- 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.international genome.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

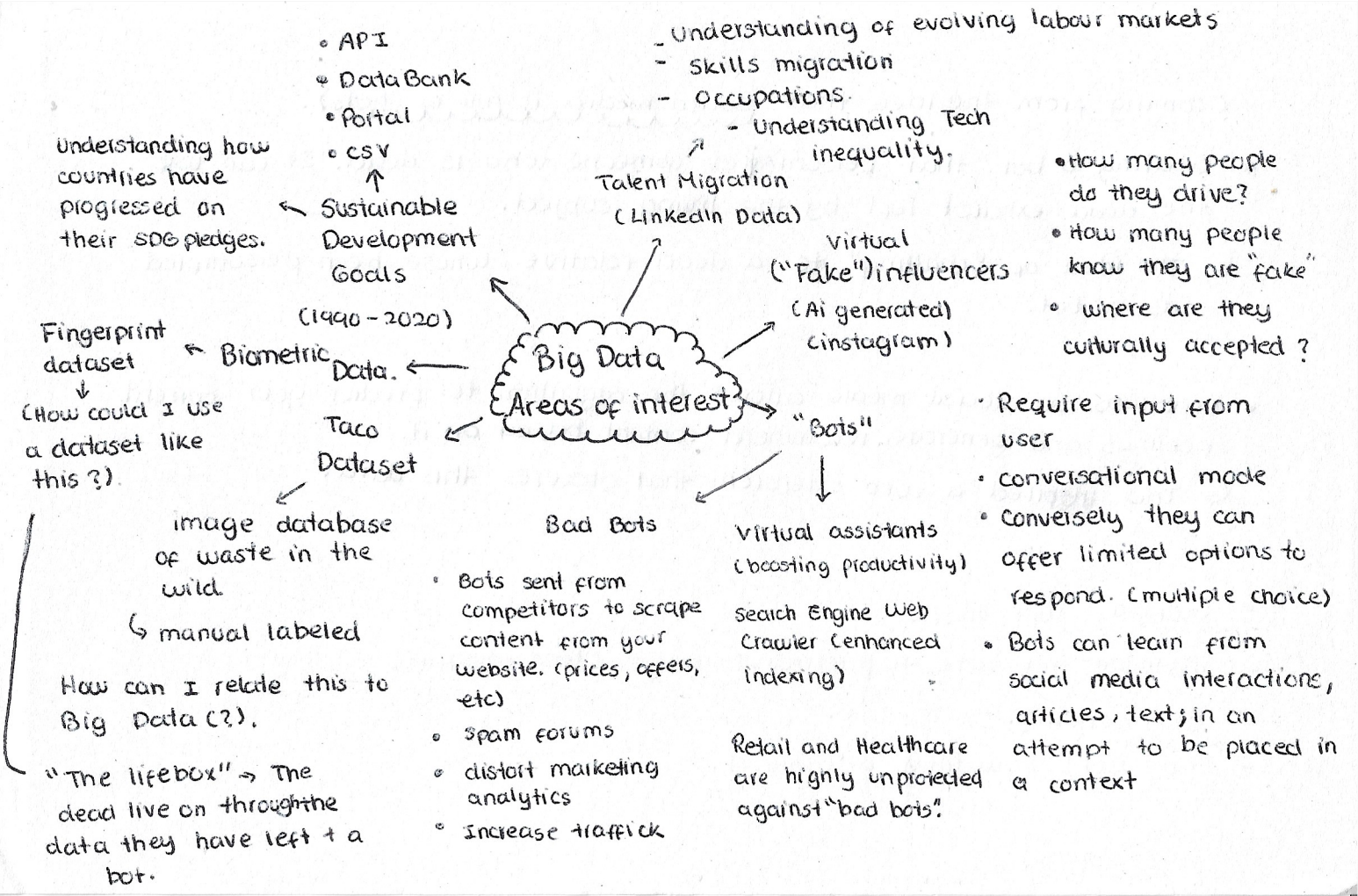
Possible Project Ideas.

→ 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
  ↓ Network style visualization


Big Data
Areas of interest
Understanding how countries have progressed on their SDG pledges.
Sustainable Development Goals (1990 - 2020)
API
DataBank
Portal
csv
Understanding of evolving labour markets
- skills migration
- occupations
- Understanding Tech inequality.
Talent Migration (LinkedIn Data)
Virtual ("Fake") influencers (AI generated) (instagram)
How many people do they drive?
How many people know they are "fake"
where are they culturally accepted?
Biometric Data.
Fingerprint dataset
(How could I use a dataset like this?)
Taco Dataset
image database of waste in the wild.
manual labeled
How can I relate this to Big Data (?).
"The lifebox" → The dead live on through the data they have left to a bot.
"Bots"
Bad Bots
Bots sent from competitors to scrape content from your website. (prices, offers, etc)
spam forums
distort marketing analytics
increase traffick
Virtual assistants (boosting productivity)
Seach Engine Web Crawler (enhanced indexing)
Retail and Healthcare are highly unprotected against "bad bots".
Require input from user
conversational mode
Conversely they can offer limited options to respond. (multiple choice)
Bots can learn from social media interactions, articles, text; in an attempt to be placed in a context