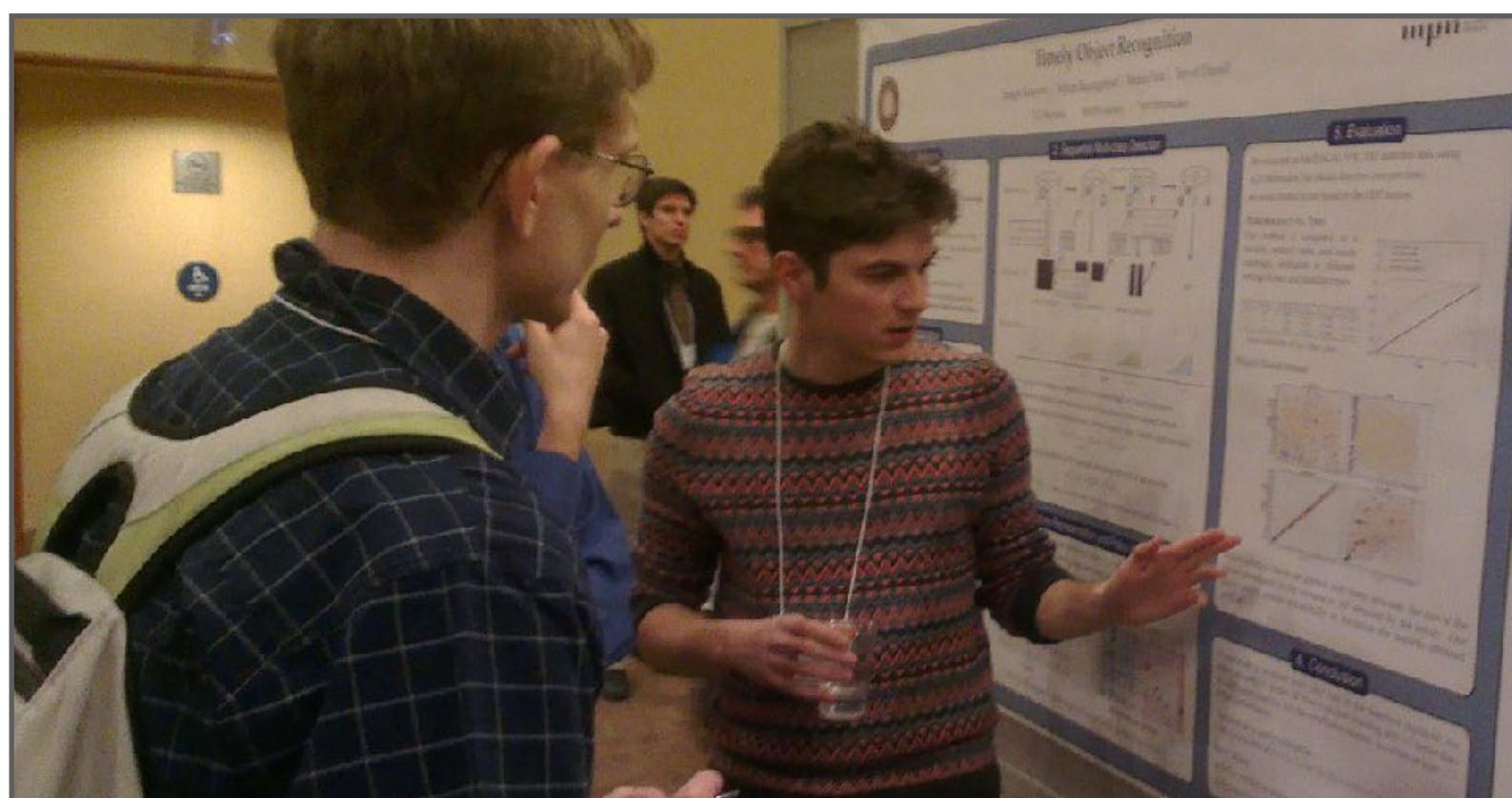


A Grading Rubric for AI Applications

Sergey Karayev
Head of AI for STEM
Turnitin, Inc.

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**



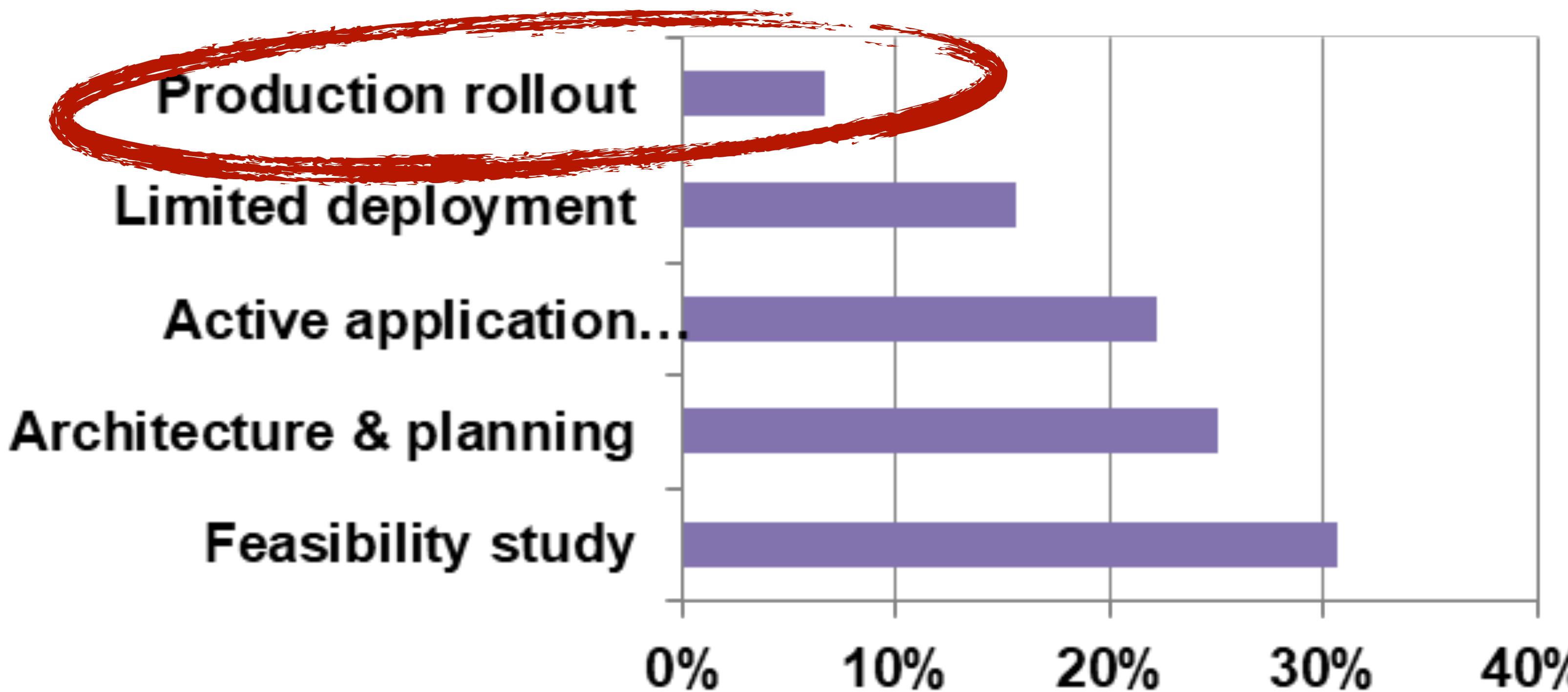
My Goal

- 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

What phase are your AI projects in today?



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

A Grading Rubric for AI Applications

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

A Grading Rubric for AI Applications

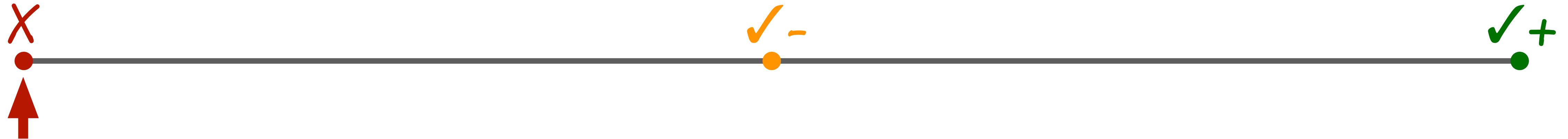
Category	X	✓ -	✓ +
1. Task Formulation			
2. User Interface / Performance Requirements			
3. Technical Difficulty			
4. Initial Data Moat			
5. Data Flywheel			

Task Formulation



- Is the problem well defined, with clear metrics?

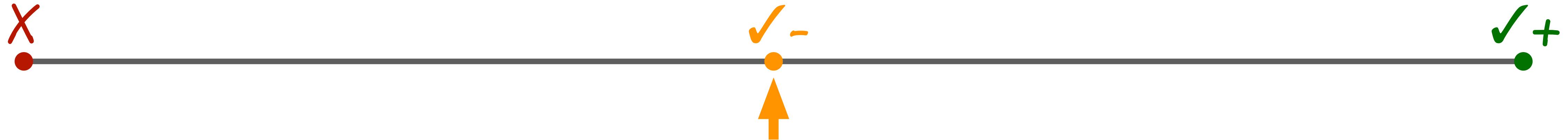
Task Formulation



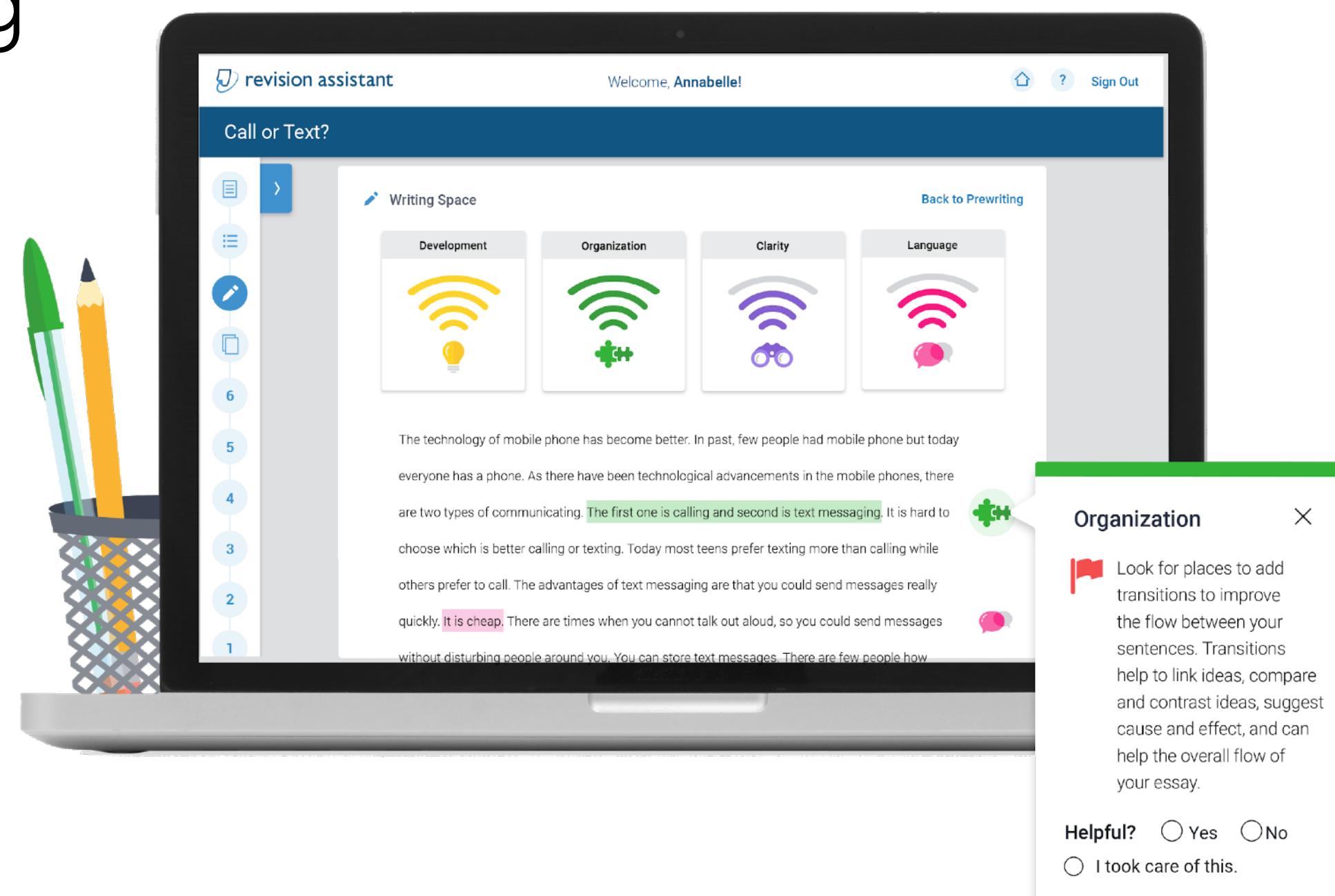
- No, AI is a magical solution to an unsolved, unmeasurable problem.
- **Example:** AI for achieving happiness

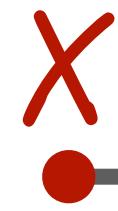


Task Formulation



- 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 important
- Example: Call or Text?



revision assistant

Welcome, Annabelle!

