

Big Data, The Self & Social Platforms

Project Documentation

Diana Milena Galindo

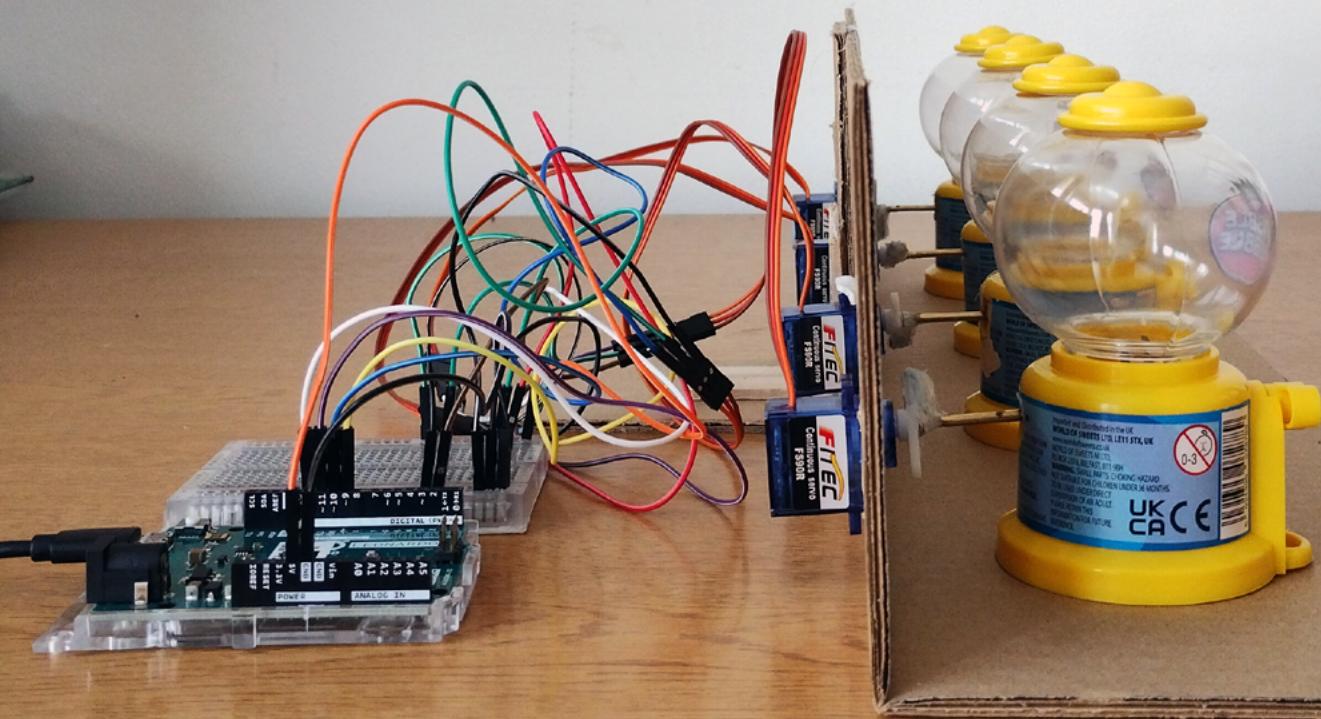


Table of contents

1. Introduction

2. Research process

2.1 Brainstorming mind map	5
2.2 Relevant readings and data resources	6
2.3 First project idea	9
2.4 Change on project idea	12

3. Design & Development - Code

3.1 Web Crawling	15
3.2 Web scraping and Data cleaning.....	16
3.3 Sentiment Analysis	17
3.4 Emotion Analysis	19
3.5 Word2Vector	23
3.6 Challenges	25

4. Design & Development - Physical output

4.1 Inspiration.....	28
4.1 Concept and wireframe	29
4.2 Prototype development	30
4.3 Final Outcome	34

5. Possible improvements

6. Bibliography

1. Introduction

The recent release of Chat GPT has gained widespread attention and has showcased the immense power of AI, revealing its disruptive nature. This sophisticated language model has generated diverse opinions and has raised concerns among many technologists.

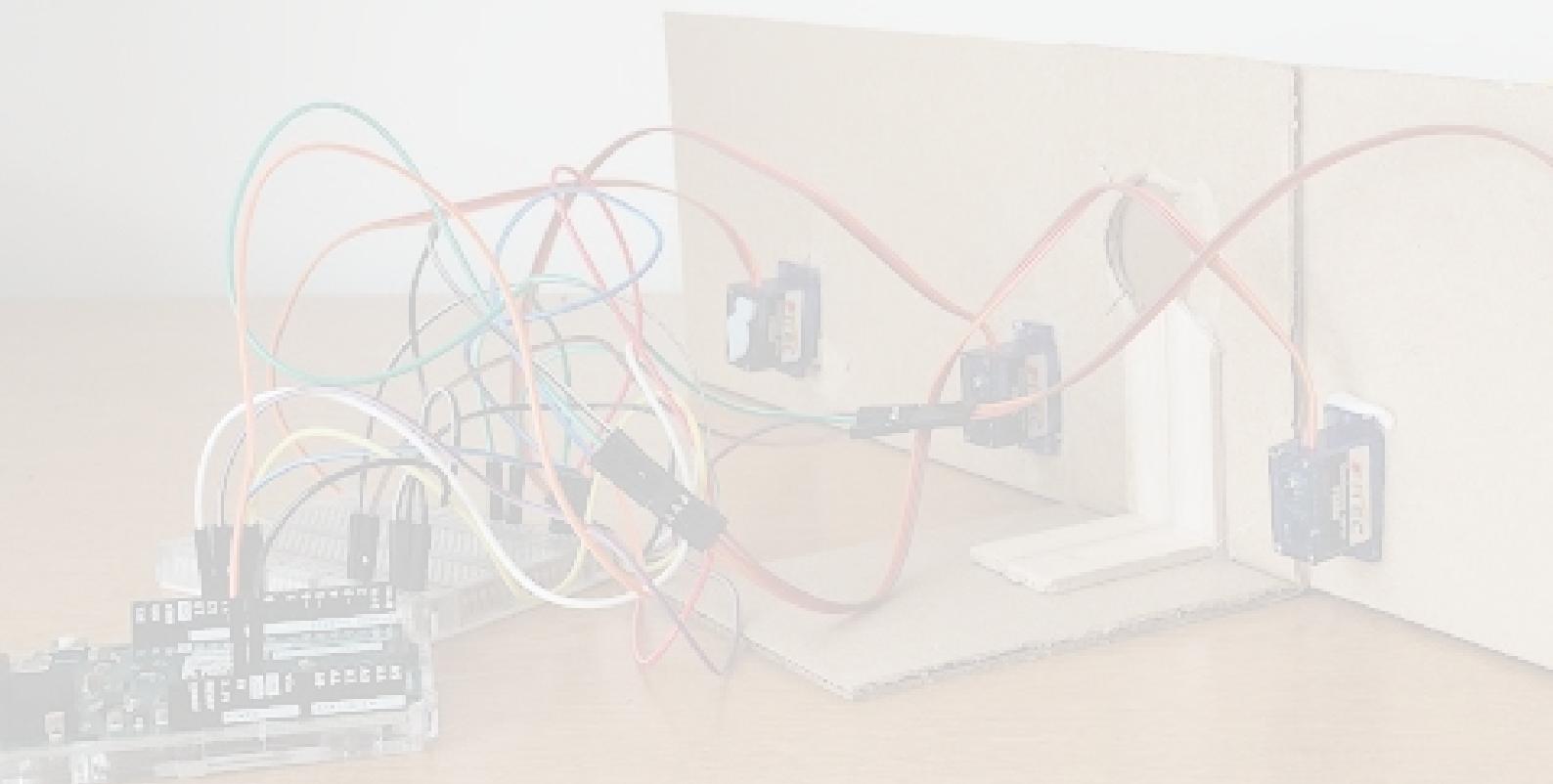
The launch of this powerful tool has ignited a race to develop even more advanced models in an industry that lacks sufficient regulation. Consequently, it has also given rise to new questions about the future of humanity. The resulting fear and apprehension led to the release of a public letter advocating for a six-month pause on large-scale AI experiments until legislation can catch up and establish better regulatory frameworks. The letter was signed by influential figures such as Elon Musk, Bart Selman, Yuval Harari, and other notable tech personalities (Future of Life, 2023).

Amidst these concerns, there is a significant discourse on how AI will impact the workplace and, specifically, the relationship between humans and work. Intrigued by this question, I embarked on a project to delve deeper into the public's opinion and gain a better understanding of the prevailing sentiments toward AI as it becomes more mainstream and a growing concern for everyone. To achieve this, I conducted sentiment and emotion analysis on a collection of web-scraped websites obtained through search queries like "**How are people feeling about AI in the workplace?**" or "**How are people feeling about AI taking over their jobs?**". After processing and cleaning the data, I extracted the general sentiment prevalent across the content.

Finally, the most prominent words identified through this analysis were imprinted on chocolates and incorporated into a physical structure designed to engage the public. This interactive structure aims to elicit individuals' real opinions on their feelings about the future of AI in the workplace. These opinions are being recorded via a dedicated Twitter account, creating a more reliable dataset.

2. Research process

- 2.1** Brainstorming mind map
- 2.2** Relevant readings and data resources
- 2.3** First project Idea
- 2.4** Change on project idea



2.1 Brainstorming mind map



2.2 Relevant readings and data resources

In my first attempt to understand what Big Data is, and what kind of data is out there to explore, I went on reading '**Confident data skills**' by Kirill Eremenko. During the reading, I was able to extract notes on examples of places that could provide open-source data. I also extracted some key steps of the data science process and started narrowing down on project ideas.

- Should we rather find a balance between the rate of technological growth and the rate that we can develop suitable ethical guidelines.

Examples of datasets / Resources:

- www.internationalgenome.org
- International dialects of English Archive.
- Note: some algorithms don't rely on data anymore but they rather "learns" from itself. Such as a program learning how to play Go.
- www.superdatascience.com
- Data governance: British Academy talks on Robotics, AI and Society. (2017)
- UK Government's Data Science Ethical Framework.
- World Bank Data
- European Union Open Data Portal
- The CIA World Factbook
-

The data science process

1) Identify the Question

- what is the goal of my analysis?
- what questions do I want to answer?

2) Prepare the Data

- Source • Clean • Prepare the data for analysis.

- Perform quality assurance verifications (?)

3) Analyze the Data

- Build models

- Perform data mining

- Run text analytics (?)

4) Visualize the Data

Translate complex insights into easy-to read visuals or animations

5) Present findings:

Translate my findings into "comprehensible" language.

i with data, we have the advantage of deriving our insights from actual evidence. we are not changing the information to suit our ideas but we are rather formulating ideas in order to derive insights.

1) Identify the Question

- Before we can prepare and analyse data, we must know what kind of data is that I need.
- To find this data I need to understand my main question and main goal.
- Such main question might need to be recalibrated in terms that the data will be able to understand.
- Some key questions to make:
 - where is the data located (?)
 - who is in charge of that data ?
 - what would success look like for the project

Possible Project Ideas.

→ web crawler (different to web scraping)

• web scraping, data scraping or content scraping is when a bot downloads the content. This might target specific pages /websites.

A web crawler systematically browses the Internet with the purpose of web-indexing.

• A web crawler that continually collects and stores data and finally visualize it.

• An installation which is based on a large dataset which I have scrapped.

(Physical manifestation of an online data source).

• A bot within a social media.

• A web browser extension (?)

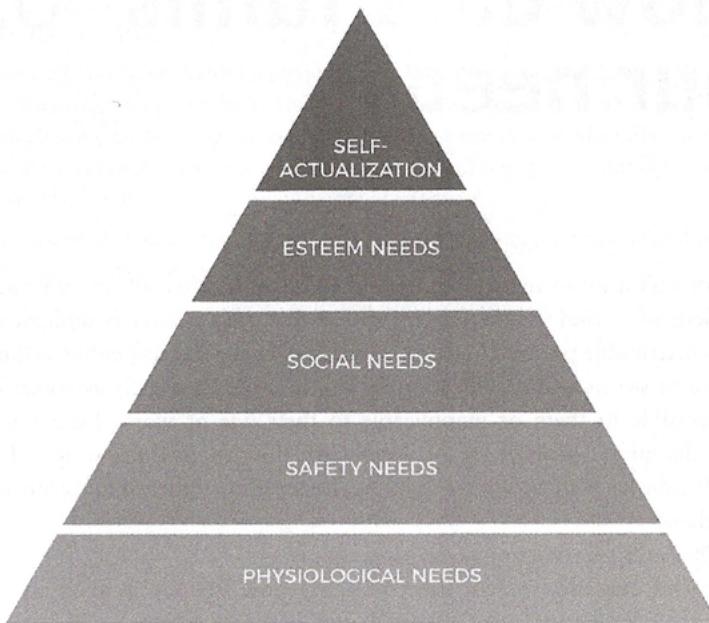
• Web crawling + VR Big data visualization ✓

How data fits Maslow's hierarchy of needs

26

Confident Data Skills

2.1 Maslow's hierarchy of needs



Data science and physiology

At the bottom of Maslow's hierarchy are physiological factors, the basic needs for humans to simply survive. How can data help with those most basic requirements? How can it improve upon them?

Let's take the air we breathe as an example. Air pollution has been a major global cause for concern ever since the Industrial Revolution of the late 18th and early 19th centuries. We might automatically imagine smog to be a phenomenon of the past, as in the case of London in the 1950s, when coal emissions regularly covered the city, but smog continues to affect a large number of places around the world, from China to Brazil.

