

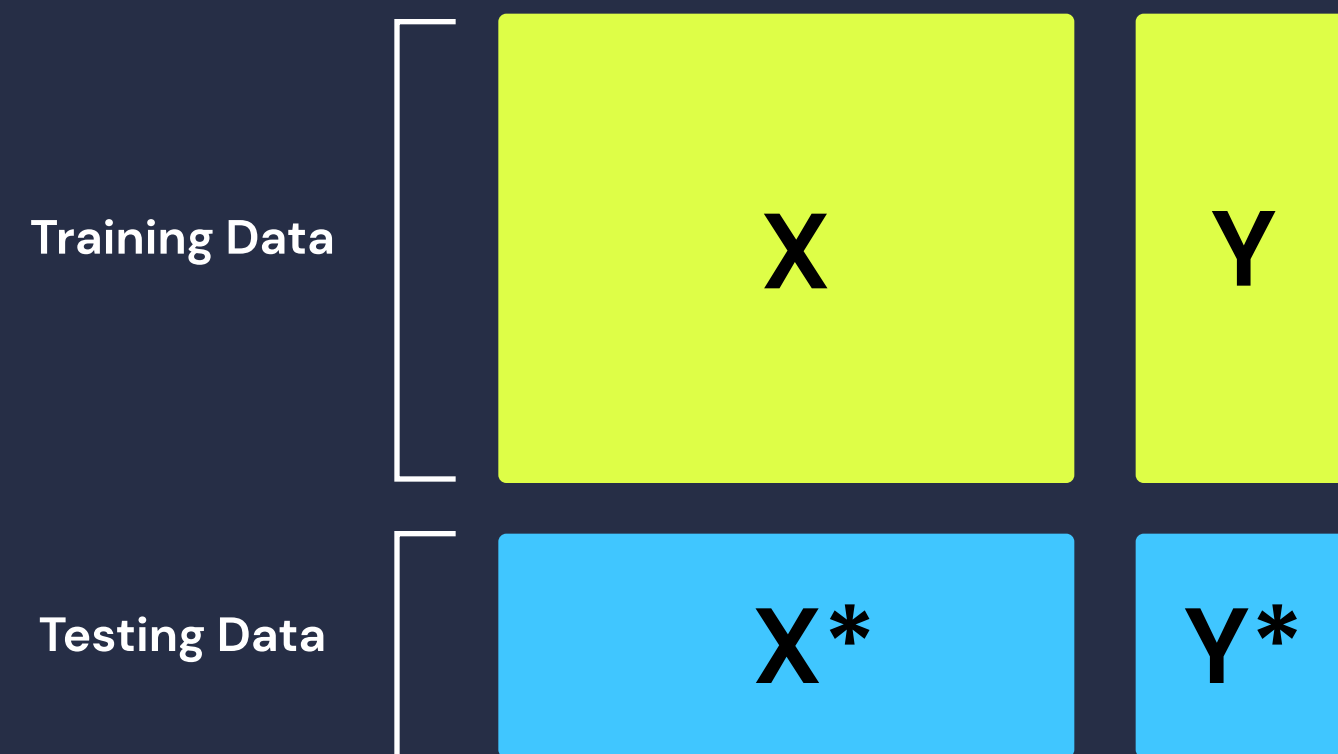


Introducing Human-Centered AI

The AI Market

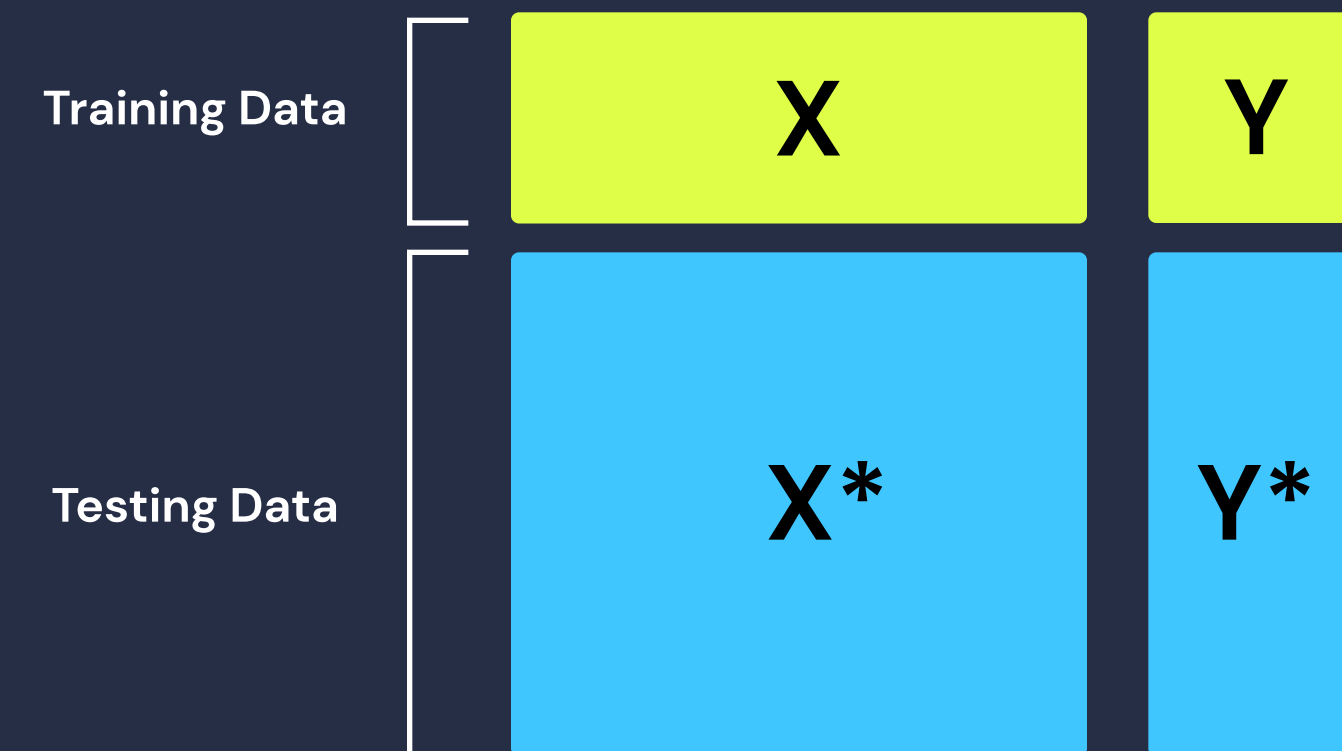
How does modern AI/ML works

there are **millions of rows of training data (X)** and **corresponding labels (Y)**. Given new testing data (X^*), the goal is to predict a corresponding label (Y^*)