Call or Text?

Writing Space

Development Organization Clarity Language

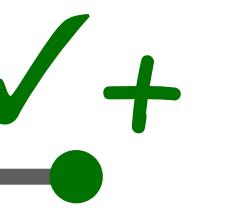
Back to Prewriting

The technology of mobile phone has become better. In past, few people had mobile phone but today everyone has a phone. As there have been technological advancements in the mobile phones, there are two types of communicating. The first one is calling and second is text messaging. It is hard to choose which is better calling or texting. Today most teens prefer texting more than calling while others prefer to call. The advantages of text messaging are that you could send messages really quickly. It is cheap. There are times when you cannot talk out aloud, so you could send messages without disturbing people around you. You can store text messages. There are few people how

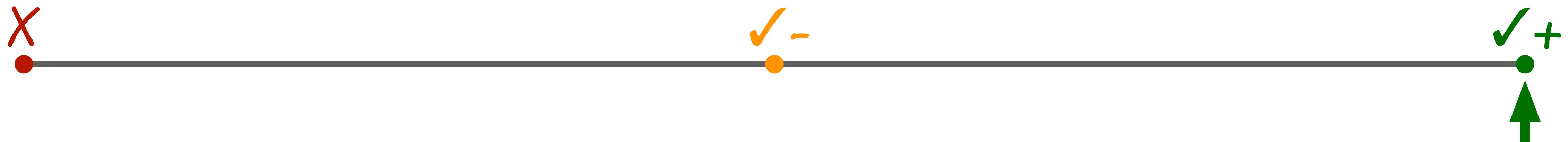
Organization

Look for places to add transitions to improve the flow between your sentences. Transitions help to link ideas, compare and contrast ideas, suggest cause and effect, and can help the overall flow of your essay.

Helpful? Yes No
 I took care of this.



Task Formulation



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



Gary Chavez added a photo you might be in.
about a minute ago ·



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

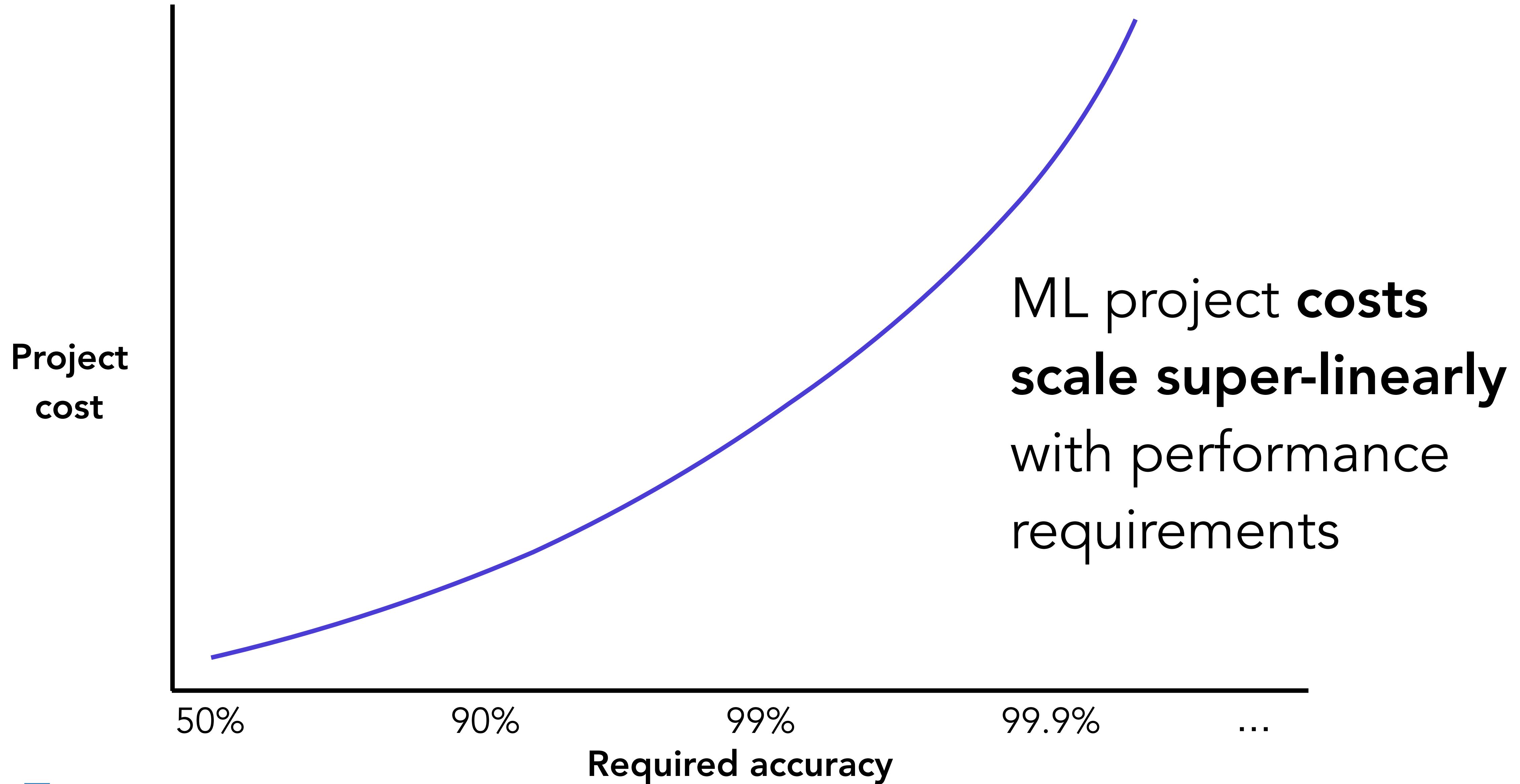
A Grading Rubric for AI Applications

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

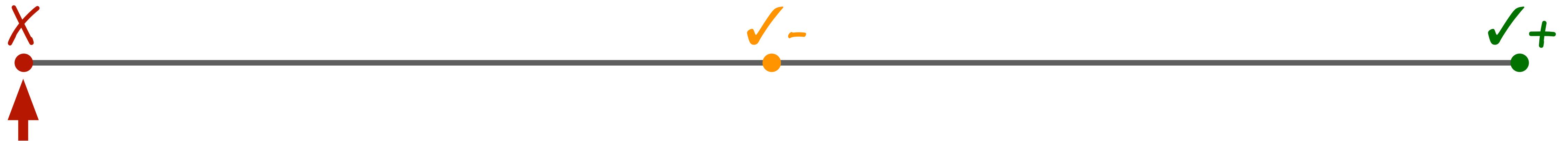
User Interface and Performance Requirements

- 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



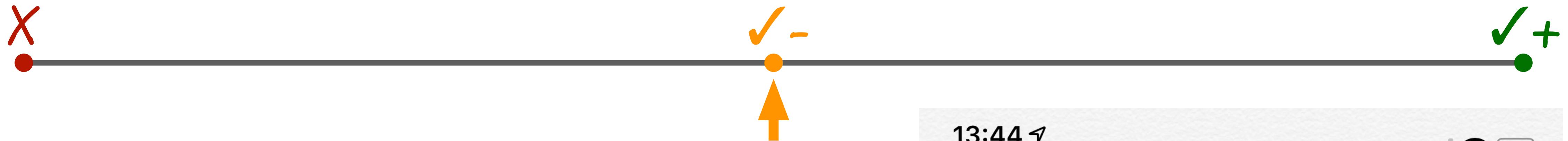
User Interface and Performance Requirements



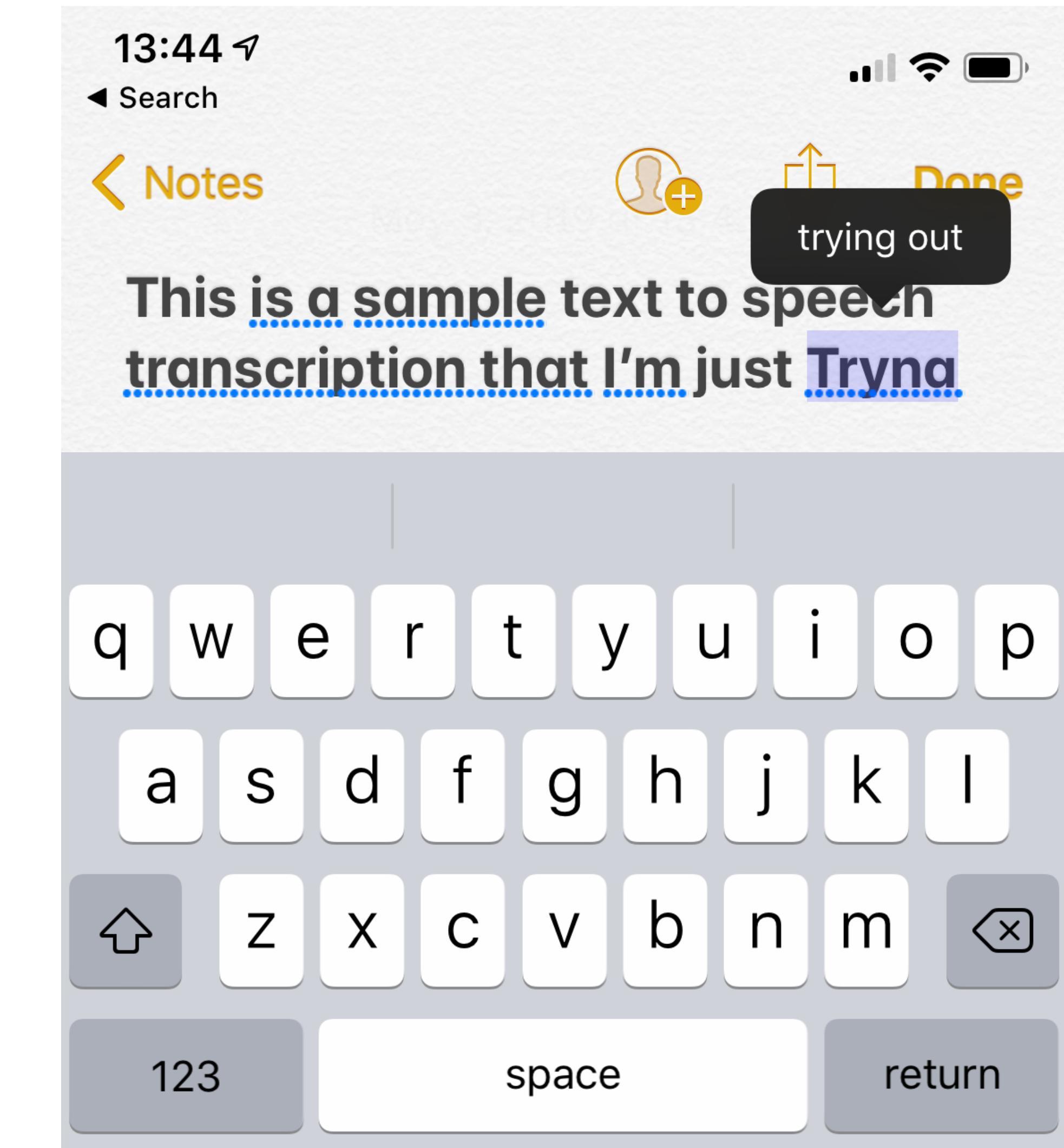
- AI must be 100% autonomous, predicting all the time, and mistakes are fatal.
- **Example:** self-driving cars



