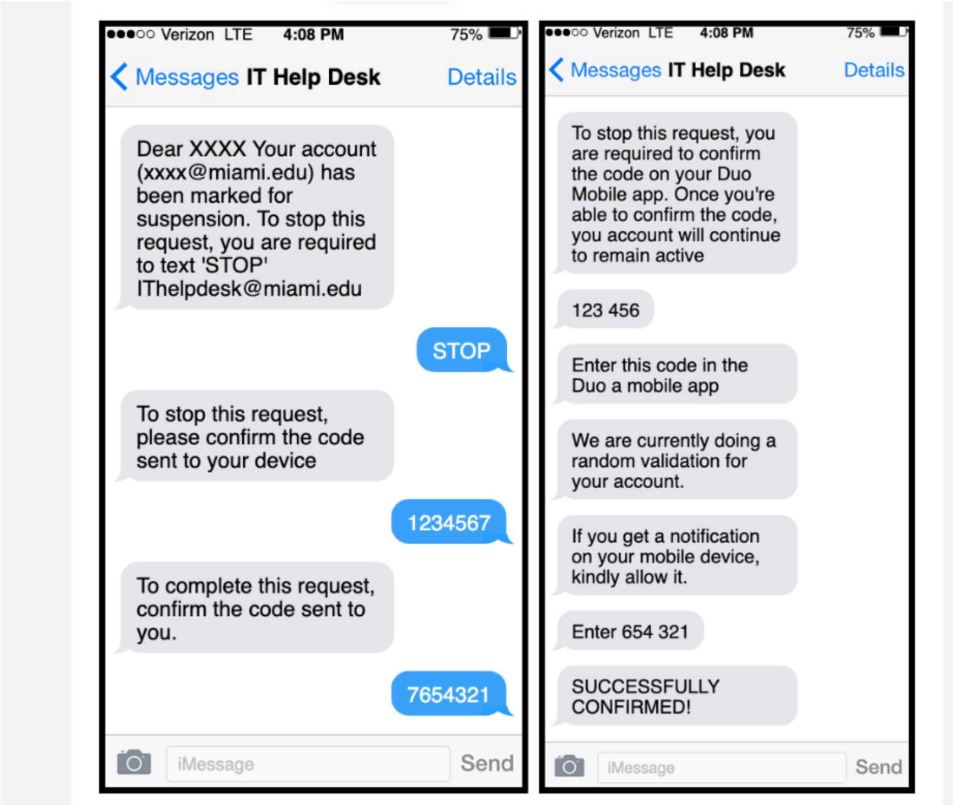
CSC 116: AI for Malicious Analysis

- Protecting Hospitals from Cyber Threats

Real Attack Case Study

- Fake email from 'IT Support'.
- Result: Student records stolen.


●●●○○ Verizon LTE 4:08 PM 75%
Messages IT Help Desk Details
iMessage
Send

Dear XXXX Your account (xxxx@miami.edu) has been marked for suspension. To stop this request, you are required to text 'STOP' IThelpdesk@miami.edu

STOP

To stop this request, please confirm the code sent to your device

1234567

To complete this request, confirm the code sent to you.

7654321

●●●○○ Verizon LTE 4:08 PM 75%
Messages IT Help Desk Details
iMessage
Send

To stop this request, you are required to confirm the code on your Duo Mobile app. Once you're able to confirm the code, you account will continue to remain active

123 456

Enter this code in the Duo a mobile app

We are currently doing a random validation for your account.

If you get a notification on your mobile device, kindly allow it.

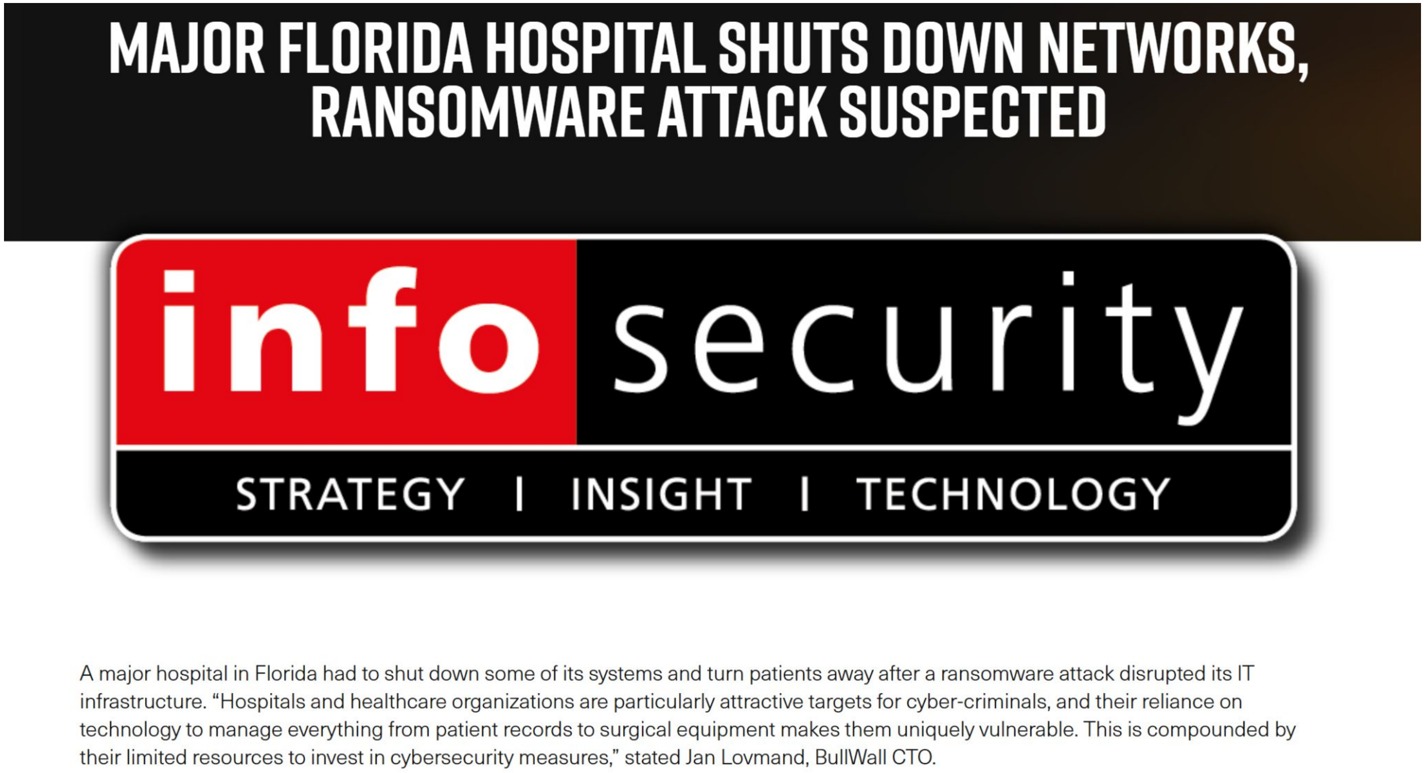
Enter 654 321

SUCCESSFULLY CONFIRMED!

the University will **never** ask you to provide sensitive information via direct phone call, text, email, and/or online form. Also, remember to **never** provide MFA access codes to anyone requesting them.

Why are Hospitals Under Attack?

- Healthcare data is highly **valuable**.
- Example: Ransomware shutting down ICU operations.



MAJOR FLORIDA HOSPITAL SHUTS DOWN NETWORKS, RANSOMWARE ATTACK SUSPECTED

info security

STRATEGY | INSIGHT | TECHNOLOGY

A major hospital in Florida had to shut down some of its systems and turn patients away after a ransomware attack disrupted its IT infrastructure. “Hospitals and healthcare organizations are particularly attractive targets for cyber-criminals, and their reliance on technology to manage everything from patient records to surgical equipment makes them uniquely vulnerable. This is compounded by their limited resources to invest in cybersecurity measures,” stated Jan Lovmand, BullWall CTO.

What Happens Without AI?

- Manual detection is slow.
- Threats are missed.
- Patient care may under suffer.

How AI Supports?

- AI scans emails, logins, and prescriptions in seconds.

What is AI?

- AI: Artificial Intelligence.
- Example: Detecting abnormal prescriptions.

Human vs. AI Detection

- Humans: Limited time.
- AI: Fast, consistent, 24/7.