Any technologies that are designed to reduce pollution in affected cities are reliant on data: to improve the condition of the air, it must first be monitored.

Figure 1. Maslow's hierarchy of needs taken from Confident data skills: Master the fundamentals of working with data and supercharge your career (Eremenko, 2018)

During my reading, something that caught my attention was the constant reminder that **data is everywhere**, and has always been. What has really changed is the way we process it and the technology we use to do so. The author was trying to explain how, the **pyramid of Maslow**, which is often studied in business courses, **can be applied to data**. Starting from the bottom as the most basic and important needs, Maslow identified the degree of importance for every 'need category'. **To each need, data could be applied**. For example, the 'safety needs' can be improved by the use of data in the healthcare sector to improve medicines and diagnoses. One sector where data is more obviously observed is 'Social needs' where social media and the data behind it 'pretends to bring us together under similar interests'.

Still feeling hesitant about the prospect of using AI in medicine?

Watson isn't the answer to all our problems, though. The machine's AI can still make mistakes. But the difference between machine doctors and human doctors is data, and as the technology to process growing quantities increases, so does the difference in ability between human and machine. After all, humans can absorb information from conferences, medical journals and articles, but we all have a finite capacity for storing knowledge. What's more, the knowledge that human doctors possess will largely be limited to their life experience. A machine doctor, on the other hand, can only get better the more data it is given. With instant access to data from other machines via the cloud, shared data can inform more accurate diagnoses and surgeries across the world. Thanks to exponential growth, these machines will have access to all manner of variations in the human body, leaving human knowledge flagging far behind.

Data science and belonging

After fulfilment of the **second stage of Maslow's hierarchy (safety)**, the need for belonging within a social environment (family, friends, relationships) will follow. It states that humans need to be part of a community of people who share their interests and outlook on life. The perceived disconnect between technology and society has been a topic of much discussion in recent years. The internet is often criticized as contributing to an increasingly isolated existence where our every whim and need is catered for. As an outdoorsy person, I won't make any case in support of socializing in the digital over the physical world. However, I do believe that the relatively democratic accessibility that the internet affords to people all over the world at all hours of the day is to my mind a great asset to human existence and experience.

What's more, what makes social networks such as Facebook, Instagram and LinkedIn successful is not the usability of the platform – it's their **data**. A badly subscribed social network is unlikely to offer the same breadth of services as a well-subscribed network **because social communication ultimately relies on relationships**. If the data isn't there to connect us to the right information, whether that means human connections, images that appeal to us, or news stories on subjects in which we are interested, the social network will not be useful to us.

Data is helping to make our world much more interconnected, and it is not only aiding us in personal ventures like finding old school friends; it is also helping scholars and practitioners who are carrying out similar projects to find each other and partner up.

CASE STUDY Forging connections through LinkedIn

I love using LinkedIn – and I think that they have really applied their data to benefit both themselves and their users. A quick visit to the business network's 'People You May Know' tab will show you an inexhaustible list of recommendations for connections with LinkedIn's other users. Some of these might be people at your current workplace, but you may also notice people from your university, and even school friends, cropping up on the system as recommended connections. To do this, LinkedIn uses the data you post to your profile – background, experience, education, existing colleagues – and matches it with the profiles of others.

LinkedIn's technology has enabled thousands of people to rebuild connections with their past. And as these connections grow, so does the network's data, thereby generating yet more connections. Whenever you connect with another user, not only do you gain what they call a 'first-degree connection' but their linked colleagues become 'second-degree connections', thereby expanding your circle much further than may be apparent.

For LinkedIn, as with any other social media channels, all that is essential is input from its users. I have found numerous friends and ex-classmates on the site, many of whom have since gone into the same field as me and, thanks to data's ability to match us, this has opened up a new dialogue between old acquaintances. Knowing that I have access to friends and colleagues online builds a sense of community and maintains it long after we have moved on, whether that be from a city or a place of work, and I find this interconnectedness comforting.

By connecting with others who share our interests, courses of study and location, LinkedIn can also give us a good insight into jobs that are relevant to us. When I was in the market for a new job, I started posting status updates to LinkedIn – the platform's data algorithms identified my needs according to key words that I had used, and this is how recruiters started to find me. What was even better was that since I was writing about subjects that interested me, LinkedIn's algorithms matched me to jobs that specifically required those branches of knowledge. It was even how this book's commissioning editor found me.

How's that for social media channels' abilities to improve happiness?

Figure 2. Annotations on 'How data fulfils our needs' taken from Confident data skills: Master the fundamentals of working with data and supercharge your career (Eremenko, 2018)

Data science and belonging.

The author of the book was forcefully trying to show the benefits of data in the realm of 'belonging' and 'communities'. Even though he does acknowledge that the Internet has been accused of making our personal relationships meaningless, he does explore some cases where social media has been useful.

Something that is very important in his view is how every person acts as a node in the network, and how interactions between those nodes can only generate more and more data. It does show the power of interconnectedness that we have, and it does help some people to find good connections. However, the question still remains as to how ethical is the use of that data and up to what point it is healthy to collect data from every interaction that we can possibly have.

2.3 First project idea - VRNetzer

Attending the Imperial Lates helped me to gauge the different programs and technologies that practitioners are using. The main focus of the event was 'Play' and it presented games as a medium for learning.

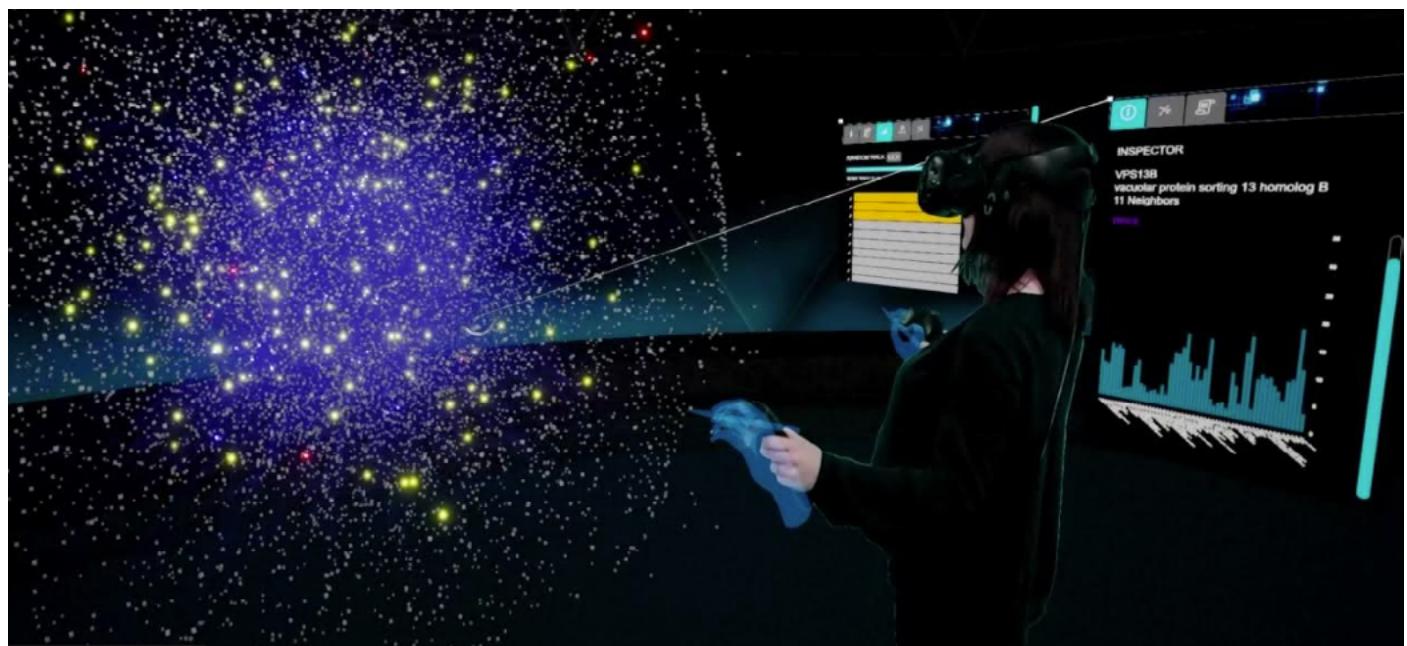
In a conversation with one of the presenters, I got introduced to the tool VRNetzer, an Open Source project that allows the visualization of Networks.

In the experience you could see different clusters and zoom in and out to select which node you were interested in.



Figure. Imperial lates event banner

This made me remember a section of a book that presented a case study for Big Data visualisation through VR. Given the nature of the dataset I have found around Talent Migration around the world, I picture a world Network, or clusters that present the movement of people across the globe.



(2021) VRNetzer Autocore. menchelab. Available at: <https://www.youtube.com/watch?v=Pd46211gc9U&t=5s> (Accessed: March 22, 2023).

During the first setup of the VRNetzer environment, I presented some issues with the visualisation of the data. I could access the environment, but I couldn't see the controllers or any preloaded data point. I used a VIVE index headset, which haven't been tested by the developers of the project. So, I will be reviewing the executable file and load some data following the github instructions (<https://github.com/menchelab/VRNetzer>) The issue might also concern the headset itself, and I should test on the HTC VIVE.

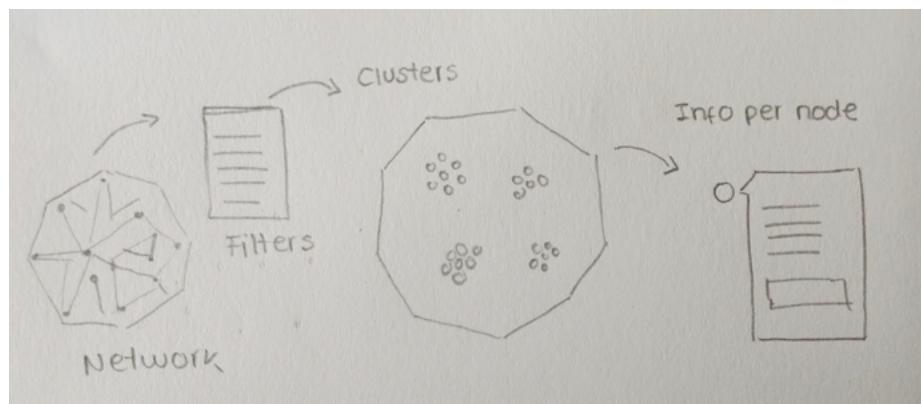
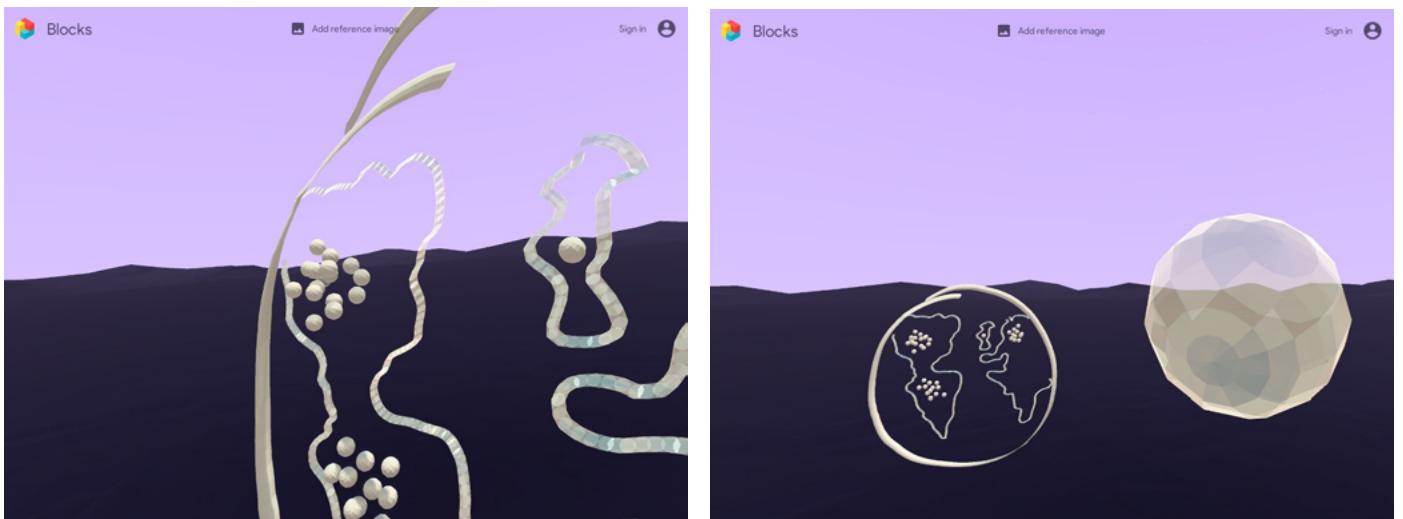
The main point of testing this early is to understand if it would be a functional medium to present my data. I want to understand how much control I have on the platform.

Mockups in Blocks-Google VR

I have also experimented with Blocks VR. The motive of these quick mockups was to interact with the space and understand how the clusters might look like. I tried different shapes and colours to differentiate the clusters.



Then I went into thinking how would those clusters look like within the earth. I clearly was thinking of it in 2D and then moved to a 3D sphere. I can picture the nodes within the sphere in groups that perhaps represent continents, and then those nodes moving from cluster to cluster according to changes in the data (triggered by the user).



