ORIGINLENS: A REAL-TIME AI-GENERATED TEXT AND PLAGIARISM DETECTION USING DEEP LEARNING AND AUGMENTED REALITY

Henry Muyang Liu and Jenny Do

Student at the Liberal Arts and Science Academy, Austin, Texas
Data Scientist, Naval Facilities Engineering Systems Command (NAVFAC)

ABSTRACT

ChatGPT has integrated itself into the academic space in an unprecedented timeframe, as the promise of hours of work done in seconds can outweigh the sense of honor and logic. This project determines whether a student is cheating or not by reducing the presence of human decision-making while simultaneously acting as a deterrent for future usage of AI technology. Utilizing databases such as Kaggle, we can procure several samples of human writingin conjunction with ChatGPT's API to generate artificial intelligence instances, which are then stored for usage in machine learning algorithms. Employing powerful Python libraries such as Sklearn and NLTK, we can utilize natural learning processing, the ability for computers to understand human writing, to yield an algorithm that can predict with approximately 96% certainty. The result is a probability ratio, with one side displaying the percentage chance of the sample being human-written, whereas the other displays the likelihood of AI-generated instances.

Furthermore, innovation lies in integrating this algorithm with wearable augmented reality technology, allowing usersto efficiently scan and assess text elements. This approach amalgamates and helps reduce the delay between text input and response, empowering users to contribute to the decision-making process in identifying academic dishonesty without any loss in efficiency. The result that is shown displays pieces of information to the user that can all play a large role when determining the possibility of cheating, granting the user a role in making the decision along with simple scanning of each text element.

1. INTRODUCTION

In the wake of the ChatGPT explosion, educators across the world found themselves facing problematic situations where academic assignments could easily be generated and thus plagiarized using artificial intelligence (AI). This issue cannot be considered isolated events, as over a quarter of recently surveyed K12 teachers have caught students utilizing ChatGPT (Klein, 2023). This project determines whether a student isindeed cheating or not by reducing the presence of human decision-making while simultaneously acting as a deterrent for future usage of AI technology. Even in top universities such as Harvard, the usage of ChatGPT to gain an advantage over fellow classmates has become a common occurrence within the campus, where several dozen anonymous interviewees admitted to using the chatbot. Even though this clearly was an issue, it took universities months to implement proper guidelines, which shows the disregard most people have for this issue. Unlike conventional applications, this project is unique due to the presence of a probability scale, allowing users to clearly access the sample's tendency towards AI generation rather than providing a definite answer, which has the possibility of harming the academic careers of innocent students. My solution to this problem is to utilize data

David C. Wyld et al. (Eds): AIBD, MLSC, ACSTY, NATP, CCCIoT, SVC, SOFE, ITCSS -2024
pp. 63-74, 2024. CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.140405

analysis along with machine learning to recognize patterns in text to determine whether something involves AIgenerated text. This tool helps bridge the gap between decision and indecision, especially with teachers who suspect the usage of chatbots. Along with this, the overall functionality can be hosted on an Augmented Reality headset, making access easy and as simple as running the application and returning a result. Though not entirely a possibility yet, this technology could be implemented in the future in a way that is feasible and functional.

Due to the often professional nature of AI-generated writing, patterns can be analyzed to form a conclusion. As such, using data analysis of existing AI writing and generating more samples is a good way ofsolving thisissue. However, the drawbacks are present, especially with extremely complicated text generations, and can limit the inputs to simpler paragraphs. Overall, this method of analysis is superior to other methods of looking at sentence patterns due to the flexibility and the adjustable success method, as more data can be fed into the algorithm, leading to higher accuracy and the introduction of a revolutionary method of integration with wearable augmented reality.