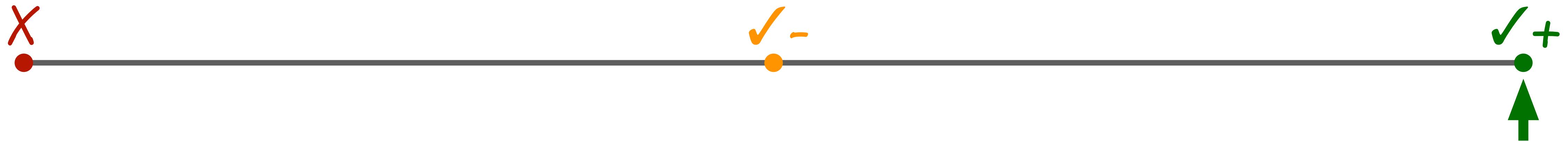
User Interface and Performance Requirements



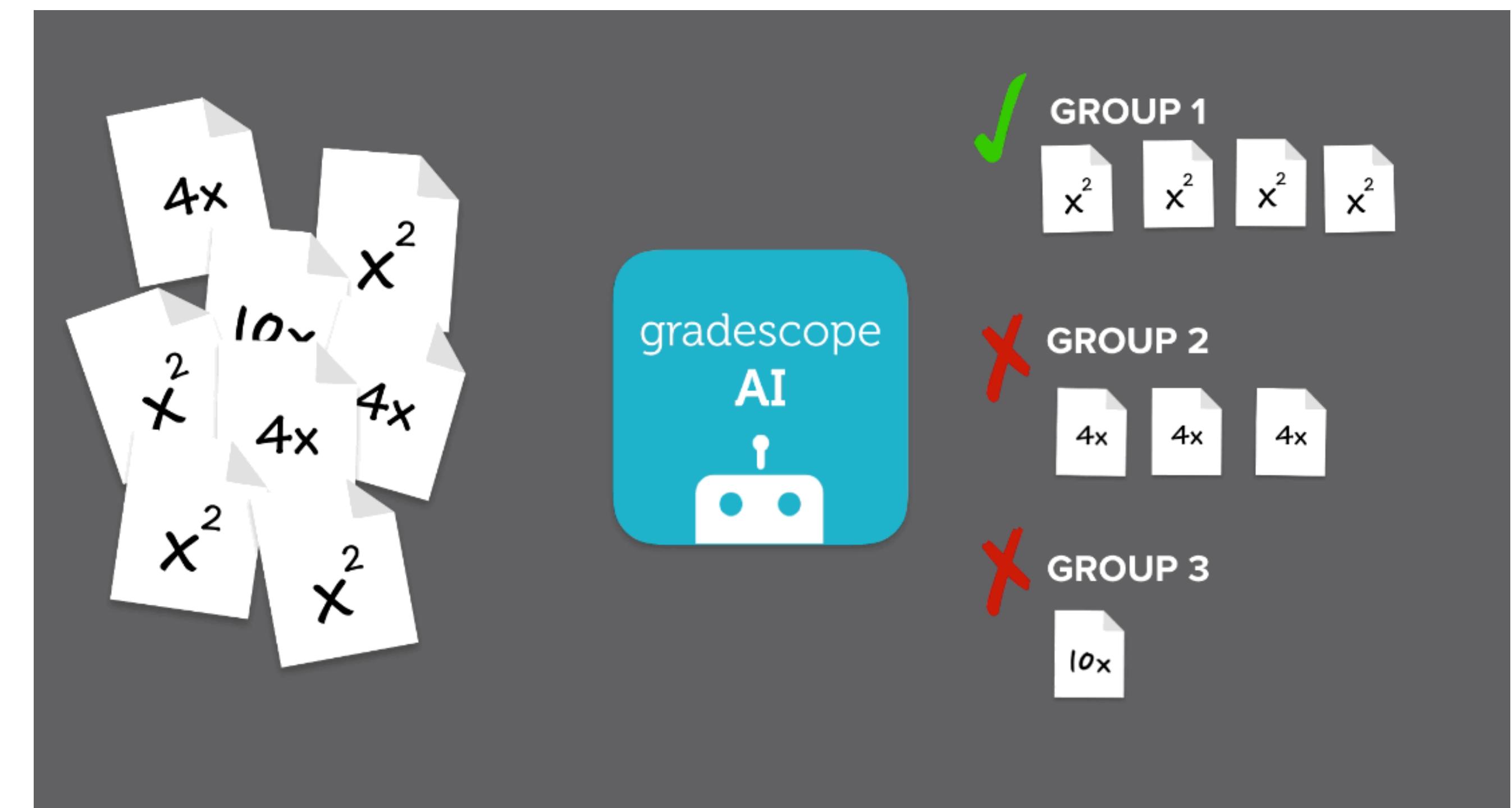
- AI always has to predict, and user resents correcting mistakes.
- **Example:** speech-to-text transcription



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

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

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

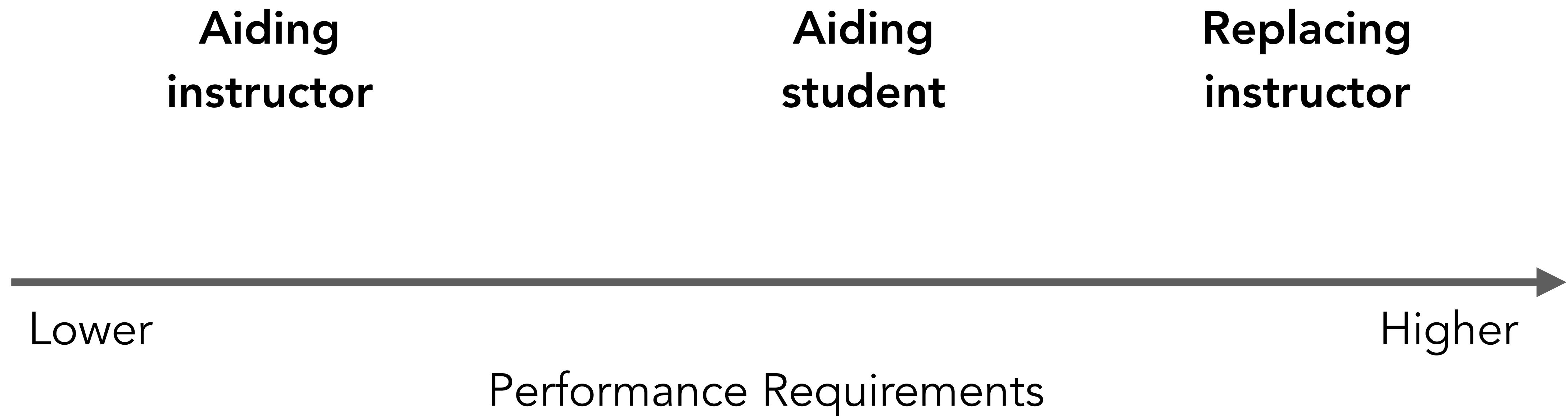
$\frac{1}{2}x^2 + C$

x^2

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

x^2

User Interface and Performance Requirements for Education Applications



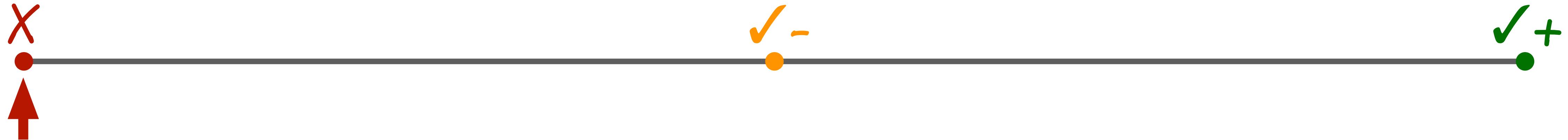
A Grading Rubric for AI Applications

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

Technical Difficulty

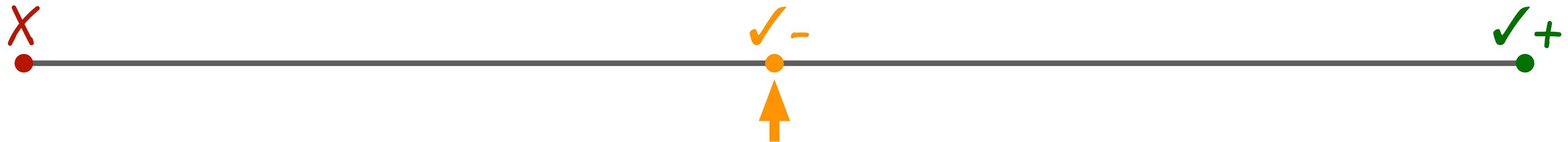
- What are the technical resources needed to solve the problem?
 - (From analogs, academic literature, and expert opinions).

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

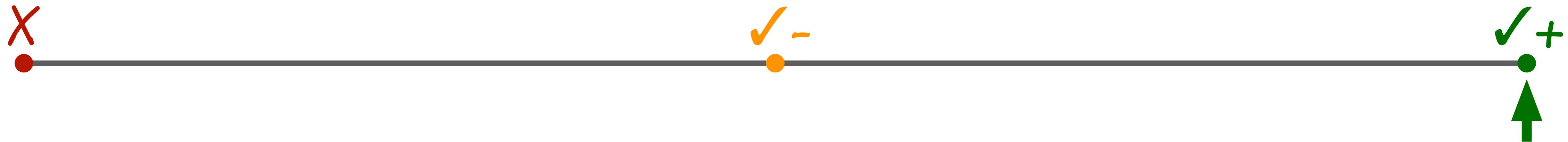
Technical Difficulty



- 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.

