AI-Generated Review Detection Using Transfer Learning

A Machine Learning Product

by Sam Middleton

Business Case

Consumer trust is of the utmost importance for online marketplaces and AI-Generated reviews undermines this trust.

We seek to affirm consumer trust be using Machine Learning to find these reviews.

The Process

01

Data
51 million book reviews from Amazon

02

Review Generation Model
Using book reviews to tune a network that generates reviews to train on.

03

AI Review Classification Model
Using real reviews and reviews generated from our first model to tune a classifier to detect the AI-Generated reviews.

04

Testing
Testing the second model on a selection of real and fake reviews that it has not seen.

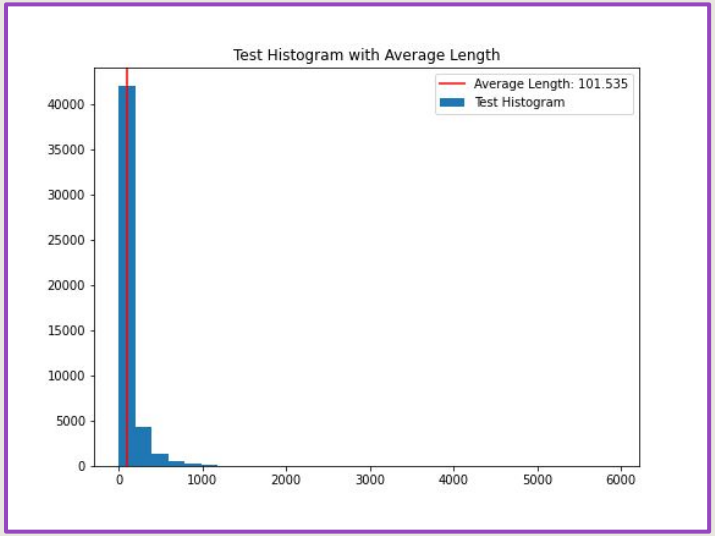
Data

Amazon Reviews

The data used to generate our AI-Generated reviews and train our second model are Amazon book reviews.

- 51 Million reviews total, 50 thousand sampled randomly for use.
- JSON Format - A file format that uses human readable text to store data in a key:value pair.
- Contains information such as:
  - Star rating
  - Text review
  - Item/Book name
  - Verified Purchaser flag

Data Length Exploration



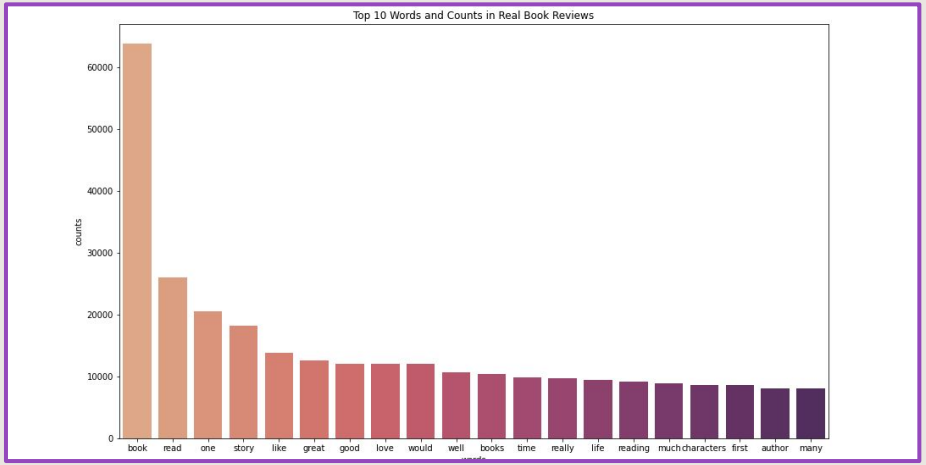
JSON Object


```
{'_id': ObjectId('5fc928d03a06ca0ca8587125'),
 'overall': 5,
 'verified': True,
 'reviewTime': '10 21, 2017',
 'reviewerID': 'AK2I3FIPVFXF6',
 'asin': '0151262276',
 'style': {'Format:': ' Paperback'},
 'reviewerName': 'Jane L. Smith',
 'reviewText': 'One of the funniest books I have ever read.  Good to give to someone who needs a laugh.  A very easy and qui
ck read.',
 'summary': 'Good to give to someone who needs a laugh',
 'unixReviewTime': 1508544000},
```

