

AMASUM

An E-commerce summarization platform for customer reviews.

23

What's Wrong?

In the past year, Amazon saw 1.5 billion reviews from 125 million customers. This deluge of feedback influences 92% of online purchasing decisions, posing challenges for consumers and businesses alike. Consumers often experience decision paralysis when faced with numerous reviews, while businesses struggle with the daunting task of manually analyzing feedback, risking misinterpretation. These challenges in the trillion-dollar e-commerce market have significant financial impacts.

Additionally, using generic AI tools and external APIs for NLP poses risks in performance, data privacy, and security. These tools often lack necessary customization and integration capabilities, leading to financial strains over time.

Companies can build proprietary Large Language Models (LLMs) using open-source models, allowing for customized, relevant AI solutions. Such models align with organizational needs, enhance data control, and maintain privacy, fostering innovation and strategic growth. This step towards technological self-reliance offers a secure, bespoke approach to leveraging AI, providing a competitive edge.

PREPARED BY
William Fussell

PREPARED FOR
BRAINSTATION

What is AMASUM?

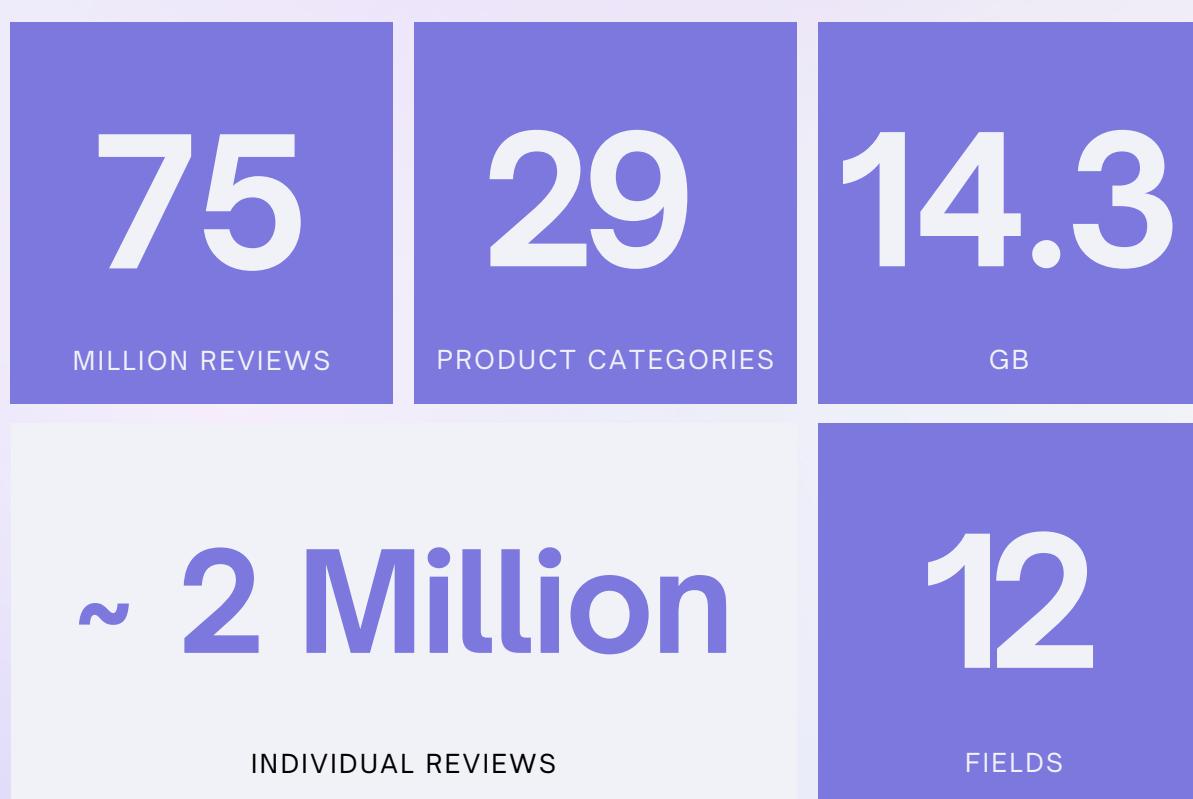
I believe my project, AMASUM is a potential solution to these challenges. AMASUM leverages state-of-the-art AI solutions to neatly summarize customer reviews as well as analyze those data trends over time, bridging the gap between abundant customer feedback and tangible, actionable insights for both purchasers and businesses.

