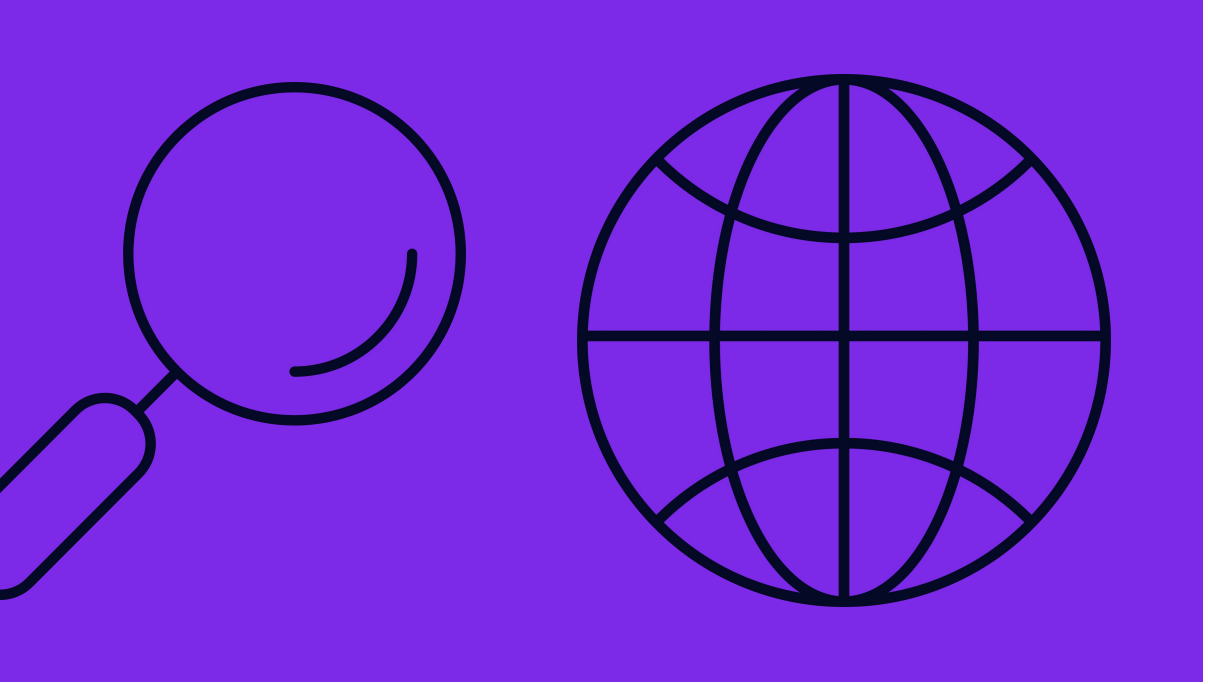
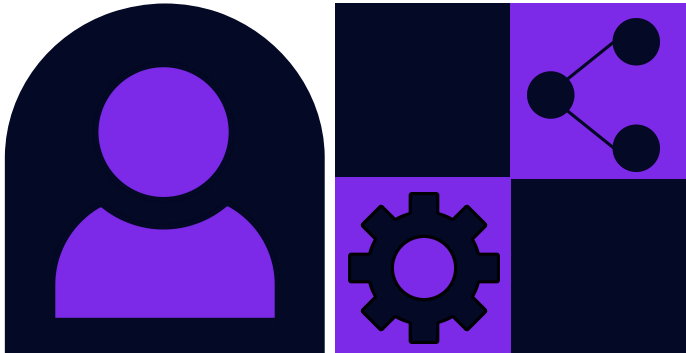