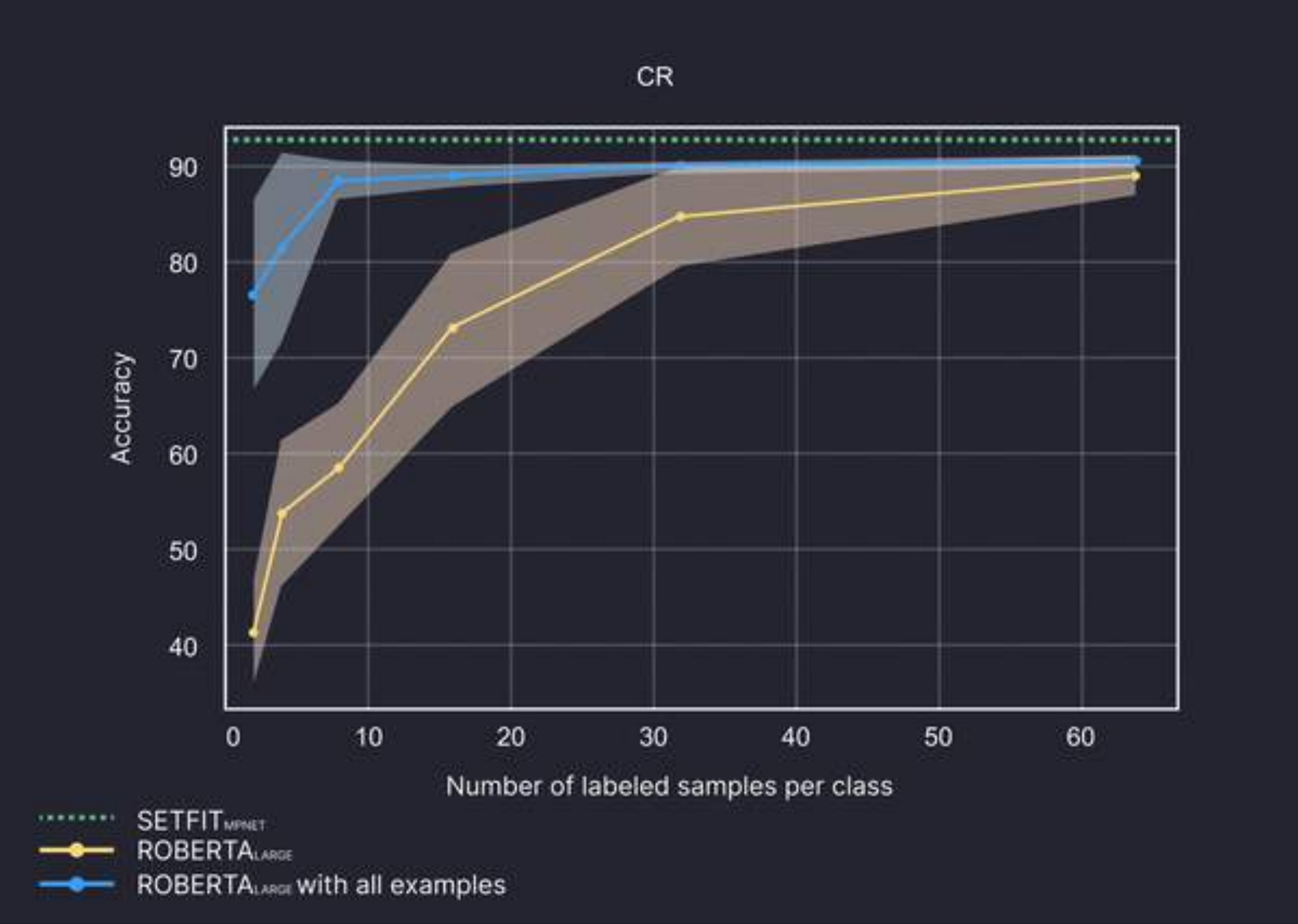


In few shot learning...

the goal is to **predict label (Y^*)** with just a **few rows of training data (X)** and corresponding labels (Y). This requires novel foundation models.