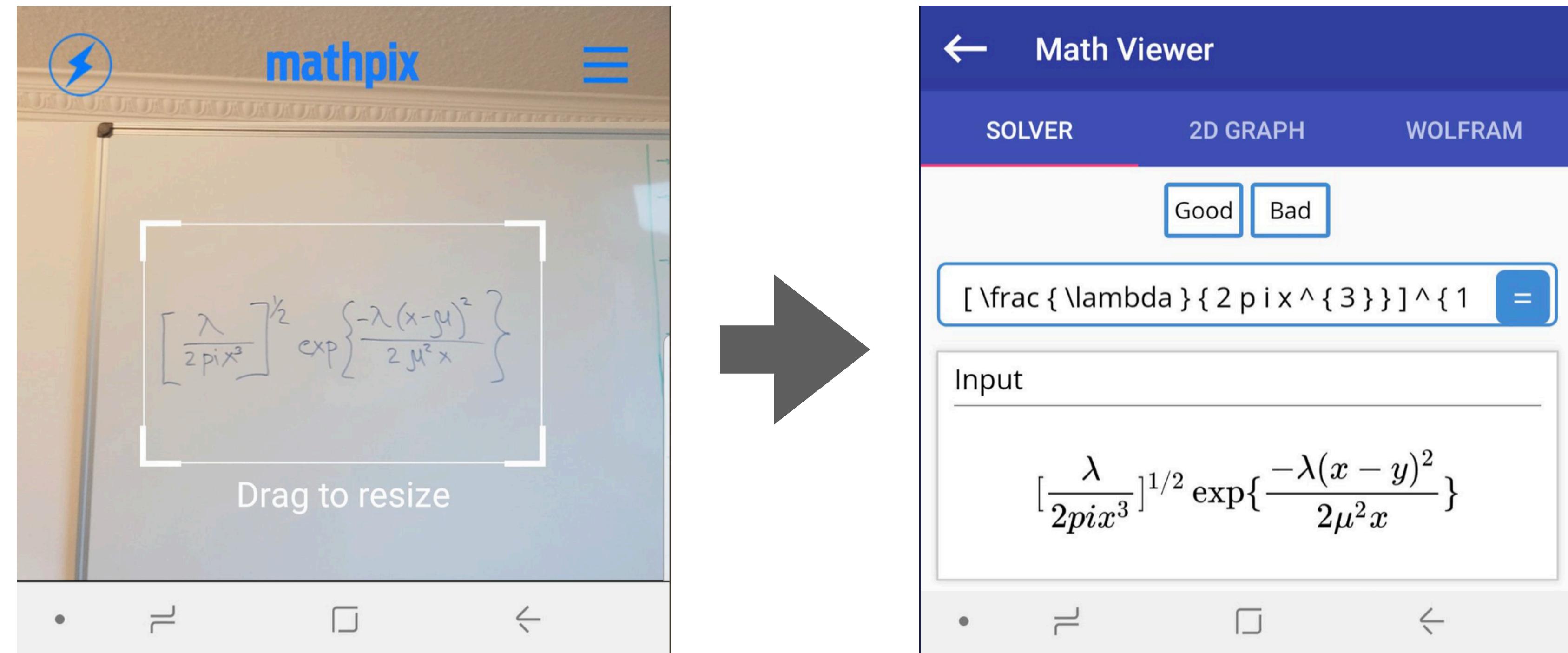


Technical Difficulty



- 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

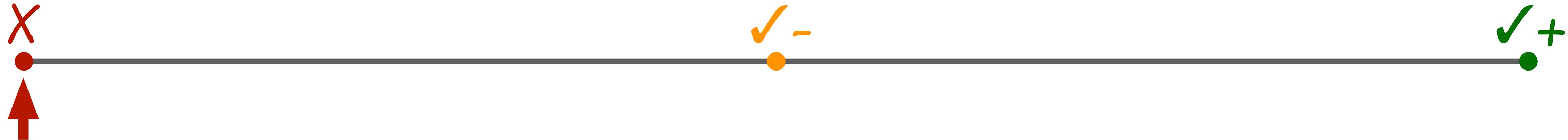
A Grading Rubric for AI Applications

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

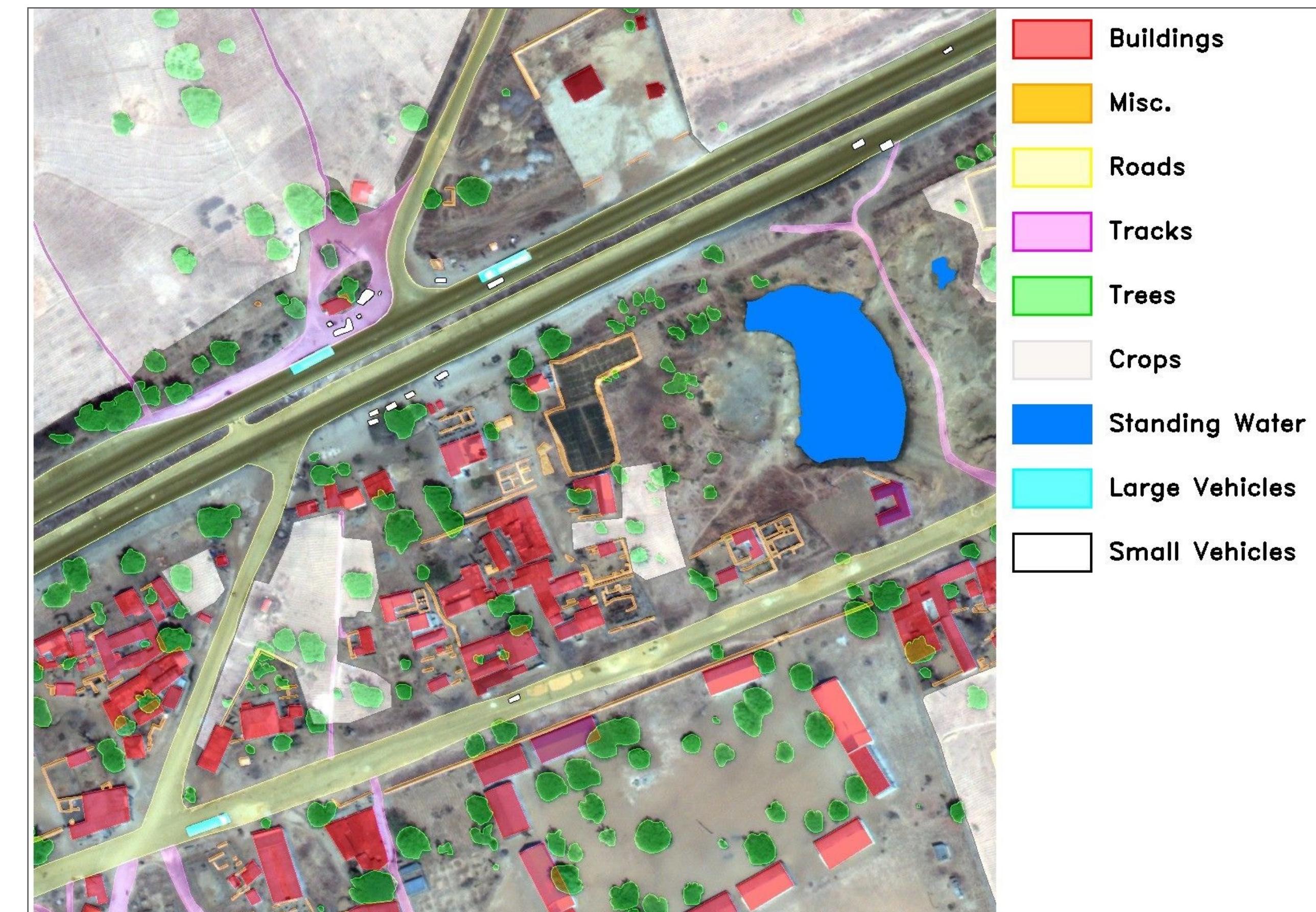
Initial Data Moat

- 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?

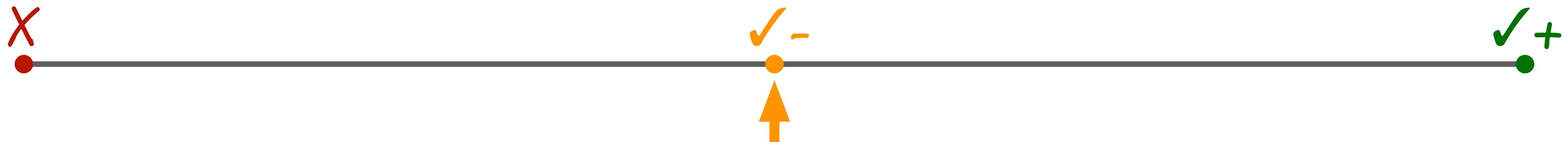
Initial Data Moat



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



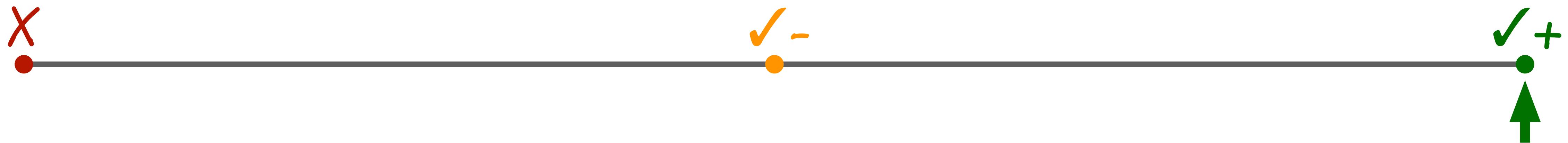
Initial Data Moat



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



Initial Data Moat



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

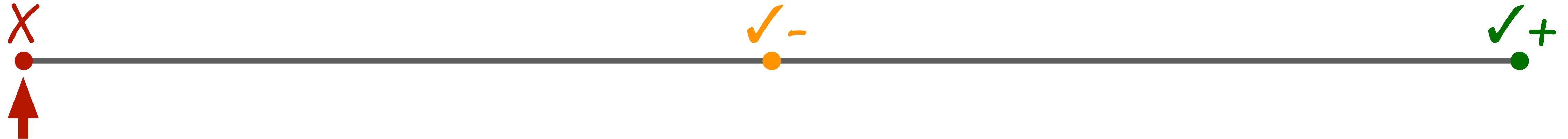
A Grading Rubric for AI Applications

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

Data Flywheel

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

Data Flywheel

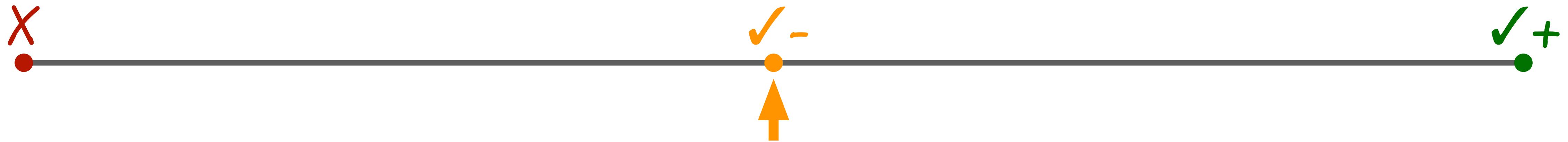


- Predictions are not monitored or corrected
- **Example:** most AI products

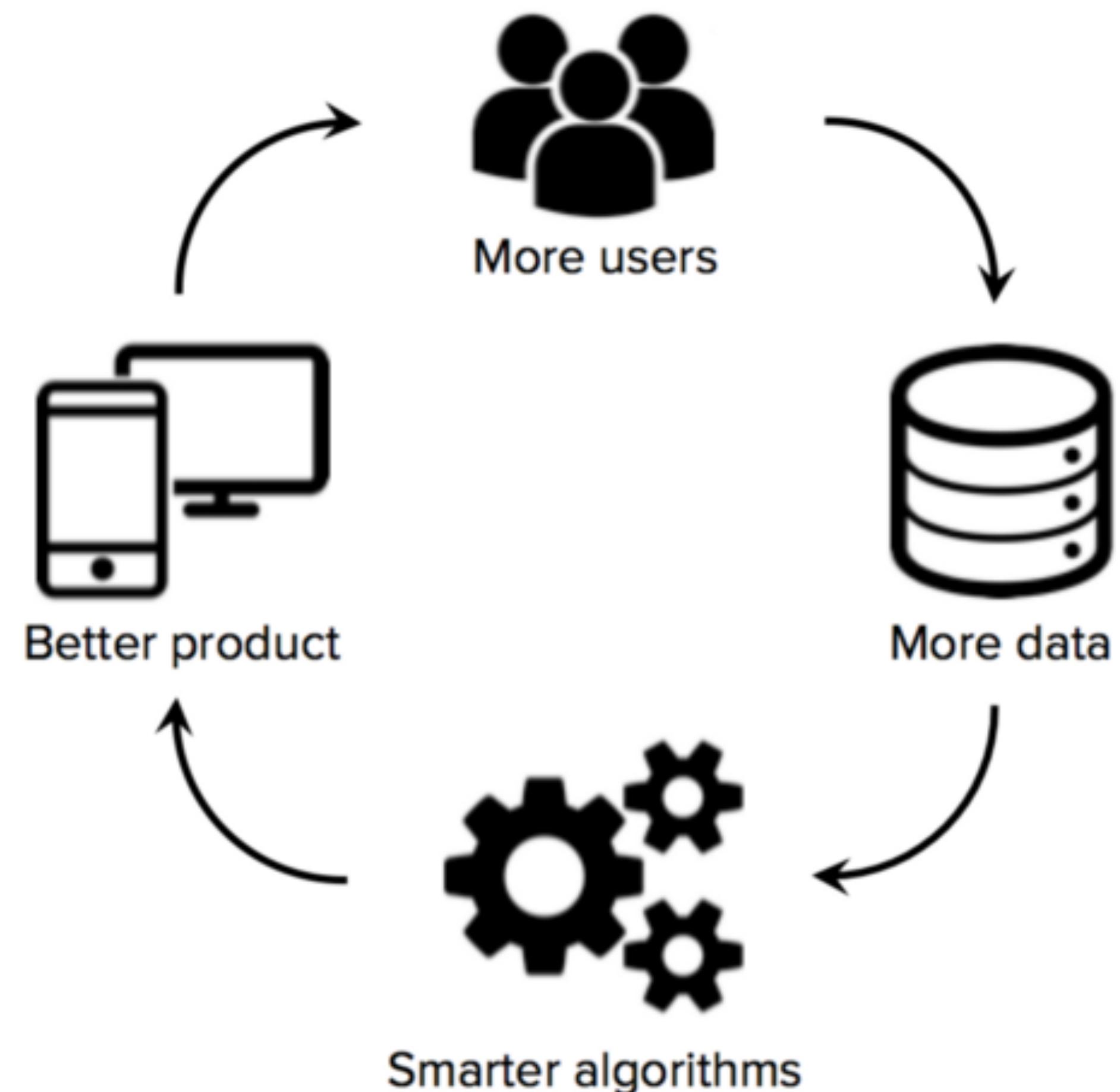


<https://rvpartners.com.au/avoid-flying-blind-what-you-can-measure-you-can-manage/>