AI in Action

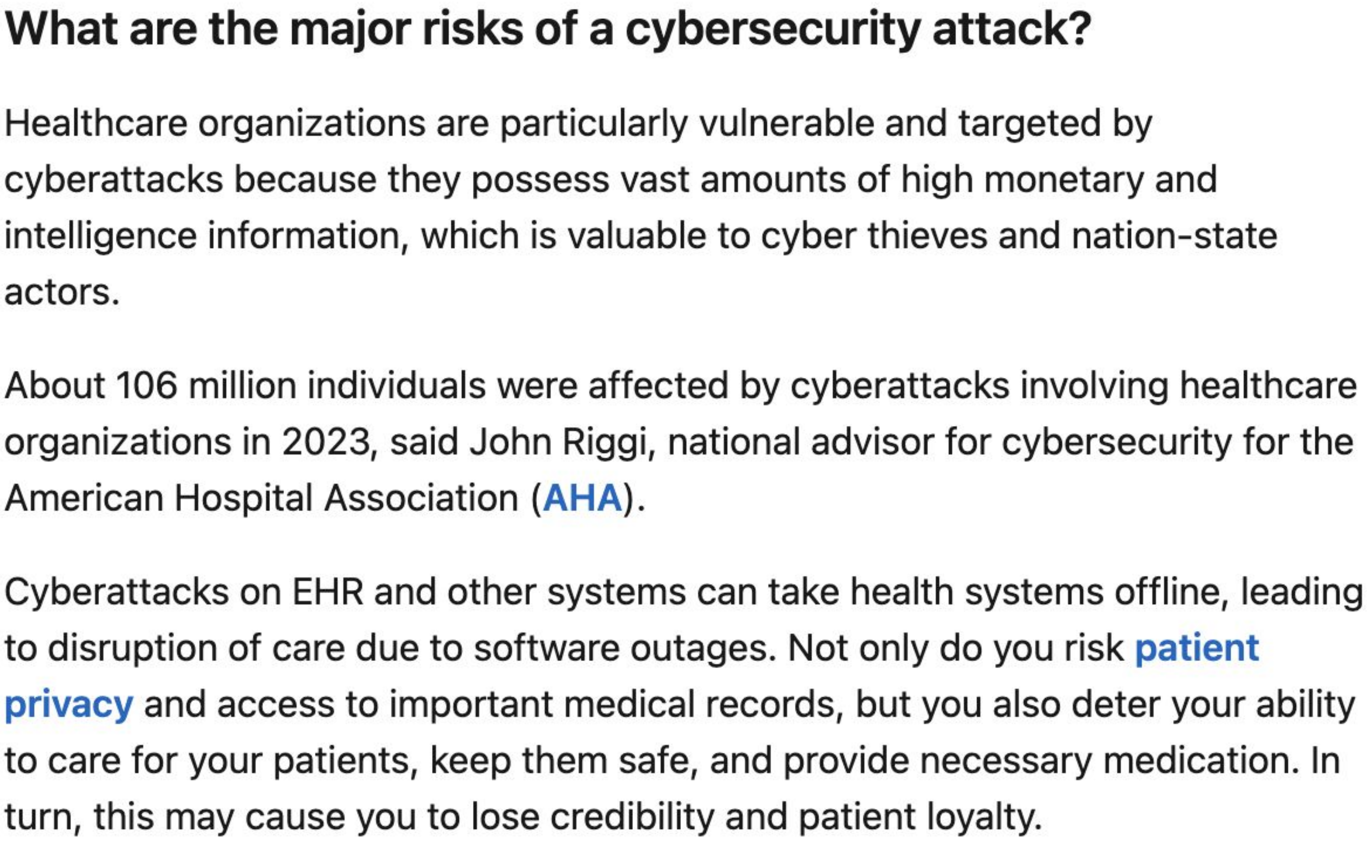
- Phishing Email → AI flags → IT reviews → Threat stopped.

Why AI in Hospitals?

- **Faster Diagnosis**: AI helps analyze medical images, lab results, and patient history quickly and accurately.
- **Predictive Analytics**: AI can forecast patient outcomes, readmission risks, and disease progression.
- **Personalized Treatment Plans**: AI tailors care based on patient data, genetics, and treatment history.
- **Virtual Assistants**: AI chatbots provide 24/7 support to patients for queries, follow-ups, or reminders.
- etc.

Malicious Activities in Healthcare

- Phishing, fake orders, unauthorized access, etc.



What are the major risks of a cybersecurity attack?

Healthcare organizations are particularly vulnerable and targeted by cyberattacks because they possess vast amounts of high monetary and intelligence information, which is valuable to cyber thieves and nation-state actors.

About 106 million individuals were affected by cyberattacks involving healthcare organizations in 2023, said John Riggi, national advisor for cybersecurity for the American Hospital Association (AHA).

Cyberattacks on EHR and other systems can take health systems offline, leading to disruption of care due to software outages. Not only do you risk patient privacy and access to important medical records, but you also deter your ability to care for your patients, keep them safe, and provide necessary medication. In turn, this may cause you to lose credibility and patient loyalty.

Phishing Email Example

- Fake request for login details from 'hospital IT'.

Unauthorized Access Case

- Login from unusual locations or other devices.

Fake Doctor Orders

- Fraudulent requests for medication changes.

Patient Record Theft

- Sensitive data stolen and sold.

Impact of Malware

- System lockdowns delaying patient care.

Insider Threat Example

- Employee leaking records.

— using blockchain for recording all the data query/download history

How IT Can Help

- Report suspicious activity.
- But it is hard, because we are humans, we are too slow.
- Too much data for humans alone.

DEMO 1: Phishing Email Detection

- **BERT Model**: Bidirectional Encoder Representations from Transformers (you don’t need to memorize this – just think of it as a smart AI language reader).

BERT is an **AI model that understands language**, just like how humans read and understand sentences.
It helps computers **“read between the lines”**, understand **meaning**, **context**, and even **medical terms** in documents or messages.

How is BERT used in hospitals?

- **Helps read medical records quickly** and extract key info like symptoms, diagnoses, or medications.
- **Sorts through thousands of documents** to find the right information for doctors and nurses.
- **Understands patient messages** in chatbots or digital forms and sends them to the right department.
- **Assists in clinical decision support systems**, by reading guidelines and matching them to patient data.

Real-world example for nurses:

A patient types: "I feel dizzy after taking the medication."

- A regular computer might just focus on "dizzy."
- BERT understands **the full meaning** and realizes that **the medication could be the cause**, helping systems alert a nurse or doctor.

Research Demo

- https://www.splunk.com/en_us/blog/security/deep-learning-in-security-text-based-phishing-email-detection-with-bert-model.html

Discussion

- Would you trust this AI?

AI may fail but if it is high accuracy, e.g., 99%, it is trusted.

If the model accuracy is 90%, then, the result may be wrong.

DEMO 2: Suspicious Login Detection

- A "Suspicious Login Detection" model is a machine learning algorithm designed to **identify and flag potential unauthorized login attempts** by analyzing user behavior patterns like **location, device, login frequency, time of day, and other relevant data**, flagging anomalies that deviate significantly from a user's typical login activity, potentially indicating a malicious attempt to access an account.

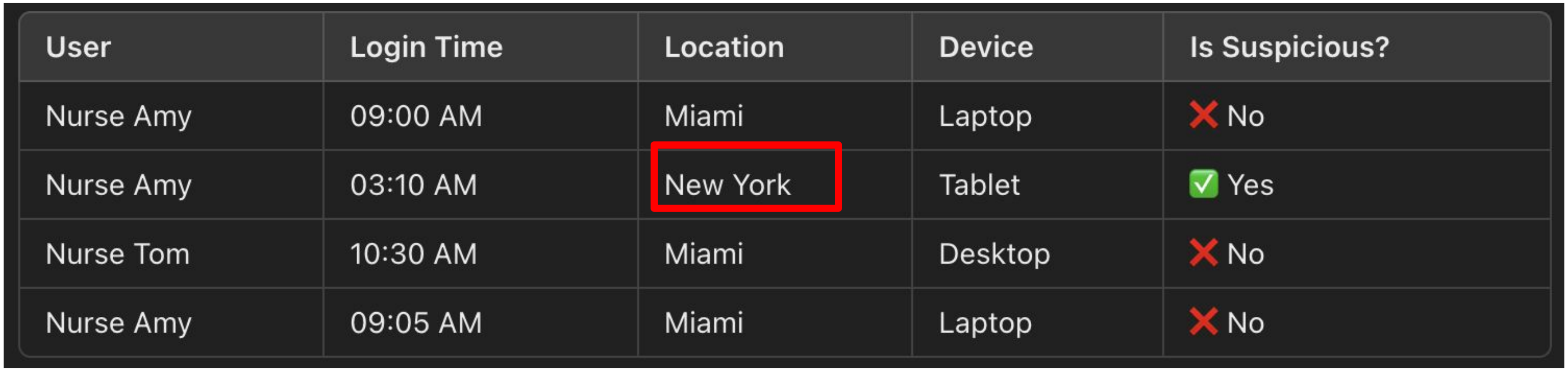
How to Improve?

Feedback loop with human review.

Human-in-the-loop feedback

Very slow as well!