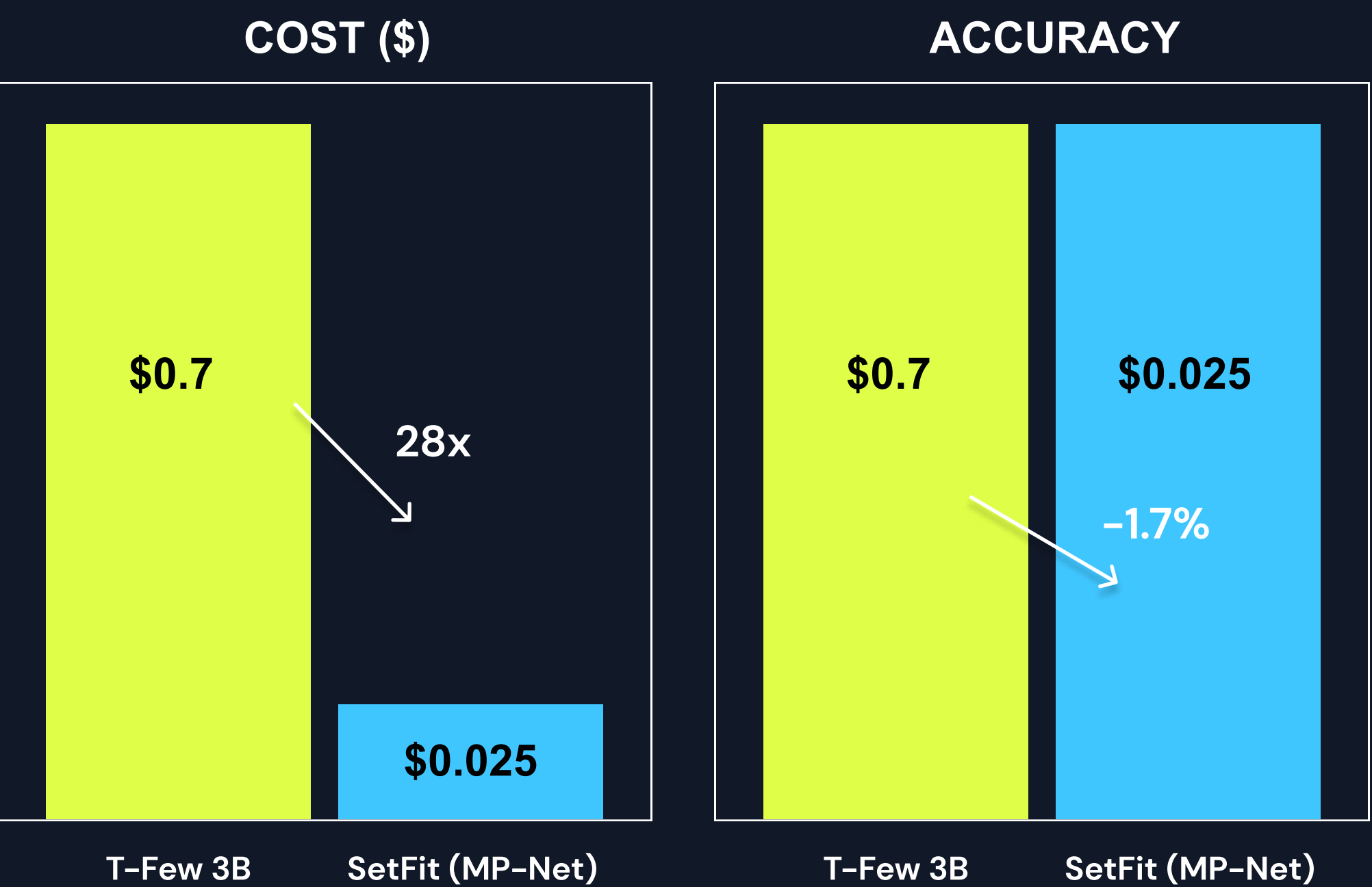


Setfit: Few Shot Learning in Practice



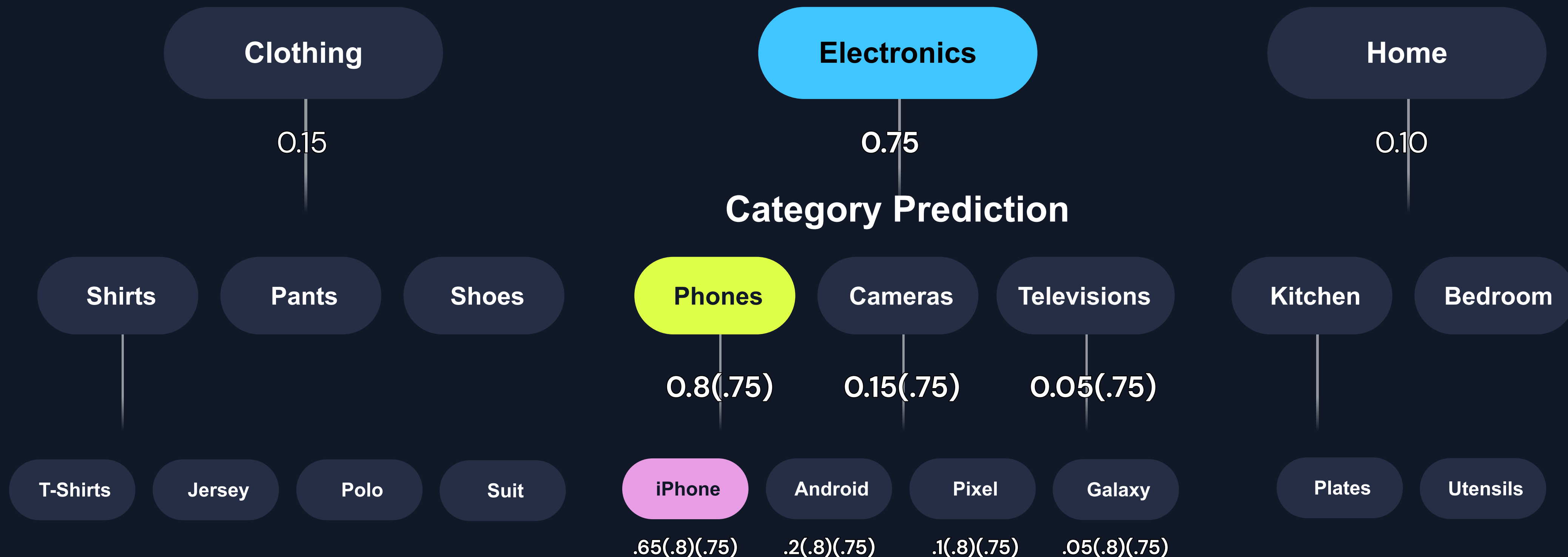
Rank	Method	Accuracy	Model Size
2	T-Few	75.8	11B
4	Human Baseline	73.5	N/A
6	SetFit (Roberta Large)	71.3	355M
9	PET	69.6	235M
11	SetFit (MP-Net)	66.9	110M
12	GPT-3	62.7	175B