Data Ethics

Isabel Heard



Agenda

01 Motivation

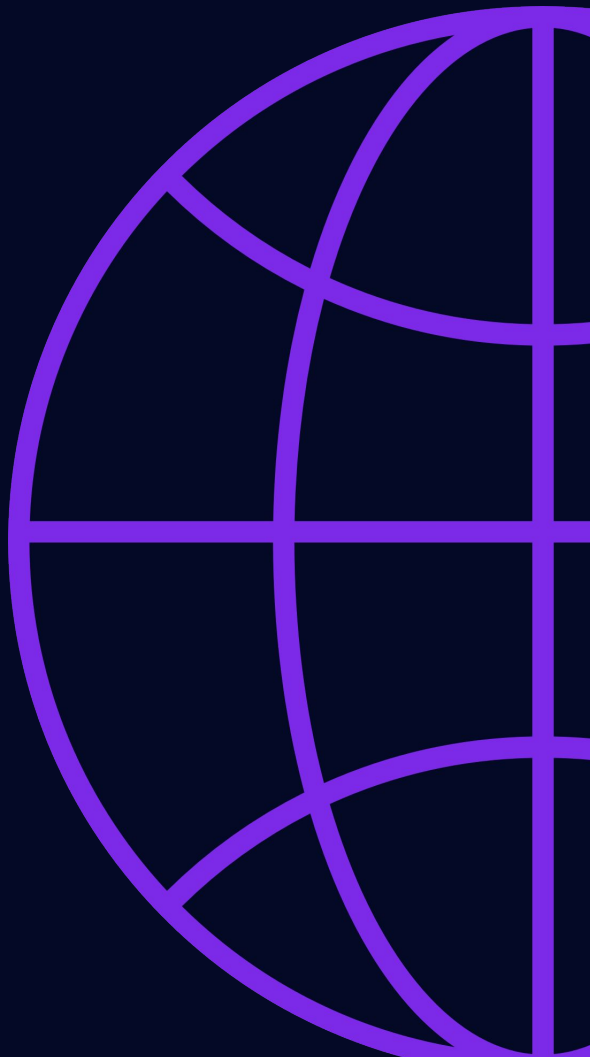
02 How Can We Define Ethics?

03 Exploring Ethical Concepts

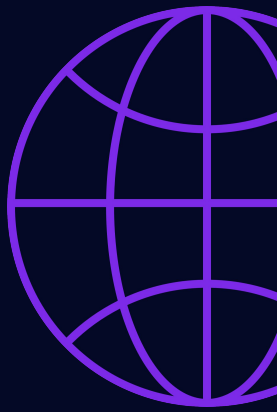
04 Examples

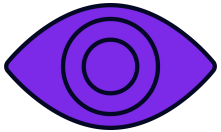
05 Conclusion

06 Discussion & Q&A



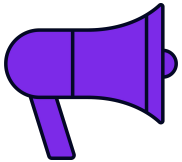
Motivation





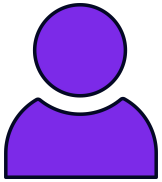
Reflection

Thinking about the code and analysis you are doing



Identification

Anticipate and identify different ethical vacuums



Professional Integrity

How can we use these value systems in the workplace?

How can we Define Ethics?

What	Why	How
Standards of right and wrong that govern human behavior	Provides a moral framework to identify and work through issues	Develop your own ethical framework

What is Ethical Data Science?

- Produces useful and reliable knowledge
- Helps drive innovation and learning
- Beneficial for society
- Minimizes bias
- Transparency
- Privacy

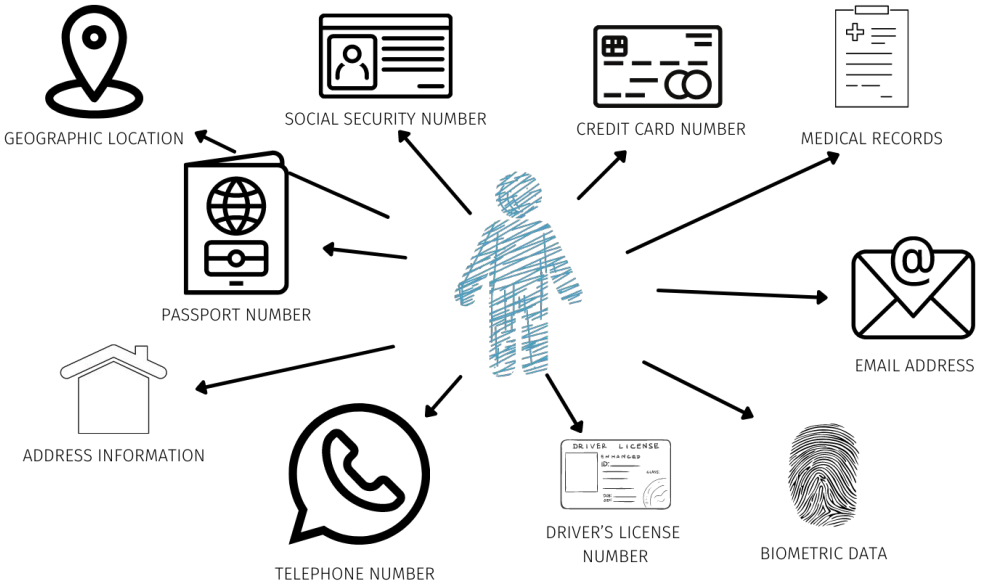
But who decides what is good?