Imagine you always check in to your hospital system from Miami at 9 AM using your laptop. But suddenly there’s a login from **New York at 3 AM using a different device**. Even if the password is correct, the system thinks: **“Hmm, that’s unusual!”**



User	Login Time	Location	Device	Is Suspicious?
Nurse Amy	09:00 AM	Miami	Laptop	❌ No
Nurse Amy	03:10 AM	New York	Tablet	✅ Yes
Nurse Tom	10:30 AM	Miami	Desktop	❌ No
Nurse Amy	09:05 AM	Miami	Laptop	❌ No

Model Selections

1. You have labeled data (normal vs suspicious)

   Random Forest or **XGBoost**

   Very effective for classification tasks, fast, interpretable, and performs well on (time, location, device, etc).

2. You don’t have labels, just want to detect unusual behavior

   Isolation Forest or Autoencoder

   Great for detecting anomalies without prior examples of “suspicious” cases.

3. You want to capture user behavior over time (sequences)

   LSTM (Long Short-Term Memory)

   Good at modeling login sequences, e.g., how often someone logs in, changes in patterns over days.

4. You want real-time detection with speed

   Logistic Regression or **Isolation Forest**

   Fast and lightweight for live systems.

DEMO 3: Fake Prescription Detection

- Spotting forged prescriptions.

– Using BERT / **Transformer** / or LLMs

the most famous one

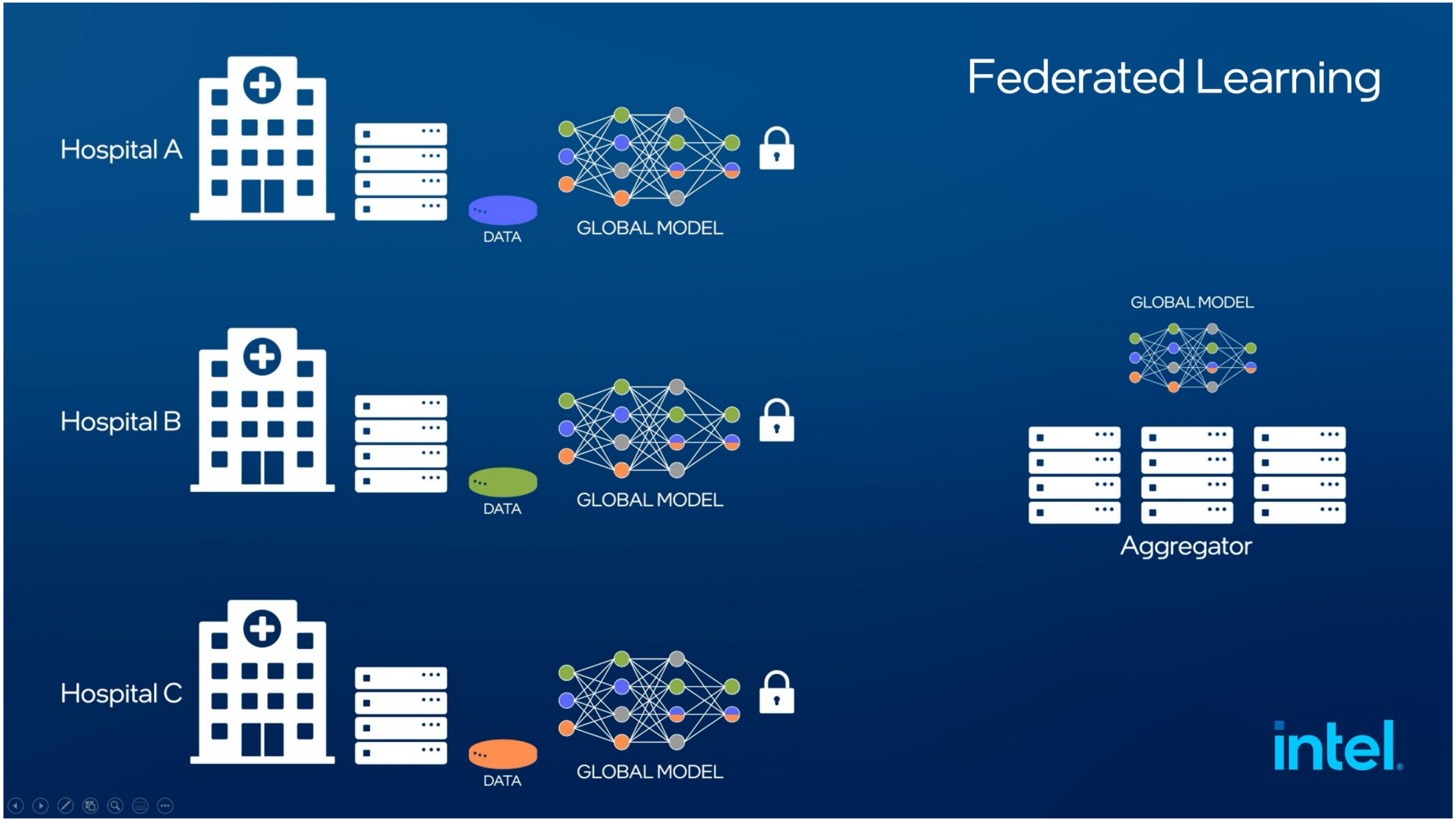
https://arxiv.org/pdf/1706.03762

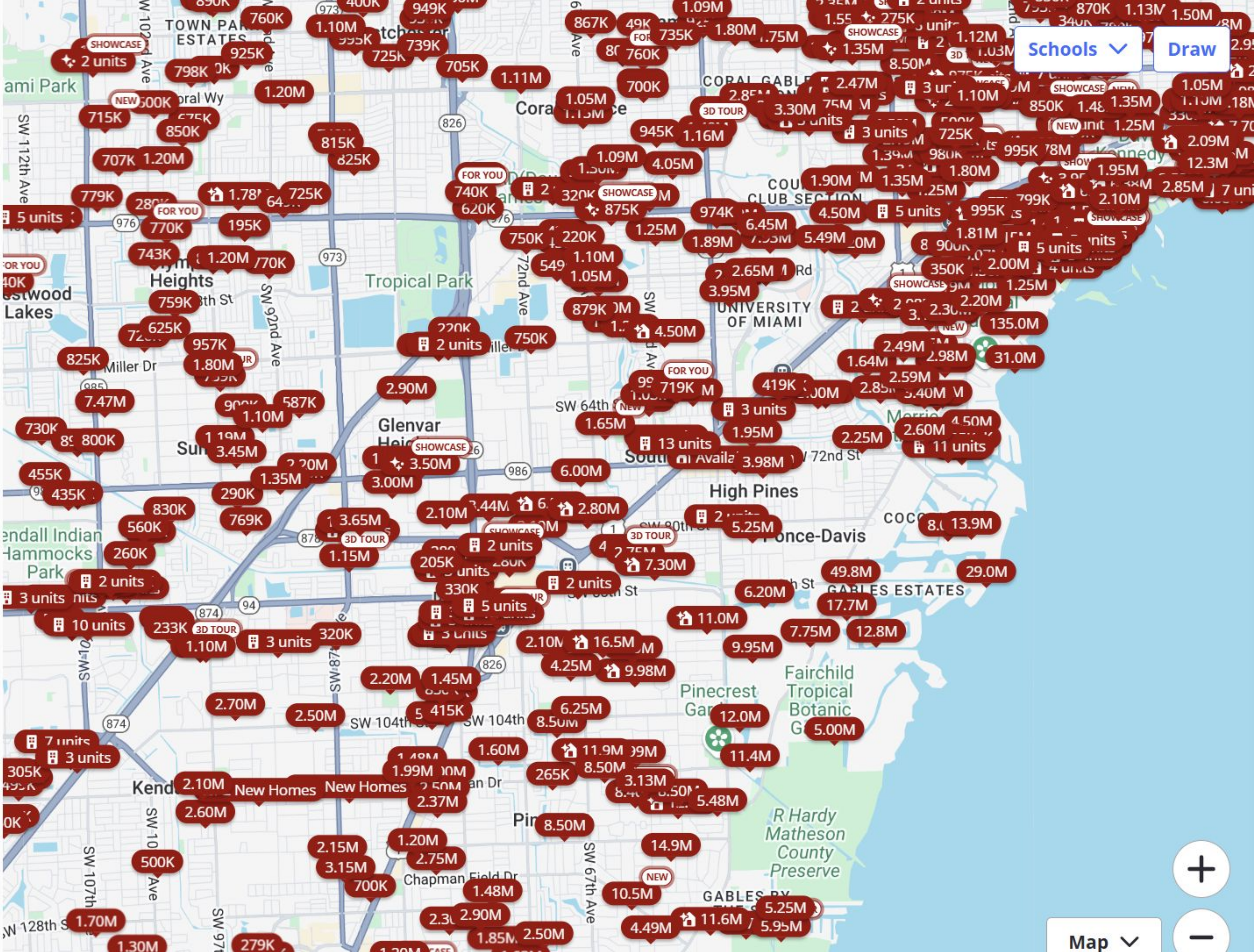
“attention is all you need”