Result



Few Shot Hierarchical Classification

I am at the Apple Store getting a **cellphone**. I admire Steve Job's vision of the **iPhone**. How can I get this **electronic device** at a good price?



Semi-Structured Prompts



AMAZON.COM ANNOUNCES SECOND QUARTER RESULTS

SEATTLE—(BUSINESS WIRE) August 3, 2023—Amazon.com, Inc. (NASDAQ: AMZN) today announced financial results for its second quarter ended June 30, 2023.

- **Net sales** increased 11% to \$134.4 billion in the second quarter, compared with \$121.2 billion in second quarter 2022. Excluding the \$0.3 billion unfavorable impact from year-over-year changes in foreign exchange rates throughout the quarter, net sales increased 11% compared with second quarter 2022.
 - North America segment sales increased 11% year-over-year to \$82.5 billion.
 - International segment sales increased 10% year-over-year to \$29.7 billion.
 - AWS segment sales increased 12% year-over-year to \$22.1 billion.
- **Operating income** increased to \$7.7 billion in the second quarter, compared with \$3.3 billion in second quarter 2022.
 - North America segment operating income was \$3.2 billion, compared with an operating loss of \$0.6 billion in second quarter 2022.
 - International segment operating loss was \$0.9 billion, compared with an operating loss of \$1.8 billion in second quarter 2022.
 - AWS segment operating income was \$5.4 billion, compared with operating income of \$5.7 billion in second quarter 2022.
- **Net income** was \$6.7 billion in the second quarter, or \$0.65 per diluted share, compared with a net loss of \$2.0 billion, or \$0.20 per diluted share, in second quarter 2022.
 - Second quarter 2023 net income includes a pre-tax valuation gain of \$0.2 billion included in non-operating expense from the common stock investment in Rivian Automotive, Inc., compared to a pre-tax valuation loss of \$3.9 billion from the investment in second quarter 2022.
- **Operating cash flow** increased 74% to \$61.8 billion for the trailing twelve months, compared with \$35.6 billion for the trailing twelve months ended June 30, 2022.
- **Free cash flow** improved to an inflow of \$7.9 billion for the trailing twelve months, compared with an outflow of \$23.5 billion for the trailing twelve months ended June 30, 2022.
- **Free cash flow less principal repayments of finance leases and financing obligations** improved to an inflow of \$1.9 billion for the trailing twelve months, compared with an outflow of \$33.5 billion for the trailing twelve months ended June 30, 2022.
- **Free cash flow less equipment finance leases and principal repayments of all other finance leases and financing obligations** improved to an inflow of \$6.7 billion for the trailing twelve months, compared with an outflow of \$26.1 billion for the trailing twelve months ended June 30, 2022.

Column 



Annotation History

what is a summary of this document?

The document provides Amazon's financial results and business metrics for the AWS segment, including net sales, operating income, and operating margin.

was the company's performance
better or worse than expected

better

worse

Confirm

Skip →

2 Annotated

3 Remaining

Model Centric AI

In model-centric AI, the training data is treated as a fixed input. Your training data is something you download as a static file. New iterations of your project result from changes to the model.

Try Different Models on Static Datasets



Text Classification

Text	Label
I like my bananas	fruit
Broccoli is not good	vegetable
We like tomatoes and potatoes	vegetable

Name Entity Recognition

Text	Label
I like my bananas	I like my bananas [FRUIT]
Broccoli is not good	Broccoli [VEGETABLE] is not good
We like tomatoes and potatoes	We like tomatoes [FRUIT] and potatoes [VEGETABLE] is not good

Data Centric AI

In data-centric AI, data quality and quantity is increasingly the key to successful results. Teams spend more time on labeling, managing, slicing, augmenting, and curating the data, with the model itself relatively more fixed. In data centric AI, you programmatically iterate on your training data.