2. METHODOLOGY

The first challenge is training the AI model. Without the set of instructions applied to the datasets, the model is practically useless and requires feature engineering in order to reach full efficiency. To achieve this, the process of filtering, which can be lengthy in order for compatibility for analysis, can possibly be used. Each individual set needs to be merged and cleaned through the removal of punctuation, stop words and other parts of writing that hinder rather than add any meaning. These all add to less computing time, which can be useful, especially in the future when datasets are expanded. The next issue isthe ability to make the user interface easily accessible and intractable in a way that isn't both overwhelming and difficult to manage. As such, it was decided that an augmented reality presence was needed in order for convenient usage, as it would be difficult to detect handwritten works along with the time it takes to manually input each piece of writing. It also could be difficult to recognize words through a camera input,but using optical OCRs like Tesseract can reduce the complexity of this issue. OCR technology has improved significantly in the past, and is almost perfect at scanning legible words from a camera. The lastmain challenge is to find usable data that is of quality and can be utilized by the model to formulate a clear response. Since this project can give two different options, the sources needed are human-written essays along with AI generated text. The difficulty of human essays is scarcity, especially when dealing with concerns of plagiarism and other countermeasures in publicly posted text, as they typically have a hefty cost for a suitable database, with free samples being of lesser quality and of improper length. To mitigate this issue, it is possible to look into large public databases such as Kaggle or Dataverse.

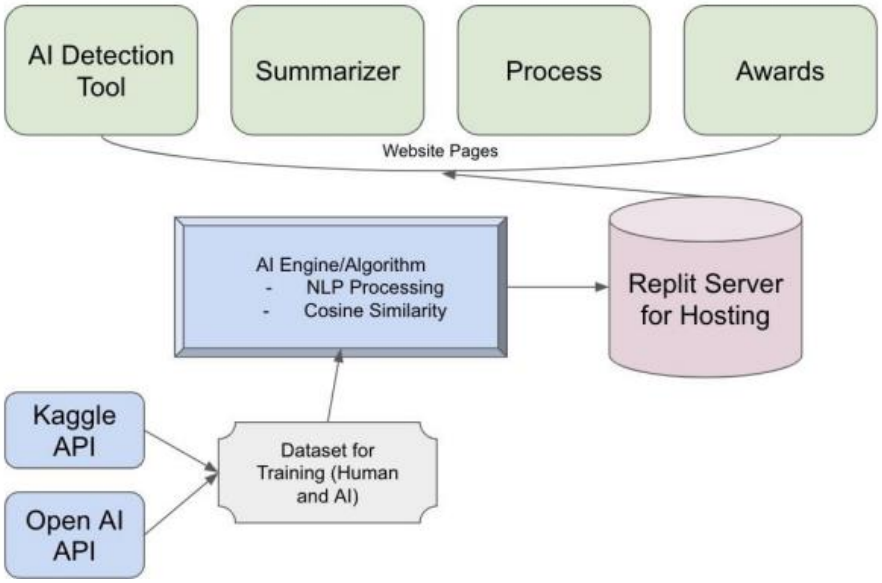
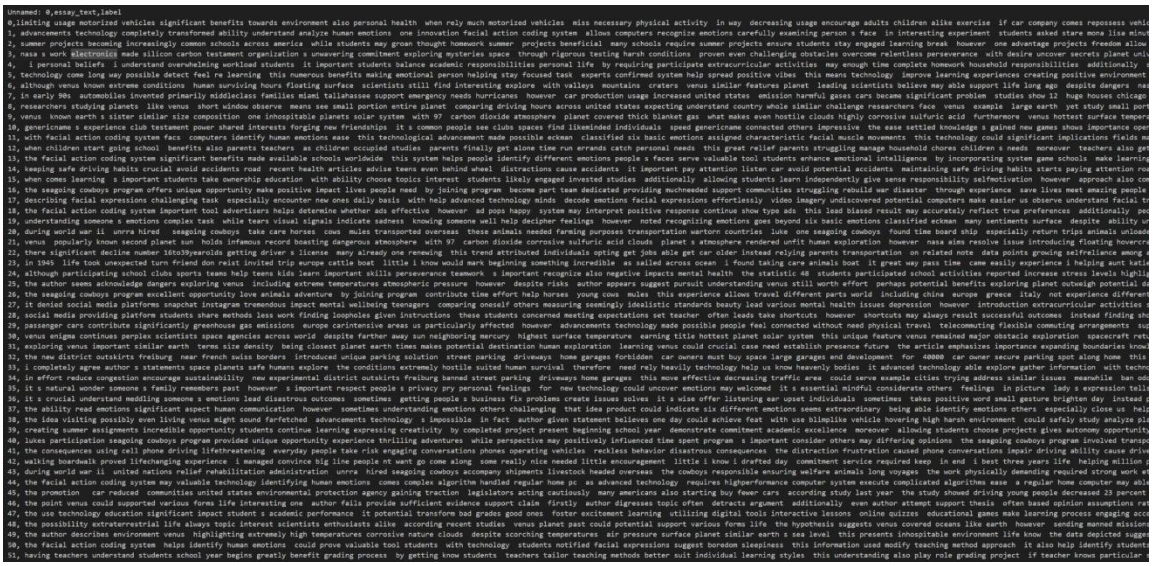