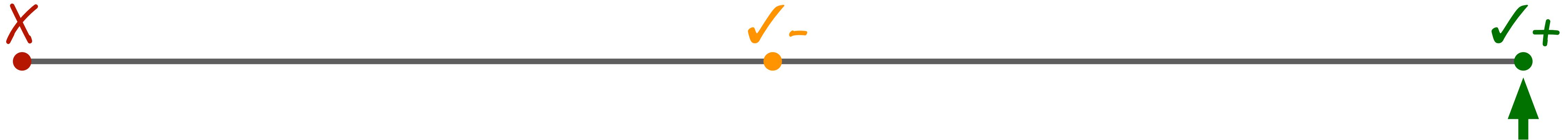
Data Flywheel



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

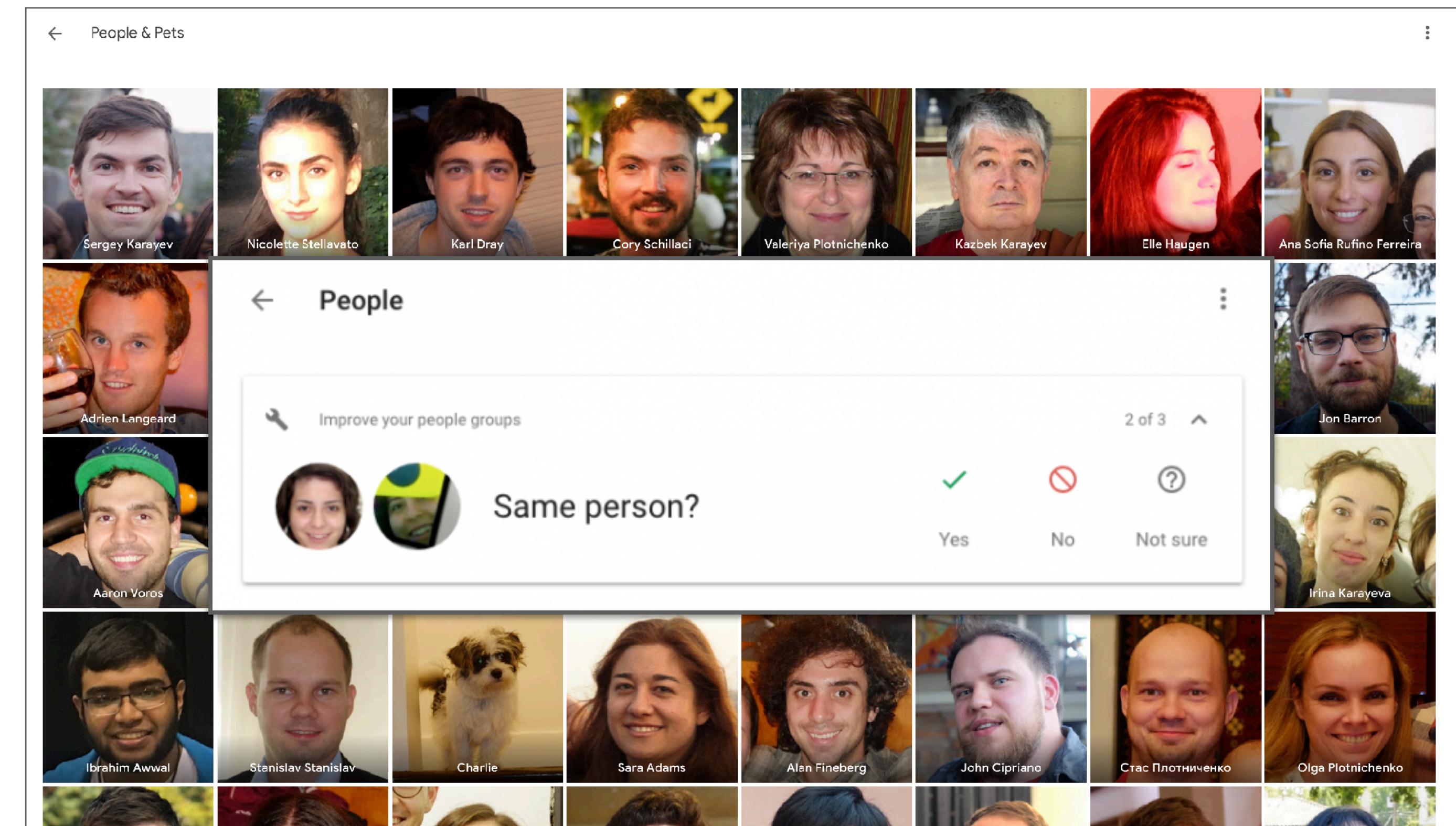


Data Flywheel



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

- **Example:** Google Photos

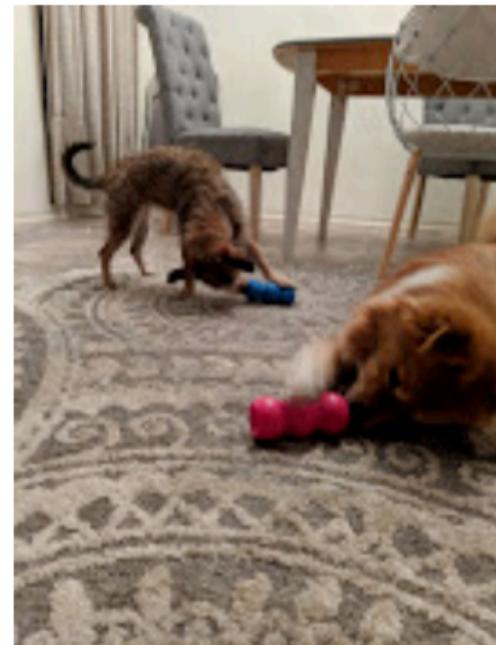




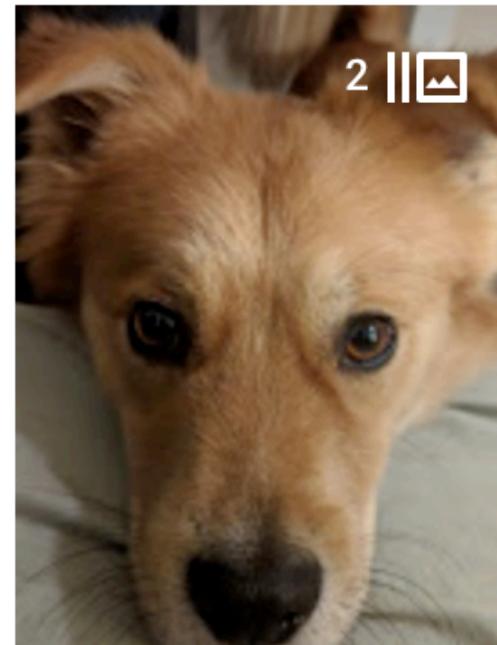
Dogs



Thu, Mar 28



Wed, Jan 9



Sun, Nov 25, 2018

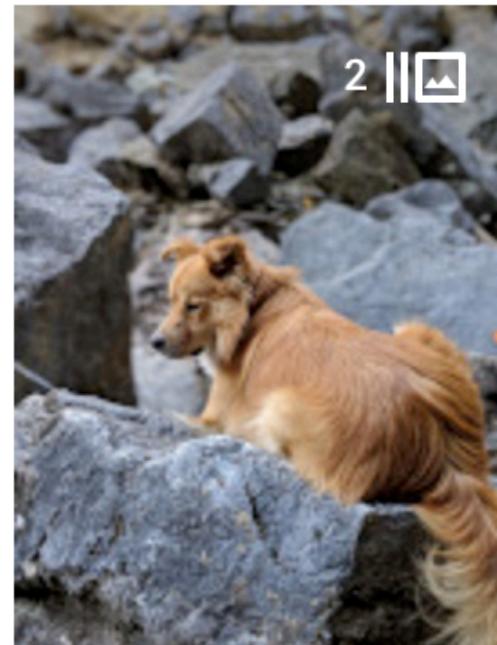


Wed, Nov 21, 2018

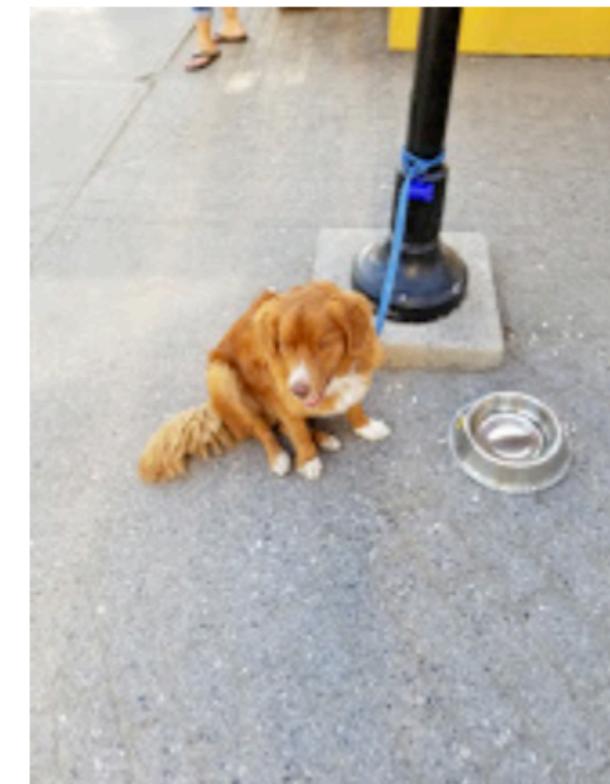


▼

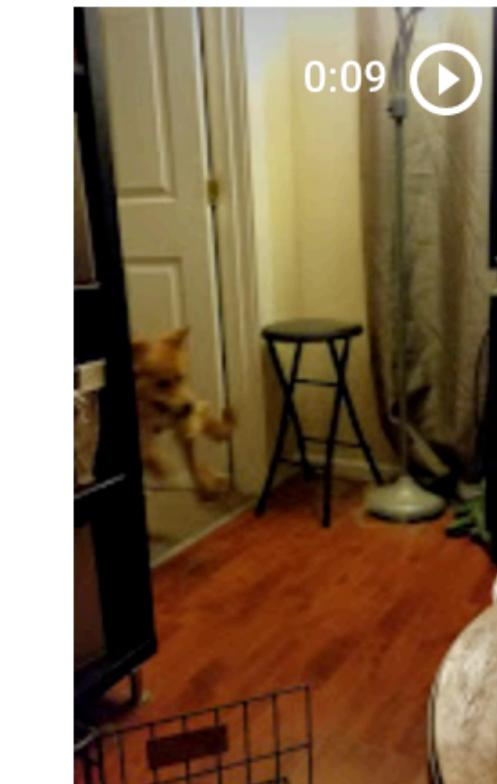
Sat, Oct 27, 2018