Using AI for detecting malicious behaviors

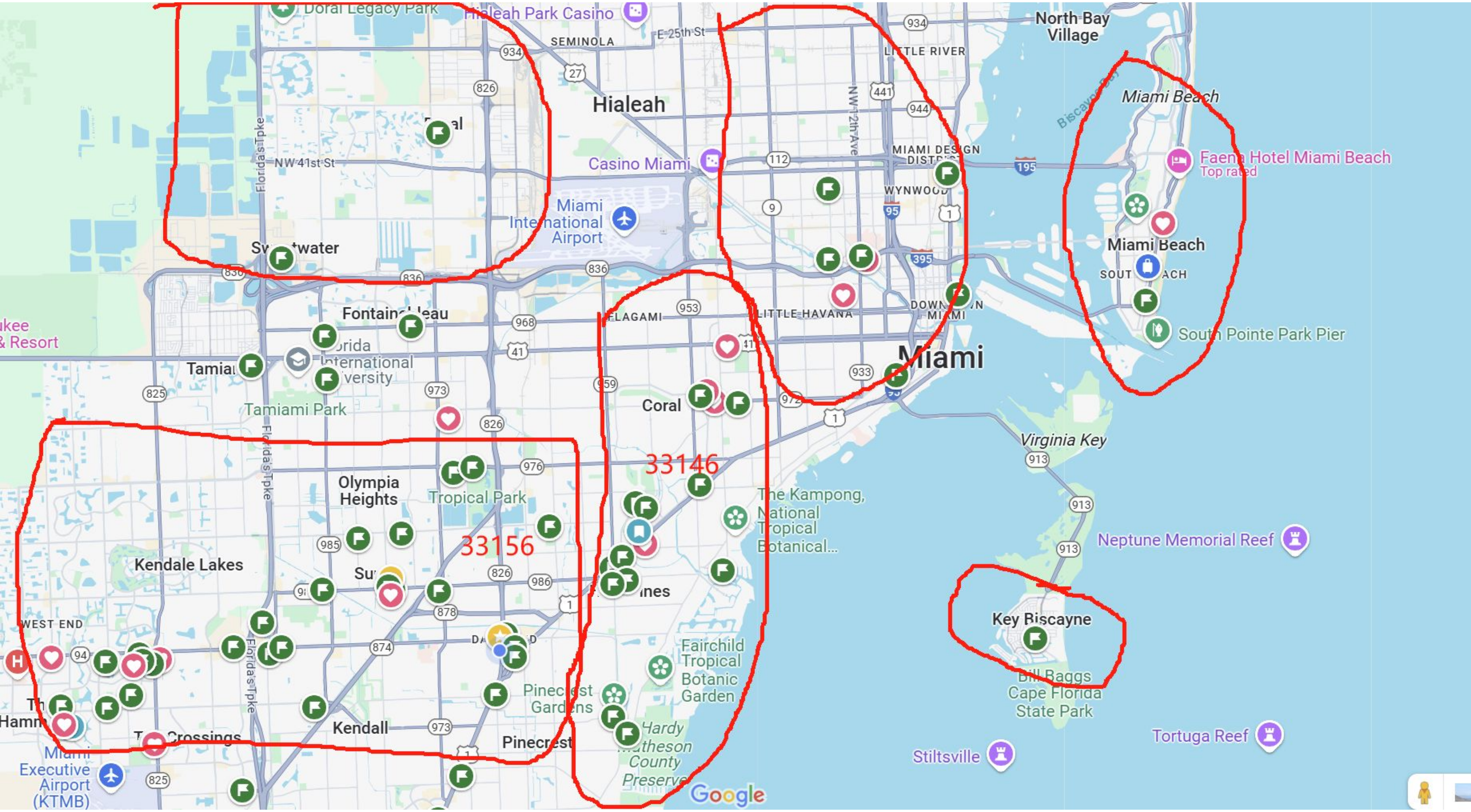
Malicious attacks in AI models

Federated Learning



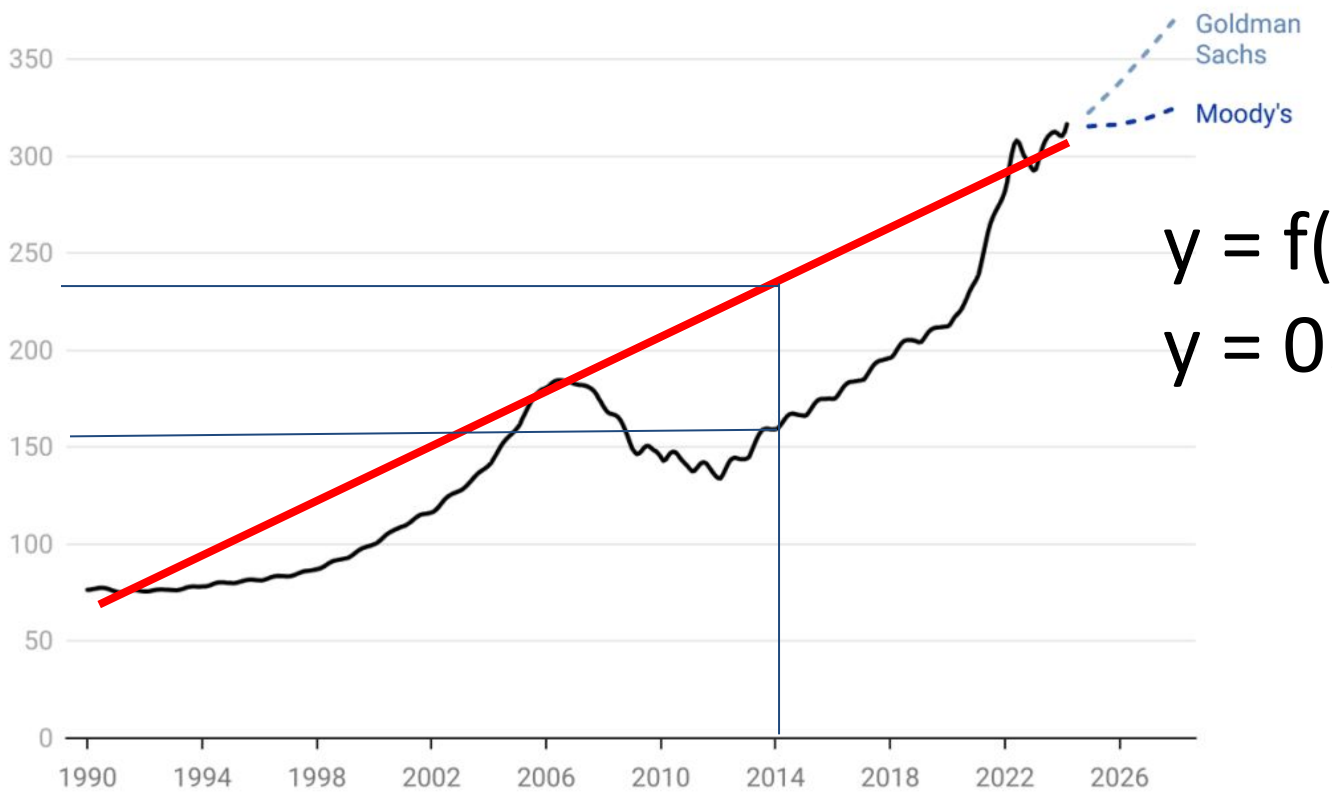

Schools
Draw
Map
+
−
TOWN PARK ESTATES
Coral Gables
Coral Terrace
COUNTRY CLUB SECTION
UNIVERSITY OF MIAMI
Tropical Park
Glenvar Heights
Olympia Heights
Westwood Lakes
Miller Dr
Sunset
South Miami
High Pines
Ponce-Davis
GABLES ESTATES
Kendall Indian Hammocks Park
Kendall
Pinecrest
Pinecrest Gardens
Fairchild Tropical Botanic Garden
R Hardy Matheson County Preserve
Chapman Field Dr
GABLES BY THE SEA
SW 112th Ave
SW 102nd Ave
SW 92nd Ave
SW 87th Ave
SW 72nd Ave
SW 67th Ave
SW 107th Ave
SW 104th St
SW 64th St
SW 72nd St
SW 80th St
SW 128th St
826
976
973
986
878
874
94
985
1
SHOWCASE
FOR YOU
NEW
3D TOUR
2 units
3 units
5 units
7 units
10 units
11 units
13 units
New Homes
760K
1.10M
949K
867K
735K
1.80M
1.75M
1.09M
925K
798K
739K
725K
705K
1.11M
1.20M
700K
1.05M
2.47M
1.10M
1.35M
1.12M
850K
715K
815K
825K
945K
1.16M
725K
995K
2.09M
12.3M
707K
1.09M
4.05M
1.90M
1.80M
1.95M
2.85M
779K
1.78M
725K
740K
620K
875K
974K
4.50M
995K
2.10M
770K
195K
750K
220K
1.25M
6.45M
5.49M
1.89M
1.81M
743K
1.20M
770K
1.10M
1.05M
2.65M
350K
2.00M
3.95M
1.25M
759K
879K
4.50M
2.20M
625K
957K
220K
750K
135.0M
825K
1.80M
2.49M
