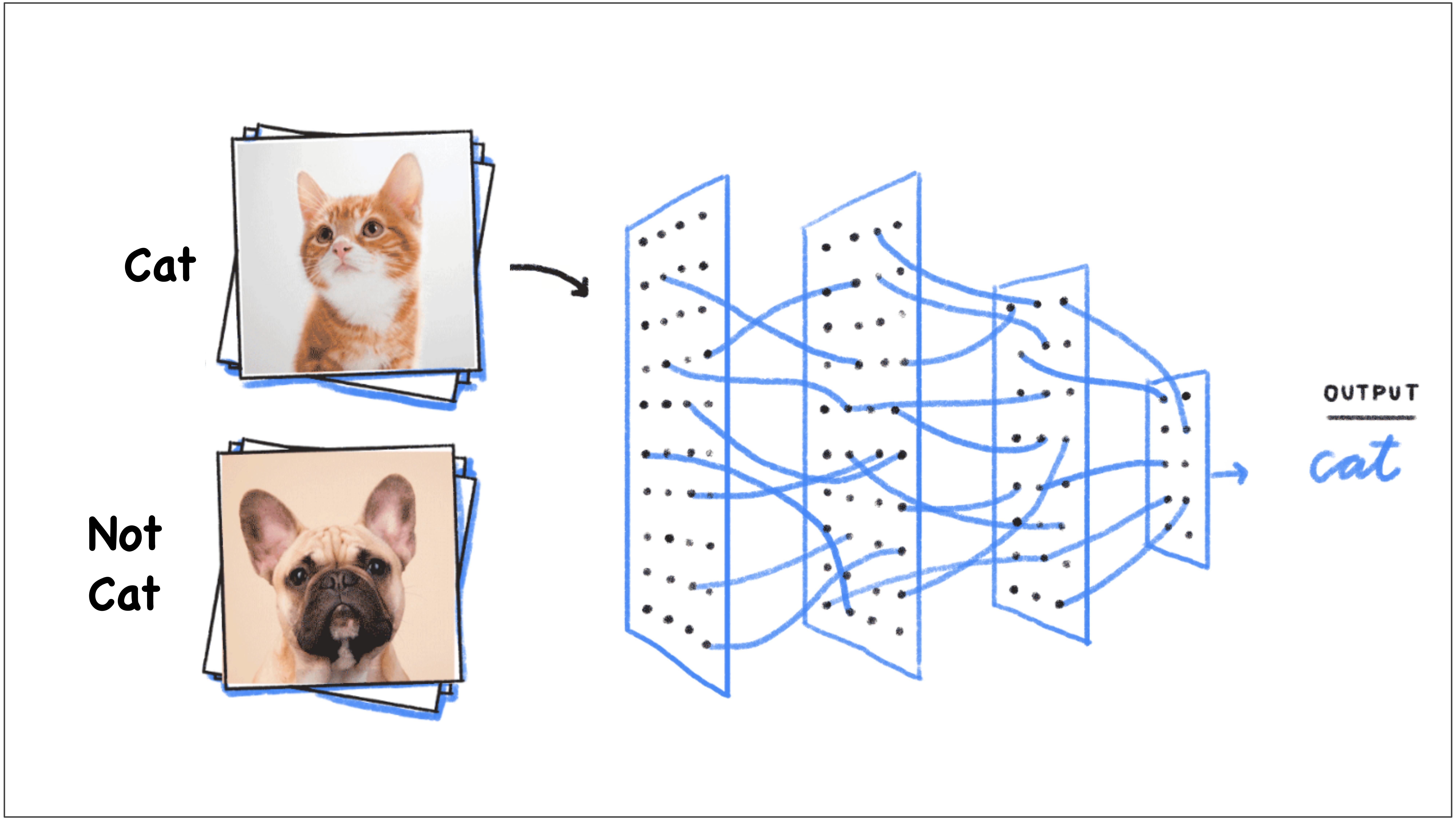
Slide courtesy of Michael Yen



Slide courtesy of Michael Yen

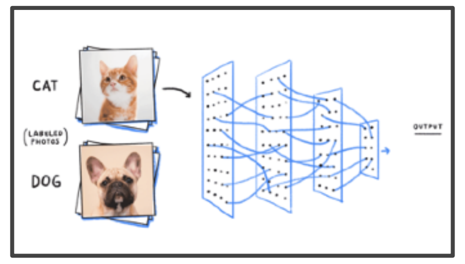
Slide courtesy of Michael Yen

The ML model



https://becominghuman.ai - [Venkatesh Tata](#)