Figure 1: System Overview

As part of the system overview, the program contains three major components: data storage, processing, and result display. The data storage section gathers information from outside sources and compiles them into a usable format through tokenization, the removal of unnecessary stop words and punctuation. After the data is filtered, the newly cleaned dataset is fed into the specified algorithm and trained in order for it to adapt to situations outside of the database. After training, the algorithm is tested for accuracy and fed to other sources to validate that the process is indeed working. After this is all complete, the algorithm is then launched to a website, ready to be used for real-world applications. The web application first takes in text input from the user. Then it sends this information through the algorithm, which returns a definite result, which the user can see in the form of a percentage ratio between human and AI writing. This way, the user can use their judgment to determine the final result, allowing for more flexibility.

2.1. Component Analysis

The first component is data gathering and filtering. A dataset suitable for the machine learning algorithm, one that does not contain wildcards or unknown symbols, is fed into the formula. To yield a better result, natural language processing tools are applied. This includes removing stop words, words that contribute no meaning, and punctuation. The removal of these two components allows for better computational power and the reduction of costs.



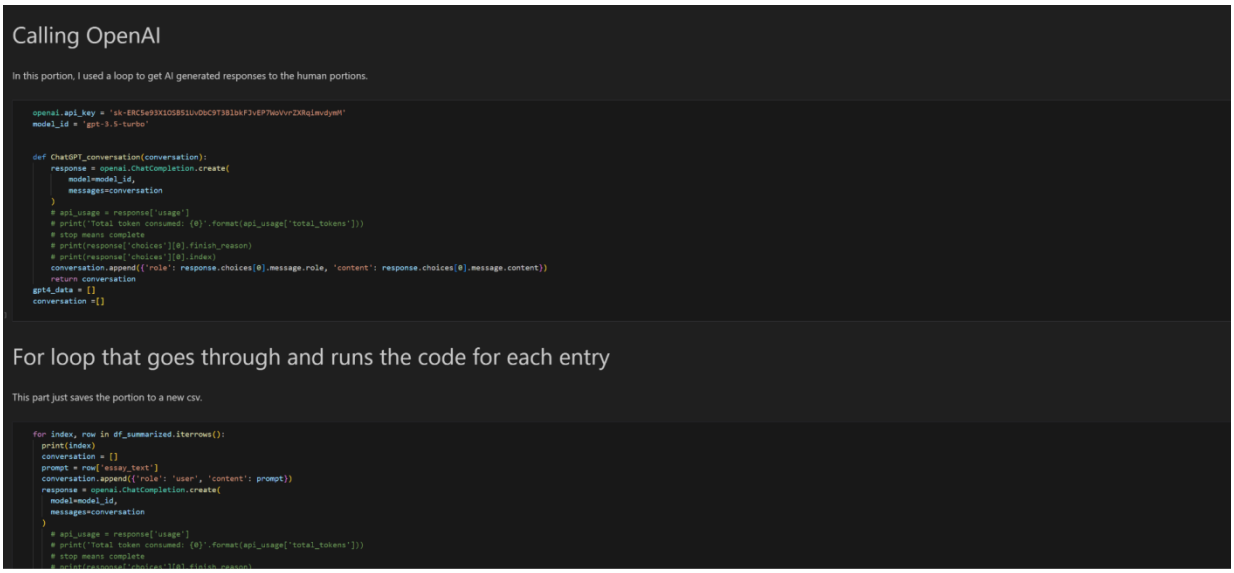


Figure 2. Data base analysis

When the application first begins, data in the form of human writing is pulled in through Kaggle's API. When this occurs, the data is considered inefficient and "dirty", therefore, needing to go through the process of filtering and cleaning before the most efficient result can be parsed. Along with this, examples of AI writing need to be generated. We can use the samples of human writing to feed into ChatGPT's API in order to generate artificial intelligence instances, which are then stored for usage in the machine learning algorithm. After each step is completed, the text is given a label and then stored in CSV files. Once sufficiently filtered, we can combine both human and AI instances into a master dataset, allowing for easier access and manipulation by the algorithm, along with easier management if more data needs to be added. It should be noted that ChatGPT's API costs money, and therefore, tools such as the summarizer library were applied to keep costs low.

2.2. Component Analysis

The purpose of the algorithm portion is to be able to process information fed in by the user, which immediately sends back information in the form of a percentage of how likely each case could be through the utilization of the database shown in the previous section.

```
              precision    recall  f1-score   support

          AI       1.00      0.83      0.91         6
       human       0.00      0.00      0.00         0

    accuracy                           0.83         6
   macro avg       0.50      0.42      0.45         6
weighted avg       1.00      0.83      0.91         6

[[0.55383634 0.44616366]
 [0.51130627 0.48869373]