1.64M
2.98M
31.0M
7.47M
587K
1.10M
2.90M
719K
419K
2.59M
730K
800K
3.45M
1.65M
3 units
1.95M
2.25M
2.60M
4.50M
455K
435K
1.35M
3.50M
3.00M
6.00M
3.98M
830K
290K
769K
560K
3.65M
2.10M
2.80M
5.25M
13.9M
260K
1.15M
205K
7.30M
49.8M
29.0M
330K
6.20M
17.7M
233K
1.10M
320K
11.0M
7.75M
12.8M
9.95M
2.10M
16.5M
4.25M
9.98M
2.20M
1.45M
2.70M
2.50M
415K
6.25M
12.0M
5.00M
1.60M
11.9M
8.50M
11.4M
305K
2.10M
1.99M
265K
3.13M
2.37M
5.48M
2.60M
8.50M
2.15M
1.20M
2.75M
14.9M
3.15M
700K
500K
1.48M
10.5M
2.90M
5.25M
1.70M
11.6M
5.95M
4.49M
1.85M
2.50M
1.30M
279K

Classification Model



Where Goldman Sachs' and Moody's forecast models expect U.S. home prices to go in 2024, 2025, 2026, and 2027

Case-Shiller National Home Price Index



y = f(x)

y = 0.8x

Goldman Sachs' forecast projects for Case-Shiller, while Moody's forecasts for their own repeat sales index.

Chart: Lance Lambert • Created with Datawrapper

ResiClub

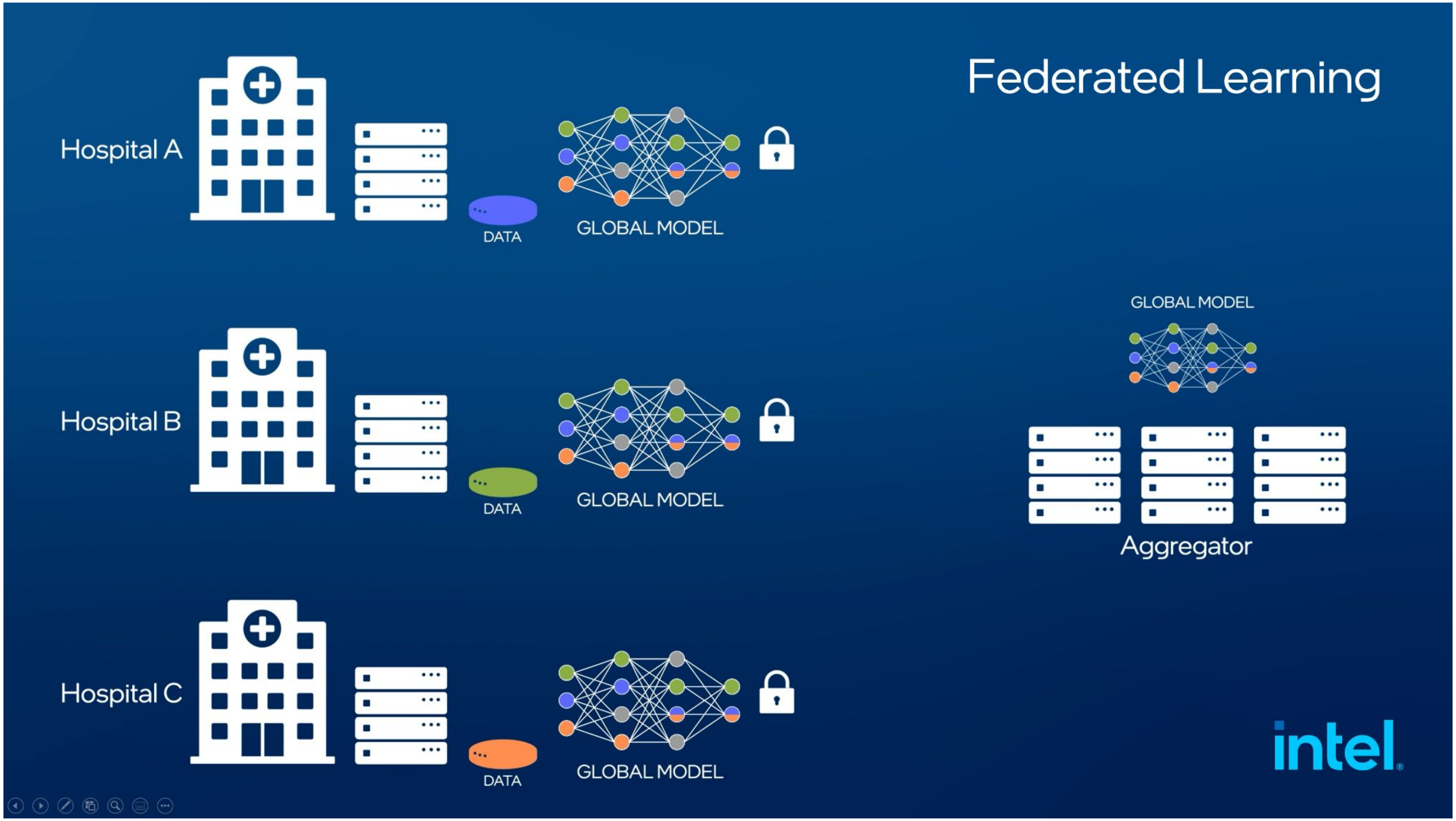
Only by the years

Linear Regression Model

Conclusion:

Building model is a way to find the best answer to your question.

Federated Learning

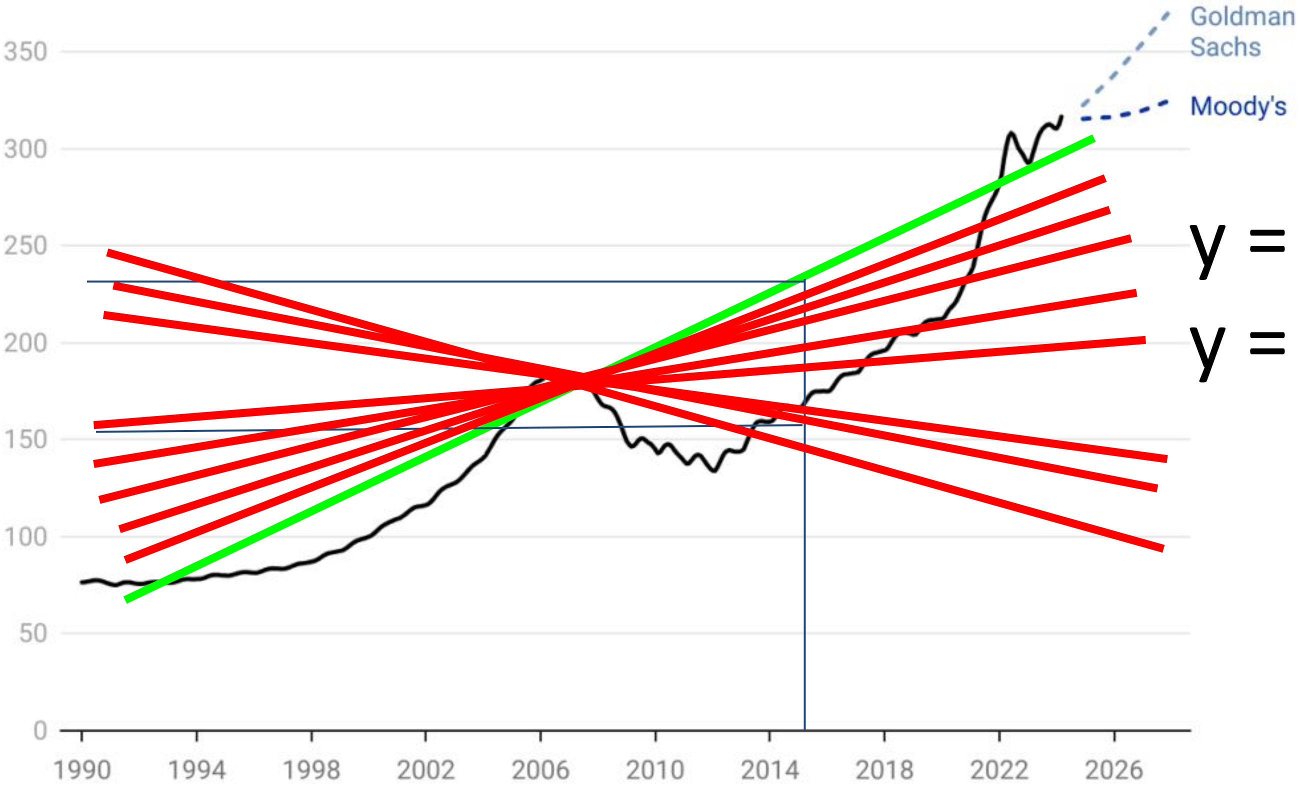


Gradient descent
is experience when model training

You want to give an injection to a patient, but you're new and unsure about the best spot to do it.