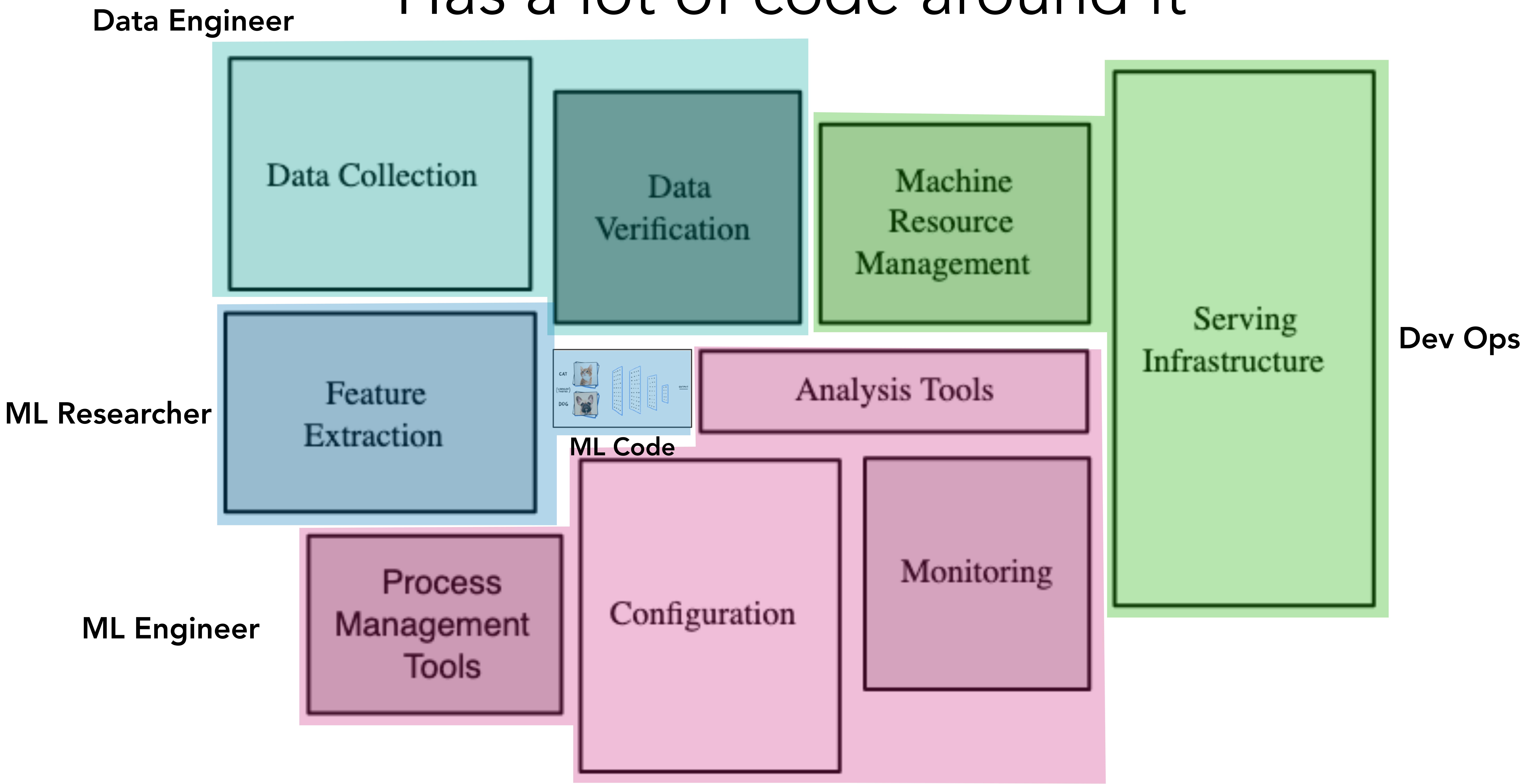
The ML model



ML Code

Slide courtesy of Michael Yen

Has a lot of code around it



https://developers.google.com/machine-learning/crash-course/production-ml-systems

Slide courtesy of Michael Yen