



# Using AI Tools Like ChatGPT to Write Code and Do Mathematics

Chris Deotte, Ph.D.  
Senior Data Scientist – NVIDIA  
7x Kaggle Grandmaster

Mar 3, 2026 - CCoM Talk

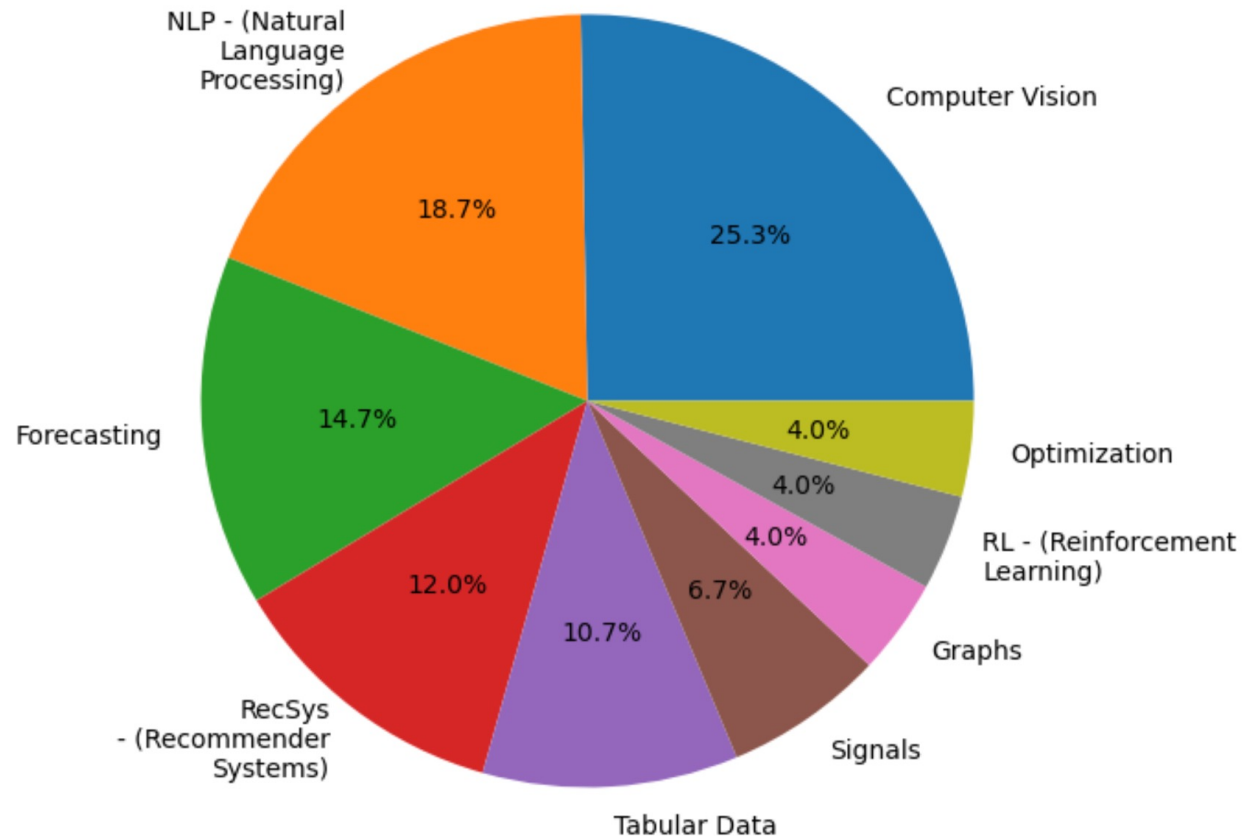
# Kaggle Grandmasters

6 out of top 8 Kaggle Grandmasters are KGMON (Kaggle Grandmasters of Nvidia)

The screenshot shows the Kaggle website's 'Rankings | Kaggle' page. The browser address bar shows 'https://www.kaggle.com/rankings/grandmasters'. The page title is 'Kaggle Rankings'. Below the title, a description states: 'This live leaderboard shows the absolute best data scientists on Kaggle. Each category on Kaggle has its own leaderboard and point system. A Kagglers must be an Expert tier or higher to be ranked for that category. Learn more'. Navigation tabs include 'Competition Rankings', 'Dataset Rankings', 'Code Rankings', 'Grandmasters', and 'Awards'. The 'Grandmasters' section is active, showing a search bar and a grid of 10 profiles. Each profile includes a circular profile picture, a name, and a badge indicating the number of times they have been a Kaggle Grandmaster (e.g., '13x', '8x', '7x', '6x', '5x'). The profiles are: Giba (13x), bestfitting (8x), Ahmet Erdem (7x), Chris Deotte (7x), CPMP (6x), Μαριος Μιχαηλιδης Kazanova (6x), Dieter (6x), Theo Viel (6x), Stanislav Semenov (5x), and Guanshuo Xu (5x). The 'KGMON' label is placed above the profile pictures of Giba, Ahmet Erdem, Chris Deotte, CPMP, Dieter, and Theo Viel. The 'former KGMON' label is placed above Ahmet Erdem's profile. The 'Level' dropdown menu is set to 'Level'. The left sidebar contains navigation options: Home, Competitions, Datasets, Models, Benchmarks, Game Arena, Code, Discussions, Learn, and More. The bottom left corner has a 'View Active Events' button.

# 100 Data Science Competitions in 6 years!

Compete against 1000s of participants to build the most accurate model to solve tasks.



## Example Competitions

Real companies pay money to participants who develop the best solutions



Computer Vision:  
Detect and  
Locate helmet  
collisions



Recommender Systems:  
Predict what a customer  
will purchase



NLP (text):  
Evaluate student writing

Tabular Data:  
Predict credit  
default



Signals:  
Classify brain waves into  
sleep states



Forecasting:  
Predict future sales.



## Demo

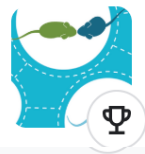
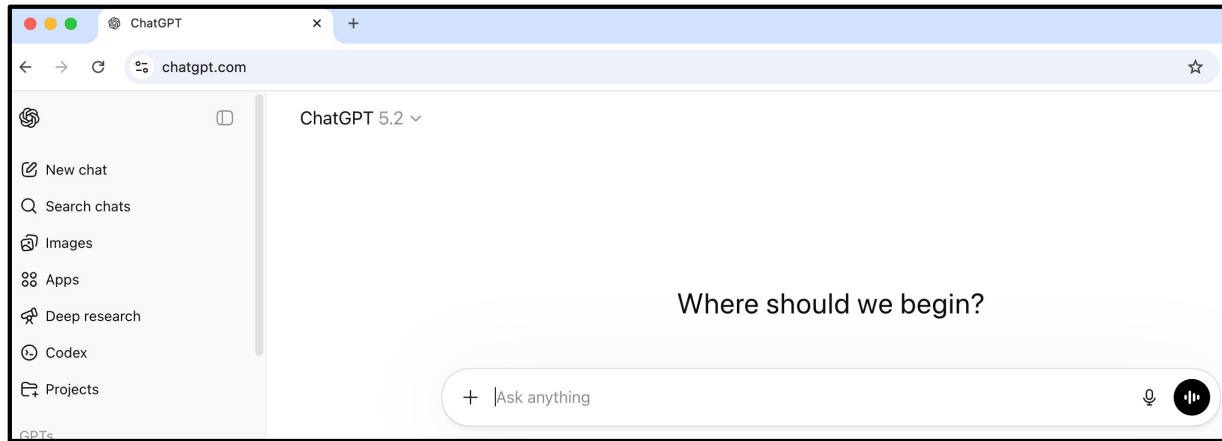
- ChatGPT
- Write Code
- Search Web
- Read Code
- Read Research
- Write Research

## Case Study

- Mouse Behavior Classification
- Experiments
- Win Gold Medal

# ChatGPT Web Interface

ChatGPT can win data science competitions!