Where Goldman Sachs' and Moody's forecast models expect U.S. home prices to go in 2024, 2025, 2026, and 2027

Case-Shiller National Home Price Index



Goldman Sachs' forecast projects for Case-Shiller, while Moody's forecasts for their own repeat sales index.

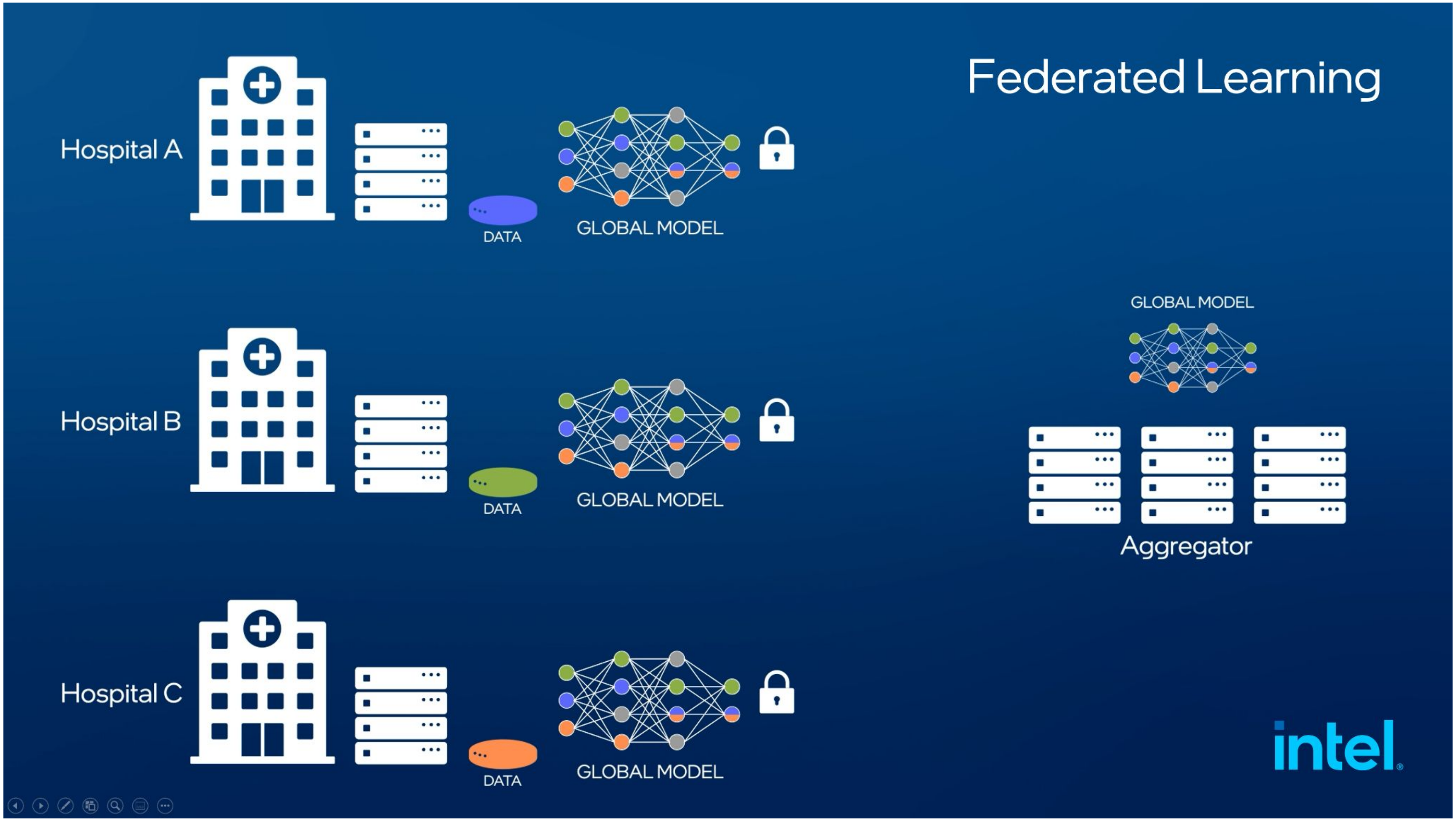
Chart: Lance Lambert • Created with Datawrapper

ResiClub

$y = f(x)$

$y = 0.8x$

Federated Learning



1 Polishing attack: data preparation phase

A "polishing attack" is a type of phishing attack that uses targeted, personalized messages to trick a specific individual or organization into revealing sensitive information or clicking malicious links.

Polishing attack

Scenario: Healthcare Records Manipulation

Imagine a hospital’s medical records database secured by an anomaly detection system that flags large, abrupt data changes as suspicious. **An attacker aiming to manipulate patient medication records**—perhaps to obtain fraudulent prescriptions—realizes direct manipulation would quickly trigger alerts.

Instead, they use a **polishing attack**, making minor, incremental adjustments over several days or weeks.

2. Patient data leakage : Data Preparation Phase

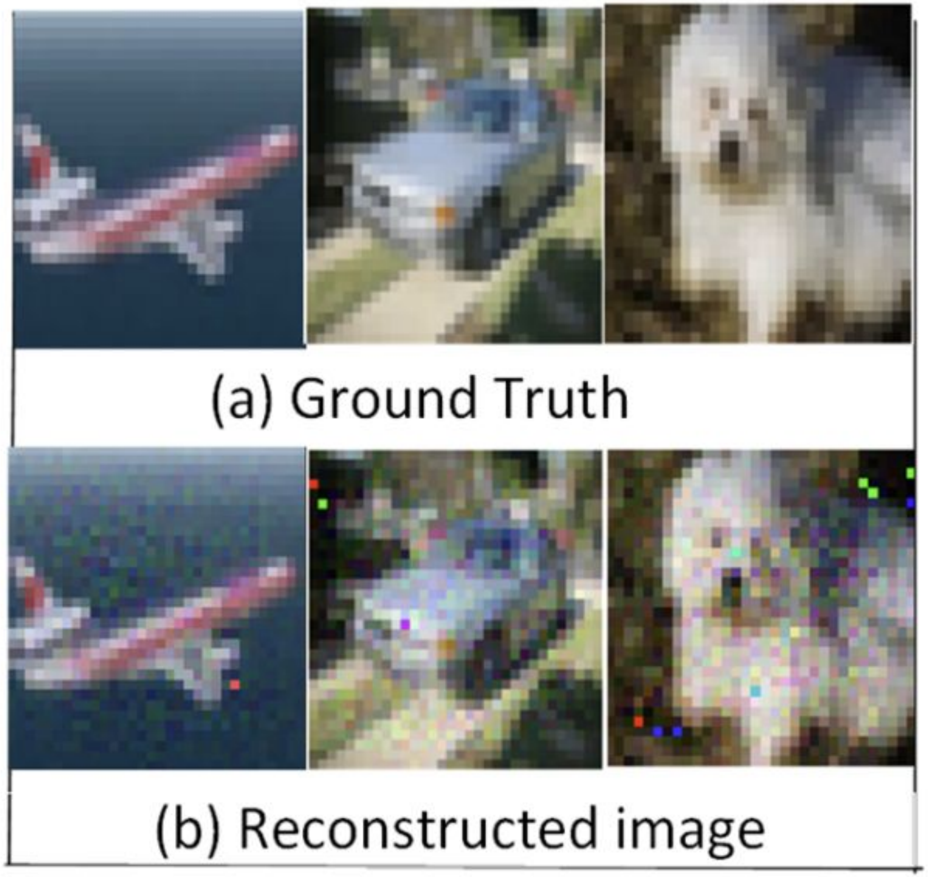
Solution: **De-Indentified**

Mayo Clinic

k health + MAYO CLINIC PLATFORM

They removed all the names, zip, SSN, address, etc.