General functioning flow idea

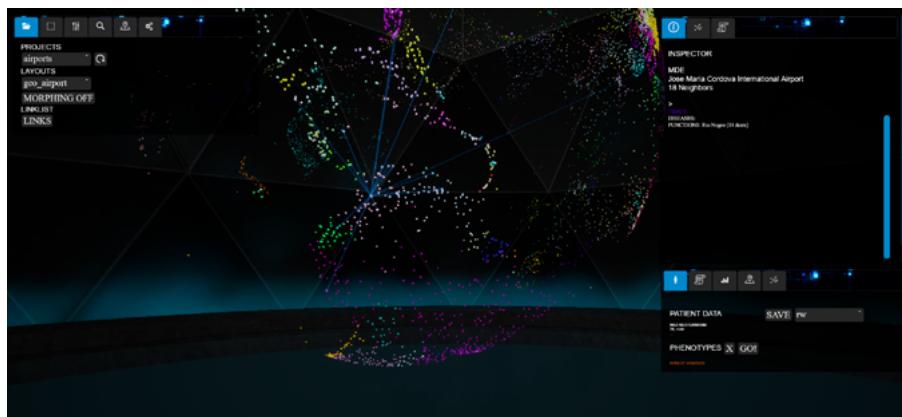
10 Initial Network -> Filters by user -> New clusters -> Information per node

First look at the VRNetzer Analysis platform

The platform uses a UI module, an Analytics module, and a VR module. To handle the data, it uses **MySQL workbench** from where we can create a new database and add new data through a **POST request** (generated in the UI module). As a first example, the platform shows a network of the world's airports. It does resemble the idea I had in mind, however, it could be expanded with interactions that would come from the LinkedIn dataset.

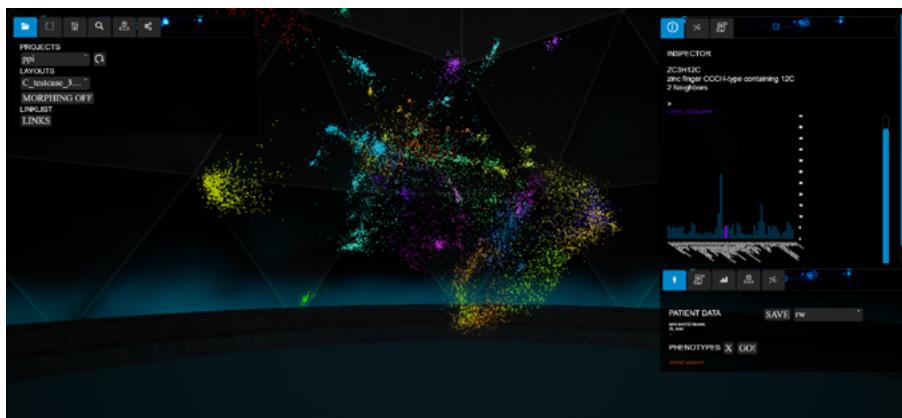


Nodes can show their immediate links



- Nodes can be clicked on, exposing visually some of the strongest links with other nodes (called neighbours).
- Clicking on the node also gives extra information about the node itself.

3D clusters. Exploring a node can give further information.



The nodes can be organized in different **layouts**. These layouts can be 2D or 3D and they can be selected by the user. The different layouts represent a change in the **CSV file**, and how data is positioned.



Challenges and Formative Assignment feedback

- Even though I had a formed project idea, it wasn't fully meeting the brief. I needed to include a 'web-based experiment' and I found that the amount of data I needed to 'embellish' the dataset I had found online was very challenging. The LinkedIn API was also very restrictive with the information we could extract, and to GET relevant data, I had to fill in an application to become a 'LinkedIn Partner'.

Furthermore, during the following classes, we got to explore '**sentiment analysis**' and how we could find the percentage of negativity or positivity of words. We also experimented with **posting to Twitter**, mainly because accessing the Twitter API to GET information had become a lot more restricted. This resulted in a new project idea that combines the different techniques we have learnt in class.

Those are the reasons why I have changed my initial idea, to a new project that still explores the subject of 'jobs' and 'talent'. However, I intend to explore how people are currently perceiving Artificial Intelligence in the workplace now, and in the future.

 [Link to Formative Assignment](#)

! 2.4 Change on project idea

How do people feel about AI, especially around the subject of job automation?

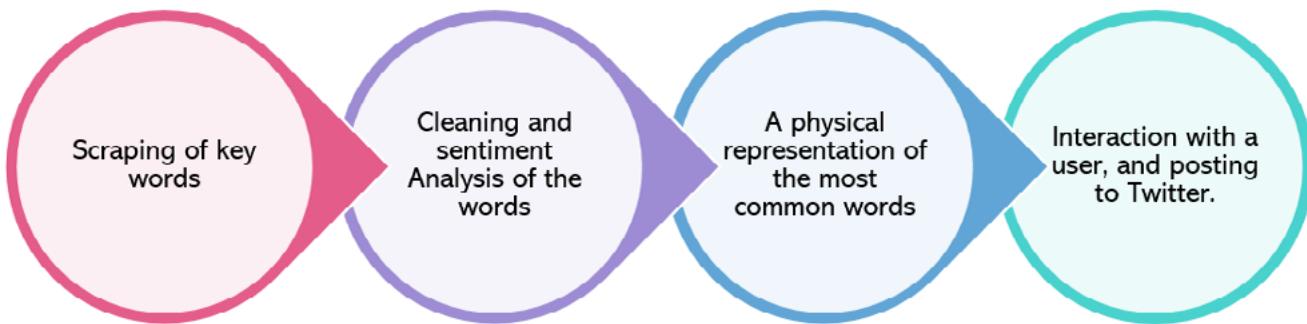


The recent release of Chat GPT has reached many users and has unveiled the power of AI and its disruptive nature. This large language model has caused many opinions and has sparked caution amongst many technologists.

The quick development of even more advanced models resulted in the publication of an 'Open Letter' which advocates for the pause of giant AI experiments for 6 months.

Such letter was signed by the likes of Elon Musk, Bart Selman, Yuval Harari, etc.

How should it work?



Scraping of key words related to feelings:

Using the '**Pause Giant AI Experiments**' open letter as a starting point to extract key words. Then, navigate to hyperlinks present within the letter to keep extracting more words related to 'feelings'

Cleaning and Sentiment Analysis:

Counting the frequency of the words previously found and assessing how positive or negative they are.

Physical Representation:

Select the 8 most prominent words and accomodate:

- 2 words per candy in 4 designs:

perform 'word to vector' to decide which words should go together in one candy.

- **Two colours:** red and blue, making reference to the iconic Matrix scene.

- **max 9 characters:** using 'Lemmatization', to reduce the word.

Interaction with an user

Using a small gum ball machine as input for an user to express how the feel about AI, using the words that were previously printed in the chocolates.

By taking out a chocolate they will be also posting their opion on Twitter.

3. Design & Development

Data and code*

3.1 Web Crawling

3.2 Web scraping and Data cleaning

3.3 Sentiment Analysis

3.4 Emotion Analysis

3.5 Word2Vector

3.6 Challenges

*Please notice that there are links included in the documentation.
It is recommended to right-click on them and open in a new tab to see their content.

3.1 Web Crawling

I have selected a series of websites (seeds) found under the search query: “**How are people feeling about AI in the workplace?**” or “**How are people feeling about AI taking up their job?**” I made sure these links had meaningful content I could use and performed further web crawling within them.

Challenges

The main aim of crawling within the selected seeds was to find more related content. However, as the crawler continued to find more links within links, it would quickly **lose track of the main subject**. To minimise this, I filtered down the links that my crawler had found to relevant sources.

Sample code for web crawling

First seed: Pause giant AI experiments

<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

Sample of links found during the crawling process.

```
#main-content
https://futureoflife.org/
http://futureoflife.org/our-mission/
#dropdown
https://futureoflife.org/cause-areas/
https://futureoflife.org/cause-area/artificial-intelligence/
https://futureoflife.org/cause-area/biotechnology/
https://futureoflife.org/cause-area/nuclear/
https://futureoflife.org/cause-area/climate/
#dropdown
https://futureoflife.org/our-work/
https://futureoflife.org/our-work/policy-work/
https://futureoflife.org/our-work/outreach-work/
https://futureoflife.org/our-work/grantmaking-work/
https://futureoflife.org/our-work/events-work/
https://futureoflife.org/project/mitigating-the-risks-of-ai-integration-in-nuclear-launch/
https://futureoflife.org/fli-area/policy/
https://futureoflife.org/project/future-of-life-award/
```

Example of filtering down to relevant links

```
filtered_first_seed = [link for link in links if link.startswith((
    'https://futureoflife.org/cause-area/artificial-intelligence/',
    'https://abcnews.go.com/',
    'https://arxiv.org/abs/',
    'https://openai.com/',
    'https://futureoflife.org/ai/',
    'https://futureoflife.org/open-letter/ai-principles/',
    'https://time.com/'))]
```

After finding all my links, I saved them into a **CSV**. This will help me keep track of the links I have crawled into and allow me to iterate over them during the cleaning process.



[Link to Web Crawling code](#)

3.2 Web scraping and data cleaning

Using the CSV I had previously created, I extracted all the relevant data from the HTML body of the links and performed some data cleaning. I wrote a series of functions that remove:

- Special characters
- Stopwords
- Uppercase
- Numbers found within String data types
- Empty spaces
- Words that have a length less than 1 and more than 12 characters.

Challenges: cleaning the data up to an acceptable point was challenging. Removing certain characters would create empty spaces, which would need to be later removed. Even with the cleaning functions in place, I was able to see some words bypassing the checks.

Sample code for cleaned data.

First seed: Pause giant AI experiments

<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

```
for individual_url in dataFrame['links'].apply(lambda row: []):
    #   dataFrame['Tokens'] = pd.array(final_clean(second_clean(clean(tokenized(individual_url)))))
    print(individual_url, final_clean(second_clean(clean(tokenized(individual_url)))))

https://futureoflife.org/cause-area/artificial-intelligence/ ['artificial', 'intelligence', 'future', 'life', 'institute', 'skip', 'content', 'mission', 'cause', 'areascause', 'area', 'overviewartificial', 'intelligencebiotechnologynuclear', 'weaponsclimate', 'change', 'workour', 'work', 'overviewpolicyoutreachgrantmakingeventscurrent', 'projectsmitigating', 'risks', 'ai', 'integration', 'nuclear', 'launchpolicyfuture', 'life', 'awardoutreachstrengthening', 'nist', 'ai', 'risk', 'management', 'frameworkpolicystrengthening', 'european', 'ai', 'actpolicyworldbuilding', 'competitionoutreacheducating', 'lethal', 'autonomous', 'weaponsoutreach', 'policyfuture', 'life', 'institute', 'podcastoutreach', 'contentarticles', 'podcasts', 'newsletters', 'resources', 'usabout', 'us', 'overviewour', 'peoplecareersdonatefundingfaqscontact', 'us', 'lake', 'action', 'search', 'tak', 'e', 'action', 'home', 'artificial', 'intelligence', 'cause', 'areaartificial', 'intelligencefrom', 'recommender', 'algorithm', 'self', 'driving', 'cars', 'ai', 'changing', 'lives', 'impact', 'technology', 'magnifies', 'risks', 'artificial', 'intelligence', 'encompasses', 'everything', 'recommender', 'algorithms', 'self', 'driving', 'cars', 'racing', 'forward', 'today', 'narrow', 'ai', 'systems', 'perform', 'isolated', 'tasks', 'already', 'pose', 'major', 'risks', 'erosion', 'democratic', 'processes', 'financial', 'flash', 'crashes', 'arms', 'race', 'autonomous', 'weapons', 'looking', 'ahead', 'many', 'researchers', 'pursuing', 'agi', 'general', 'ai', 'perform', 'well', 'better', 'humans', 'wide', 'range', 'cognitive', 'tasks', 'ai', 'systems', 'design', 'smarter', 'systems', 'may', 'hit', 'intelligence', 'explosion', 'quickly', 'leaving', 'humanity', 'behind', 'could', 'eradicate', 'poverty', 'war', 'could', 'also', 'eradicate', 'us', 'risk', 'comes', 'ai', 'potential', 'malevolence', 'consciousness', 'competence', 'words', 'feels', 'humans', 'could', 'instance', 'lose', 'control', 'high', 'performing', 'system', 'programmed', 'something', 'destructive', 'devastating', 'impact', 'even', 'ai', 'programmed', 'something', 'beneficial', 'could', 'still', 'develop', 'destructive', 'method', 'achieve', 'goal', 'ai', 'need', 'consciousness', 'pursue', 'goals', 'heat', 'seeking', 'missiles', 'equally', 'danger', 'robots', 'per', 'se', 'intelligence', 'requires', 'nothing', 'inter
```

After finding all the relevant words per link, I added them into the existing CSV I had previously created with the purpose of keeping everything organised.



[Link to Cleaning code](#)

3.3 Sentiment Analysis

For this analysis I have used the CSV I previously created (Cleaned_data.csv). From this CSV I extracted the 'tokens' field and created a function that would remove the commas and the quotation marks. This step caused me a lot of trouble since I was working with a pandas data frame. I later learnt that I needed to define the 'tokens' field as an array, to be able to work with it.

Sample code:

The 'stringToken' field is the result of having a list of strings and then applying the function freedTokens to have a long string. The purpose of having a long string is to have the data in the right format to use the nltk sentiment functions.

```
##converting my DataFrame column into a numpy array
tokenList = DataFrame['tokens'].to_numpy()
#print(tokenList)

def freedTokens(tokenList):
    modifiedList = []
    for token in tokenList:
        #print(token)
        remove = re.sub(r'[,\n]', ' ', token)
        modifiedList.append(remove)
    return modifiedList
```

Main Challenge: as mentioned, in order to apply str functions to the values of the 'tokens' field, I had to first convert the DataFrame['tokens'] into a numpy array. Then, this new list of type numpy array could go into the function freedTokens and then output the required string.