## MABe Challenge - Social Action Recognition in Mice

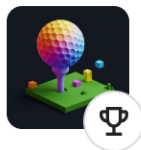
Detect unique behaviors from pose estimates of mice.

Research · Code Competition · 1412 Teams · 2 months ago

7/1412 🏆



ChatGPT wrote 99% of my code and won Gold and Silver medals!



## NeurIPS 2025 - Google Code Golf Championship

Implement a variety of programs using the fewest number of characters!

Research · 1142 Teams · 3 months ago

18/1142 🥈





# ChatGPT Demo - UC Irvine Cleveland Heart Disease Dataset

## UCI Cleveland Heart Disease Dataset

Age

Gender

Chest pain type

Blood pressure

Cholesterol

Fasting blood sugar > 120

EKG Results

Max heart rate

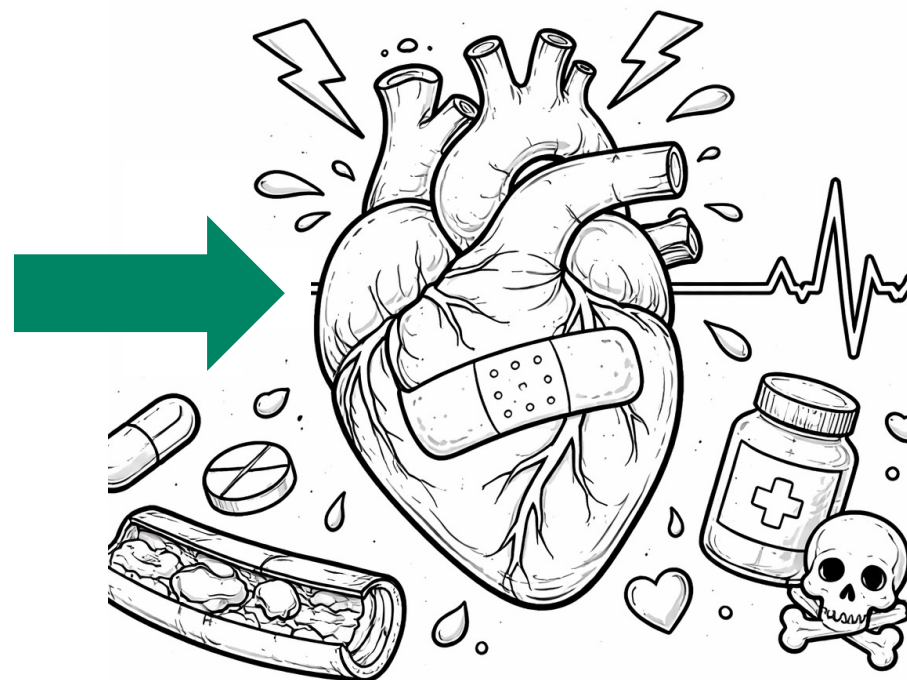
Exercise angina

ST depression

Slope of ST

Number of vessels fluro

Thallium



**Heart Disease?**

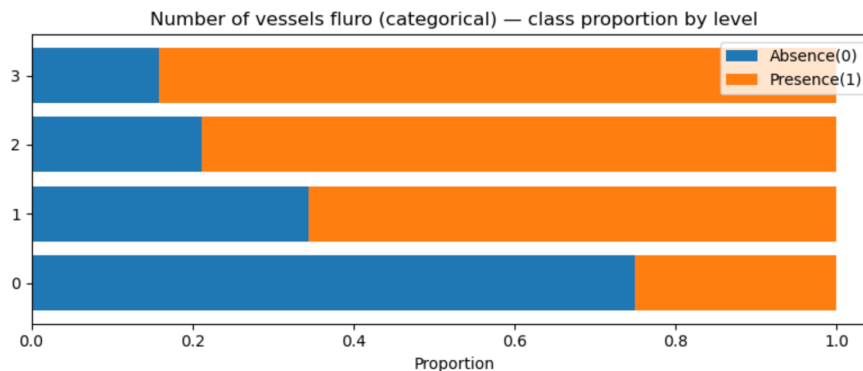
<https://archive.ics.uci.edu/dataset/45/heart+disease>

# ChatGPT Writes Code

Explain clearly what you want

## Prompt:

```
Please write code to read the CSV
"Heart_Disease_Prediction.csv". The target
columns is "Heart Disease" with labels
"Absence" and "Presence", the other columns are
features. Treat all features with less than 10
unique values as categorical. First plot EDA of
each feature. Next train a logistic regression
model to predict target. Use 5Kfold and display
AUC and ACC metric each fold and overall.
Afterward, display the logistic regression
equation explicitly and make a plot of most
important feature importance.
```



```
--- Logistic Regression Formula
(EXPLICIT) ---
Target: p = P(Heart Disease = Presence)
p = 1 / (1 + exp(-logit(p)))

--- Top 20 terms by |coefficient| ---
logit(p) ≈ -0.077246
-1.496077 * Number of vessels fluro_0
+1.146542 * Chest pain type_4
+0.837993 * Thallium_7
-0.830627 * Chest pain type_1
+0.828624 * Number of vessels fluro_2
-0.757889 * Sex_0
+0.680643 * Sex_1
-0.520431 * Thallium_3
-0.514537 * Chest pain type_3
+0.499815 * ST depression
+0.473501 * Slope of ST_2
-0.429996 * Max HR
+0.422904 * BP
-0.394808 * Thallium_6
-0.368591 * Slope of ST_1
-0.332160 * EKG results_0
+0.306903 * Number of vessels fluro_3
-0.299585 * Exercise angina_0
```

## ChatGPT Searches Web

### Prompt:

```
Please read all the discussion posts in Kaggle's Heart Disease competition forum. Summarize what the top models and strategies are to create the most accurate models. The forum URL is https://www.kaggle.com/competitions/playground-series-s6e2/discussion
```

### Prompt:

```
Please write complete code to read dataset and train an XGBoost model on our 5Kfold. Display ACC and AUC metrics.
```

Here are the top 10 concise takeaways from Kaggle forum discussions for the Heart Disease Playground S6E2 competition:

- Gradient boosting models dominate — LightGBM, CatBoost, and XGBoost are the strongest single models.
- CatBoost often performs best due to many low-cardinality categorical features.
- Logistic regression is a very strong baseline ( $\approx 0.93$ – $0.94$  AUC) because relationships are mostly linear.
- Treat features with  $<10$  unique values as categorical — this is one of the most impactful preprocessing steps.
- Minimal feature engineering is needed since the synthetic dataset is clean and low-noise.
- Simple interaction features (e.g., Age $\times$ Cholesterol, risk counts) provide small but consistent gains.
- Stacking works best — boosting models at level-1 with logistic regression as meta-model.
- Weighted averaging of multiple boosting models is commonly used by top leaderboard solutions.
- Neural networks and deep learning do not outperform tree models on this dataset.
- OOF cross-validation blending is critical to avoid overfitting and achieve top scores ( $\sim 0.96$  AUC).