Exploring Ethical Concepts Within...

Personally Identifiable Information	Location Data
Cookies & Scraping	Institutionalized Bias
Biometrics	Data vs. Truth

Personally Identifiable Information

- Any data that can be used to identify a specific individual
- As technology advances, so does PII
- You determine the value of your data
- **Is privacy an inalienable right?**
  - At home?
  - At work?
  - On the street?
  - In a store?



Personally Identifiable Information

Shiru Café

- Exchanges PII for free coffee
- Data then goes to corporate sponsors
- Used as a recruiting tool for companies

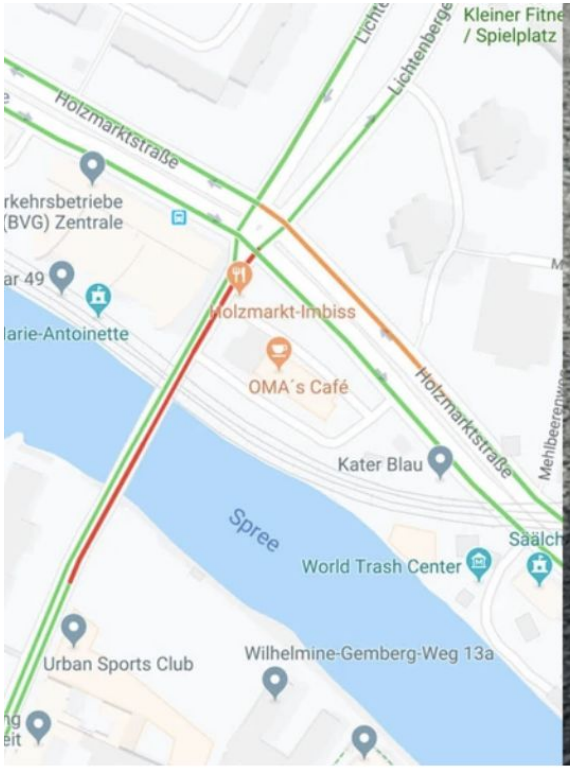
Muslim Pro

- App that helps users with Islamic daily rituals
- The app's compass orients devices toward Mecca
- BitsMedia sells user location data to a broker called X Mode
- X Mode then sells that info to contractors

Location Data

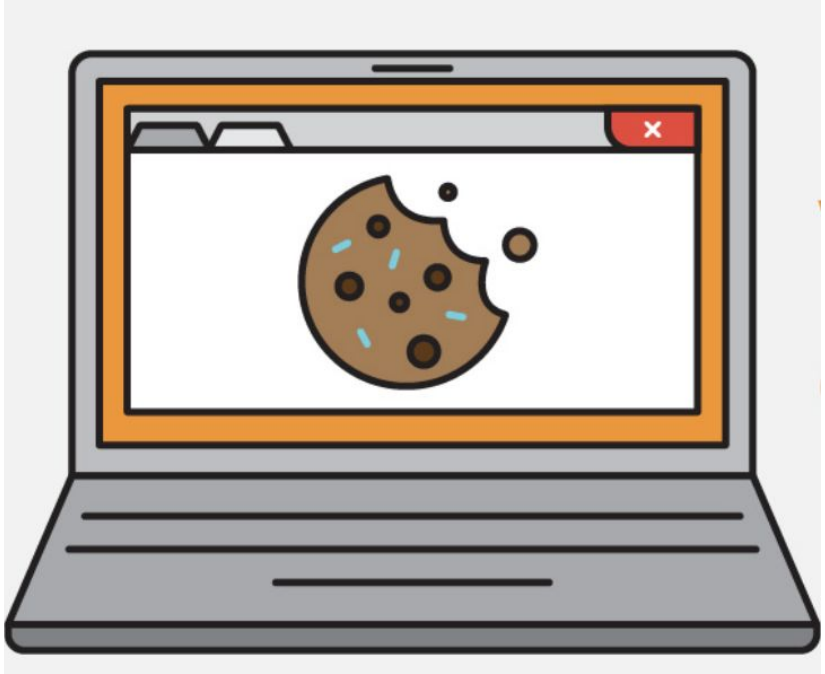
- Information about specific geographic whereabouts
- Can get location data from
  - GPS
  - WiFi
  - Carrier data/cell towers
- All associated with a specific deceive
- Device is then tied to the owner via a unique deceive identifier
- Over time, location data can determine
  - Where you live
  - Where you work
  - How often you go to the gym