Data frame ready for sentiment analysis

	web_id	links	tokens	stringToken
0	0	https://futureoflife.org/open-letter/pause-gia...	['pause', 'giant', 'ai', 'experiments', 'open'...]	[pause giant ai experiments open letter future...]
1	1	https://futureoflife.org/cause-area/artificial...	['artificial', 'future', 'life', 'institute', ...]	[artificial future life institute skip content...]
2	2	https://futureoflife.org/open-letter/ai-princi...	['ai', 'principles', 'future', 'life', 'institu...']	[ai principles future life institute skip cont...]
3	3	https://openai.com/blog/planning-for-agi-and-b...	['planning', 'agi', 'beyond', 'closesearch', '...']	[planning agi beyond closesearch submit skip m...]
4	4	https://futureoflife.org/ai/faqs-about-flis-op...	['faqs', 'fli', 'open', 'letter', 'calling', '...']	[faqs fli open letter calling pause giant ai e...]
...
108	108	https://www.mckinsey.com/featured-insights/fut...	['ai', 'automation', 'future', 'work', 'ten', ...]	[ai automation future work ten things solve te...]
109	109	https://www.bbntimes.com/technology/advantages...	['advantages', 'ai', 'workplace', 'advantages'...]	[advantages ai workplace advantages ai workpla...]
110	110	https://en.wikipedia.org/wiki/Workplace_impact...	['workplace', 'impact', 'artificial', 'wikiped...']	[workplace impact artificial wikipedia jump co...]
111	111	https://www.akerman.com/en/perspectives/hr-def...	['hr', 'def', 'â\x80\x9csmartâ\x80\x9d', 'usin...']	[hr def â\x80\x9csmartâ\x80\x9d using artifici...]
112	112	https://planenergy.com/blog/how-ai-is-transformi...	['forbidden', 'forbidden', 'nginx']	[forbidden forbidden nginx]



[Link to Sentiment Analysis code](#)

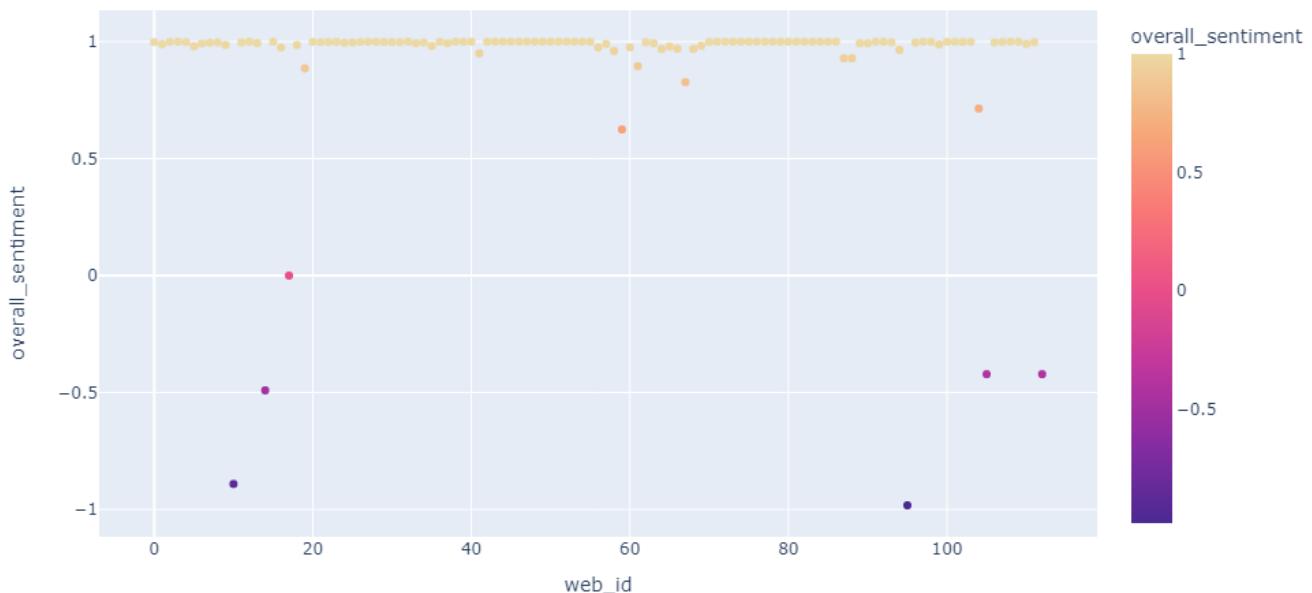
For the sentiment analysis itself, I used the `nltk.sentiment.vader` library and applied the `sentimenta.polarity_scores` function to the 'stringToken' field. I found the positive, negative and neutral sentiment for each weblink. To be able to visualise the result I also found the overall sentiment score (average between positive, negative and neutral scores), and gave an id to each weblink.

Sample code for sentiment analysis results

web_id		links	tokens	stringToken	positive sentiment	negative sentiment	neutral	overall_sentiment
0	0	https://futureoflife.org/open-letter/pause-giant-ai-experiments/	['pause', 'giant', 'ai', 'experiments', 'open'...]	[pause giant ai experiments open letter future...]	0.180	0.096	0.724	0.9980
1	1	https://futureoflife.org/cause-area/artificial-intelligence/	['artificial', 'future', 'life', 'Institute', ...]	[artificial future life Institute skip content...]	0.167	0.126	0.708	0.9897
2	2	https://futureoflife.org/open-letter/ai-principles/	['ai', 'principles', 'future', 'life', 'institu...'...	[ai principles future life institute skip cont...]	0.224	0.078	0.699	0.9992
3	3	https://openai.com/blog/planning-for-agi-and-best-practices/	['planning', 'agi', 'beyond', 'closesearch', '...'...	[planning agi beyond closesearch submit skip m...]	0.261	0.062	0.677	0.9996
4	4	https://futureoflife.org/ai/faqs-about-fliis-open-letter/	['faqs', 'fli', 'open', 'letter', 'calling', '...'...	[faqs fli open letter calling pause giant ai e...]	0.185	0.088	0.727	0.9978
...
108	108	https://www.mckinsey.com/featured-insights/future-of-work/ten-things-to-solve-the-workplace-automation-challenge	['ai', 'automation', 'future', 'work', 'ten', ...]	[ai automation future work ten things solve te...]	0.207	0.047	0.746	0.9998
109	109	https://www.bbntimes.com/technology/advantages-of-ai-in-the-workplace/	['advantages', 'ai', 'workplace', 'advantages'...]	[advantages ai workplace advantages ai workpla...]	0.218	0.047	0.735	0.9995
110	110	https://en.wikipedia.org/wiki/Workplace_impact	['workplace', 'impact', 'artificial', 'wikiped...'...	[workplace impact artificial wikipedia jump co...]	0.120	0.106	0.775	0.9905

Data visualisation of the results

This scatter plot shows how positive, negative or neutral is the content of the weblinks I had previously scraped. There's a great amount that tend to be positive (to my surprise)



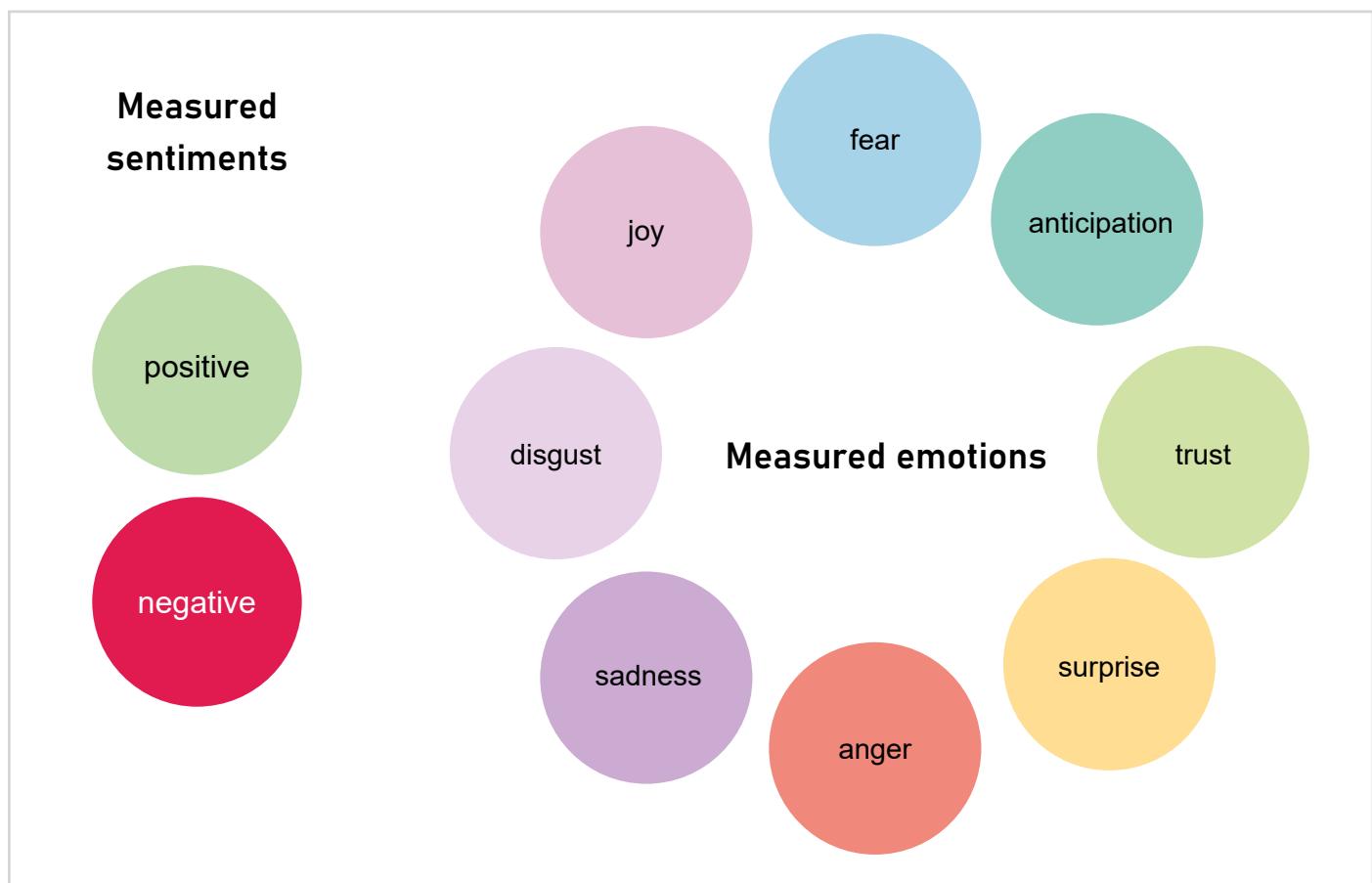
[Link to Sentiment Analysis code](#)

3.4 Emotion Analysis

My initial idea was to extract from the token list, every possible word that could be classified as a feeling/emotion. The purpose was to find the frequency (if any) of those words, and therefore understand the **general feeling** on all these articles about AI. However, extracting certain words might not communicate the real ‘sentiment’ that the text is intending to communicate.

I understood that by extracting very specific words, I wasn’t taking into account the general context of those words, and therefore I wasn’t conveying the overall sentiment of the text. While I was hoping for negative sentiment on almost every weblink, I found that many were neutral or positive. The ‘weight’ of meaning in the words must balance all these articles towards positiveness, even if they contain a few strong negative words such as ‘anxiety’, ‘fear’ and ‘apocalyptic’.

This is how I encountered the **NRCLex python project** which measures **emotional affect** from a body of text. The affect dictionary contains approximately 27,000 words and is based on the National Research Council Canada **affect lexicon** and the NLTK library’s WordNet synonym sets. (According to their official documentation: <https://pypi.org/project/NRCLex/>)



Text can be measured within 8 emotions and 2 sentiments (interesting concept separation)
The text analysed will have a score for each emotion and sentiment.