 [0.49734406 0.50265594]
 [0.7901712  0.2098288 ]
 [0.52760536 0.47239464]
 [0.60655273 0.39344727]]
```

```
print(classification_report(y_test,prediction))
#gets the report
cm = confusion_matrix(y_test, prediction, labels=model.classes_)
#uses the confusion matrix
disp = ConfusionMatrixDisplay(confusion_matrix = cm, display_labels = model.classes_)
#displays on a chart
disp.plot()
#shows it
plt.show()
```

```
              precision    recall  f1-score   support

          AI       0.96      0.96      0.96       139
       human       0.97      0.97      0.97       161

    accuracy                           0.97       300
   macro avg       0.97      0.97      0.97       300
eighted avg       0.97      0.97      0.97       300
```

Figure 3. Component

After the datasets are combed through and filtered, the AI uses analysis to determine the result, which visually outputs a confusion matrix. This confusion matrix determines how accurately the computer can evaluate results if they were randomly fed into the algorithm. Based on the depiction above, both have been correctly identified with an accuracy of roughly 97%. Then, to measure individual accuracy, more experiments were done with only human examples, utilizing unused parts of the original dataset. The external dataset was then fed into the algorithm containing the trained dataset. The results show that when 400 external data points were used, the algorithm predicted human writing correctly 377 times compared to 23 false categorizations. It is shown that the model had roughly 94% percent accuracy after the analyzation phase, which was up to the predicted standard.

2.3. Component Analysis

The data display portion acts as a linkage between every layer of the project. Using a Flask server, we can easily transmit data from the AI portion to a website, which is then displayed. Using dynamic text boxes and buttons, the Flask server can take in an input, and then evaluate using the algorithm.

```
app = Flask(__name__)

def load_data():
    final_data = pd.read_csv('final.csv')
    final_data = final_data.loc[:, ~final_data.columns.str.contains('^Unnamed')]
    return final_data

def train_model():
    final_data = load_data()
    X_train, X_test, y_train, y_test = train_test_split(final_data['essay_text'], final_data['label'], test_size=0.15)

    pipe = Pipeline([
        ('vect', CountVectorizer()),
        ('tfidf', TfidfTransformer()),
        ('model', LogisticRegression())
    ])

    model = pipe.fit(X_train, y_train)
    return model

def predict(model, input_text):
    prediction = model.predict([input_text])
    prob = model.predict_proba([input_text])
    return prediction[0], prob[0]

@app.route("/", methods=["GET", "POST"])
def index():
    if request.method == "POST":
        input_text = request.form.get("input_text", "")
        if input_text:
            model = train_model()
            prediction, prob = predict(model, input_text)
            return render_template("index.html", input_text=input_text, prediction=prediction, prob=prob)
        else:
            error_message = "Please enter some text."
            return render_template("index.html", error_message=error_message)
    return render_template("index.html")

if __name__ == '__main__':
    app.run('0.0.0.0',8080)
```