# ChatGPT Reads Code

### Prompt:

```
Read the following IPYNB file.  
Reproduce the XGB model with the same  
feature engineering and model  
parameters on our KFold scheme.  
Display the AUC and ACC metrics.
```

### Prompt:

```
Read the following IPYNB file. What  
are the top 5 strongest feature  
engineering ideas for their XGB  
model?
```

## Predicting Heart Disease

Overview Data **Code** Models Discussion Leaderboard

Search notebooks

All Your Work Shared With You Bookmarks



### Heart|XGB/LightGBM/CatB|Baseline|K-Fold

Updated 7d ago  
Score: 0.9539 · [6 comments](#) · Predicting Heart Disease +1



### The best solo model so far | RealMLP | LB: 0.95397

Notebook copied with edits from a private notebook · Updated 2d ago  
Score: 0.95397 · [12 comments](#) · Predicting Heart Disease +1



### Logistic Regression | PS: S6E2

Notebook copied with edits from a private notebook · Updated 7d ago  
Score: 0.95371 · [20 comments](#) · Predicting Heart Disease

## ChatGPT Reads Research

### Prompt:

Find research papers on best way to model the Cleveland heart disease dataset. Please give me a list with links to download the papers.

### Prompt:

Read this paper and summarize ideas that can make our models more accurate

Here are 5 key papers (concise list) — just title + direct PDF/download link:

- Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning <https://pmc.ncbi.nlm.nih.gov/articles/PMC8266441/pdf/main.pdf>
- Prediction of Heart Disease Based on Machine Learning Techniques <https://pmc.ncbi.nlm.nih.gov/articles/PMC10378171/pdf/healthcare-11-01894.pdf>
- A Proposed Technique for Predicting Heart Disease Using Feature Selection <https://pmc.ncbi.nlm.nih.gov/articles/PMC11458608/pdf/s41598-024-69014-9.pdf>
- Early Prediction of Heart Disease Using Feature Engineering and Machine Learning <https://pmc.ncbi.nlm.nih.gov/articles/PMC11471268/pdf/healthcare-12-02023.pdf>
- Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm <https://arxiv.org/pdf/2101.00183.pdf>

# ChatGPT Writes Research

## Prompt:

Please read our IPYNB experiments and write an outline of a latex paper. Thank you

## 🍓 Heart Disease Prediction Pipeline

Feature Engineering + Robust ML Modeling

### 🔗 Problem Setup

We model heart disease prediction as a binary classification task:

$$y \in \{0, 1\}$$

where:

- $y = 1 \rightarrow$  presence of heart disease
- $y = 0 \rightarrow$  no disease

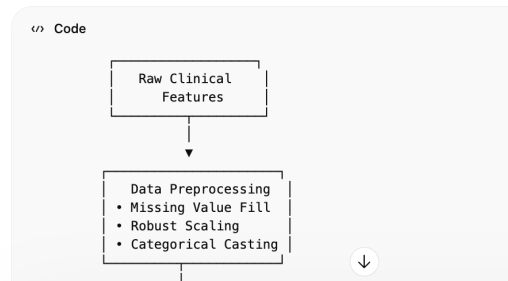
Goal:

$$\hat{y} = f(X)$$

Maximize predictive performance:

$$AUC = P(\hat{y}_{positive} > \hat{y}_{negative})$$

### ⚙️ Modeling Pipeline



## 🧠 Logistic Regression Model

Probability prediction:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}}$$

Loss function:

$$\mathcal{L} = - \sum_{i=1}^n [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

## 🌲 Gradient Boosting Model

Ensemble prediction:

$$\hat{y} = \sum_{m=1}^M \gamma_m h_m(x)$$

where:

- $h_m$  = decision tree
- $\gamma_m$  = learned weights

Optimization via gradient descent on loss.