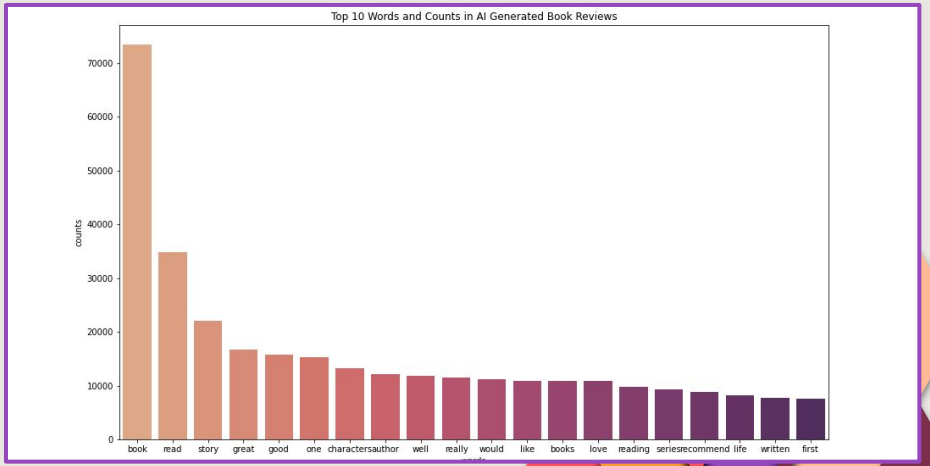
Review Exploration

These bar charts show the 20 most common words in both kinds of reviews and their actual counts within the reviews.

Real Review Wordcloud



AI- Review Wordcloud



Review Generation Model

OpenAI's GPT2

GPT2 is the second iteration of the GPT architecture, which simply put a neural network already structured and trained specifically for language processing.

The model takes a 4 word sequence as a prompt, and here is an example of what it can do:

```
Real review prompt:Thorough reporting of the long ago crime.  Gives insight into the lives of those affected.  Especially liked the bonus chapter.

Fake review from prompt: Thorough reporting of the political, economic, and cultural influences which have shaped this country. This is a great, well researched book that covers important issues.
```

Training and Testing loss represent how well the model fared in learning the material presented.

Training Statistics

Neural Networks:

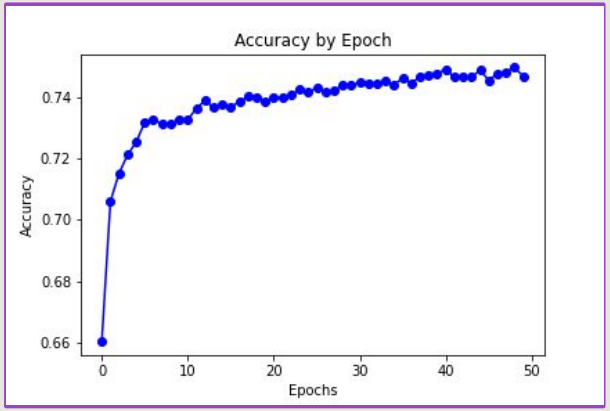
A system of mathematical operations that seek to simulate how the human brain learns and develops knowledge of context.

Fake Review Classifier

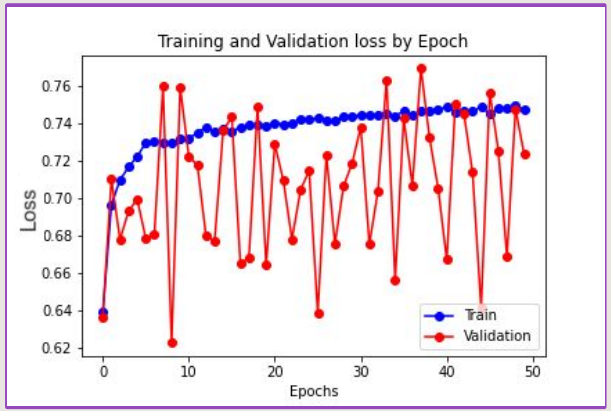
Google AI's BERT

BERT is very similar to GPT, but it is more specialized towards classification of text. We used our generated and obtained reviews to fine-tune BERT for classifying book reviews.