Figure 4. Code

The website has two main selections: AI detection and a summarizer. As stated above, the summarizer tool was vital to the success of this project, as it contributed heavily to the overall cost of running ChatGPT's API and shortened phrases which would reduce the tokens of each API call. The primary function, however, is still the AI portion, which uses the database and analytical formulas to gather a result and send it to an HTML frontend. CSS is then used to determine the style and overall appearance ofthe website, along with a portion of Javascript for display purposes. This website can handle a variety of text input and can be considered accurate in many scenarios. The hosting service used was repl.it, as it is cheaper and more costefficient than running on a dedicated service.

The augmented reality portion was developed solely as a way to host the application in an area other than a website, which utilizes optical character analysis through a camera to deliver real-time information. This was developed to be a proof-of-concept and is more of the second step in this project, which has not yet been fully developed. The main experimentation done so far is utilizing the source code and adapting it to an augmented reality base. By doing so, we can test real-life examples of texts more rigorously in the form of paper-based writing samples. This project also raises some significant questions about ethics. Most, if not all schools contain warnings against cheating and academic dishonesty, making the scanning of text
in order to analyze data to be ethical. Along with this, it's definitely possible for teachers to raise concerns before running text through this code.

3. EXPERIMENTATION

Accuracy in the AI's results is paramount in a project with this scope, as it can have severe repercussions if done incorrectly. This issue can be negated through high amounts of training and accuracy testing. To do this, a collection of paragraphs written by humans and AI was tested to see if the accuracy would be an issue. This data would be gathered from sources all across the internet, along with niche test cases that the author writes as a form of edge case. Also included are examples of AI-generated writing gathered from curated prompts from ChatGPT. The two categories are then separated and fed into the data as a blind source to observe how the AI would react. The more arbitrary data the algorithm can encounter, the greater the likelihood of success in the analysis phase.