Try different datasets on static model



Text Classification

Text	Label
I like my bananas	fruit
Broccoli is not good	vegetable
We like tomatoes and cake	dessert
Cookies and milk are great	dessert

Name Entity Recognition

Text	Label
I like my bananas	I like my bananas [FRUIT]
Broccoli is not good	Broccoli [VEGETABLE] is not good
We like tomatoes and cake	We like tomatoes [FRUIT] and cake [DESSERT]
Cookies and milk are great	Cookies [DESSERT] and cake [DESSERT] are great

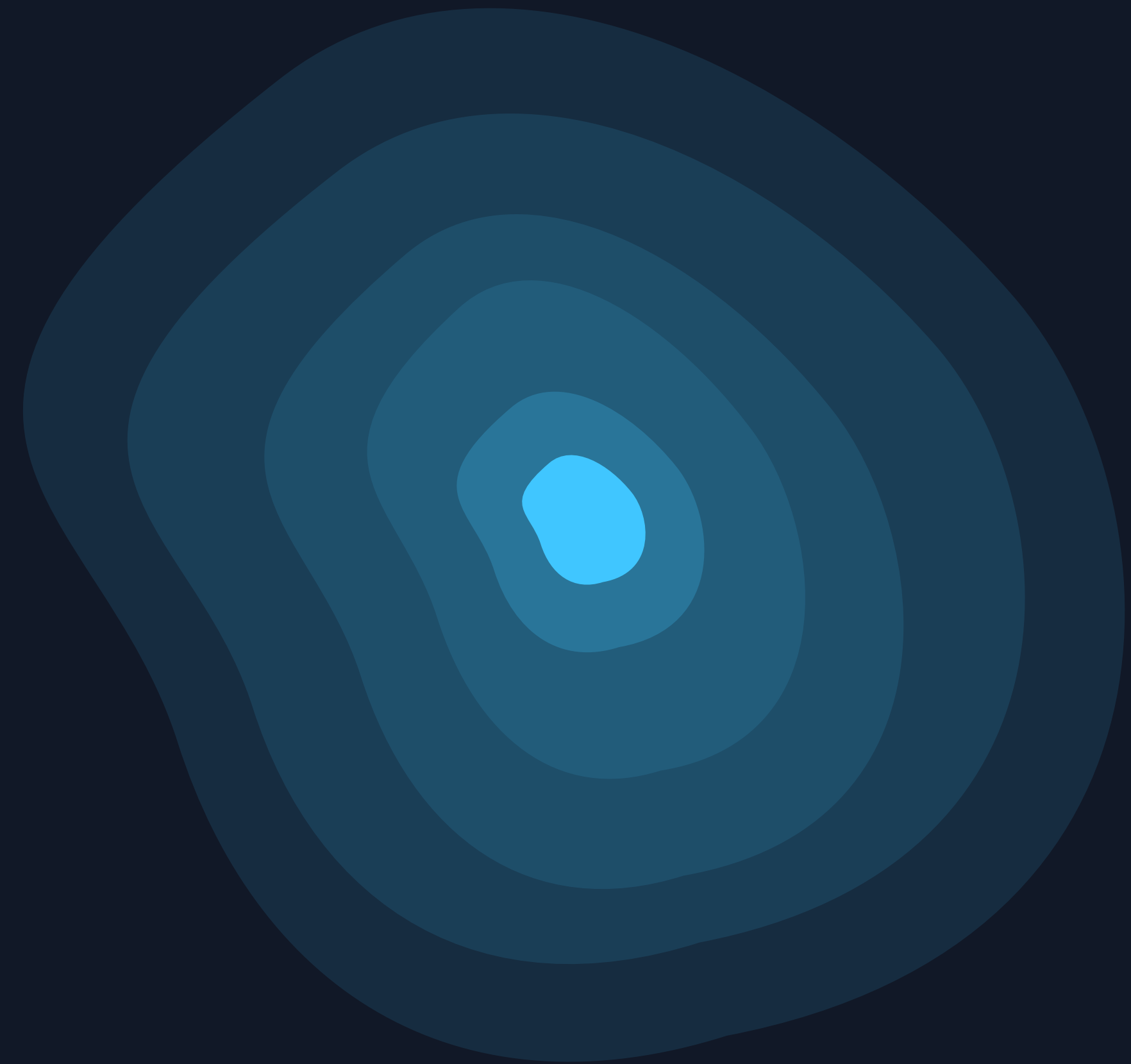
The “Scale”-ability Issue

Manual Labeling

As we label data, new classes of data are added due to changing business requirements. Oftentimes, this requires manually relabeling all of the data again, from scratch.