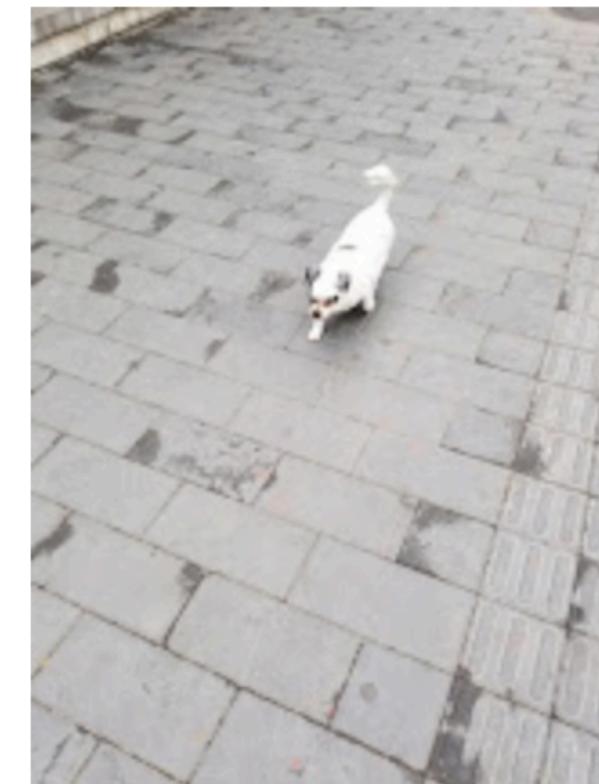
Sat, May 26, 2018



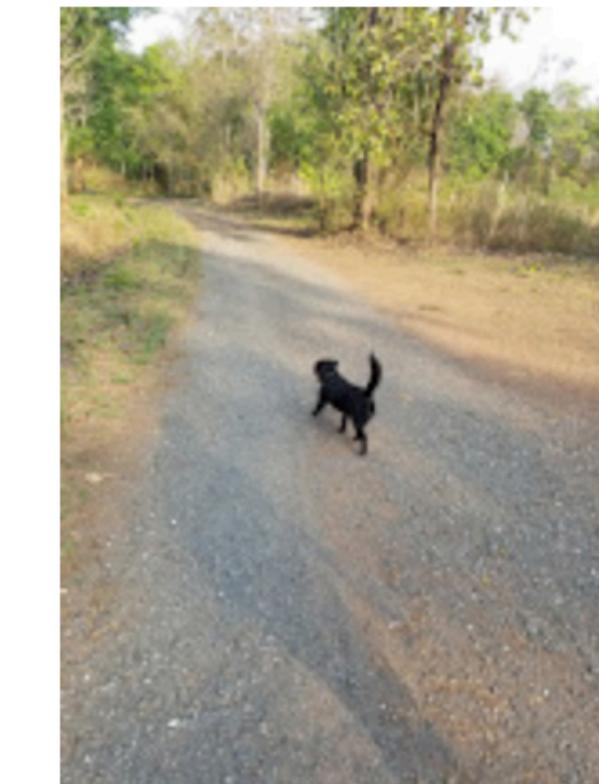
Wed, May 2, 2...



Mon, Apr 23, 2018



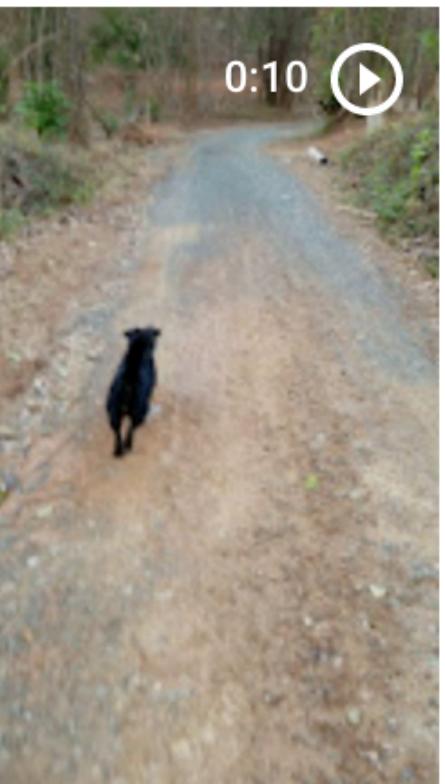
Thu, Apr 19, 2018



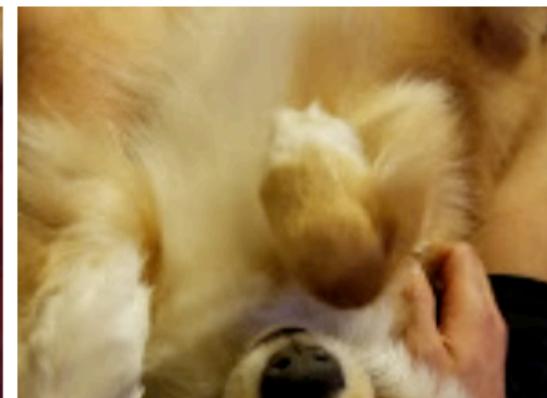
Wed, Apr 18, ...



Tue, Apr 17,



Tue, Mar 20, 2018



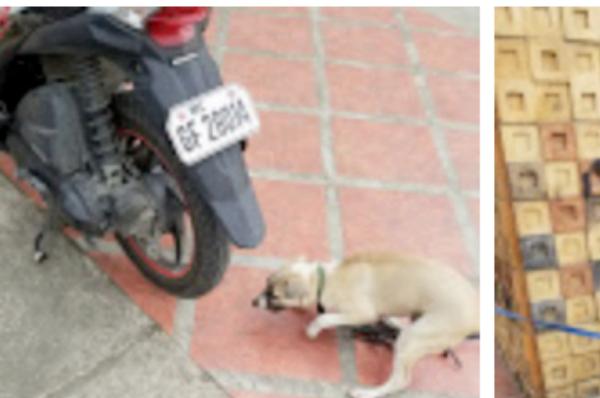
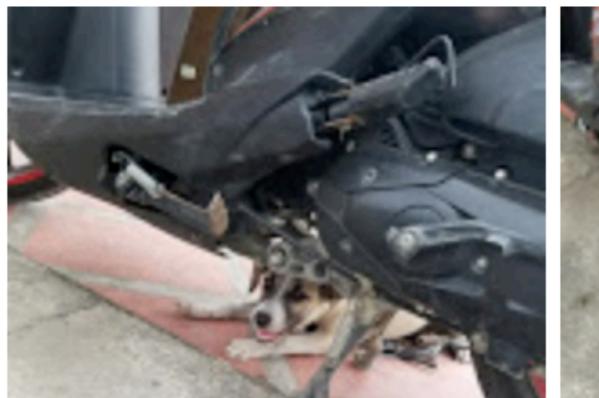
Wed, Feb 14, 2018



Tue, Feb 13, 2018



Mon, Feb 12, 2018

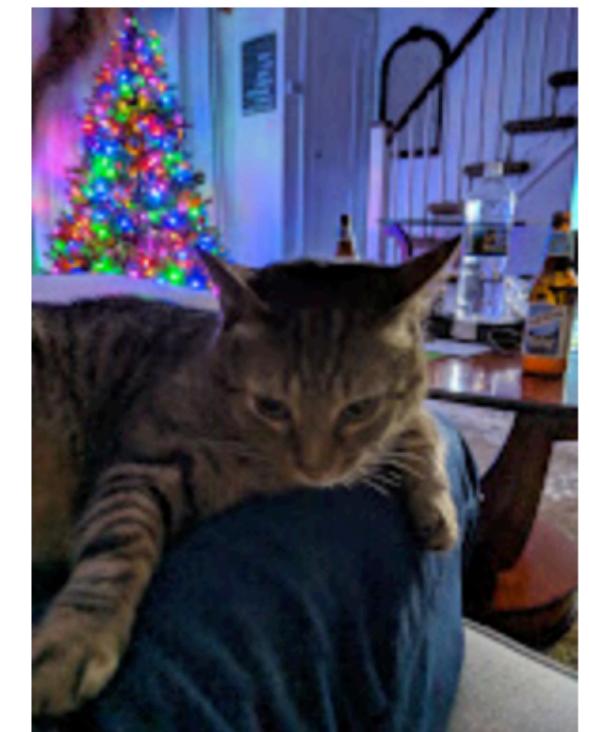




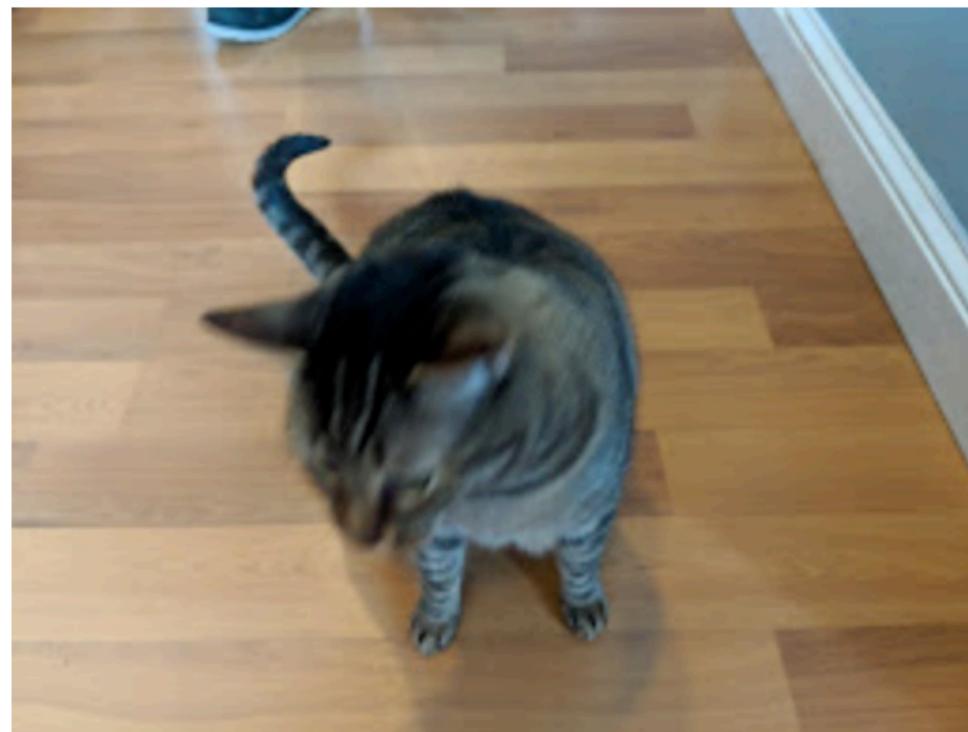
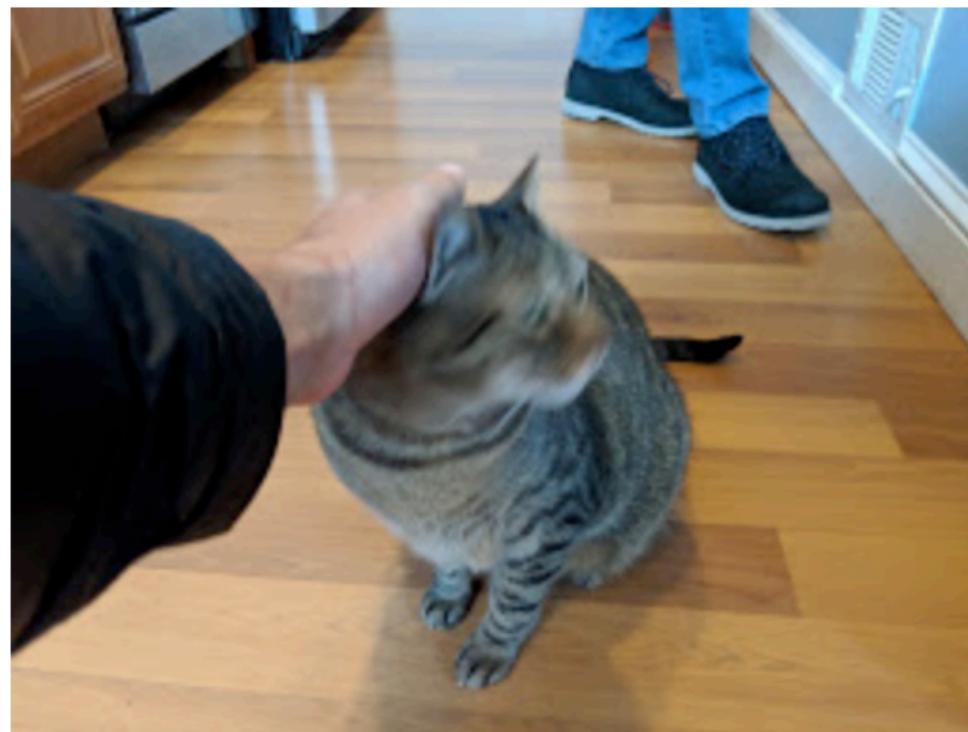
Cats