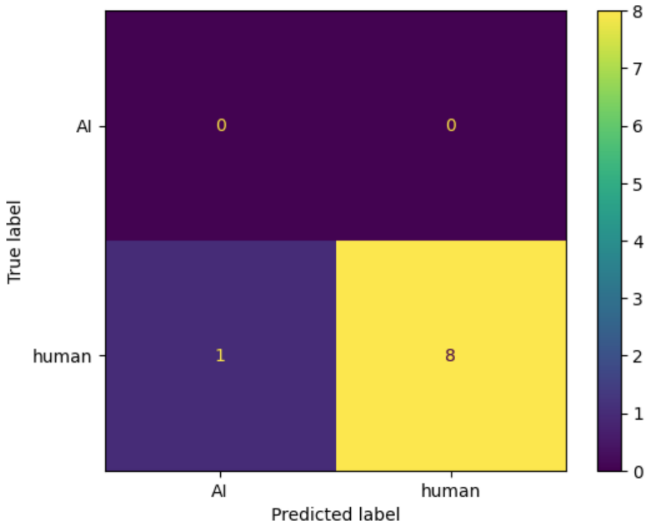


Figure 5. First Confusion Matrix

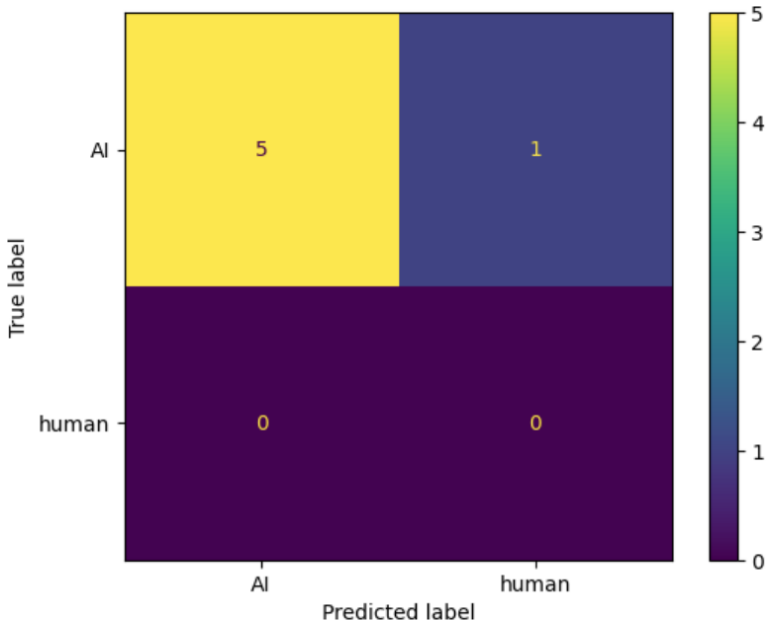
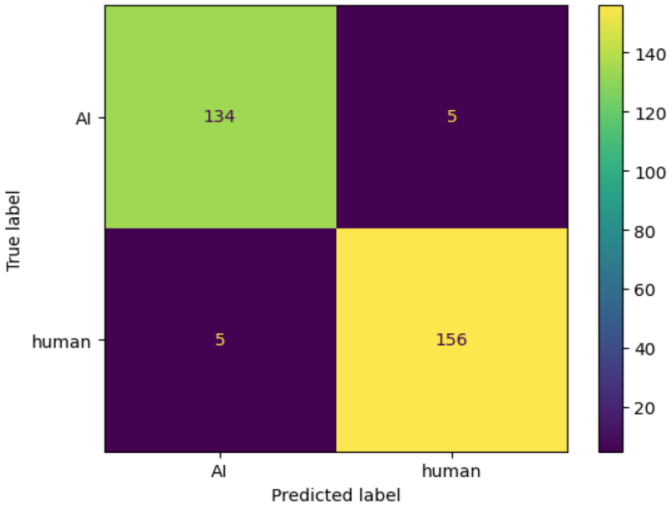


Figure 6. Second Confusion Matrix

The data shown is the two samples we fed into the algorithm. The first chart shows the accuracy versus human writing, and the second graph displays the results from samples of AI writing. On average, the algorithm can return an accurate result, with some outliers likely having sections that could be flagged as being an improper estimate. Due to the smaller nature of this dataset, this could be a random error as previous experiments had a more significant percentage of 97% accuracy. This is unsurprising, as more fine-tuning is necessary for positive results. Not shown in these graphs are the percentages, which are bordering 90% for both tests. With such a small dataset, it can be difficult to gather a conclusion based on the results, showing that outliers could occur periodically. However, the overall positive ratio shows the proper functionality of the algorithm.

3.1. Experiment 2

To test the functionality, 150 samples of AI and human text were pulled from random sources, with all labeling removed and fed into the program in order to discern the result. As shown below, the accuracy of each section can be shown.



	precision	recall	f1-score	support
AI	0.96	0.96	0.96	139
human	0.97	0.97	0.97	161
accuracy			0.97	300
macro avg	0.97	0.97	0.97	300
weighted avg	0.97	0.97	0.97	300

Figure 7. Combined Confusion and Statistical Analysis

This experiment uses a similar method of using data and analyzing sources but differs as it uses a mixed dataset of unused human and AI writing from the original Kaggle request that hasn't been used for experimentation. This is done to discern whether the algorithm can handle various unrelated inputs. This would be the most realistic scenario and experiment since data would never be consistent and is typically variable and fluid, which this experiment addresses.

3.2. Experiment 3

Taken on February 2, 2023, by Henry Liu

The item being presented is a perfect prototype for next-generation wearable technology. The premise behind this device is a simple optical system to reflect a microOLED onto the user's field of view. The display is then hooked up to a Raspberry Pi via HDMI, not much unlike a television. Using the terminal, it is possible to apply the AI detection software using several layers of development and adaptation. The primary way of doing so currently is using the website on the glass, but a more focused approach of writing code specifically for the glasses can be possible in the future.

3.3. Future Improvements

As discussed in the abstract, a possible future improvement would be implementing this into an augmented reality device. This review will discuss the possibilities and how viable this switch could be asopposed to a website.