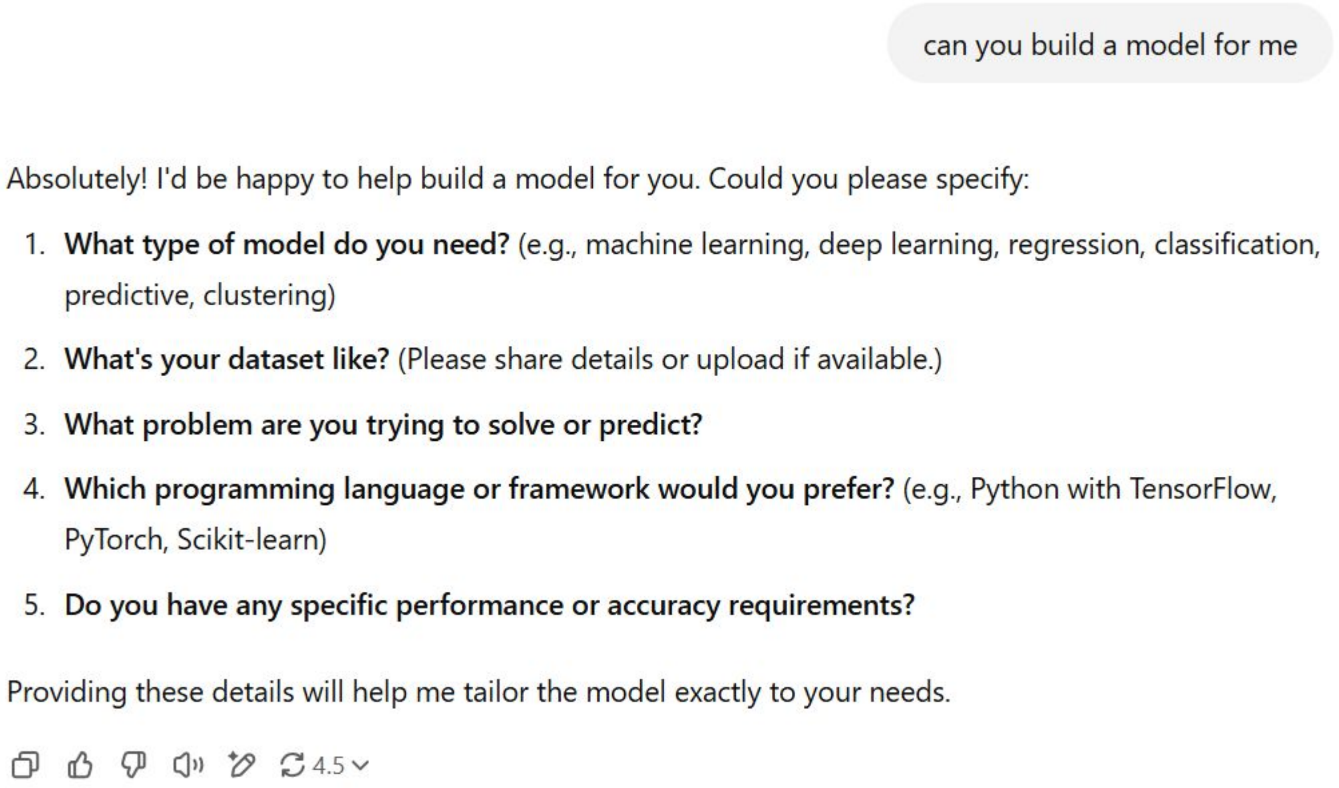
3. Gradient Attack： Gradient Transmission Phase


(a) Ground Truth

(b) Reconstructed image

Differential Privacy： add noises into the gradients

LLMs for model building

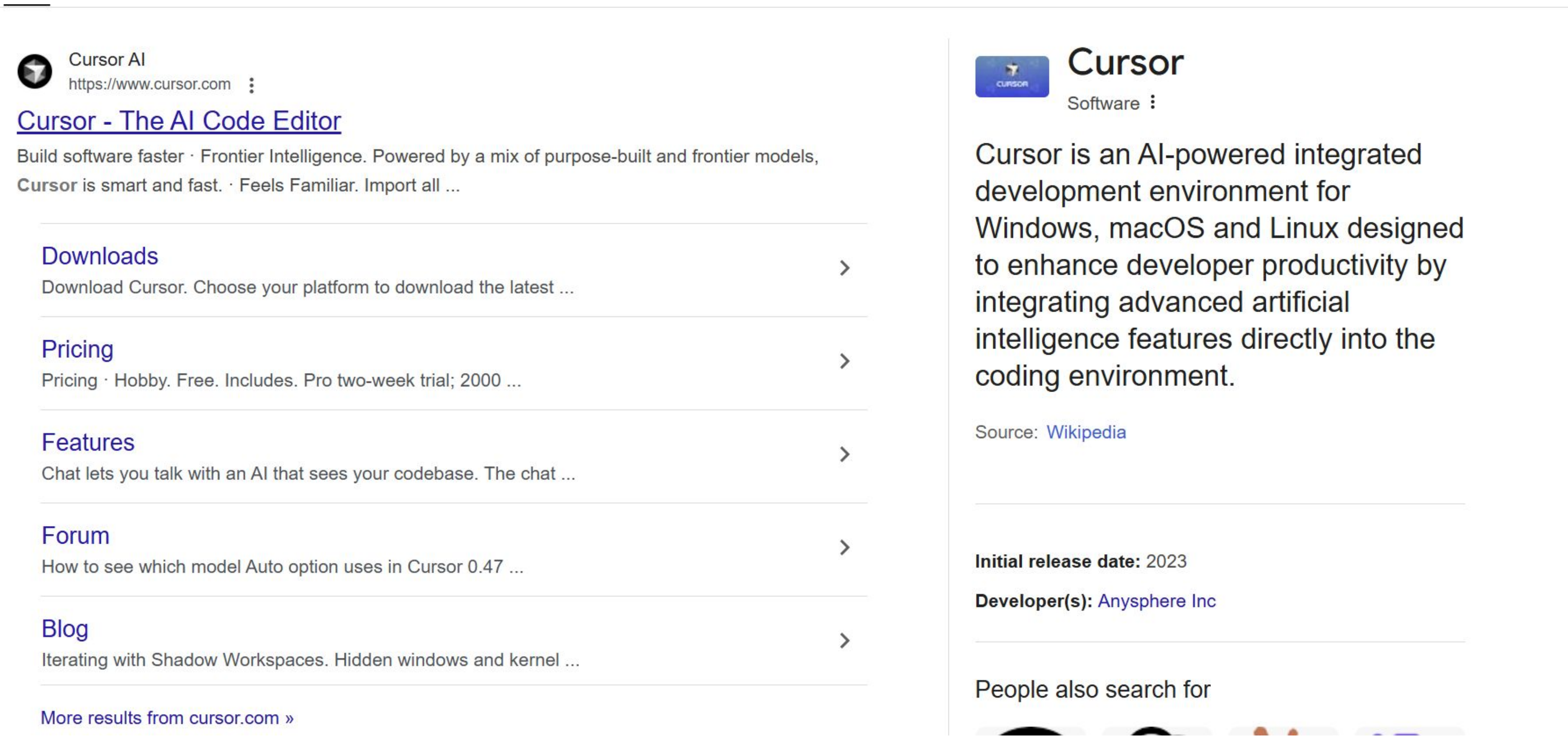


can you build a model for me

Absolutely! I'd be happy to help build a model for you. Could you please specify:

1. **What type of model do you need?** (e.g., machine learning, deep learning, regression, classification, predictive, clustering)
2. **What's your dataset like?** (Please share details or upload if available.)
3. **What problem are you trying to solve or predict?**
4. **Which programming language or framework would you prefer?** (e.g., Python with TensorFlow, PyTorch, Scikit-learn)
5. **Do you have any specific performance or accuracy requirements?**

Providing these details will help me tailor the model exactly to your needs.



Cursor AI
https://www.cursor.com

Cursor - The AI Code Editor

Build software faster · Frontier Intelligence. Powered by a mix of purpose-built and frontier models, **Cursor** is smart and fast. · Feels Familiar. Import all ...

Downloads
Download Cursor. Choose your platform to download the latest ...

Pricing
Pricing · Hobby. Free. Includes. Pro two-week trial; 2000 ...

Features
Chat lets you talk with an AI that sees your codebase. The chat ...

Forum
How to see which model Auto option uses in Cursor 0.47 ...

Blog
Iterating with Shadow Workspaces. Hidden windows and kernel ...

More results from cursor.com »

Cursor
Software

Cursor is an AI-powered integrated development environment for Windows, macOS and Linux designed to enhance developer productivity by integrating advanced artificial intelligence features directly into the coding environment.