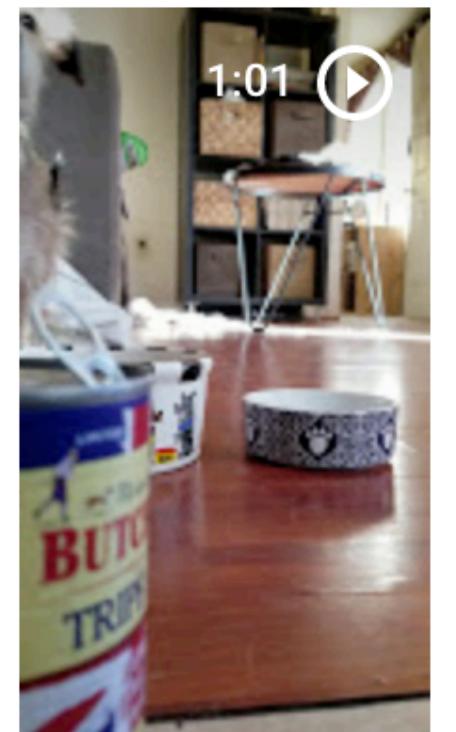
Fri, Dec 28, 2018



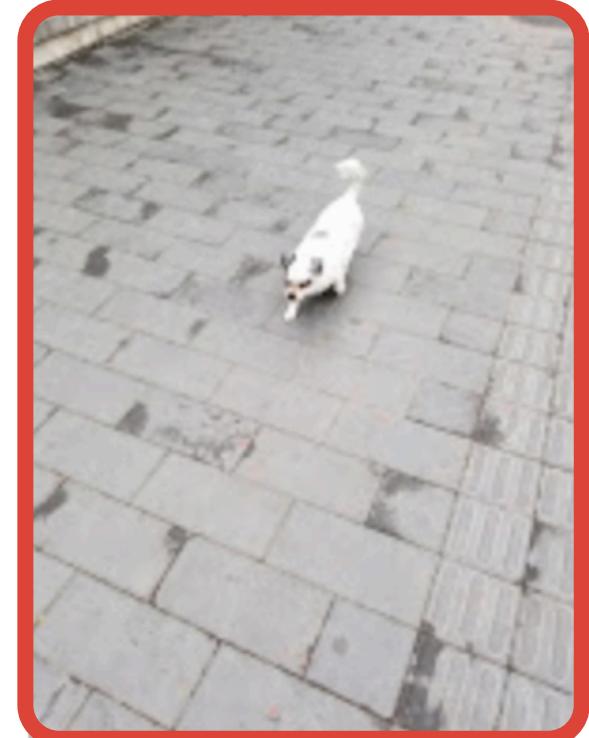
Sat, Dec 22, 2018



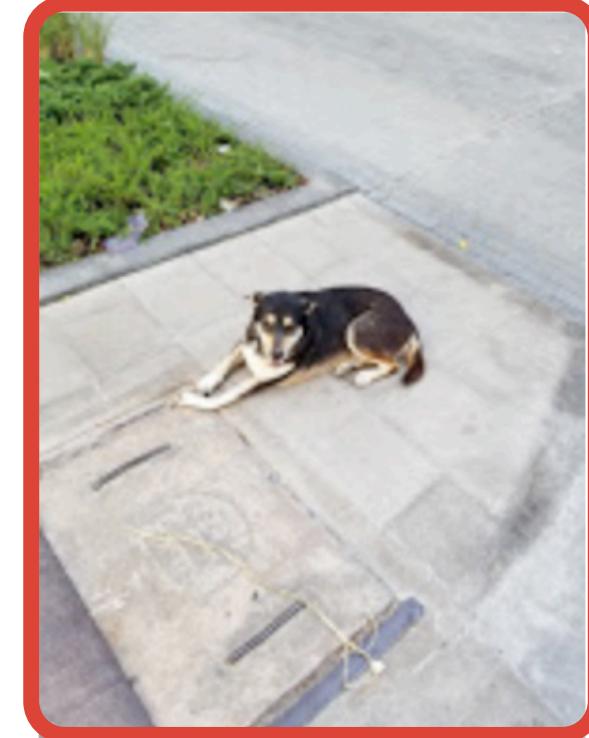
Mon, Sep 2...



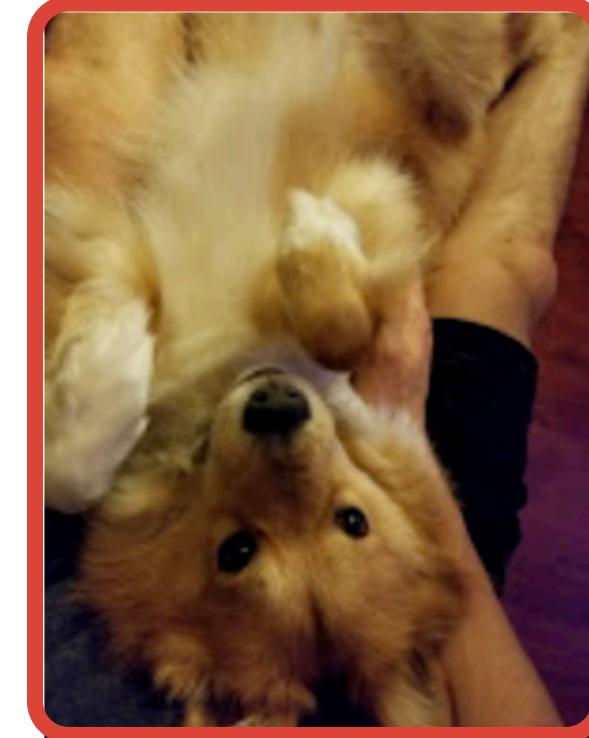
Mon, Apr 23, 2018



Sat, Apr 14, 2018

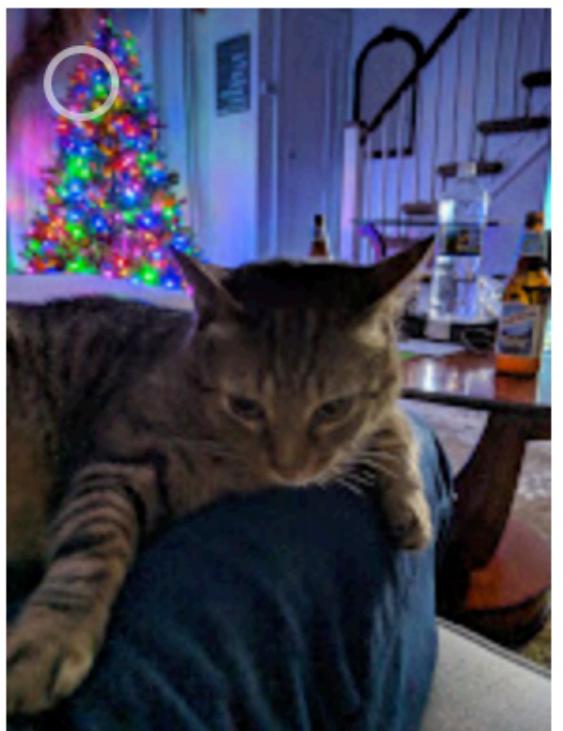


Tue, Mar 20, 2018

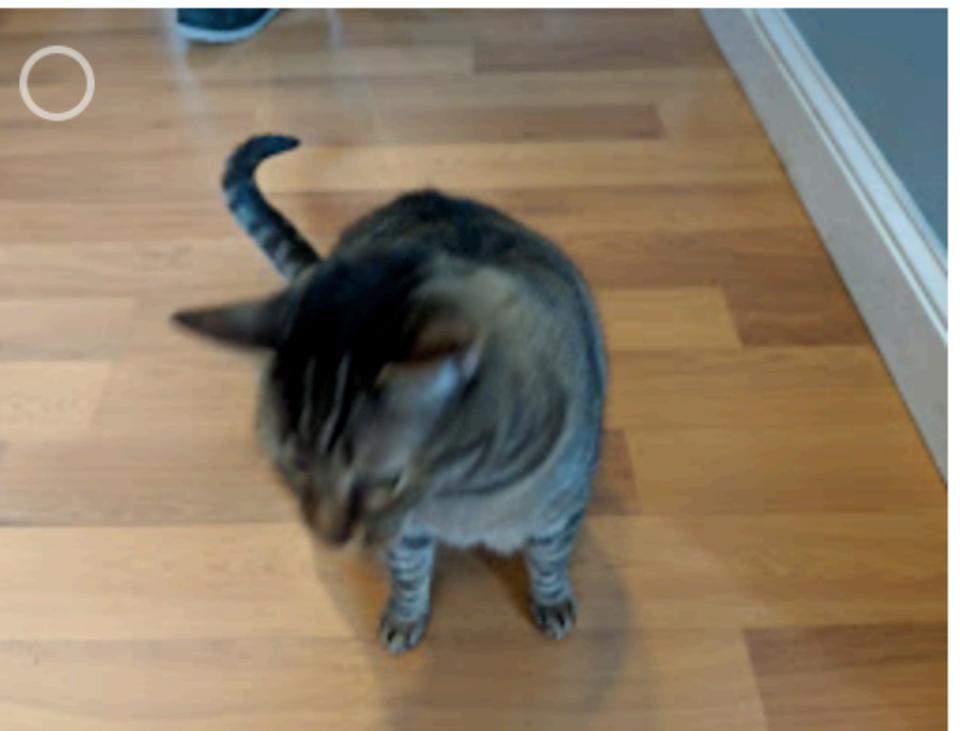
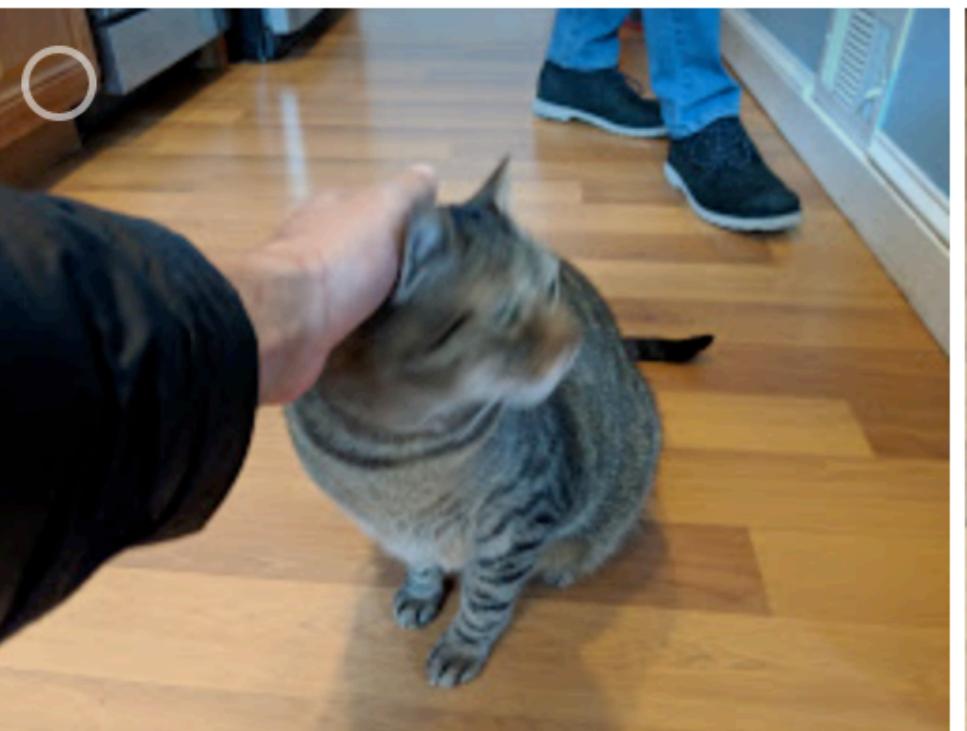
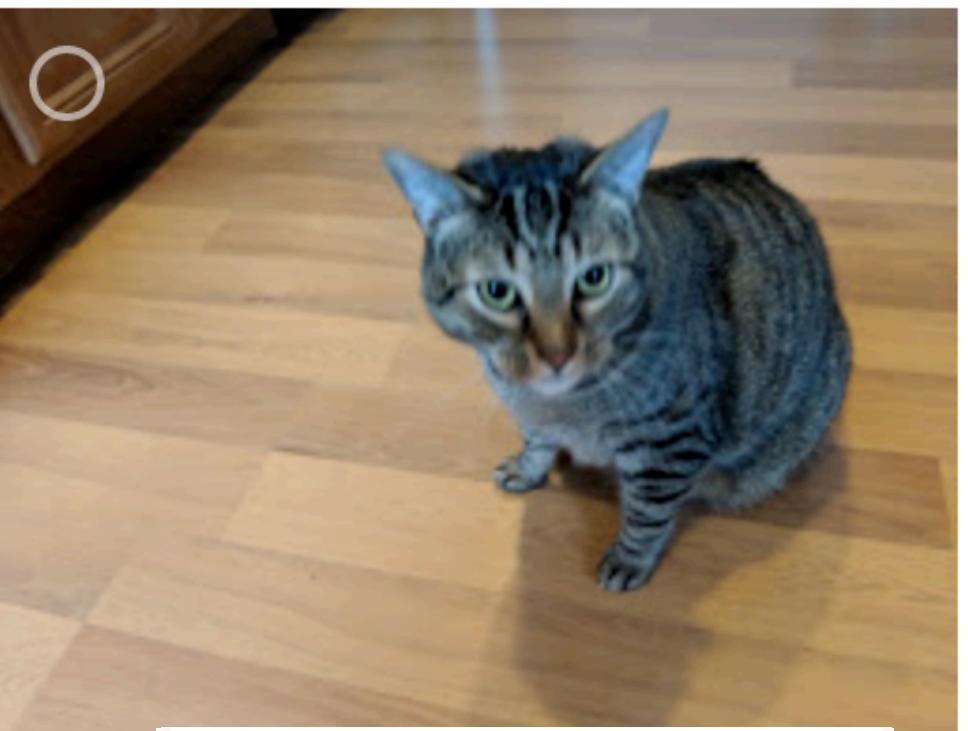