Sample code for emotional analysis results

After joining my list of tokens back into a whole string, I was able to apply the raw_emotions_scores function to the text. The result is an overall view to the different 'emotion scores' that I get for all the data I have collected across all the different weblinks. **Similar to the sentiment analysis** done with the NLTK library, we get a prominent positive sentiment, while negative sentiment is considerably low.

```
emotions = text.raw_emotion_scores  
print(emotions)  
  
{'fear': 3845, 'anticipation': 7220, 'trust': 11595, 'negative': 6444, 'joy': 4110, 'positive': 20930, 'disgust': 1075, 'sadness': 2768, 'anger': 2427, 'surprise': 2004}
```

Results in descending order

Emotion_Classification	Emotion_Count	
5	positive	20930
2	trust	11595
1	anticipation	7220
3	negative	6444
4	joy	4110
0	fear	3845
7	sadness	2768
8	anger	2427
9	surprise	2004
6	disgust	1075

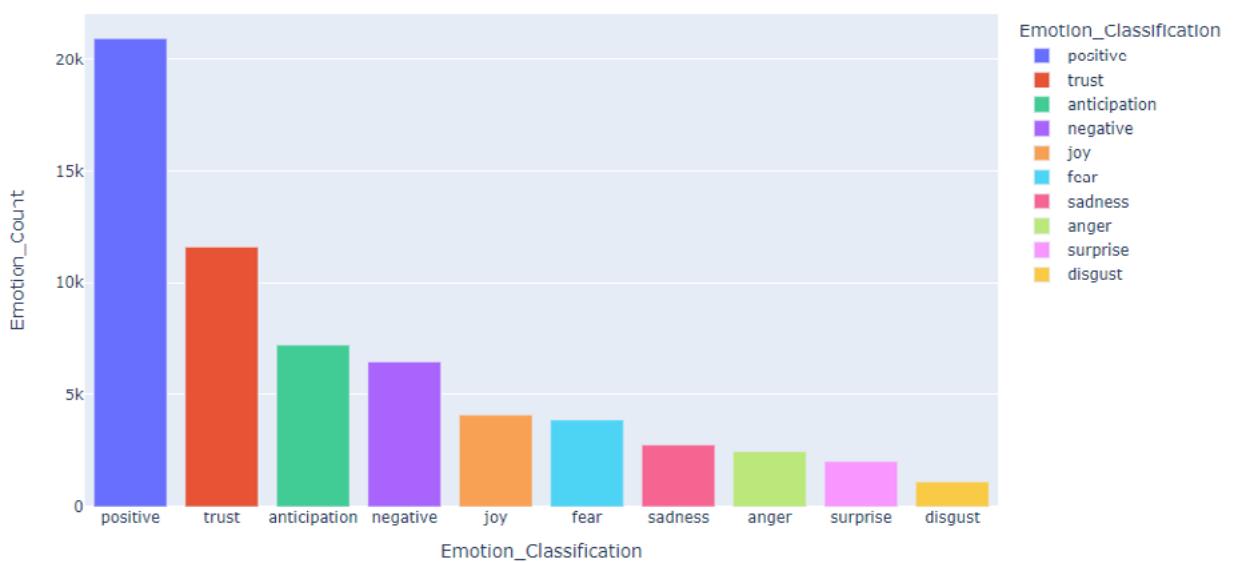
What is really powerful about these scores is the extra information that they offer on the distribution of 'emotions' that the text contains. It is a lot richer than the 'neutral', 'positive' and 'negative' labels.

What really surprised me, was to see the high scores for 'trust' and 'positive'. **Let's remember that the main search sentences were:**

- How are people feeling about AI in the workplace?"
- How are people feeling about AI taking up their job?"

Emotion Analysis

Perceived **spectrum of emotions** on the subject 'How do people feel about AI in the workplace?'



[Link to Emotion Analysis code](#)

Results Analysis

Initial hypothesis: the general sentiment in the articles should have leaned towards a negative sentiment, since the search queries indicated a negative connotation of AI in the workplace. Some of the titles I found helped to feed this hypothesis:

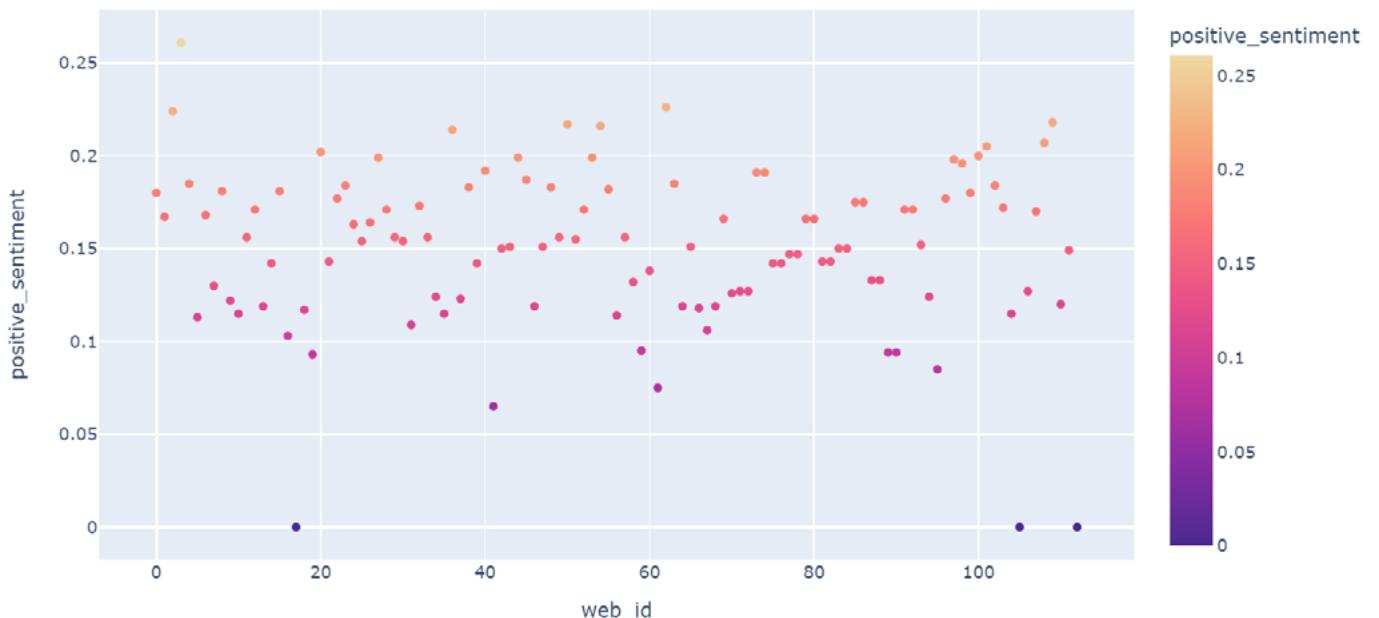
- [Welcome to the age of ‘AI-nxiety,’ in which anxiety about AI taking over our jobs and lives is at the top of everybody’s minds](#)
- [Machine anxiety: How to reduce confusion and fear about AI technology](#)
- [How has public sentiment towards AI evolved, and how should we inform/educate the public?](#)
- [Artificial Intelligence in the workplace: what is at Stake for workers?](#)

However, under the same search queries, I found a lot of articles that were trying to ease the general fear and negative opinions towards AI. Many of them were news articles or services whose main product had some AI integration. This made me consider how the google algorithm will be inevitably affected by companies trying to reach the top 10 search results, implying their already **‘embedded bias’ towards a positive sentiment about AI**. It is in their interest to switch that negative mentality of the consumer and give them confidence and comfort.

I also noticed that many of the articles highlighted not only the negatives but also the positives of AI, pushing the final results towards ‘positive’.

Sentiment Analysis

Perceived **positive sentiment** on the subject ‘How do people feel about AI in the workplace?’



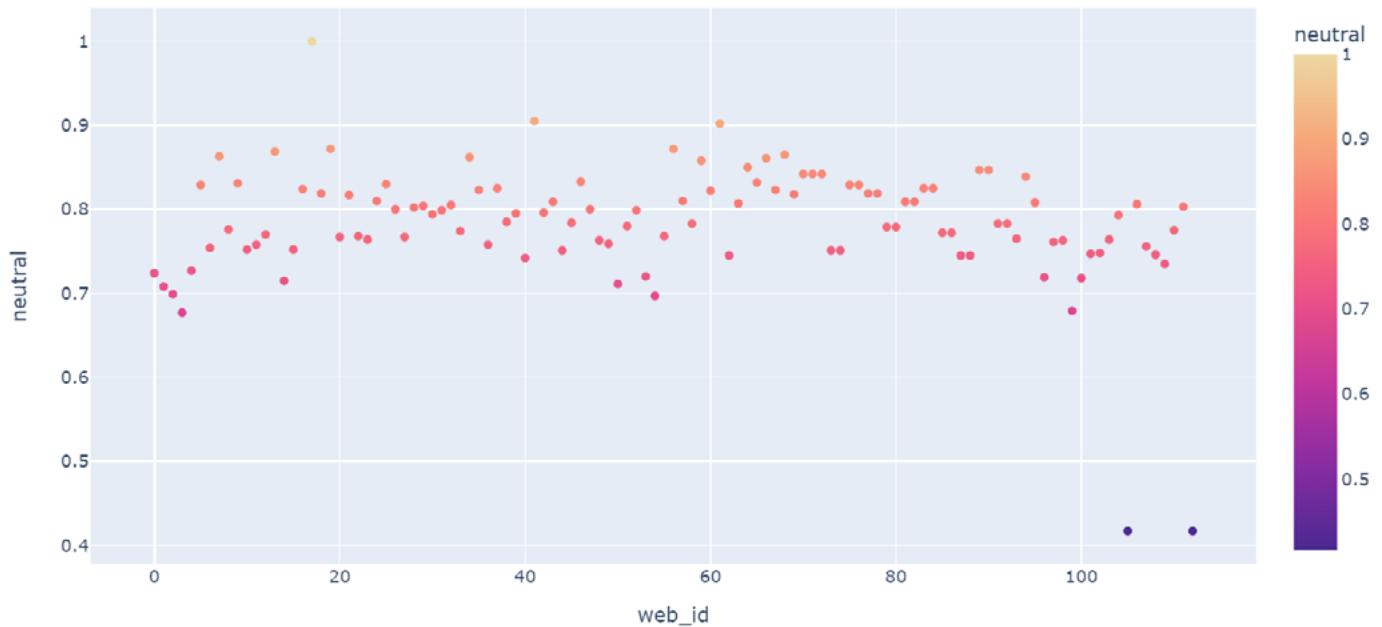
The general sentiment about AI, and AI in the workplace moves mostly within the positive range, although only up to score 0.3 (approximately)

There's relevant dispersion between 0 (neutral) and 0.3 (positive).

However, when performing the sentiment analysis alone, I could notably identify ‘neutral’ sentiment score to be higher compared to ‘negative’ and ‘positive’ scores on almost every weblink. This neutrality makes me think that only very strong words, and its frequency pushes the general sentiment up.

Sentiment Analysis

Perceived **neutral sentiment** on the subject ‘How do people feel about AI in the workplace?’



The general sentiment about AI, and AI in the workplace moves mostly within the positive range. Neutral sentiment moves within this range, and we can notice how the **relevant dispersion moves between 0 (neutral) and 1 (positive)**. As opposed to the overall positive sentiment, neutral does have higher scores.

This makes us think that the neutral scores do contribute to the overall sentiment, pushing ‘positive’ higher.

! New hypothesis:

The general sentiment around AI can't be accurately defined by the scraping of weblinks, since it is not a reflection of how people really feel and think about a subject matter. It rather gives us an insight on how the mass media and AI companies want to shift our impressions, and almost **sell a positive view** around the future of AI. But **how do people really feel about AI?** I think this is where my physical piece will help to get input from people.

3.5 Word2Vector

This final stage of the treatment I have given to my data was rather challenging and it took me through some understanding of what was the real purpose of the words I had extracted.

Let's remember that there is a physical constraint with the amount of words I can print.

My first instinct was doing some word-to-vector analysis to the top 8 emotions that described the general mood of my extracted content. Making use of the **spacy library** to find what words were similar by looking at their **cosine similarity**, I found that words like Positive and Negative were very close to each other.

Cosine similarity with a value over 0.6 - Spacy

Using the Spacy library, we get the words that are close in vector space yet, not in meaning.

```
cosine similarity between positive and positive is: 1.0
cosine similarity between positive and negative is: 0.8790249824523926
cosine similarity between trust and trust is: 1.0
cosine similarity between anticipation and anticipation is: 1.0
cosine similarity between negative and positive is: 0.8790249824523926
cosine similarity between negative and negative is: 1.0
cosine similarity between joy and joy is: 1.0
cosine similarity between joy and sadness is: 0.6499127149581909
```

Gensim - Word2Vec

I tried using a different library for word-to-vector called Gensim, mainly to understand if the training data of Spacy was causing the issue. However, I got a similar behaviour with words that are 'meant to' be far away from each other.

Cosine similarity - Gensim

Using the Gensim library, we get a similar result. (The range seems to change from [0-1] to be a lot more granular, with values like 0.002)

```
similarity= model.wv.most_similar("negative", topn=3)
for i in similarity:
    print(i)

('positive', 0.0040717096999287605)
('trust', 0.0024236778263002634)
('sadness', 0.0012841515708714724)
```

Vector size

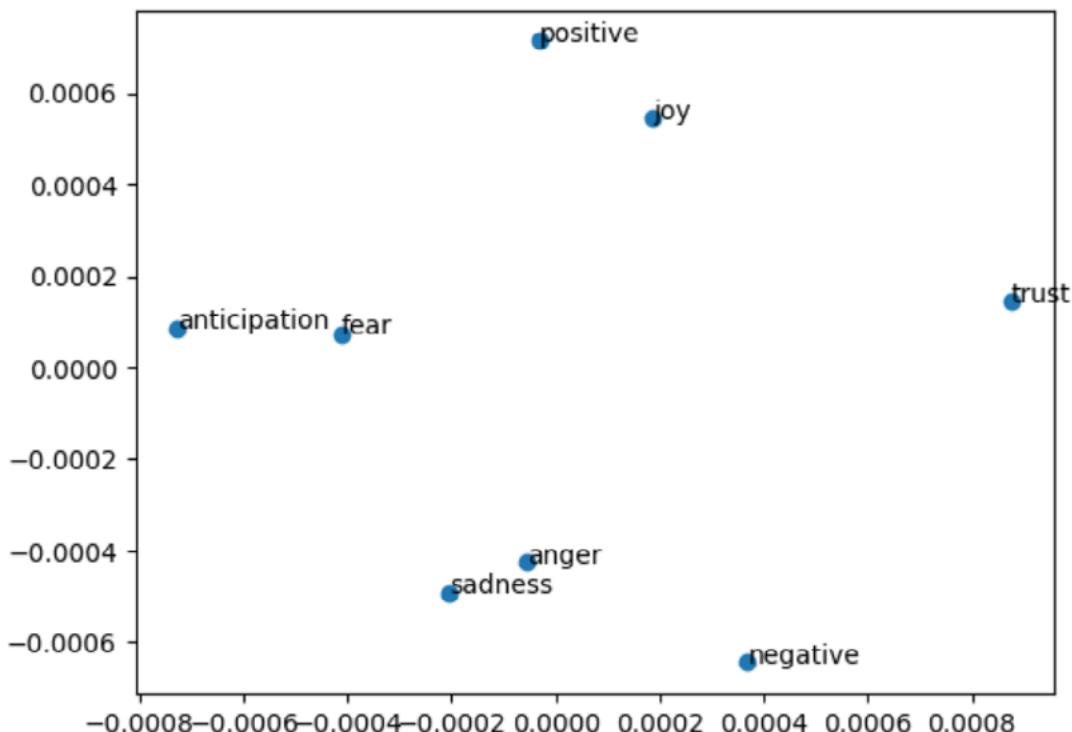
The key variable that I changed to get a (almost) desired output was the 'vector_size'. By changing the vector size I was changing the length and direction of the vectors in the word embedding space. By increasing the value, I would increase the level of detail in the word embedding, which is how I get to pair up the words.