As well, AMASUM is the first product in a project aimed at advancing the proliferation of individual proprietary AI language models (LLMs) for businesses across various sectors. The ethos of the project is “A personal Large Language Model for every business.” This initiative is a cornerstone in championing Data Democratization, Privacy, and Security, offering Customization, and ensuring astute Cost Management and ROI. It embodies a transformative approach to leveraging technology for competitive advantage, offering scalable and future-proof solutions. The significance of this project lies not just in its immediate e-commerce benefits but also in its ethos of democratizing state-of-the-art AI solutions custom tailored for any business, marking an invaluable contribution to the evolving landscape of global commerce and digital innovation.

What's Inside?

At the core of this platform is the fine-tuned **Llama 2** model(**Meta AI**'s open source state-of-the-art LLM), which serves as a springboard for training a proprietary Large Language Model (LLM).

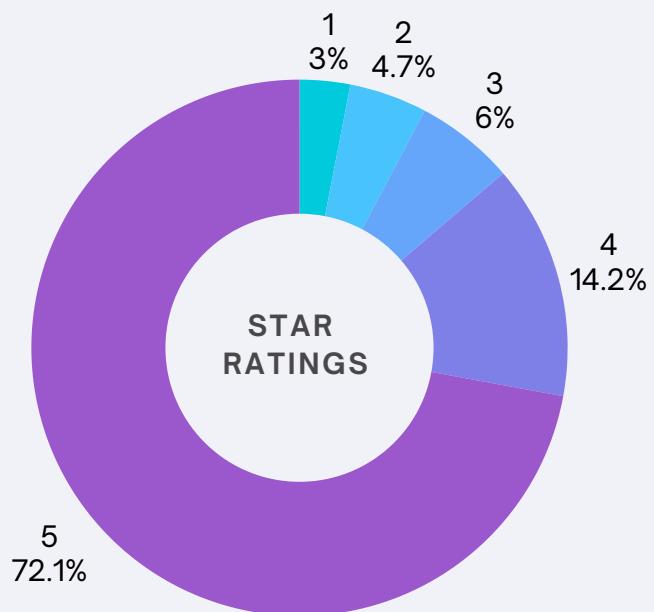
The platform's foundation is built upon an extensive, open-source dataset of Amazon Customer Reviews sourced from [UCSD](#) open-source GitHub Repository. Compiled by researchers **Jianmo Ni**, **Jiacheng Li**, and **Julian McAuley** for their 2019 **EMNLP** research, this dataset was created to support studies in natural language processing, particularly in understanding the impact of customer reviews on product recommendations. [Source](#)



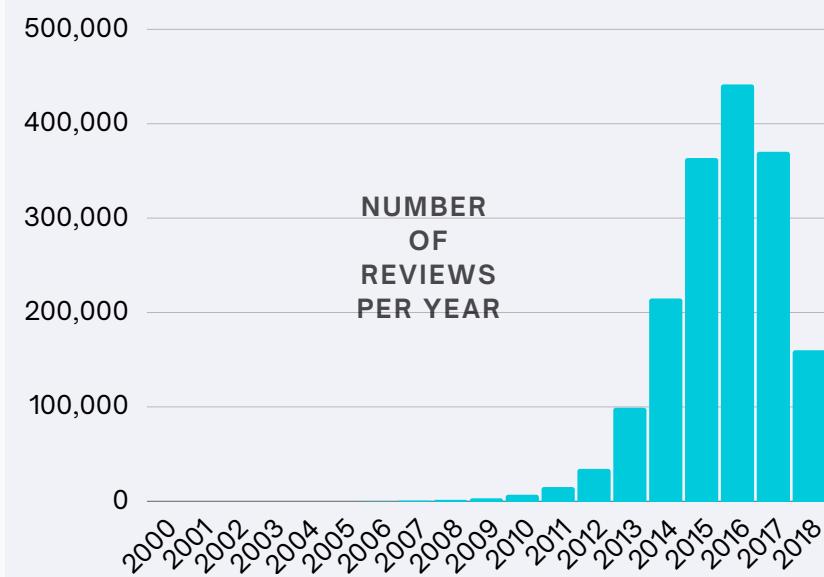
DATA SET SPECIFICATIONS

What did we find?