Location Data



Cookies & Scraping

- Tracking online user activity
  - Page views
  - Click path
  - Visit duration
  - Search queries
- First party vs third party
- Try and block cookies?
- Fingerprinting
- Web Scraping



Cookies & Scraping

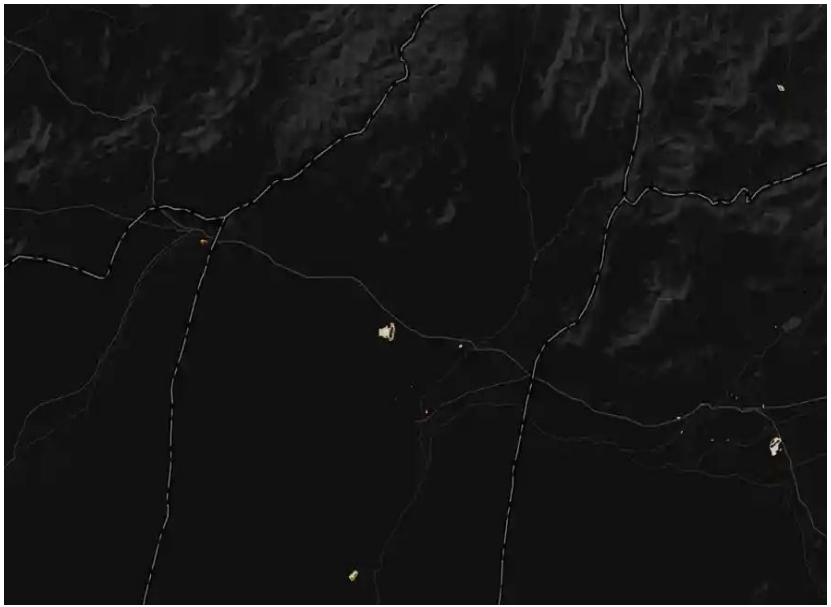
- Facial recognition tool used by law enforcement to identify perpetrators and victims of crime
- Scrapes pictures from public sites such as:
  - Facebook
  - YouTube
  - Venmo
- The data is publicly available, but is this right?

Biometrics

- Physical identifiers are immutable and device independent
  - Fingerprints
  - Facial recognition
  - Voice
  - DNA
  - Iris/Retina scans
- Behavioral identifiers are less reliable, but are often used in conjunction with physical
  - Typing patterns
  - Physical movements
  - Navigation patterns
  - Engagement patterns
- Is this data collection too intimate?