3.3.1. An Emerging Market

A common theme that most successful companies utilize is the social culture, which drastically affects markets. Wearable wellness monitors have become mainstream through careful marketing and playing on the ethos of interconnectivity (Brophy et al., 2021). Devices such as Fitbits and Apple Watches give app developers a platform that can affect the user directly through fitness and well-being monitors. Like many online forum users who have struggled to lose weight after pregnancy, the Apple Watch fitness app helped them lose over eight lbs in under three weeks, a task which is a lot easier said than done (LinkedIn, 2022). Wearable augmented reality is still considered a relatively new field, and a portion of the population is now aware of the magic behind such devices through promotion and past devices such as Google Glass. According to Yin (2021), this movement has garnered attention and effort after seeing how computergenerated images could be combined with the natural world; some first-generation AR devices are being pushed out with greater practicality and efficiency.

These sources heap praise on wearable technology while also providing its shortcomings. Brophy (2021) describes the problems that users face when using such devices through privacy concerns and the everpresent public perception. However, these sources need a collection of general data and how a real-life setting can damper wearables' potential but prove an extensive collection of data based on projections. This is not necessarily worse but can increase variabilities related to data.

3.3.2. The Advancement of Technology

Reports say the AR technology market will surpass 800 billion dollars by 2032. As a result, various companies are willing to pour large amounts of money as capital, knowing that whoever gains dominance in this market will write the next chapter in computing. In the world of economics, technology is one of the critical drivers of growth within modern-day society (Hausmann, 2018). Technology companies can afford more investments in newer areas, and that has been augmented reality in recent years. The sources I have analyzed contain information regarding the broad overview of technology rather than focusing directly on augmented reality and indirectly tied into the development of wearables.

3.3.3. The Separation Between Augmented Reality and Virtual Reality

To the average user, it becomes difficult to distinguish the separation between the definitions of virtual and augmented reality. Although both are similar, the truth is that both sectors have their advantages and disadvantages. The benefits of virtual reality include easier-to-build devices and a more accepted market, with penalties including devices unable to connect the real world with the virtual.

Augmented reality solves this issue by allowing users to see the real world while seeing virtual images (Billinghurst et al., 2018). Virtual reality's primary use is limited to online tasks such as games. For AR, daily tasks such as cooking, walking, and biking can be easily accomplished with virtual images such as a recipe or maps. Virtual reality devices can be easily achieved without pooling as many resources as augmented reality, meaning that mature devices already exist in the market along with a market that has a positive view. However, this isn’t necessarily the case with AR since a more extended R&D cost means most companies stay away from devices until better technology reduces costs.

In conclusion, although similar, AR and VR differ heavily in the practical world. The sources I analyzed throughout this time explained the underlying differences. Still, it does include several weaknesses, including a disability, to prove which side is “superior” to others. As with most papers, there are several gaps within the literature review, which hinders the readers' understanding. This paper will address the “uses” instead of the more technical information that most journals give.

3.3.4. Issues

With most technological devices, several issues exist within this sector. The key is addiction and problematic surveillance. Addiction is a reasonably large issue within devices such as phones and other portable electronics, and more steps are needed to prevent these issues. With powerful software such as Screen Time, users can either restrict their children or themselves to improve their lives in a reasonable fashion. With surveillance, there definitely will be several issues when AR devices increase in popularity. As with the iPhone, several hackers jumped on the opportunity to have access to millions of users, but as the generations improved, more flaws were filled and eventually were eradicated.

4. METHODOLOGY COMPARISON

Hu et al. proposed a method of using databases to detect AI summarizers. However, this format has several unforeseen issues, mostly having to do with the difficulty of separating summarized texts, as shown by the results of their data. It has ignored the ability to compare different samples of summarized text, and even then, it will be difficult to discern a result, as typically, elements of human writing are involved when using an AI-summarized version. Overall, it doesn't seem very viable when analyzing the result section of this paper, especially when the most prominent testing can't reach around 60% accuracy when using sources of data that weren't the original database.