I found A notable imbalance in review ratings, with 86% of reviews having 4 stars or above...this surprising finding gives a project like AMASUM real meaning...by helping find the sentiment of all aspects of a product including the bad ones... even when most of the feedback is positive



95% of reviews occurred post-2013, despite data spanning from the year 2000, this insight shows that more and more reviews are coming each year.



Extractive vs Abstractive Summarization

When tackling a summarization problem with Natural Language Processing (NLP) there are 2 main techniques: **Extractive** and **Abstractive** Summarization

Extractive

Involves identifying key sentences or phrases from a text, based on criteria such as similarity or term frequency, and then compiling them to form a summary.

(This is what will serve as my baseline model that I will judge the final product off of.)

1

2

Abstractive

Is the process of creating a summary by interpreting and paraphrasing the main ideas of a text, rather than simply extracting key phrases or sentences.

What's the Standard?

The **baseline model** for this tool employs an extractive approach, leveraging Google's **BERT** word embeddings for its implementation. **BERT** (Bidirectional Encoder Representations from Transformers) provides a sophisticated vectorized representation of words in a document. These embeddings have shown superior performance in various NLP tasks compared to traditional methods like TF-IDF.

The process involves using **BERT**'s transformer architecture to generate word embeddings, which are then applied to calculate the cosine similarity between sentences.

Cosine similarity measures the closeness of these embedding vectors, indicating the degree of similarity between sentences. The model selects sentences that are closest to the average embedding of all sentences in the reviews, forming the base of the summary.

This is then compared to a reference summary usually generated by human writing but for time and clarity sake I opted to use **OPEN AI GPT 3.5 TURBO**. Lets take a look below.

Extractive

"- Flimsy cheap mirror that has an extremely distorted view"
"- Not worth the price.ther Cheesy! Really small
Very dissatisfied with mounted it on my
pontoon boat drove to the lake when I got to
the lake did not realize it the glass part of it was
missing now all I have is a plastic frame you
wood thinking would last a lot longer then one
trip than that very very dissatisfied No good,
rusted out after one season"

"- Boooooool great price, looks good, but hard to
see any thing with it. Cheap plastic with a mirror
finish"

Reference

- The product is described as flimsy and cheap,
with an extremely distorted view, making it not
worth the price.

- Many customers expressed dissatisfaction
with its quality, mentioning issues like rusting
after one season, difficulty in seeing clearly, and
the plastic frame not lasting long.

- Several users recommended avoiding this
product and suggested investing in a higher-
quality mirror for better visibility and durability.

Is that good enough?

ROUGE - 2

8%

ROUGE - L

20%

Next we will need to evaluate the quality of the extractive summary compared to a reference summary. To do this we will use the **ROUGE Evaluation** (Recall-Oriented Understudy for Gisting Evaluation). It measures the overlap in content between the generated summary and a reference summary.

To understand these metrics, you can think of it as a **percentage** of how much the extractive summary's content is shared with the reference summary.

At the most, only 20% of the representational reference summary is generated through extractive methods, this seems to need the hands of a more state-of-the-art option.

What's the Endgame?

Pre-Processing AMASUM

Llama 2

The less than satisfactory findings of the extractive summary, lead us to the more robust solution. **AMASUM** is the answer to this problem as it integrates an abstractive summarization model, fine-tuned using **Meta AI's Llama 2**, a state-of-the-art, open-source Large Language Model (LLM).

Prompts

Next we need an instruction for the model to use as guideline to know what to do. These are called prompts. I developed a custom **Alpaca** style prompt, combining an instruction, concatenated product reviews, and a reference summary.



Labels

For effective **Llama 2** utilization, pre-processing was essential. Reference summaries, crucial for training, typically derive from human-generated content. These are essentially the models labels that help show the model what to produce. However, to expedite the process, I used **OpenAI's API** for generating these references, resulting in summaries comparable to **human-crafted ones** and likely better than humans when dealing with products with thousands of reviews.