Biometrics

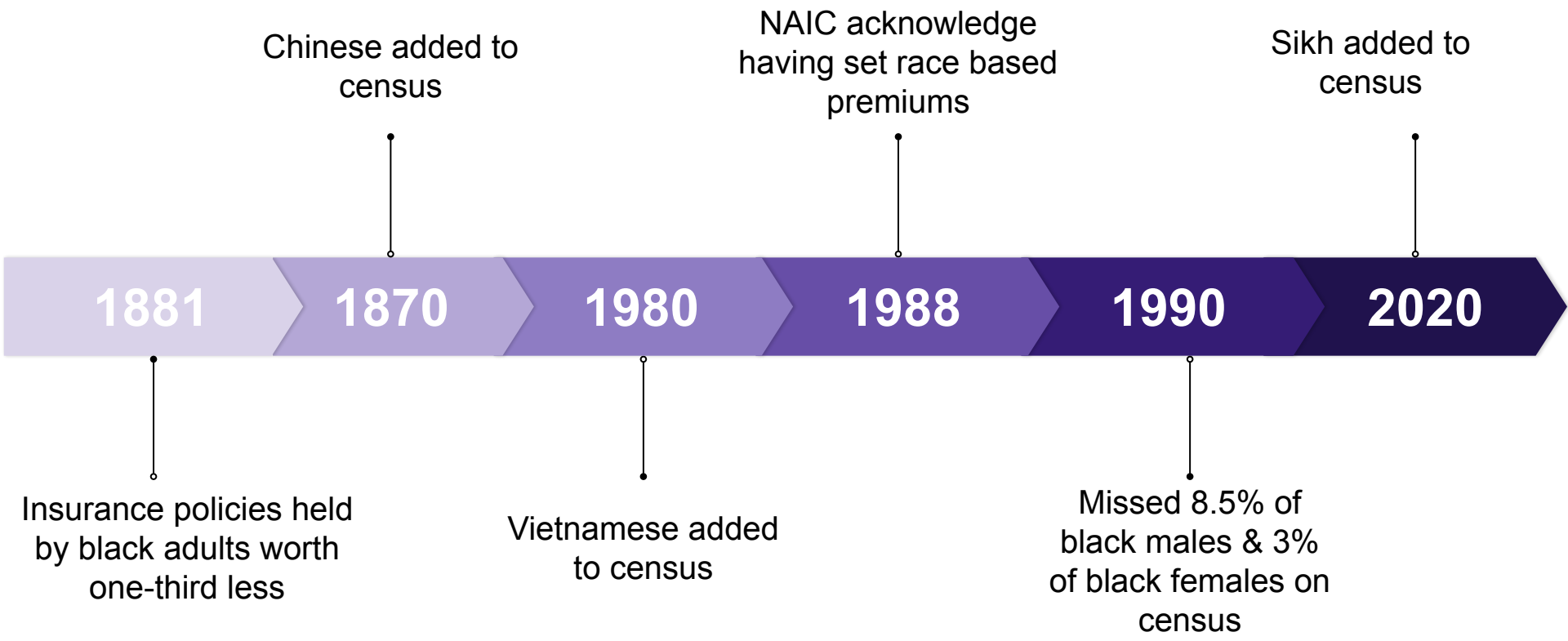
Strava’s wearable technology: A map released in 2017 with more than 3 million GPS data points -> Found remote trails with unusually high concentration of foot traffic



Institutionalized Bias

- Occurs when:
  - Decisions are based on biased or incomplete data
  - Benefit certain groups or agendas
  - Results in perpetual systemic advantages or disadvantages
- It might not be:
  - Overt
  - Intentional
  - Obvious
  - Delibate
    - But it is inherent in our systems
- Institutionalized bias can become implicit bias
  - Attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner

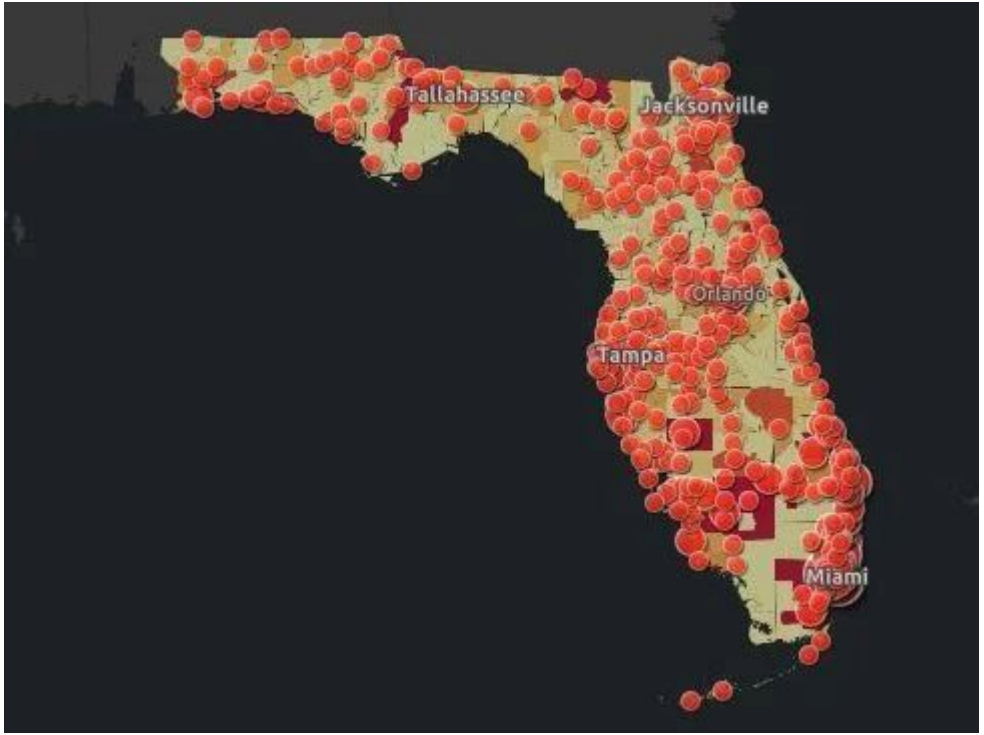
Institutionalized Bias - Insurance + Census