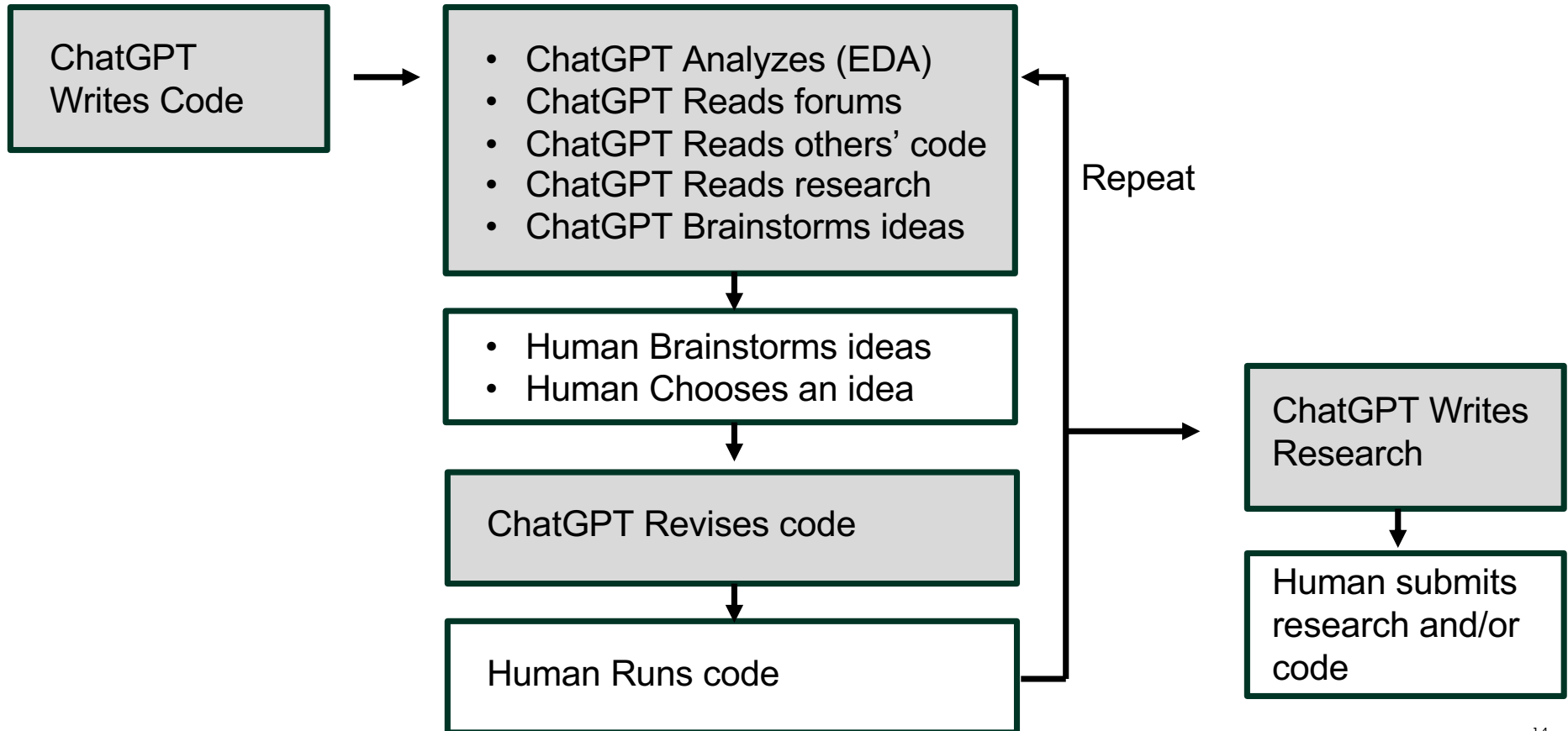
## 📊 Feature Engineering Strategy

Frequency Encoding

$$FE(x) = \frac{\text{count}(x)}{N}$$

Captures category prevalence.

## Summary ChatGPT Demo



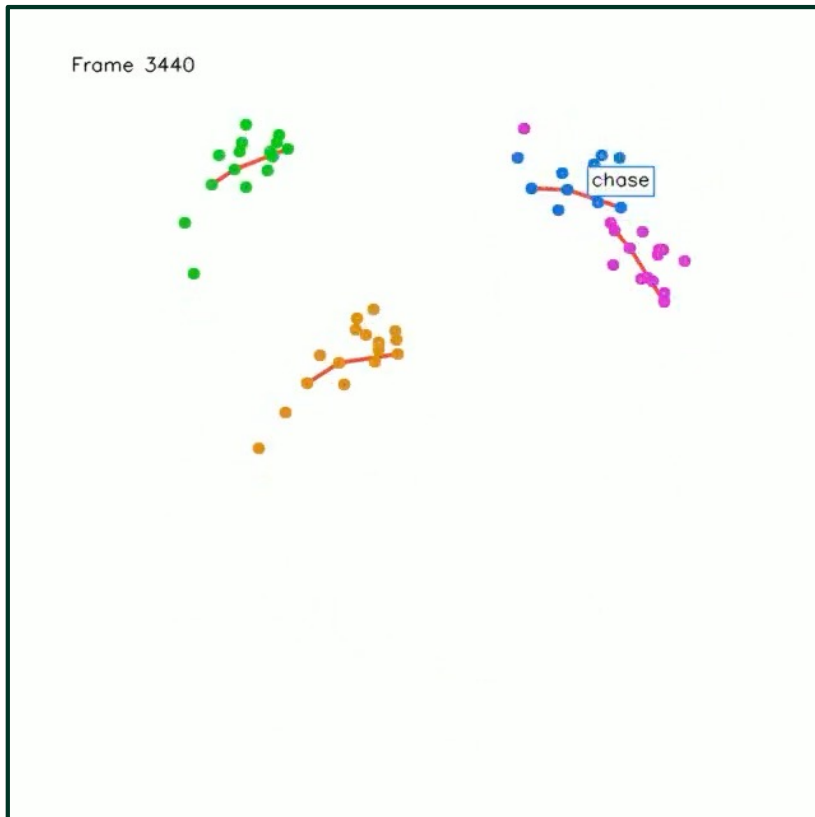


Kaggle Competition  
Classify Mouse Behavior  
7<sup>th</sup> Place Gold by ChatGPT!



# Kaggle Competition - Classify Mouse Behavior

Classify mouse behavior into 37 classes



- **Classify** Mouse behavior into 37 actions
- **Example Prediction:**

video_id	agent_id	target_id	action	start_frame	stop_frame
438887472	mouse1	mouse2	sniff	100	150

- **Metric** is **F1 Score** (harmonic mean of precision and recall)

- **Precision:**  $\frac{TP}{TP + FP}$  (True Positives / Predicted Positives).

- **Recall:**  $\frac{TP}{TP + FN}$  (True Positives / Actual Positives).

- **Combined Formula:**  $F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$ .

## Mouse Competition - Input Data

### 9.6k Files, 2.8GB of Tabular Data!

**train.csv** columns:

- lab\_id
- video\_id
- pix\_per\_cm\_approx
- frames\_per\_second
- body\_parts\_tracked
- behaviors\_labeled
- 32 more meta data columns

1 file - shape (8789,38)

**"train\_tracking/{lab\_id}/{video\_id}.parquet"**

- video\_frame
- mouse\_id
- bodypart
- x
- y

8789 files

**"train\_annotation/{lab\_id}/{video\_id}.parquet"**

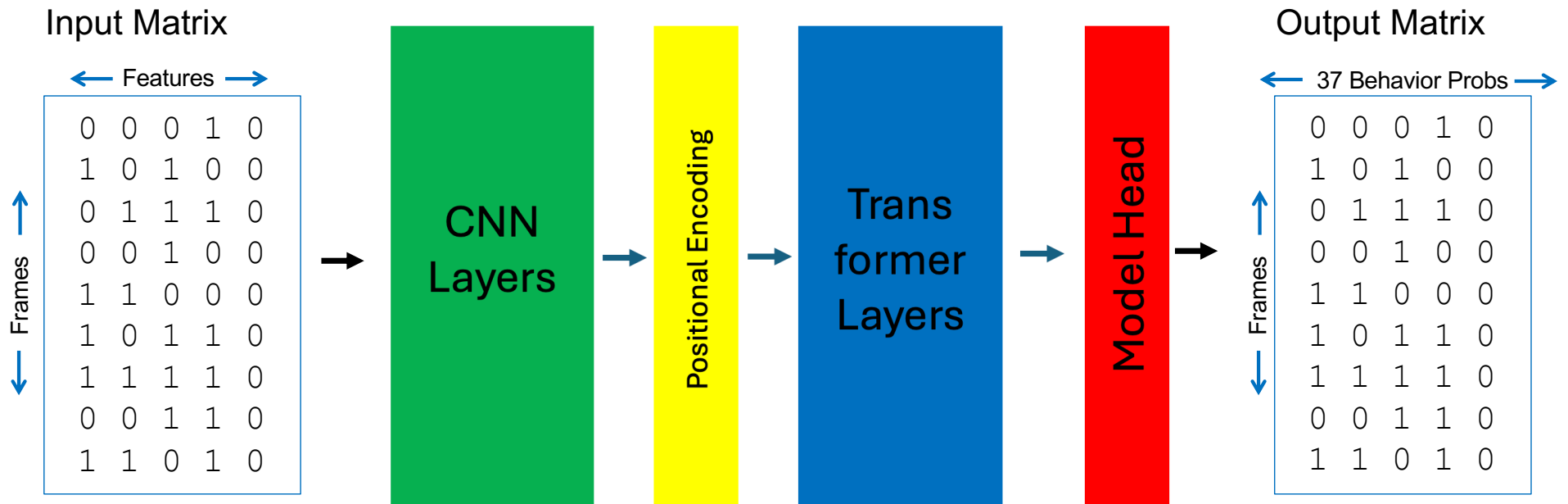
- agent\_id
- target\_id
- action
- start\_frame
- stop\_frame