Semantic Relationship

Word embedding for the emotions and sentiments on the subject

'How do people feel about AI in the workplace?'

```
: pyplot.scatter(result[:, 0], result[:, 1])
for i, word in enumerate(tokensvtwo):
    pyplot.annotate(word, xy=(result[i, 0], result[i, 1]))
pyplot.show()
```



designs Final word 'pair-up'

Positive
Joy

Negative
Fear

Anger
Sadness

Anticipation
Trust



[Link to Word2Vec code](#)

3.6 Challenges

Websites blocking web crawling.

During the process of sentiment analysis I was able to spot that some of the websites had a very negative score. By looking at those very specific links, I noticed that the negative score came from the fact that the website was being blocked, and the sentiment analysis was catching the negative wording used ("denied", "error", "prevent", etc). This meant that I couldn't get content out of them and therefore I had to remove those links.

Sample code for wording used in blocked websites

```
https://cybernews.com/editorial/chatgpt-takes-your-job-experts-give-their-verdict/ ['access', 'denied', 'access', 'denied', 'error', 'code', 'access', 'cybernews', 'com', 'the', 'site', 'owner', 'may', 'set', 'prevent', 'accessing', 'site', 'error', 'details', 'provide', 'site', 'owner', 'information', 'got', 'error', 'visiting', 'cybernews', 'com', 'editorial', 'chat gpt', 'takes', 'your', 'job', 'experts', 'give', 'their', 'verdict', 'error', 'code', 'ray', 'id', 'cddbdbab', 'country', 'gb', 'data', 'center', 'lhr', 'ip', 'cc', 'bca', 'timestamp', 'utc', 'click', 'copy', 'page', 'helpful', 'yes', 'thank', 'feedback', 'performance', 'security', 'cloudflare']
```

```
where', 'enter']
```

```
https://planergy.com/blog/how-ai-is-transforming-the-workplace/ ['forbidden', 'forbidden', 'nginx']
```

Sample data frame of sentiment analysis catching blocked websites

web_id	links	tokens	stringToken	positive sentiment	negative sentiment	neutral	overall_sentiment
0	0 https://futureoflife.org/open-letter/pause-gia...	['pause', 'giant', 'ai', 'experiments', 'open'...]	[pause giant ai experiments open letter future...]	0.180	0.096	0.724	0.9980
1	1 https://futureoflife.org/cause-area/artificial...	['artificial', 'future', 'life', 'institute', ...]	[artificial future life institute skip content...]	0.167	0.126	0.708	0.9897
2	2 https://futureoflife.org/open-letter/ai-princi...	['ai', 'principles', 'future', 'life', 'Institu...'	[ai principles future life institute skip cont...]	0.224	0.078	0.699	0.9992
3	3 https://openai.com/blog/planning-for-agi-and-b...	['planning', 'agi', 'beyond', 'closesearch', ...]	[planning agi beyond closesearch submit skip m...]	0.261	0.062	0.677	0.9996
4	4 https://futureoflife.org/ai/faqs-about-flis-op...	['faqs', 'fli', 'open', 'letter', 'calling', ...]	[faqs fli open letter calling pause giant ai e...]	0.191	0.086	0.723	0.9983
...
99	99 https://www.nytimes.com/2023/05/01/technology/...	['nytimes', 'complease', 'enable', 'js', 'disa...'	[nytimes complease enable js disable ad blocker]	0.000	0.000	1.000	0.0000
100	100 https://foundation.mozilla.org/en/blog/we-ask...	['moment', 'enable', 'javascript', 'cookies', ...]	[moment enable javascript cookies continue]	0.000	0.000	1.000	0.0000
101	101 https://www.nytimes.com/interactive/2022/12/26...	['pioneer', 'really', 'fear', 'new', 'york', '...	[pioneer really fear new york times skip phone...]	0.219	0.095	0.686	0.9993

Other pattern that helped me to identify blocked websites was the **sentiment score for neutral (1) and overall_sentiment (0)**. These values drew my attention, and by navigating to the links I understood that these websites were not going to allow scraping. These websites are usually articles that need to be paid for.

Connections forcibly closed by remote host

When I was running my jupyter notebook to scrap the data from the weblinks, I had some issues with connections being closed. It was very likely the result of **websites blocking my request for scraping**. This would stop my scraping and it wouldn't allow me to continue. However, it seems that after re-starting the notebook after a few minutes would allow me to scrap all my weblinks.

```
File c:\users\diana\appdata\local\programs\python\python39\lib\site-packages\requests\adapters.py:547, in HTTPAdapter.send(self, request, stream, timeout, verify, cert, proxies)
  544         raise
  546 except (ProtocolError, OSError) as err:
--> 547     raise ConnectionError(err, request=request)
  549 except MaxRetryError as e:
  550     if isinstance(e.reason, ConnectTimeoutError):
  551         # TODO: Remove this in 3.0.0: see #2811

ConnectionError: ('Connection aborted.', ConnectionResetError(10054, 'An existing connection was forcibly closed by the remote host', None, 10054, None))
```

IOPUB data rate exceeded

When I got to the stage of doing sentiment analysis, my jupyter notebook would struggle to process all the data I wanted to process. I found that jupyter notebook does cap the RAM limit, therefore crashing my program and the overall notebook.

```
IOPub data rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_data_rate_limit`.

Current values:
NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
NotebookApp.rate_limit_window=3.0 (secs)
```

To solve the issue of handling big amounts of data with jupyter notebook, I found that the solution was to increase the data rate limit with the value being in bytes. I increased it to 120000000000.

```
jupyter notebook --NotebookApp.iopub_data_rate_limit=120000000000
```

4. Design & Development

Physical Output *

4.1 Inspiration

4.2 Concept and wireframe

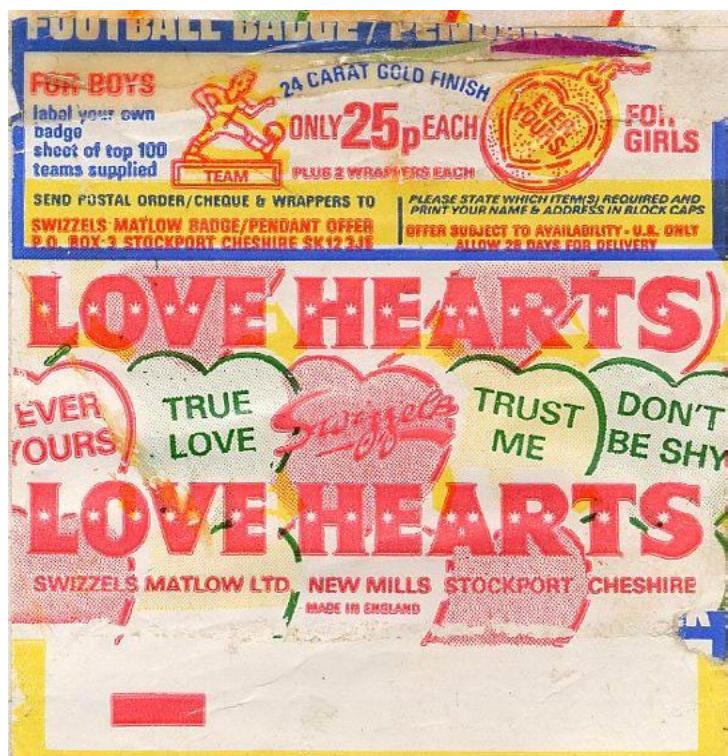
4.3 Prototype development

4.4 Final outcome

*Please notice that there are video links included in the documentation. It is recommended to right-click on them and open a new tab to see their content. Furthermore, increase the quality of the video. YouTube usually reduces quality level, but it can be increased up to 1080p.

4.1 Inspiration.

Love Hearts - Communication through Sweets



Love Heart Sweets

These heart-shaped sweets with messages printed on them have been circulating in the UK since 1933. They have been a reflection of how British society has evolved into new mannerisms and the use of language. The sweets have removed expressions that are out of date and introduced new ways of communication, such as emojis or new words that come with society changes. They have been partly an inspiration for my own project which pretends to communicate keywords about our current climate, through sweets.

Love Heart Sweets written by AI (Shane, 2021)

There are some projects that have explored the use of words in sweets. One that came to my mind was Love Hearts/ Conversational Hearts, which are sweets in the form of hearts that contain messages that relate to love. In 2018, Janelle Shane collected the messages that are currently printed in the sweets and fed them to a neural network. The main purpose was to generate similar messages, that could be used in the sweets. The results were varied, but the principle of using text in sweet remains. I found very interesting how words imprinted in sweets can be playful yet that playfulness combined with words is very powerful

Accurate word output



Less accurate word output

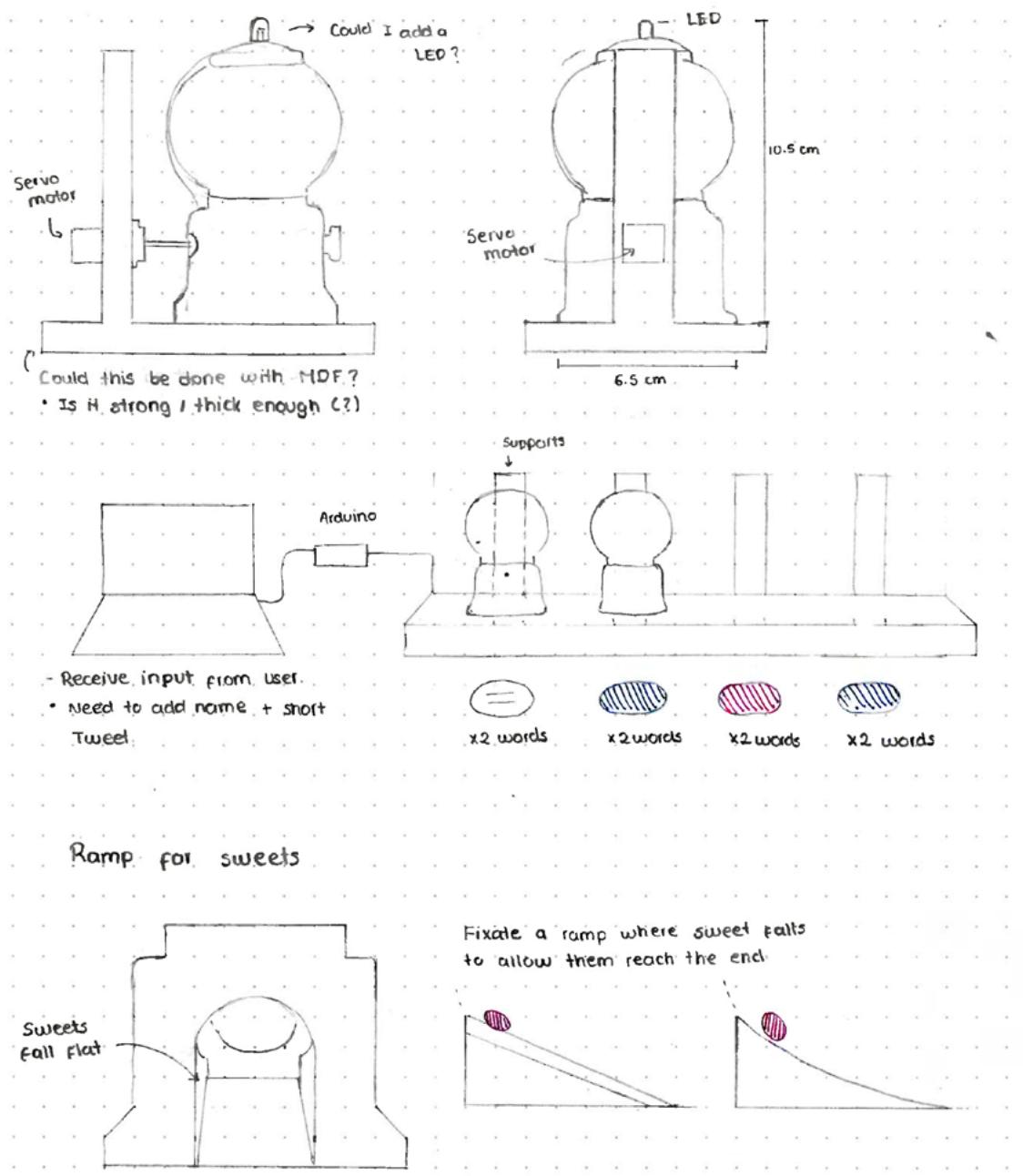


4.2 Concept and wireframe

Concept

I will be representing the findings from my **data through sweets**. The extraction of words that represent sentiment and emotions will be used to print keywords in chocolates. The words will be grouped according to how close they are in meaning (rather than vector space), and they will be stored in gum ball machines. Triggered by user input, and primarily driven by how they feel about AI in the workplace, the gum ball machine will provide a sweet with a printed word, while at the same time, it will post such word to Twitter. **The purpose is to feed an account that pretends to reflect how people really feel about AI.**

Wireframe



4.3 Prototype development.

To watch the video documentation please right click on the '[Watch video](#)' button, and open on a new tab.