Data vs Truth

- Data is not truth
  - Data points represent single instances of a fact
  - Data is only a subset of reality
  - Data can be interpreted to support multiple conclusions
- Data can be supportive, but not necessarily conclusive
  - Humans collect data, and choose what to collect
- Simpson’s Paradox
- Humans can introduce imperfection into data
  - Random errors
  - Systematic errors
  - Erroneous assumptions
  - Exclusion errors

Data vs Truth



Conclusion

- Think about what data you are collecting or analyzing
- Continue to learn and adjust value system
- Be transparent about the methods and data sources you access
- Don’t just mindlessly complete tasks
- Just because a company has a privacy policy, does not equal privacy
- Algorithms can only calculate, not interpret.
- Just because something looks right, does not mean that it is

“Data are viewed as the world itself, forgetting that the numbers are only representing a model of the world.”

Sources

https://www.latimes.com/business/technology/story/2020-11-23/muslim-pro-data-location-sales-military-contractors

https://www.npr.org/sections/thesalt/2018/09/29/643386327/no-cash-needed-at-this-cafe-students-pay-the-tab-with-their-personal-data

https://www.vice.com/en/article/9393w7/this-man-created-traffic-jams-on-google-maps-using-a-red-wagon-full-of-phones

https://www.youtube.com/watch?v=GeIc-yhGmx4

https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

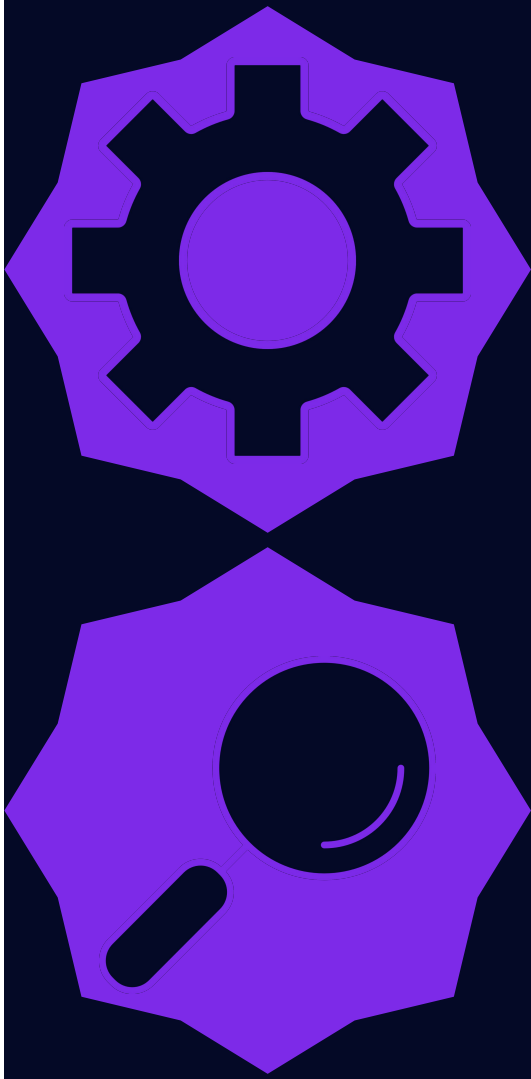
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1497788/

https://daily.jstor.org/how-insurance-companies-used-bad-science-to-discriminate/

https://www.pewresearch.org/short-reads/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race/

https://www.npr.org/2020/06/14/876584284/fired-florida-data-scientist-launches-a-coronavirus-dashboard-of-her-own

Questions?



Biometrics