863 files

# Mouse Competition - ChatGPT Writes Pipelines

## Prompt:

```
Please write code to read data and train a KFold time series deep learning model to classify mouse behavior. The data is saved in files named blah blah and has format blah blah. Predict the target class for each time step blah blah. Compute and display F1 metric. Save the model and OOF predictions to disk.
```

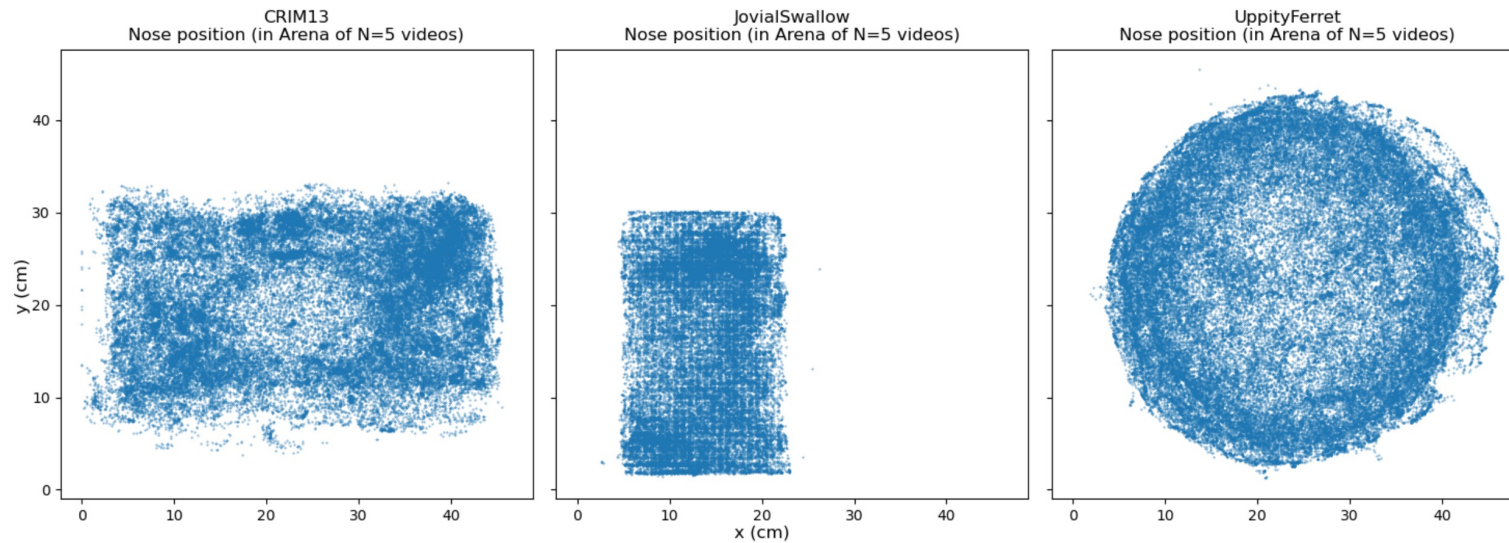


# Mouse Competition - ChatGPT Does EDA

## Prompt:

```
Write code to display EDA. For each lab, load data for 5 videos and display a scatter plot showing the location of mouse's nose for mouse with id 1.
```

## Agent Mouse Nose Location:

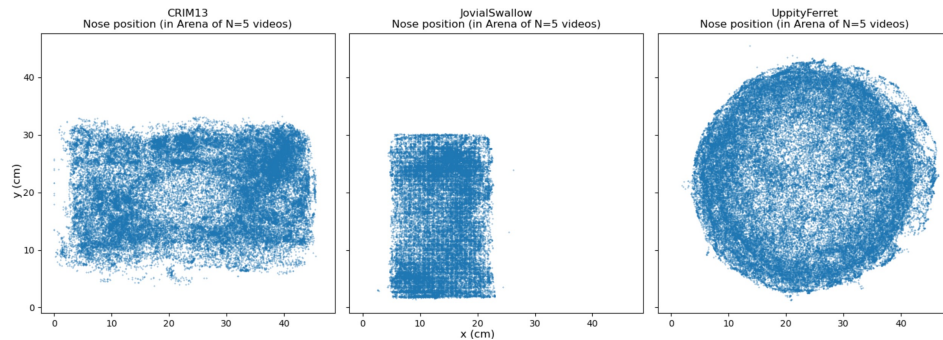


# Mouse Competition - ChatGPT Does Mathematics!

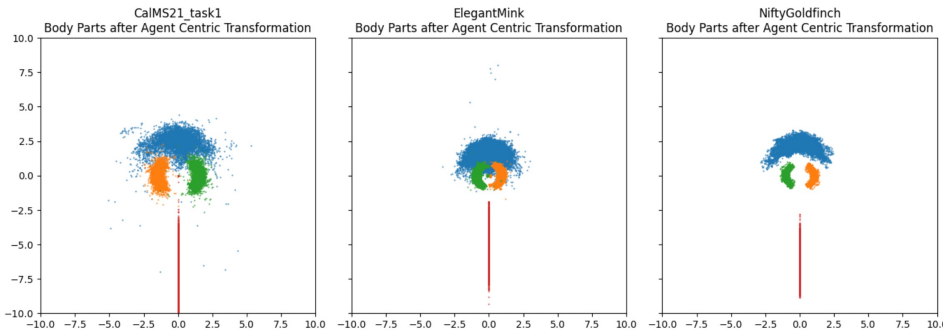
## Prompt:

Write code to read all the position data files one by one. For each one, transform the  $x, y$  coordinates in each frame so that the agent mouse's nose is at  $0, 0$  and agent's tail base points downward in direction of negative  $y$  axis. Use this transform on the target mouse too. Afterward write the result to a new file named old filename plus "transformed".