Why a Gum Ball machine ?

This small toy machine was randomly bought by my sister. I thought it was quite playful and also brought back memories of the uncertainty one can feel when turning the knob. You can see sweets but can't choose which one you will get. **It is a metaphor for the uncertainty that AI can cause and has been causing to people.** We can turn the knob to AI, but we don't know what we will get.

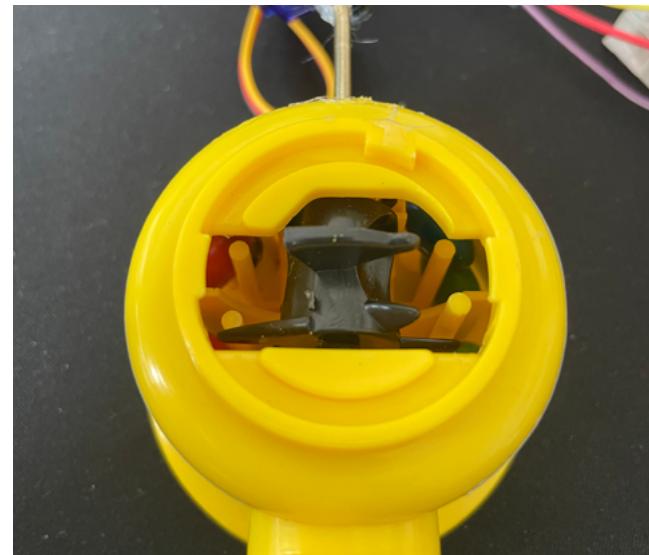
I also like the idea of '**eating words**' that will be printed on the chocolates.

Testing the Gum Ball machine internal mechanism.

The machine is fully made of plastic and relies on user input. This meant that I could add movement through the back by connecting a servo motor.

The shape of the chocolate also matters, as some of them get stuck internally.

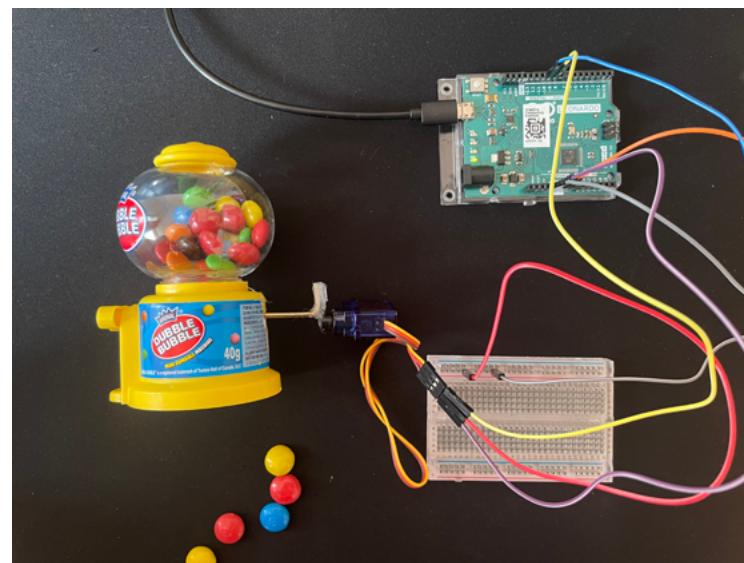
 [Watch video](#)



Testing the Gum Ball machine with a 180 ° Servo.

One of the key aspects of using the machine is how the user will interact with it, and how such interaction allows for posting on Twitter. My first thought was to use a 180 servo to observe how well it would release the chocolates.

 [Watch video](#)



Using the 180 ° Servo to post on Twitter

This idea came from the usage of the Tweepy library to post on Twitter. By creating rotation conditions with the Servo I could trigger Tweepy to create a new Tweet.

However, to be able to work within Python, I had to use pyfirmata and the Firmata library in Arduino. Firmata is a generic protocol for communicating with microcontrollers from software on a host.computer

```
import tweepy
import config
from pyfirmata import Arduino, SERVO, util

while True:
    x = input("input:")
    if x == "1":
        response = client.create_tweet(text='i love me')
        print(response)
        for i in range(0,180):
            rotateservo(pin, i)

        for i in range(180,1,-1):
            rotateservo(pin,i)
```

Home

For you



What's happening?



DMG @diana_mgc_25 · Apr 29

i love me



Testing the Gumball machine with a 360 ° Servo.

The continuous servo was a challenging component, since it works completely differently from a 180 servo. However, it does do a full spin and allows the sweets to fall faster, which at the same time reduces the number of sweets getting stuck.

[Watch video](#)



Checkpoint

- The **continuous servo** will be used to control the gumball machine. However, it needs to start and end at the same point. (*code improvement*).
- The servo motors will need to be held, as well as the gumball machine so that the only thing spinning is the knob (*physical structure improvement*).
- Twitter **can't accept** the same Tweet twice. Perhaps, a way to solve this, is allowing the user to generate a new input (*code improvement*).

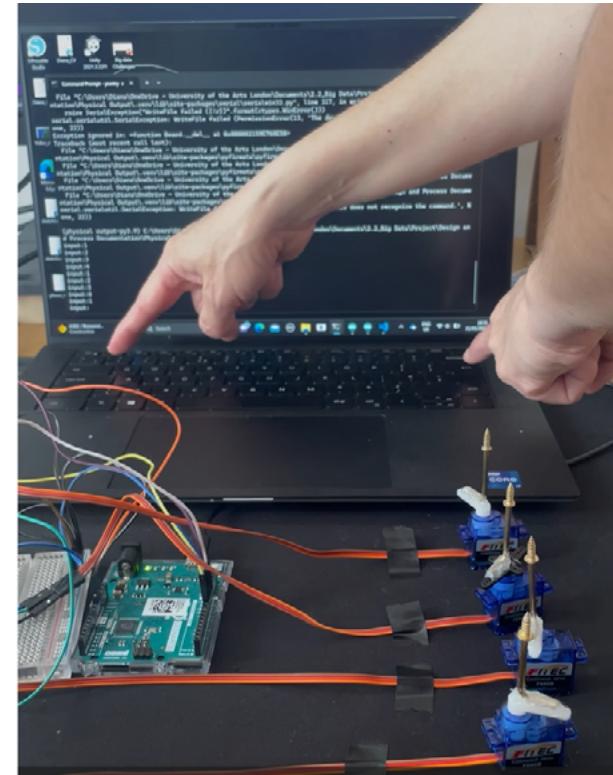
Scaling up the use of continuous servos

After understanding that every servo could be triggered by user input (keyboard numbers), I was able to replicate the behaviour of one servo to four. The servo does a full 360 spin and then waits for user input to activate again. Each servo motor corresponds to the number 1, 2, 3 and 4 respectively.

```
#0 indicates rotation clockwise to a 360 rotation  
#180 indicates rotation anti-clockwise to a 360 rotation  
board.digital[pin].write(0)  
sleep(0.5)  
board.digital[pin].write(90) #90 indicates stop  
  
while True:  
    x = input("input:")  
    if x == "1":  
        rotateContinuoServo(pin)  
    elif x == "2":  
        rotateContinuoServo(pin2)  
    elif x == "3":  
        rotateContinuoServo(pin3)  
    elif x == "4":  
        rotateContinuoServo(pin4)
```

360 servo input code.

 [Watch video](#)



Structure and functionality Prototype

This first prototype embodies the view I first wireframed:

- The gumball machines are positioned side by side, and the servo motors are hidden (and attached) to a wall.
- The machines are **triggered by user input** through my computer's keyboard.
- The sweets have a lower risk of getting stuck because the servo motor does a 360 spin. I also **padded the inside** with cotton and attached a **cardboard ramp** in the inside, so that the sweets are driven outside of the machine.

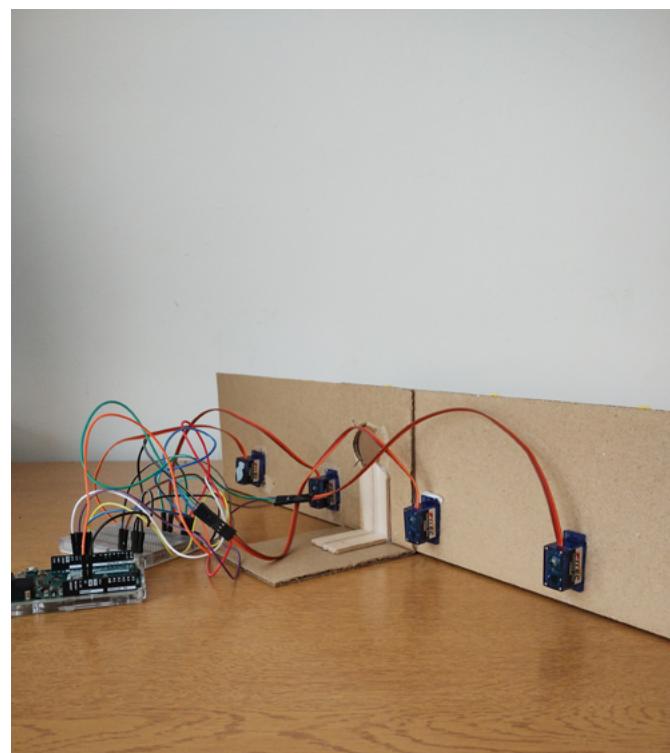
 [Watch video](#)



First full working prototype.