Progammatic Labeling

A solution that works in a specific vertical (i.e. finance) may not translate well to a different vertical (i.e. healthcare). Many labeling applications can't be solved programmatically



Less than 1/1000th of data that businesses require to be labeled is currently labeled today. As **Carl Sagan** said, we are just a pale blue dot.

A New Way of Doing AI

As we iterate on our data, we also iterate on our ensemble model
The ensemble model output is a function of the data input
As we label more, the ensemble model learns which local model is best

Try different datasets on static model



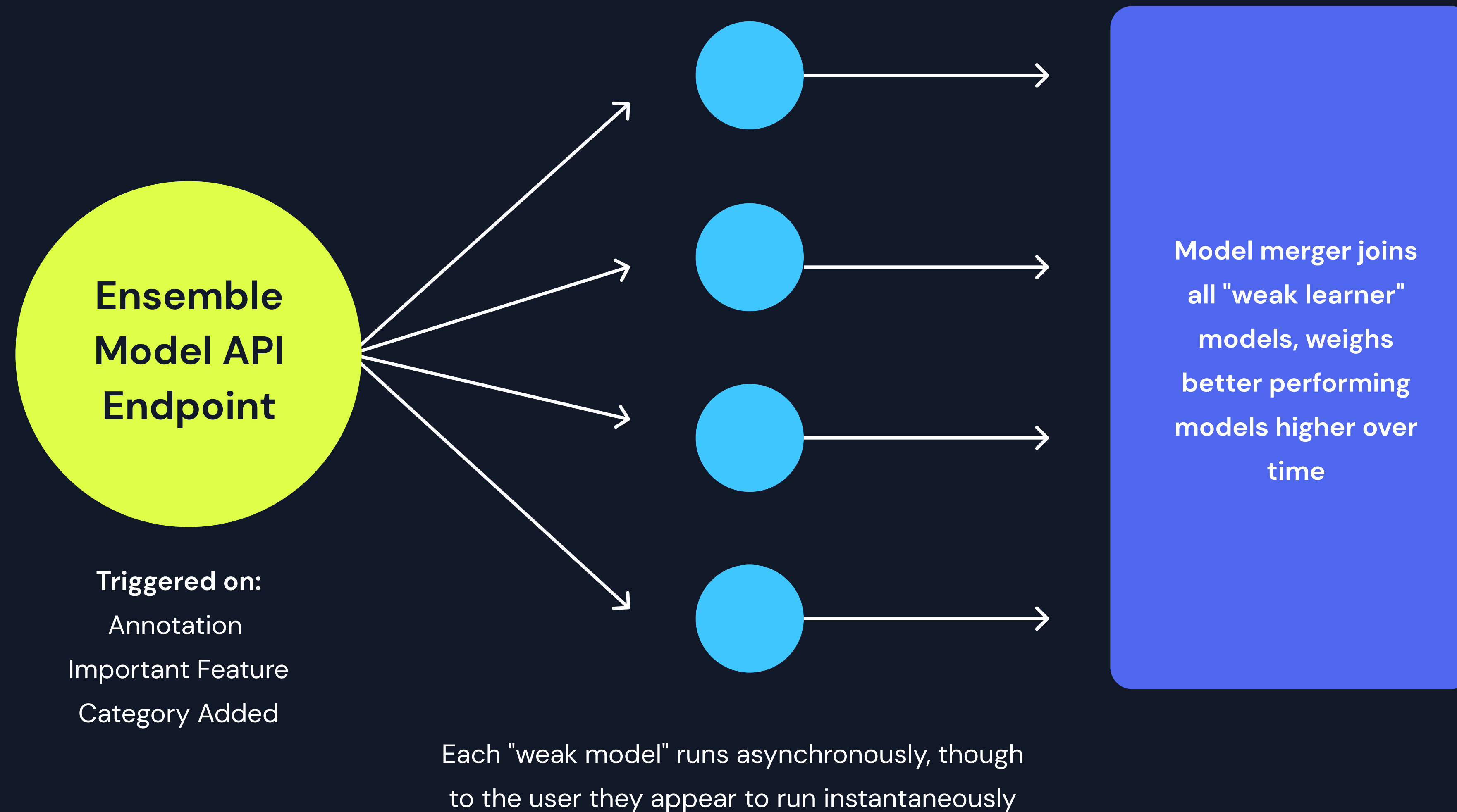
Text Classification

Text	Label
I like my bananas	fruit
Broccoli is not good	vegetable
We like tomatoes and cake	dessert
Cookies and milk are great	dessert

Name Entity Recognition

Entity_Text	Entity_Label
I like my bananas	I like my bananas [FRUIT]
Broccoli is not good	Broccoli [VEGETABLE] is not good
We like tomatoes and cake	We like tomatoes [FRUIT] and cake [DESSERT]
Cookies and milk are great	Cookies [DESSERT] and cake [BEVERAGE] are great

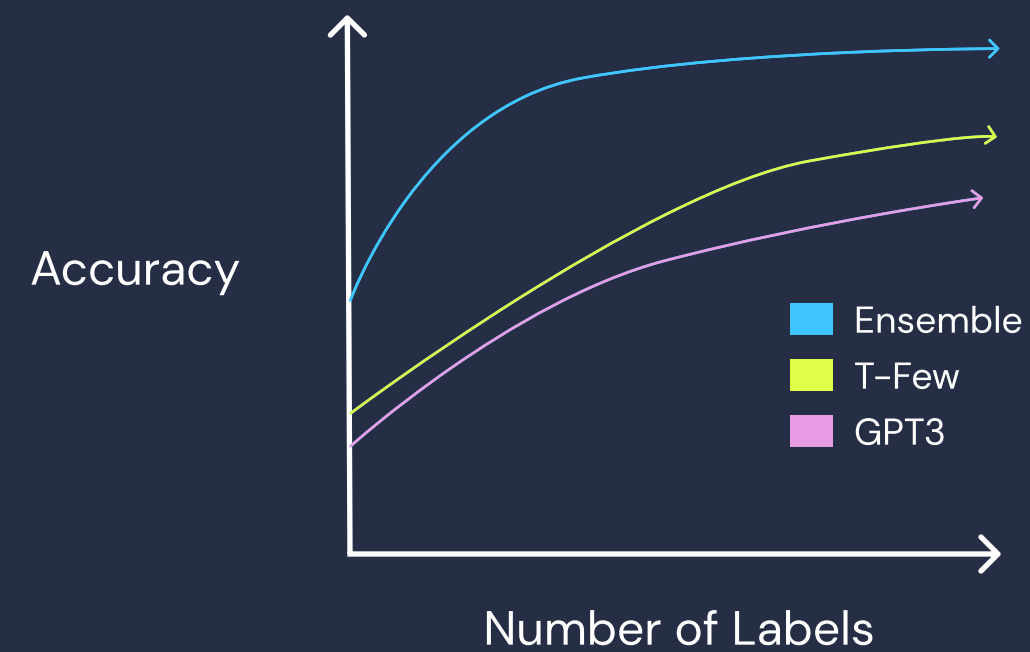
Enabling a Scalable Data Labeling Solution



Features

Few Shot Learning

After you label just a few row of data (the edge cases), we label the rest, using state of the art transformer models



Programmatic Labeling

Heuristics such as key words, entities and regex expressions are fed into the model

John flew to **Brazil**
to play **soccer** and
eat **steak**



Decomposition

Convert unstructured text data to a spreadsheet



Synchronous

Get your data labeled in real time, and receive immediate feedback on the model's performance as you label data

Index	Text	Label	Probability
0	Click this link to win \$500	SPAM	.98
1	This is a normal sentence	NOT SPAM	.89
2	You just won a free iPhone	SPAM	.93
3	Hope you are doing well	NOT SPAM	.87
4	Enter your credit card info	SPAM	.97

Human in the Loop

We actively learn from subject matter experts, who provide input into our model

Click this **link** to
win **\$500**



Contextual

Our large language models extract the context of words in sentences, not just the individual words themselves

I robbed a **bank** to make money

I went to the river **bank** to relax