Old Agent  
Nose  
Location:



New Agent  
Nose  
Location:  
(in blue)



Transform Equation:

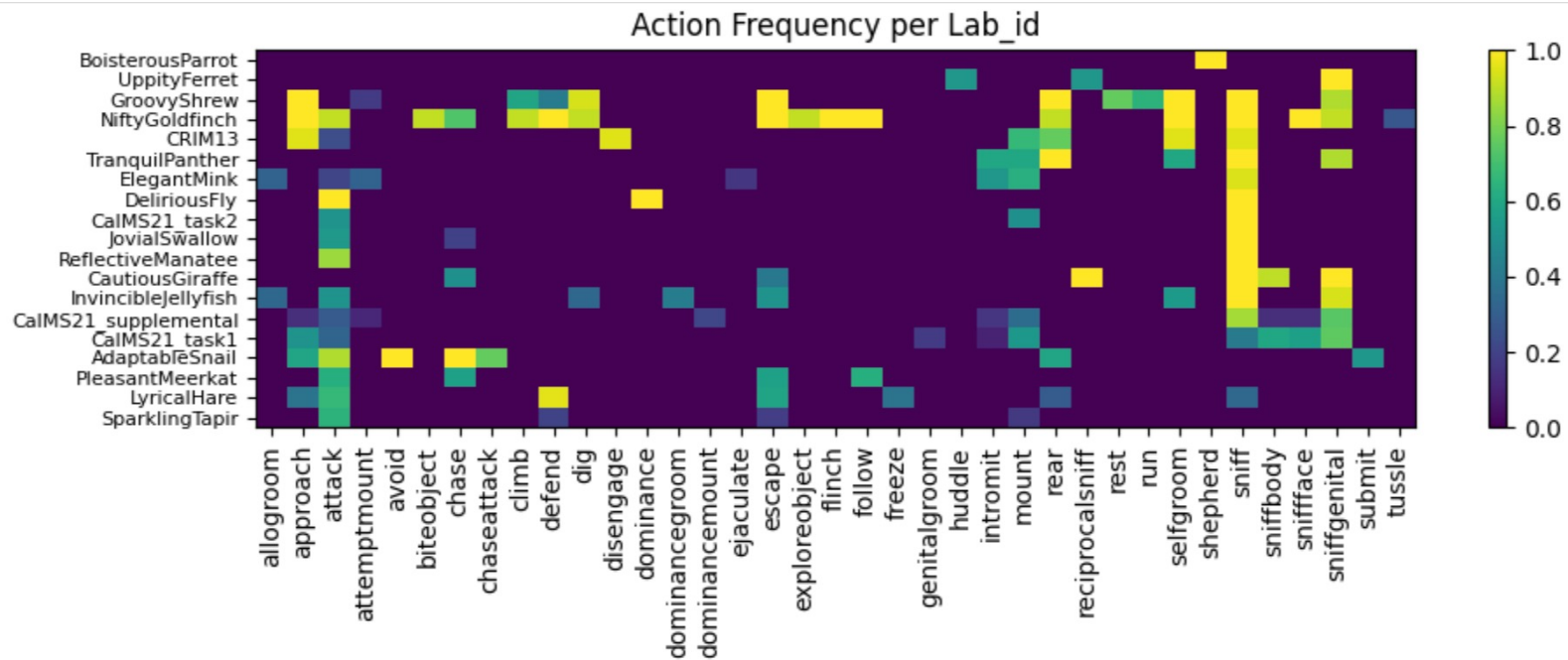
$$\mathbf{p}' = \mathbf{R}\left(-\frac{\pi}{2} - \text{atan2}((\mathbf{t} - \mathbf{n})_y, (\mathbf{t} - \mathbf{n})_x)\right) (\mathbf{p} - \mathbf{n})$$

where  $\mathbf{R}(\alpha) = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$

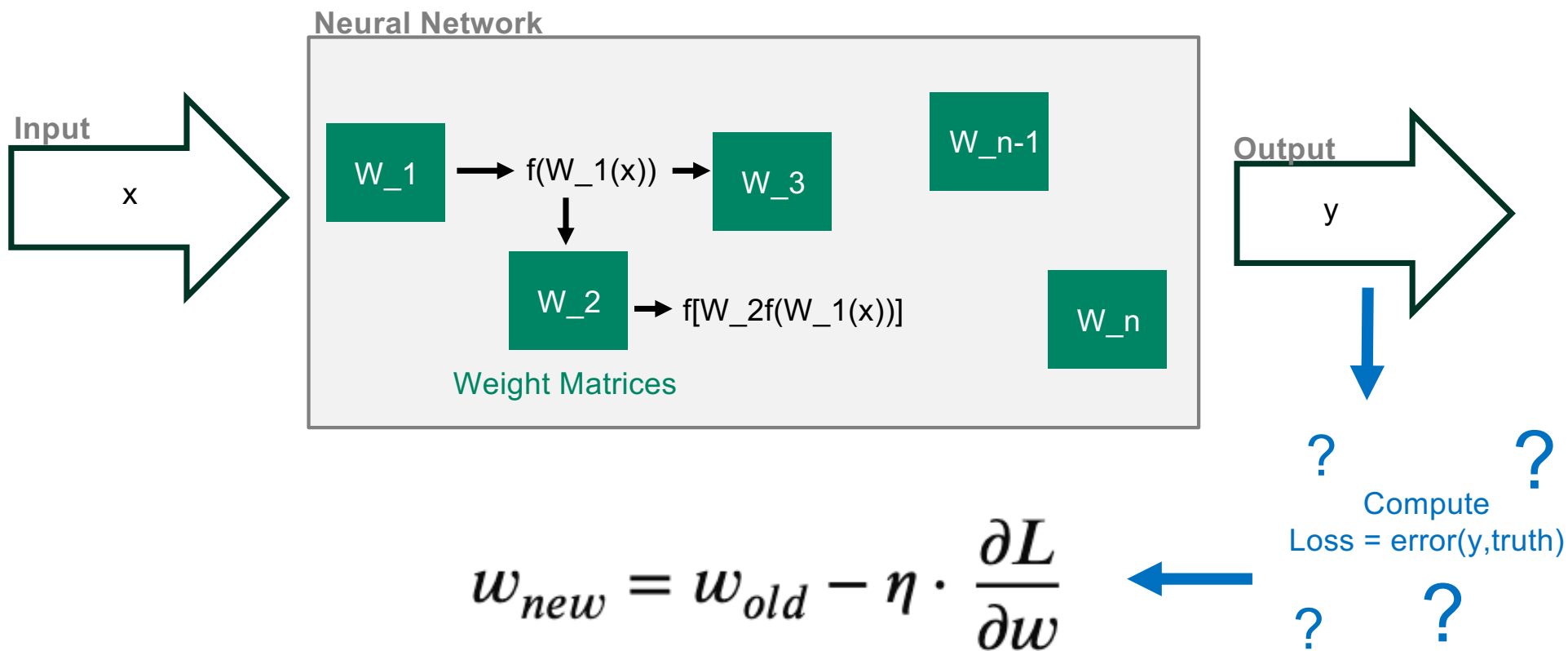
# Mouse Competition - ChatGPT Does EDA

## Prompt:

```
Write code to display EDA. For each lab, load all the data and determine which behavior targets are present. Then create a heatmap with labs and behaviors.
```