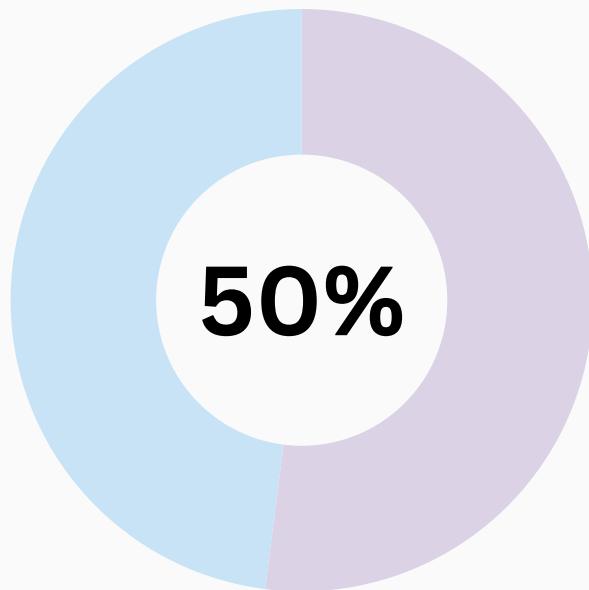


QLoRA & Training Arguments

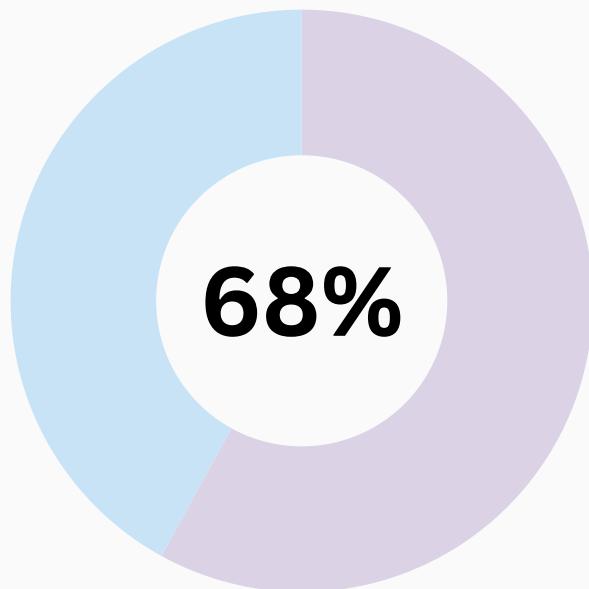
Next, I Utilized QLoRA, a technique for quantizing LLM parameters allowing me to train this massive model on a single Nvidia A100 GPU in a google collab notebook.

Final Evaluation

ROUGE - 2



ROUGE - L



The **ROUGE** scores were promising, commendable even.

In fact, the performance of **AMASUM**, when creating summaries for unseen reviews, reached approximately **68%** of the effectiveness demonstrated by **OPEN AI's GPT 3.5 Turbo**. This indicates a substantial proficiency of **AMASUM** in this task.

However, the real task is to compare them side by side so lets do that now.

How did we do?

AMASUM

A final comparison of AMASUM measured up against CHAT GPT...

GPT 3.5 TURBO

- Brightness and compatibility issues: Many customers found that these license plate lights were not bright enough and did not fit their specific car models, leading to disappointment and the need for returns.

- Light output: The light output was a major concern for customers, with some mentioning that the lights appeared blue instead of white and were not as bright as expected.

- Desire for brighter options: Some customers expressed a desire for brighter license plate lights, indicating a preference for more powerful options.

- Many customers found that these license plate lights were not bright enough and did not meet their expectations.