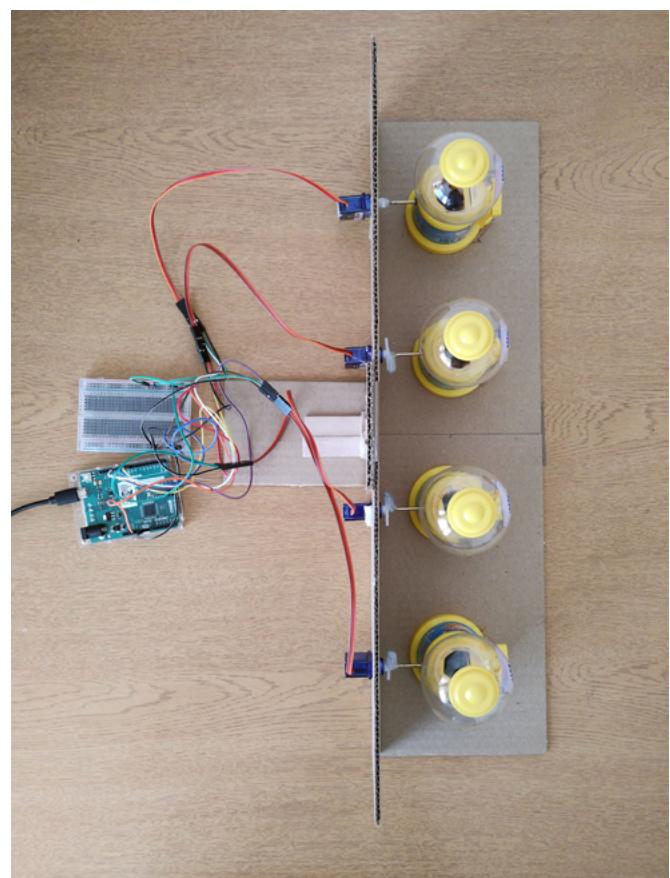
Front view



Back view



Servo assembly



Top view

4.4 Final Outcome - physical structure



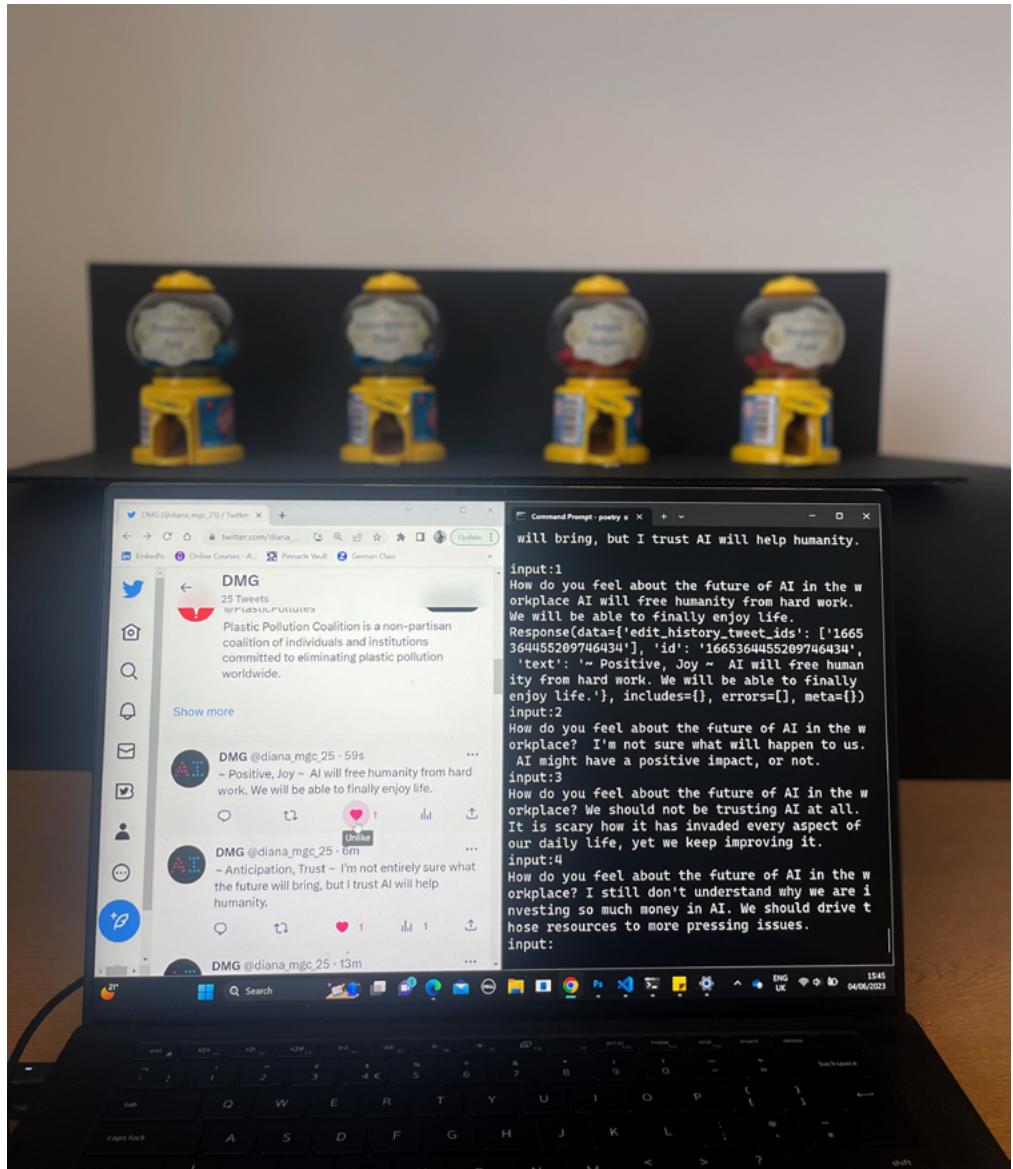
After conducting tests on the prototype using four machines, I decided to enhance the visual contrast of the sweets by covering the surface with black cardboard. To facilitate the selection process for users, I designed a set of labels featuring the words I had extracted during my research. These labels indicate the types of sweets available for selection. Additionally, I ensured that the sweets themselves were imprinted with the corresponding words.



[Watch video of the final physical outcome](#)



Final Outcome - Code and machines interaction



This represents the final outcome of the project, showcasing the extensive research conducted on the data extracted from various websites that talk about AI in the workplace. The findings are presented to the user in a captivating and interactive manner. The interaction begins with the user facing four machines, and the primary question posed by the computer is: **“How do you feel about the future of AI in the workplace?”** The user choose from the words that the machines hold and is encouraged to share (type) their personal opinion as well. After the user hits enter, they receive a chocolate from the machine, and simultaneously, their opinion is posted on a dedicated Twitter account.

In the following two videos, I showcase a **full interaction** with the 4 machines where the comment inputs have been premade, and a **shorter version** where I type a single opinion and trigger only one machine. Please read the descriptions on the videos..

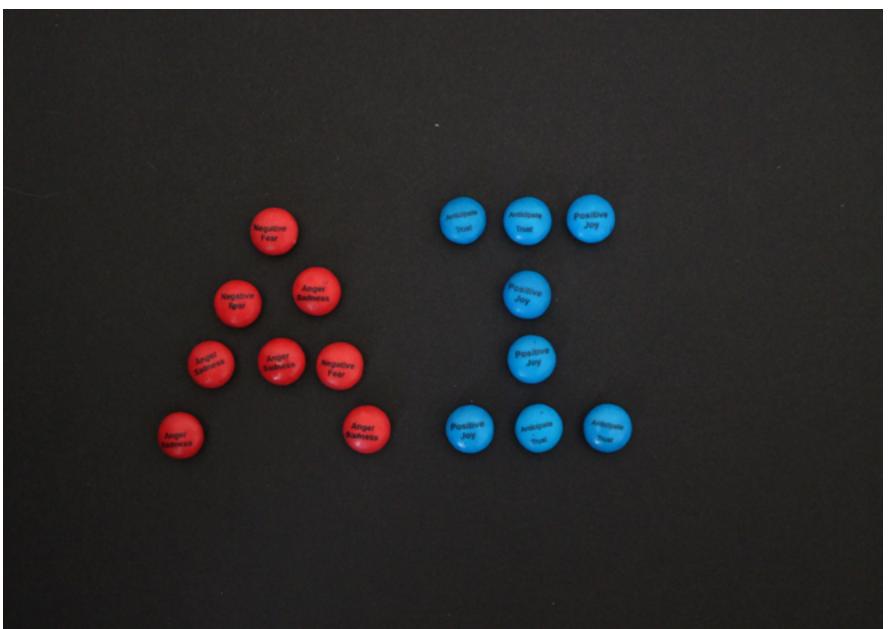
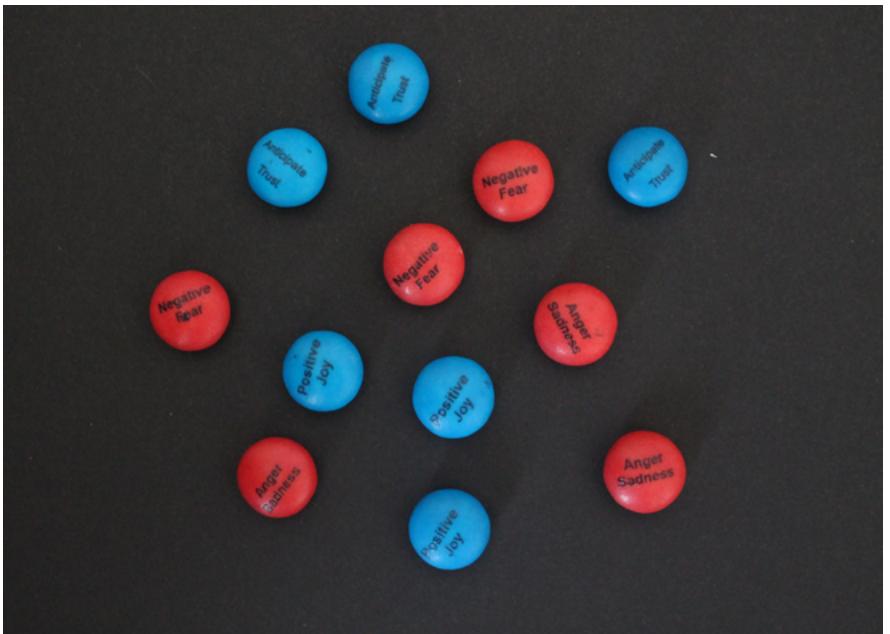
[Watch full interaction video](#)

[Watch short interaction video](#)

Final Outcome: Sweets, emotions and sentiment.

This is a close-up view of the sweets, where I have strategically assigned negative emotions/sentiments to the red-coloured sweets, while positive words are represented by the blue ones. To create a visually appealing arrangement, I arranged the sweets to spell out “AI.” This image served as the profile picture for the Twitter account that hosts the comments provided to the machines.

The image shows a screenshot of a Twitter profile page. The header image is a photograph of a candy machine with various colored candies spelling out the letters 'A' and 'I'. The profile picture is a circular icon containing a similar arrangement of red and blue candies. The user's name is 'DMG' and their handle is '@diana_mgc_25'. Below the handle is a bio: 'How do we really feel about AI in the workplace?'. The user joined in April 2023 and has 0 following and 2 followers. The 'Tweets' tab is selected, showing one tweet from the user: 'DMG @diana_mgc_25 · 3h ~ Anger, Sadness ~ ~ Is there any regulation around AI? I don't feel safe in the current climate. In fact, I feel anxious ~'. The tweet has 1 like. Navigation icons for Twitter are visible on the left side of the page.



Final Outcome - Twitter account

Very recently, we have faced new changes and restrictions to the Twitter API. It was my initial idea, to extract the general feeling of people from Tweets, but since we can't easily extract information anymore, I thought that perhaps creating an account that holds this information might be **a way around gathering new data.**

The code has been modified so that every comment keeps the **set words** that came from the research. It also allows the user to freely type their own opinion. This functionality takes away the issue of generating duplicate comments, as different people have different opinions, and therefore the wording and style will be different every time. The purpose of this account is to hold comments and opinions from the public that interact with the piece, and from which I could further analyse. It might also be a **more reliable resource** that truly reflects people's opinions on the matter.



[View Project's Twitter Account](#)

The screenshot shows a Twitter interface with the 'Tweets' tab selected. There are three tweets from the user 'DMG @diana_mgc_25' posted 3 hours ago:

- DMG @diana_mgc_25 · 3h**
~ Anger, Sadness ~ ~ Is there any regulation around AI? I don't feel safe in the current climate. In fact, I feel anxious ~
- DMG @diana_mgc_25 · 3h**
~ Negative, Fear ~ ~ We should not trust AI or the people who build it. I feel scared it will take over my job ~
- DMG @diana_mgc_25 · 3h**
~ Anticipation, Trust ~ ~ I'm not against AI, but I'm also aware of its risks. It might benefit us, or not ~

Below the tweets is a block of Python code:

```
while True:  
    x = input("input:")  
    if x == "1":  
        userInputFeeling = input("How do you feel about the future of AI in the workplace")  
        finalTweet = ("~ Positive, Joy ~ {}").format(userInputFeeling)  
        rotateContinuoServo(pin)  
        response = client.create_tweet(text= finalTweet)  
        print(response)  
    elif x == "2":  
        userInputFeeling = input("How do you feel about the future of AI in the workplace")  
        finalTweet = ("~ Anticipation, Trust ~ {}").format(userInputFeeling)  
        rotateContinuoServo(pin2)  
        response = client.create_tweet(text= finalTweet)
```

5. Possible improvements

Gathering of more reliable data

One of the big surprises during the data analysis process, was to see how positive sentiment overcame negative sentiment, especially over a disruptive subject such as AI. One of my hypothesis is that, given the nature of the websites, where many were news articles or websites promoting AI based products, the sentiment would lean towards positivity and expectation. This means that the **media might be already biased with regards to the public sentiment around AI.**

Therefore the analysis might benefit from the gathering of data coming from different resources other than websites, for example, social platforms or forums. Something that I also noticed was the amount of websites that are blocking content, meaning that users have to pay to access the articles. This also stopped me from getting more information, and it reduced considerably the body of content.

Further development with new body of data

As I have expressed previously, the changes in the **Twitter API** made the retrieval of public data a lot more difficult. This was the main reason I thought of creating my own account, where people can input their feelings about AI in the workplace. Even though it is a single account, it collects shared thoughts, and because it is an account I manage, retrieving information is a lot more permissive. If I were able to collect many different opinions, I would be able to develop a **better-informed analysis** of the real sentiment of people towards AI. Let's remember that many of the articles from the websites were biased by companies that were pushing their AI products.

Physical prototype improvement

There are certain elements of the physical structure that could be improved to make the experience smoother. The internal mechanism can change, so that the sweets don't get stuck when the user triggers the machine. The machines could be bigger and perhaps allow for other kinds of sweets with different shapes. I would also like to improve the way the user inputs the information, for example, using arcade buttons, and seeing the text on a screen.

6. Bibliography

Bird, S., Klein, E. and Loper, E. (201AD) 1. Language Processing and python. Available at: <https://www.nltk.org/book/ch01.html#ref-freq-dist-call> (Accessed: 05 June 2023).

Eremenko, K. (2018a) Confident data skills: Master the fundamentals of working with data and supercharge your career. London: KoganPage.

Kirenz, J. (2022) Text mining and sentiment analysis with NLTK and pandas in Python, Jan Kirenz. Available at: <https://www.kirenz.com/post/2021-12-11-text-mining-and-sentiment-analysis-with-nltk-and-pandas-in-python/text-mining-and-sentiment-analysis-with-nltk-and-pandas-in-python/> (Accessed: 11 May 2023).

LZP, J. (2021) Simple emotion classification in python, Medium. Available at: <https://medium.com/geekculture/simple-emotion-classification-in-python-40fb24692541> (Accessed: 15 May 2023).

Pause giant AI experiments: An open letter (2023) Future of Life Institute. Available at: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (Accessed: 27 March 2023).

Shane, J. (2021) Candy heart messages written by a neural network, AI Weirdness. Available at: <https://www.aiweirdness.com/candy-heart-messages-written-by-a-18-02-09/> (Accessed: 05 May 2023).

Figures:

Eremenko, K. (2018) ‘How data fulfils our needs’, in Confident data skills: Master the fundamentals of working with data and supercharge your career. London: KoganPage, p. 26.

Eremenko, K. (2018) ‘How data fulfils our needs’, in Confident data skills: Master the fundamentals of working with data and supercharge your career. London: KoganPage, pp. 28–29.

Wyawahare, C. (2020) Leveraging the power of Jupyter Notebooks, Medium. Available at: <https://towardsdatascience.com/leveraging-the-power-of-jupyter-notebooks-26b4b8d7c622> (Accessed: 14 May 2023).