## Mouse Competition - Backpropagation



# Mouse Competition - ChatGPT Does Mathematics!

## Prompt:

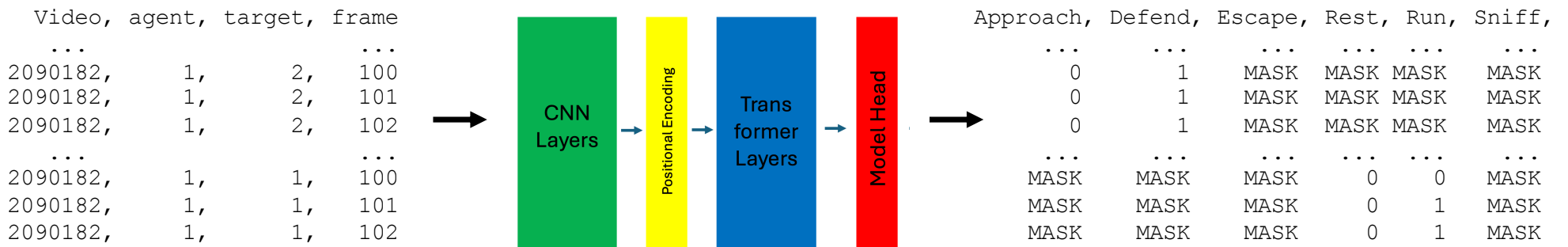
Write code to compute and back propagate masked BCE loss. For each input time step, create 37 targets defined to be 0, 1, or masked.

Let  $K = 37$ . For example  $i$ :

- $y_{ik} \in \{0, 1\}$  is the target
- $m_{ik} \in \{0, 1\}$  is the mask (1 = observed, 0 = ignored)
- $z_{ik}$  is the logit
- $\sigma(u) = \frac{1}{1+e^{-u}}$

Masked BCE loss (per example, normalized by observed labels)

$$\mathcal{L}_i = \frac{1}{\sum_{k=1}^K m_{ik}} \sum_{k=1}^K m_{ik} [-y_{ik} \log \sigma(z_{ik}) - (1 - y_{ik}) \log (1 - \sigma(z_{ik}))]$$

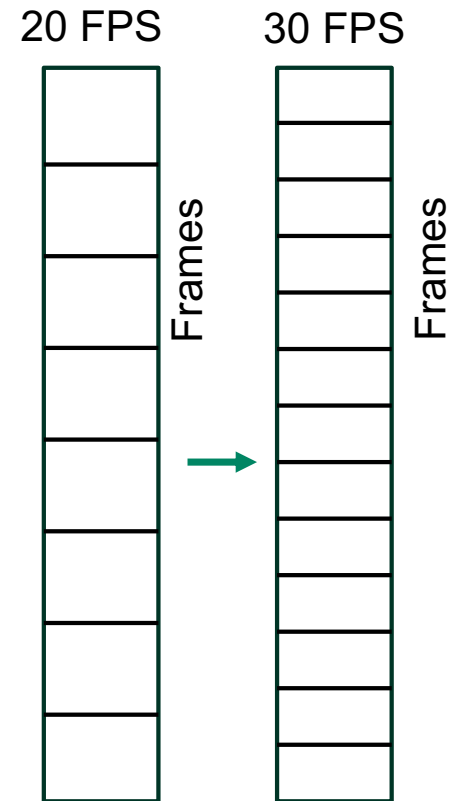
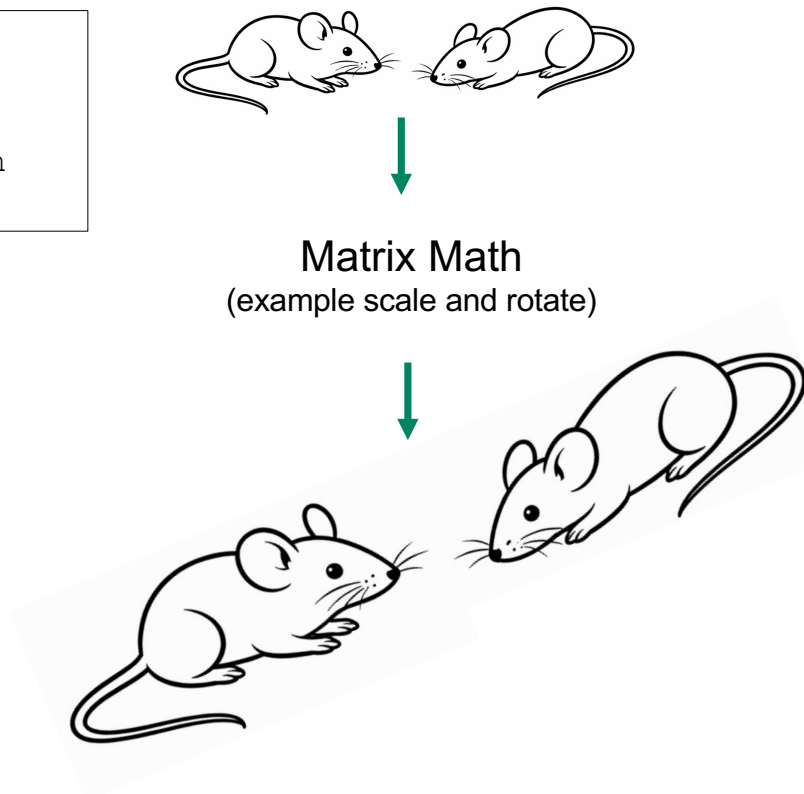


# Mouse Competition - ChatGPT Does Mathematics!

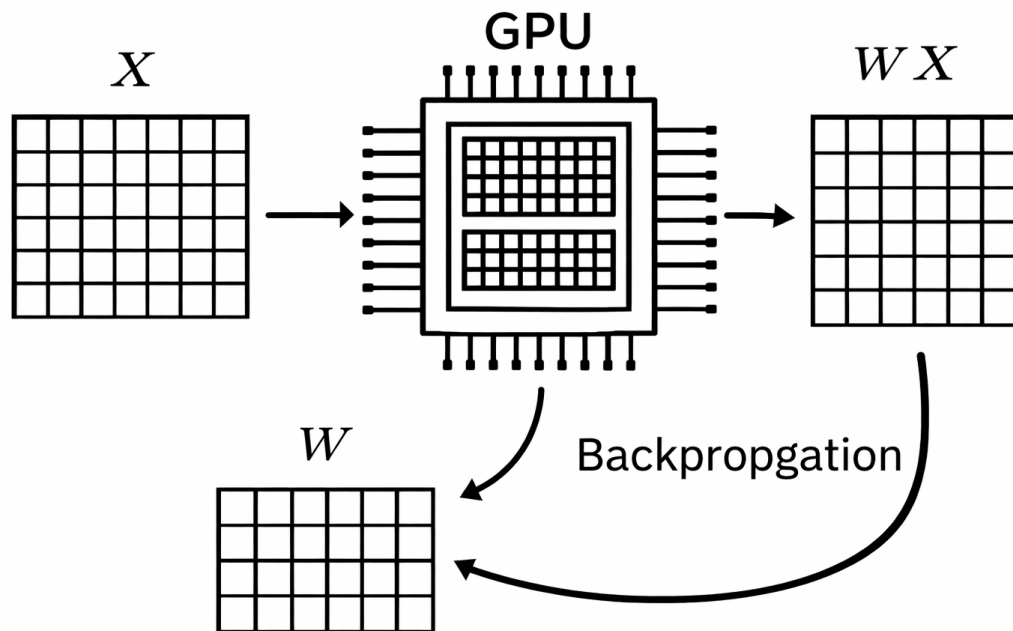
## Prompt:

```
Write code to apply data augmentation and randomly modify the data during training. Please apply blah blah.
```