Yan et al. present a system of using datasets in a similar fashion to this project. However, instead of building an algorithm, this paper took a text analyzer known as roBERT. This approach is decent at using a reliable base but lacks the ability to discern slight changes within the inputs. It also doesn't include a portion where interactivity is possible, only showing theoretical experiments done in a restricted environment. In my project, it is possible for those who have no key understanding of code to input examples of writing while having the AI portion return the result.

5. CONCLUSION

The limitations of my project can be the improbability of certain samples, especially within the field of text analysis. In general, the dataset needs to be adapted for certain cases, which can alter the results of such an algorithm. Certain methods of disguising AI-generated text, such as summarizing or using existing human samples and altering them using the help of AI. By enriching the dataset, it can be hoped that this issue will eventually be fixed and improved upon through more efficient methods of data cleaning. Along with this, the difficulty of making alterations to the front end could hamper any simple possibility of improvements in the future. This is due to a variety of issues, but mostly due to the fixed nature of Flask servers and the way they communicate with Python. However, improvements can be madeto allow for a more dynamic front end, by using a different method of sending data to HTML. Academic honesty isfacing pressure from writing generated from artificial sources. Establishing this app can alleviate this issue by helping people have a resource that produces the percentage likelihood, allowing for user-based decisions. In the future, this project can have an easier time detecting AI, thus decreasing the amount of academic dishonesty that occurs.

6. REFERENCE

[1] Billinghurst, M. (2002). Augmented reality in education. New horizons for learning, 12(5), 1-5. Retrieved November 26, 2022, fromar_edu-libre.pdf (d1wqtxts1xzle7.cloudfront.net

[2] Boyce, P. (2019, August 18). Schools Are Outdated. It's Time For Reform. Foundation for Economic Education.Retrieved February 18, 2023, from https://fee.org/articles/schools-are-outdated-its-time-for-reform/

[3] Brophy, K. (2021). The future of wearable technologies. Imperial London College. Retrieved December 11, 2022, from https://www.guderesearch.com/wp-content/uploads/2021/12/IMSJ8878-Wearable-Tech-B riefing-Paper-210601.pdf

[4] Burnham, J. P. (2018). Using wearable technology to predict health outcomes: a literature review. JAMIA. Retrieved November 9, 2022, from https://academic.oup.com/jamia/article/25/9/1221/5047137

[5] Chen, H., & Tan, G. (2017). Ambient contrast ratio of LCDs and OLED displays. Optica. Retrieved 2 18, 2023, from https://opg.optica.org/oe/fulltext.cfm?uri=oe-25-26-33643&id=380371

[6] Consumer Electronics Market Size to Surpass US$ 1.13 Trillion by 2030. (2022). GlobeNewswire. Retrieved January 5, 2023, from https://www.globenewswire.com/en/news-

release/2022/03/28/2411214/0/en/Consumer-El ectronics-Market-Size-to-Surpass-US-1-13-Trillion-by-2030.html
[7] Godfrey, A. (2018). From A to Z: Wearable technology explained. ScienceDirect. Retrieved November 11, 2022, from https://doi.org/10.1016/j.maturitas.2018.04.012
[8] Iqbal, M. H. (2012). A review of wearable technology in medicine. Journal of the Royal Society of Medicine. Retrieved December 11, 2022, from https://doi.org/10.1177/0141076816663560
[9] McCann, J., & Bryson, D. (Eds.). (2009). Smart Clothes and Wearable Technology. Elsevier Science.
[10] What Is Augmented Reality? (n.d.). The Franklin Institute. Retrieved November 27, 2022, from https://www.fi.edu/what-is-augmented-reality
[11] Xiong, J. (2022). Augmented reality and virtual reality displays: emerging technologies and future perspectives. Retrieved November 27, 2022, from https://www.nature.com/articles/s41377-021-00658-8
[12] Yeung, A., Tosevska, A., & Klager, E. (2021). Virtual and Augmented Reality Applications in Medicine: Analysis of the Scientific Literature. Journal of Medical Internet Research. Retrieved November 27, 2022, fromhttps://www.jmir.org/2021/2/e25499
[13] Yin, K., He, Z., & Xiong, J. (2021). Home. Journal of Physics. Retrieved November 27, 2022, fromhttps://iopscience.iop.org/article/10.1088/2515-7647/abf02e/meta