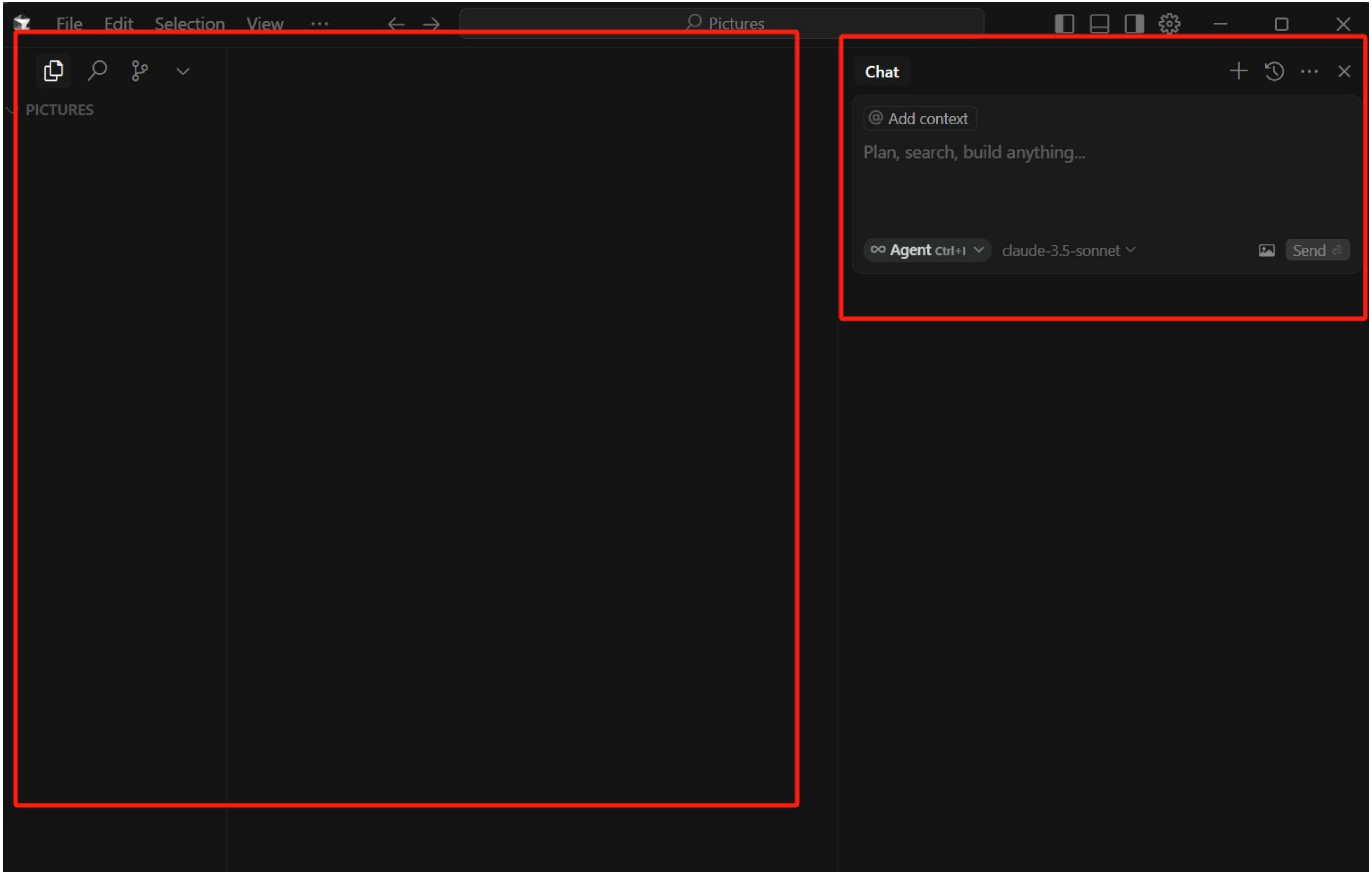
Source: Wikipedia

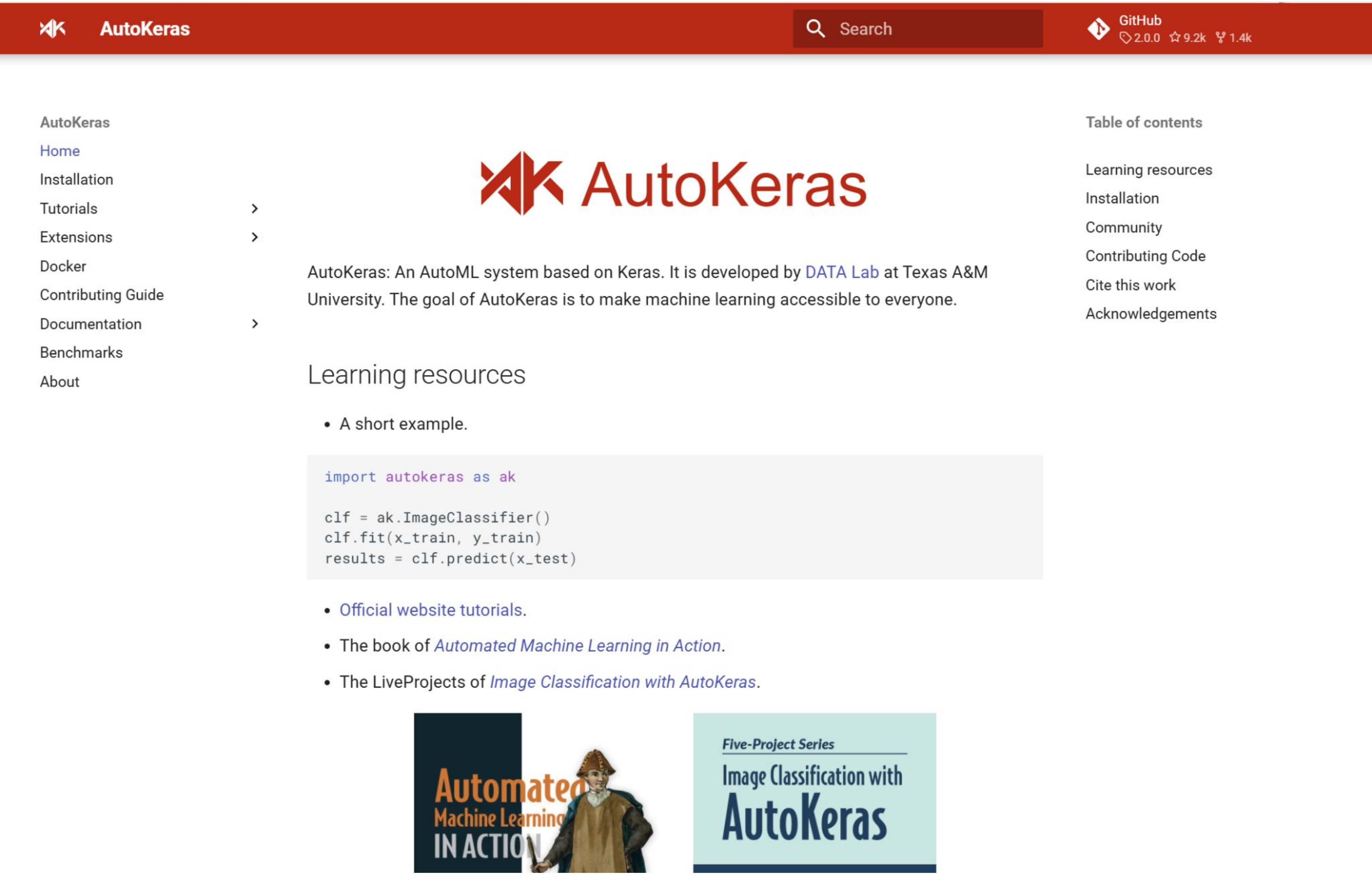
Initial release date: 2023

Developer(s): Anysphere Inc

People also search for

Cursor - can write papers, latex





AutoKeras

Search

GitHub
2.0.0 ☆ 9.2k 1.4k

AutoKeras

- Home
- Installation
- Tutorials ›
- Extensions ›
- Docker
- Contributing Guide
- Documentation ›
- Benchmarks
- About

Table of contents

- Learning resources
- Installation
- Community
- Contributing Code
- Cite this work
- Acknowledgements

AutoKeras

AutoKeras: An AutoML system based on Keras. It is developed by DATA Lab at Texas A&M University. The goal of AutoKeras is to make machine learning accessible to everyone.

Learning resources

- A short example.

```
import autokeras as ak

clf = ak.ImageClassifier()
clf.fit(x_train, y_train)
results = clf.predict(x_test)
```

- Official website tutorials.
- The book of *Automated Machine Learning in Action*.
- The LiveProjects of *Image Classification with AutoKeras*.

Terms:

Model: A program trained to recognize patterns in data.

Training: The process of teaching the model using data.

Feature: Input variables used by the model to learn.

Accuracy: How often the model predicts correctly.

Thanks