- time augmentation - (change fps)
- rotation augmentation
- x scale augmentation
- y scale augmentation
- horizontal flip
- relabel right with left body parts
- body part drop out



## Mouse Competition – GPUs Run Code!



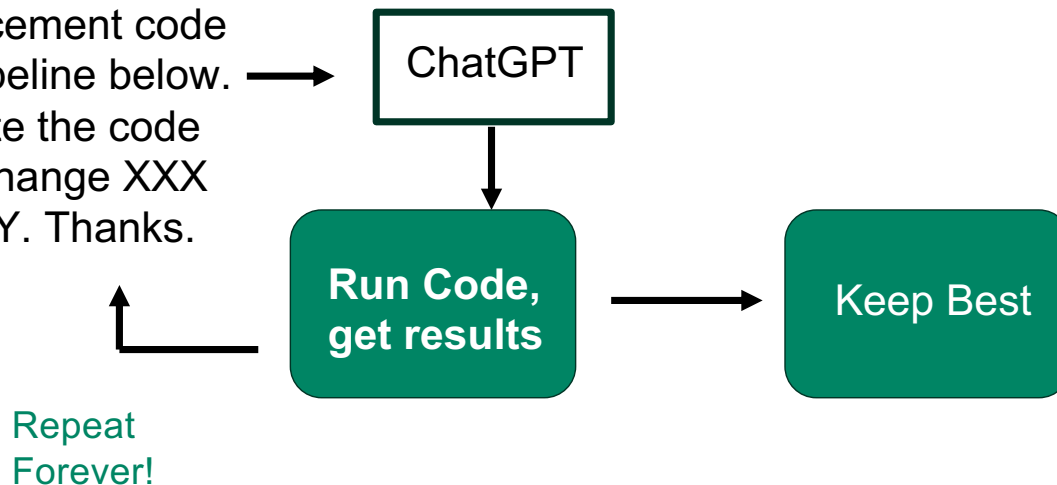
- Run training code on A100 GPU 80GB VRAM
- 2-3 hours to train 20 epochs
- Afterward, displays local validation F1 metric score
- 20 to 600 TFLOPS (trillion floating point operations per second!)

## Mouse Competition - ChatGPT Does Experiments!

### Prompt Experiment 1:

Please update the code below. Instead of inputting  $x, y$  coordinates into the model, use distances instead of absolute positions. Each mouse has 6 body parts, provide the model with all distance pairs from the combined 12 body parts of agent and target mouse. Also add some key angles.

- Please give me complete drop in replacement code for pipeline below. Update the code and change XXX to YYY. Thanks.

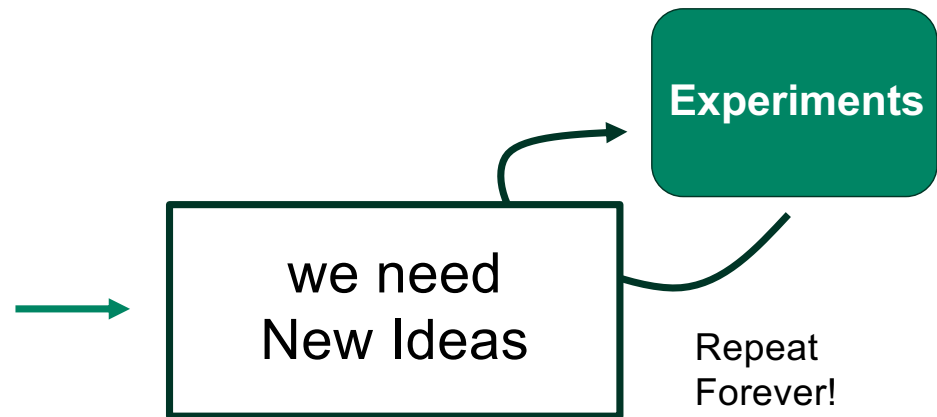


### Prompt Experiment 2:

Please update the code below. Instead of inputting  $x, y$  coordinates into the model, use min max scaling. For each  $x$ , subtract min  $x$  for given video and divide by max minus min from that video.

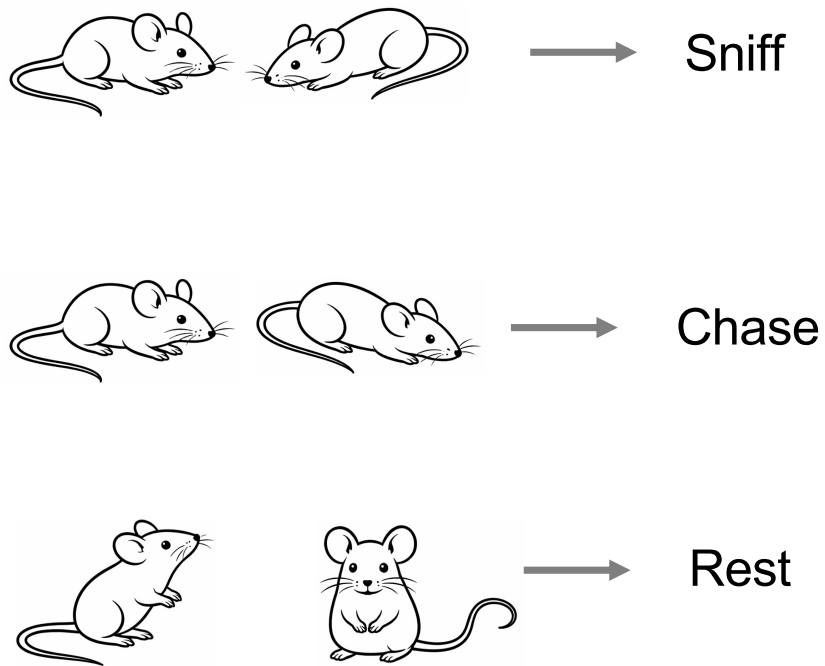
## Mouse Competition - ChatGPT Generates Ideas

- We need new diverse ideas to repeatedly run experiments.
- Ask ChatGPT for ideas
- Upload research papers (PDF) to ChatGPT to extract ideas
- Upload Kaggle public notebooks (IPYNB) to ChatGPT to extract ideas
- Show ChatGPT EDA (JPEG) and ask for ideas based on EDA
- Mention buzz words and ask for ideas.
- Brainstorm our own ideas



## Mouse Competition - ChatGPT Wins Gold Medal

7<sup>th</sup> Place out of 1500!



**F1 Score = 0.52311**