Training Accuracy



Training Statistics

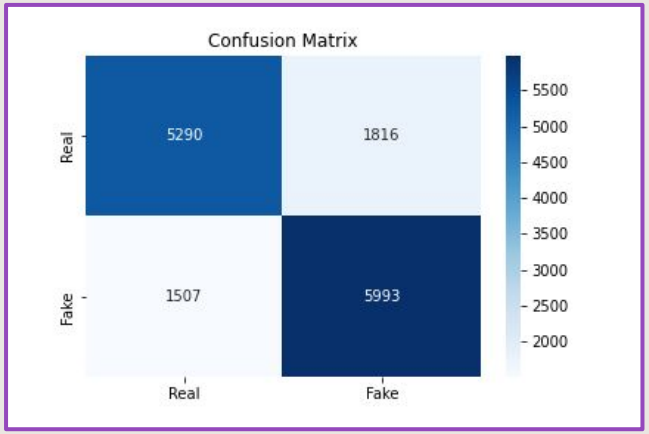


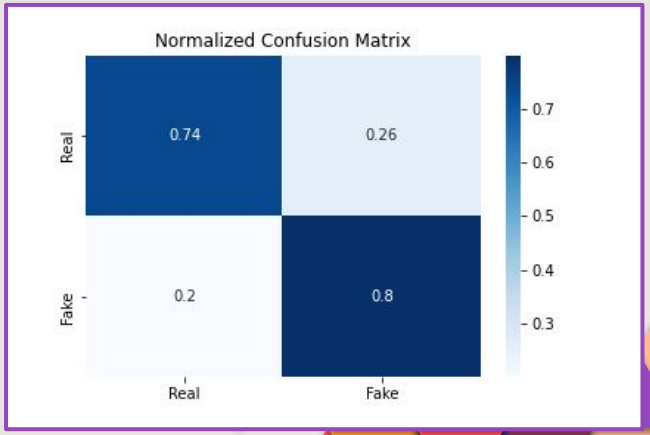
Fake Review Classifier Cont.

Google AI's BERT

Confusion matrices just show how many the model truly classified as AI-Generated (True Positive), how many it classified as AI-Generated but were Real (False Positive), how many it truly classified as Real (True Negative), and how many it classified as Real but were AI-Generated (False Negative).
A normalized confusion matrix just shows the same data but as percentages

We are predicting AI-Generated reviews at an **80% Accuracy**!





Recommendations

Transfer Learning

As we have done here, transfer learning is the process where we use a pretrained model as a base.

In order to stay ahead of those that generate these reviews **we recommend maintaining the latest in transfer learning models.**

Benefits:

- Cost Efficiency - These models are general released free and open source.
- Time Efficiency - Without the pretrained models developing and training would take significantly longer and use much more resources.

Training

Training the BERT classifier for a much longer time is **HIGHLY RECOMMENDED**.

The paper BERT is based on recommended training for 1 million steps, or approximately 40 epochs. As proof of concept we have trained for 5 epochs, or approximately 4 hours.

Deployment

Much of the focus we put into the project was so it could be setup as a marketplace backend, however **we recommend investing in the development of a consumer facing product that utilizes this technology.**

This consumer facing tech could be deployed as a Software as a Service (SaaS) operation.

Future Works

Deployment

The biggest goal and the next step of the process in developing this into a full machine learning product would be to deploy this model to a market backend or customer facing app.

Tuning

Our next step after deployment would be further tuning, so our model can head towards 99%+ accuracy at classifying a review as bot generated.

THANKS!

Name	Sam Middleton
LinkedIn	https://www.linkedin.com/in/samuel-middleton-a8533491/
Github	https://github.com/emperorner0
EMAIL	samuelmiddleton93@gmail.com

CREDITS

This is where you give credit to the ones who are part of this project.

Did you like the resources on this template? Get them for free at our other websites.

- Presentation template by Slidesgo
- Icons by Flaticon
- Infographics by Freepik
- Author introduction slide photo created by Freepik
- Text & Image slide photo created by Freepik.com