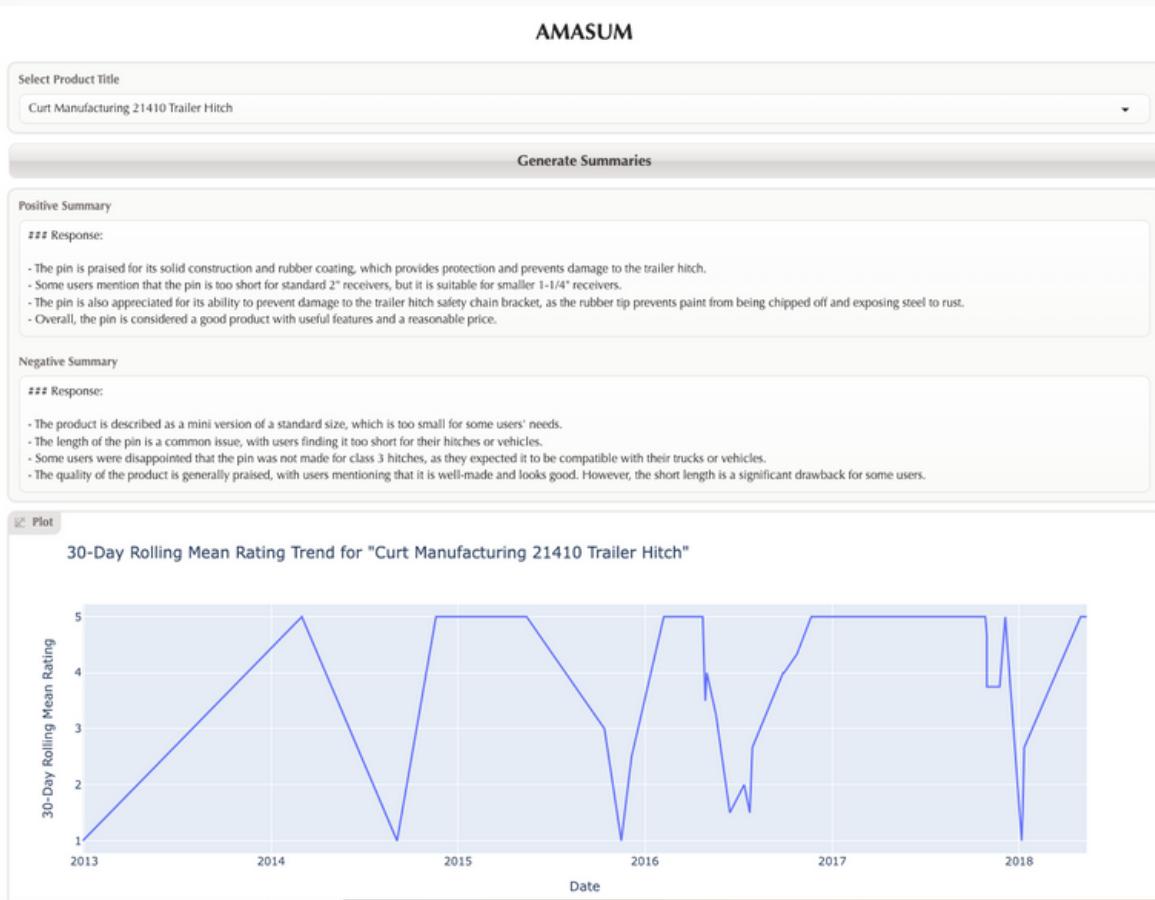
- Compatibility was a major issue, with customers reporting that the lights were not suitable for their specific car models, leading to disappointment and the need for returns.

- Some customers were particularly dissatisfied with the light output, with one customer mentioning that the lights appeared blue instead of white, and others expressing a desire for brighter options.

The summaries from AMASUM from a personal glance seem to capturing competing answers to that of the current industry leader OPEN AI. Proving that open-source models and data can compete now with industry standard and the potential for proprietary AI is in the reach for many businesses.

THE AMSUM APP

If you follow through to the github link in the next page you can guide yourself to the Amasum App Jupyter Notebook and try out the results for yourself!



What's Next?

The initial version of this project demonstrates the potential of leveraging open-source language models (LLMs) to create safe, cost-efficient, and potentially profitable AI applications fine-tuned with proprietary data, aiming for near-industry-leading performance. To enhance this model, I believe the following steps can be taken:

01 **Larger Training Dataset**

Initially, I filtered 20K high-quality data points from a pool of 2 million. Expanding the dataset for training would result in more robust models.

02 **Increased GPU Power**

The model was constrained by the GPU on Google Colab. Utilizing more powerful GPUs would enable me to handle larger datasets, improve training parameters for the Transformer Model, and employ larger model versions. This extended training period, compared with previous successes of fine-tuned Llama2 models, can further enhance performance metrics.

03 **Efficiency Optimization**

I will allocate more time to fine-tune inference parameters, allowing the model to run efficiently on CPU, making it more accessible for various deployment scenarios

Thank you!

Thank you for taking the time to read this report. If you have any questions or would like to discuss my findings further, please don't hesitate to reach out.

 416-570-8821

 wefussell@gmail.com

 <https://github.com/williamfussell/amazum>