X 3 selected

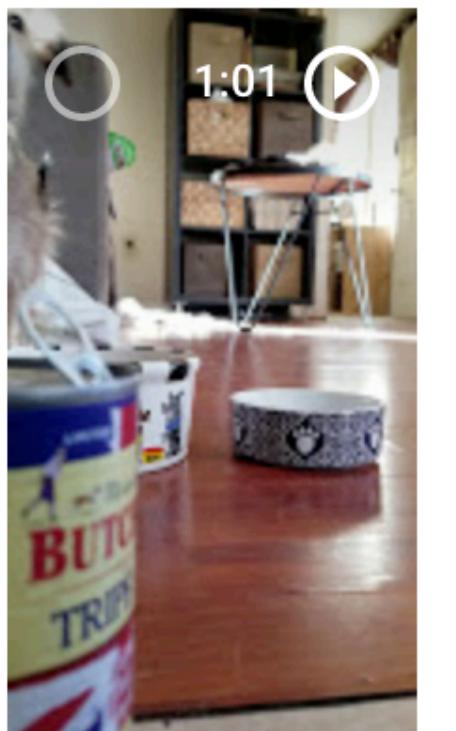
✓ Fri, Dec 28, 2018 ▾



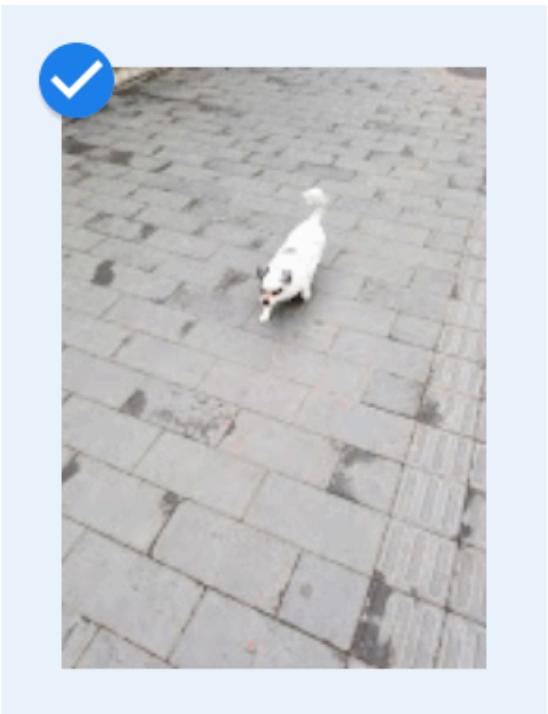
✓ Sat, Dec 22, 2018 ▾



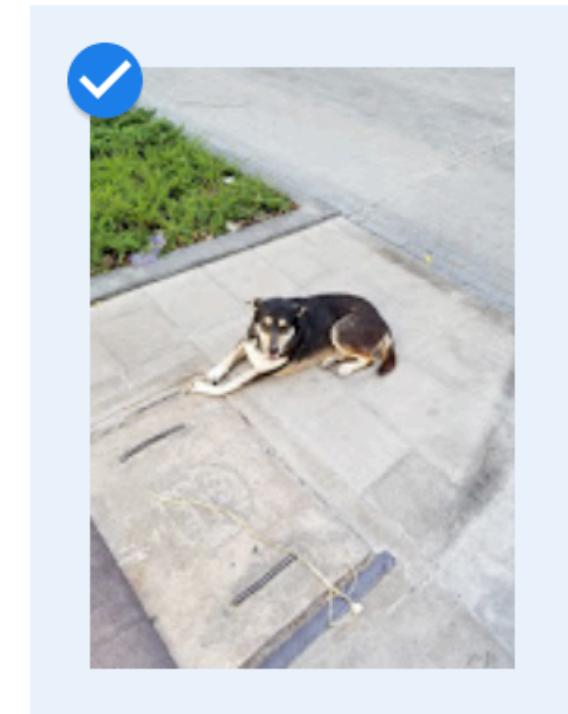
Mon, Sep 2... ▾



Mon, Apr 23, 2018 ▾



Sat, Apr 14, 2018 ▾



Edit date & time

Edit location

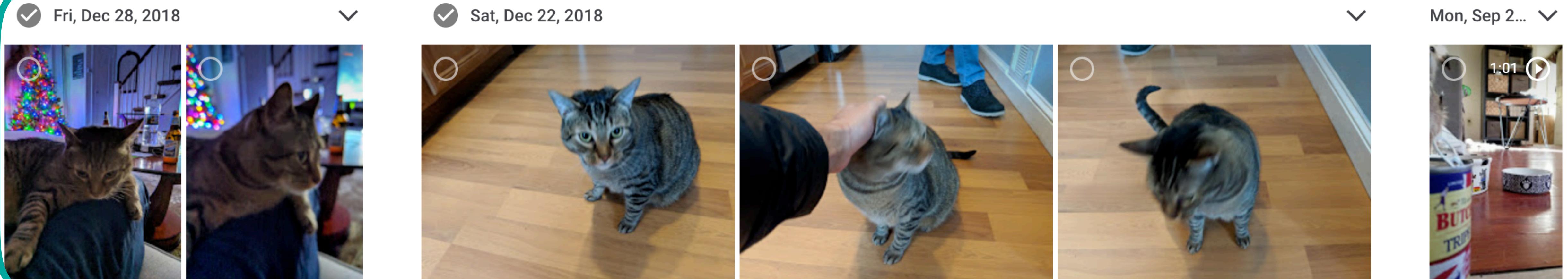
Download Shift+D

Archive Shift+A

Remove results



X 3 selected



Mon, Apr 23, 2018

Sat, Apr 14, 2018

Edit date & time

Edit location

Download Shift+D

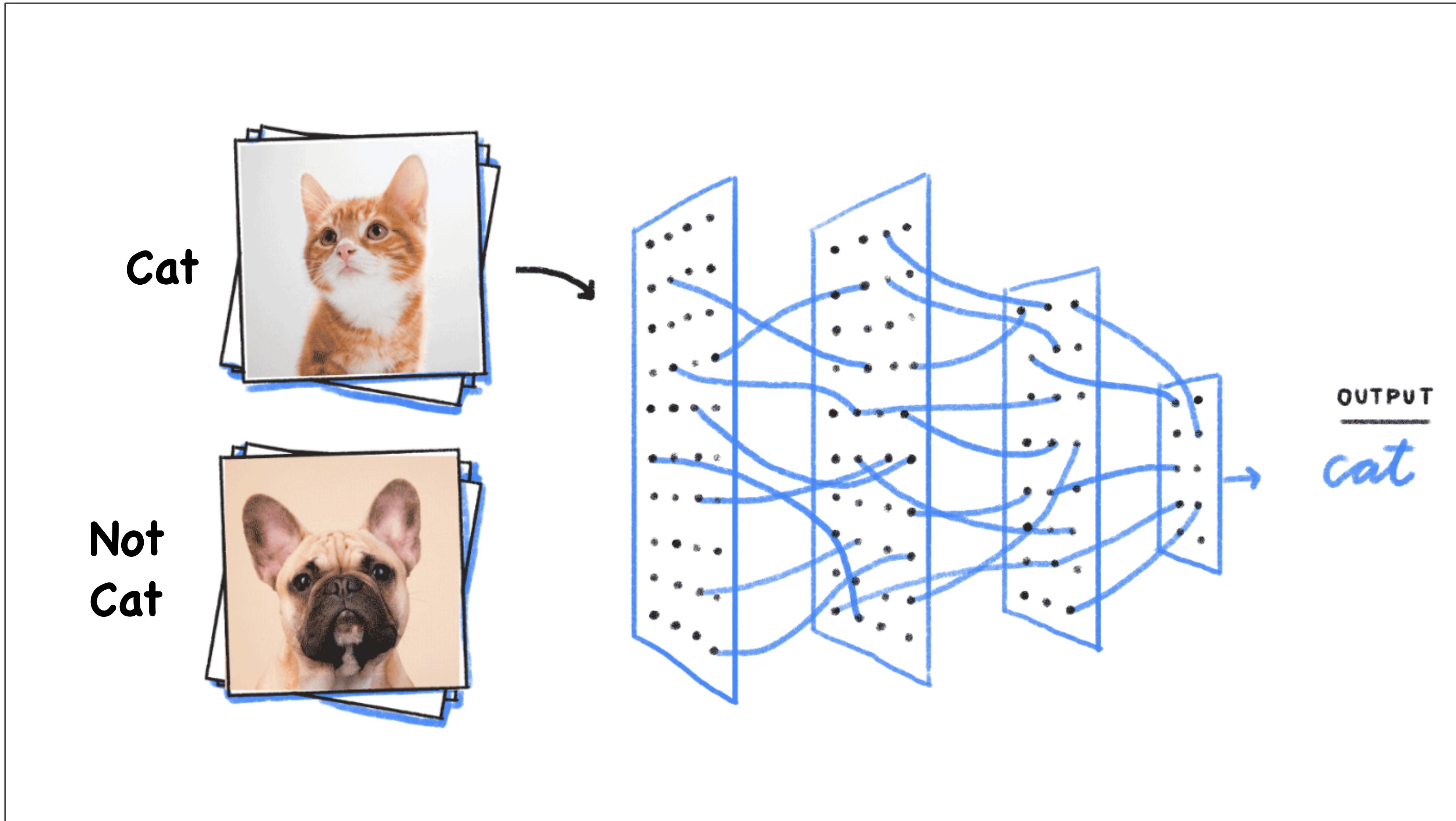
Archive Shift+A

Remove results

Not Cat

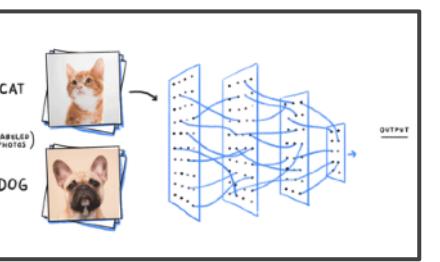
Cat

The ML model



<https://becominghuman.ai> - Venkatesh Tata

The ML model



ML Code

Has a lot of code around it

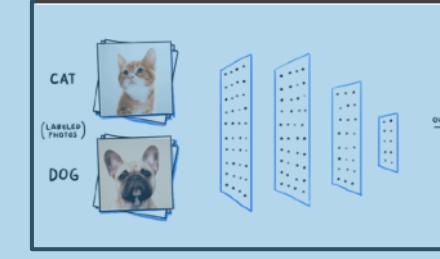
Data Engineer

Data Collection

Data
Verification

Machine
Resource
Management

Feature
Extraction



ML Code

Analysis Tools

Monitoring

ML Engineer

Process
Management
Tools

Configuration

Serving
Infrastructure

Dev Ops