GPT wasn't trained for this (quickly hit the API's token limit), it was guessing purely off of what it is by default

The whole dataset was fed into GPT to guess if it was AI or human, but logistic_regression.py and svm.py had an 80/20 split between training and testing data.

logistic_regression.py and svm.py were able to be trained on the sets individually as well as combined.

All numbers are rounded to one decimal place.

POETRY

COST FOR THIS:

gpt-3.5-turbo-0301: $0.37
gpt-4-turbo-2024-04-09: $1.76

gpt-3.5-turbo-0301:

Total correct: 198 / 500
Accuracy: 39.6%
Guessed AI 12% of the time, human 88% of the time

gpt-4-turbo-2024-04-09:

Total correct: 427 / 500
Accuracy: 85.4%
Guessed AI 38.2% of the time, human 61.8% of the time

logistic_regression.py (avg of 100 trials):

Accuracy: 89.1%
Guessed AI 50.2% of the time, human 49.8% of the time

svm.py (avg of 100 trials):

Accuracy: 98.2%l
Guessed AI 50.1% of the time, human 49.9% of the time

ESSAYS

COST:

gpt-3.5-turbo-0301:	$0.18
gpt-4-turbo-2024-04-09:	$2.53

gpt-3.5-turbo-0301:

Total correct:	161/400
Accuracy:	40.3%

Guessed AI 17.5% of the time, human 82.5% of the time

gpt-4-turbo-2024-04-09:

Total correct:	377/400
Accuracy:	94.3%

Guessed AI 47.5% of the time, human 52.5% of the time

logistic_regression.py (avg of 100 trials):

Accuracy:	96.2%

Guessed AI 51.3% of the time, human 48.7% of the time

svm.py (avg of 100 trials):

Accuracy:	99.2%

Guessed AI 50.0468% of the time, human 49.9532% of the time

Only one not rounded to 1 decimal place because Wow that's close to perfect

WIKI INTROS

COST:

gpt-3.5-turbo-0301:	$0.28
gpt-4-turbo-2024-04-09:	$1.85

gpt-3.5-turbo-0301:

Total correct:	339/600
Accuracy:	56.5%

Guessed AI 27.7% of the time, human 72.3% of the time

gpt-4-turbo-2024-04-09:

Total correct:	254/600
Accuracy:	42.3%

Guessed AI 37.2% of the time, human 62.8% of the time

logistic_regression.py (avg of 100 trials):

Accuracy:	65.7%

Guessed AI 46.6% of the time, human 53.4% of the time

svm.py (avg of 100 trials):

Accuracy:	73.8%

Guessed AI 49.5% of the time, human 50.5% of the time

COMBINED DATA

logistic_regression.py (avg of 100 trials):
Accuracy: 74.7%
Guessed AI 57.2% of the time, human 42.8% of the time

svm.py (avg of 100 trials):
Accuracy: 82.6%
Guessed AI 51.7% of the time, human 48.3% of the time