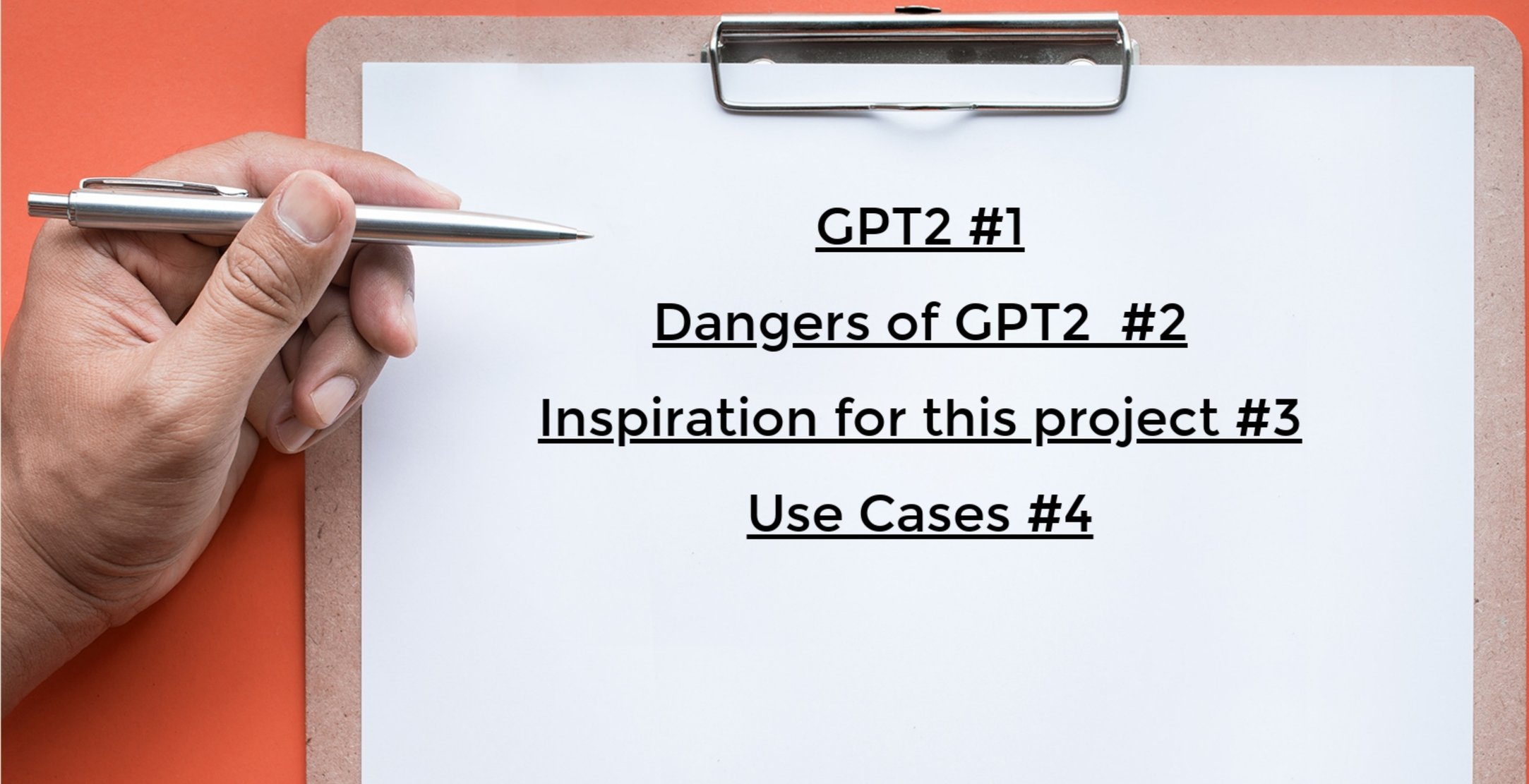
AI Written Text API

by Spencer Holley

what is it?

- AI Written Text API is a web app that estimates the probability that a ~2,000 word essay was written by an AI

what we'll cover



GPT2 #1

Dangers of GPT2 #2

Inspiration for this project #3

Use Cases #4

GPT2

GPT2 is a Language generator that was released by OpenAI in 2018.

- uses an encoder/decoder architecture that outperforms the recurrent models traditionally used for NLP
- can generate any kind of text when finetuned on texts of similar length
- GPT2 was slow to release it due to safety concerns

Dangers
of GPT2

This technology is amazing, but in the wrong hands it can be dangerous!

- Anyone can publish massive amounts fake news and articles on the internet.
- GPT2 generated text can be very convincing, despite holding no truth.
- Students can cheat on their essays without plagiarizing, therefore unlikely to get caught.

Inspiration

I was interested in creating something that could help Wikipedia users and contributors flag AI written articles.

- Wikipedia articles are very easy to create
- they are often the first thing that come up in a google search

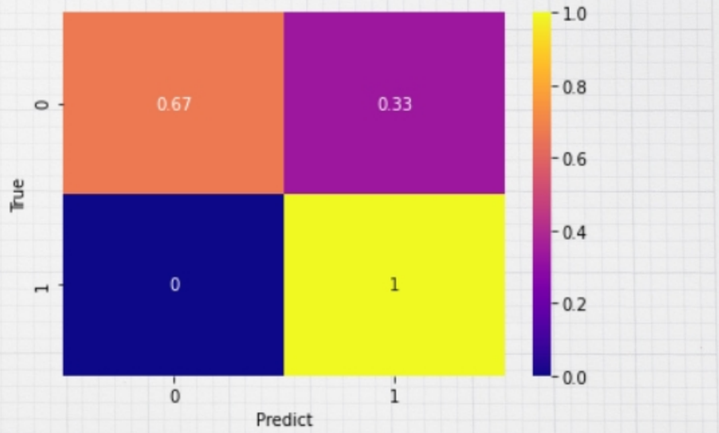
How I did it...

Generation

1. I retrained GPT2 on 1,000 Wikipedia articles on random topics
2. The tuned model generated 180 articles

Classification

1. I put the generated articles in a dataset with 180 randomly selected real articles into a dataset
2. I trained another model to classify the articles appropriatley



other use cases

Education

Teachers and Professors can flag AI written Essays

News

News companies can make sure that all the stories they get were written by their reporters

Future Works

A great continuation of this project would be to do it bigger

1. generate more articles and create a bigger dataset overall.
2. Use Google Cloud Platform or AWS to help with the extra computation.


THANK YOU