
ICS 491: Human-Centered Data Science

Course Review

Q&A Session: Any final questions about Human-Centered Data Science?

Can be out any course material, final project, jobs/careers, general data science questions, etc.

Zerin

MOOD SWING DETECTION IN POLYCYSTIC OVARY SYNDROME (PCOS) PATIENTS USING MACHINE LEARNING



Presented By
Zerin Nasrin Tumpa
University of Hawaii at Manoa

Joshua



What is Human-Centered Data Science?

Python Overview

```
def
```

```
    some_complicated_function(data):
```

```
        data = data[data['column'] != 'wrong']
```

```
        data = data.groupby('date').apply(lambda x: comp)
```

```
        data = data[data['value'] < 0.9]
```

```
    return data
```

Python for Data Science

 **matplotlib**

 **pandas**

 **Keras**

 **SciPy**

 **python**[™]

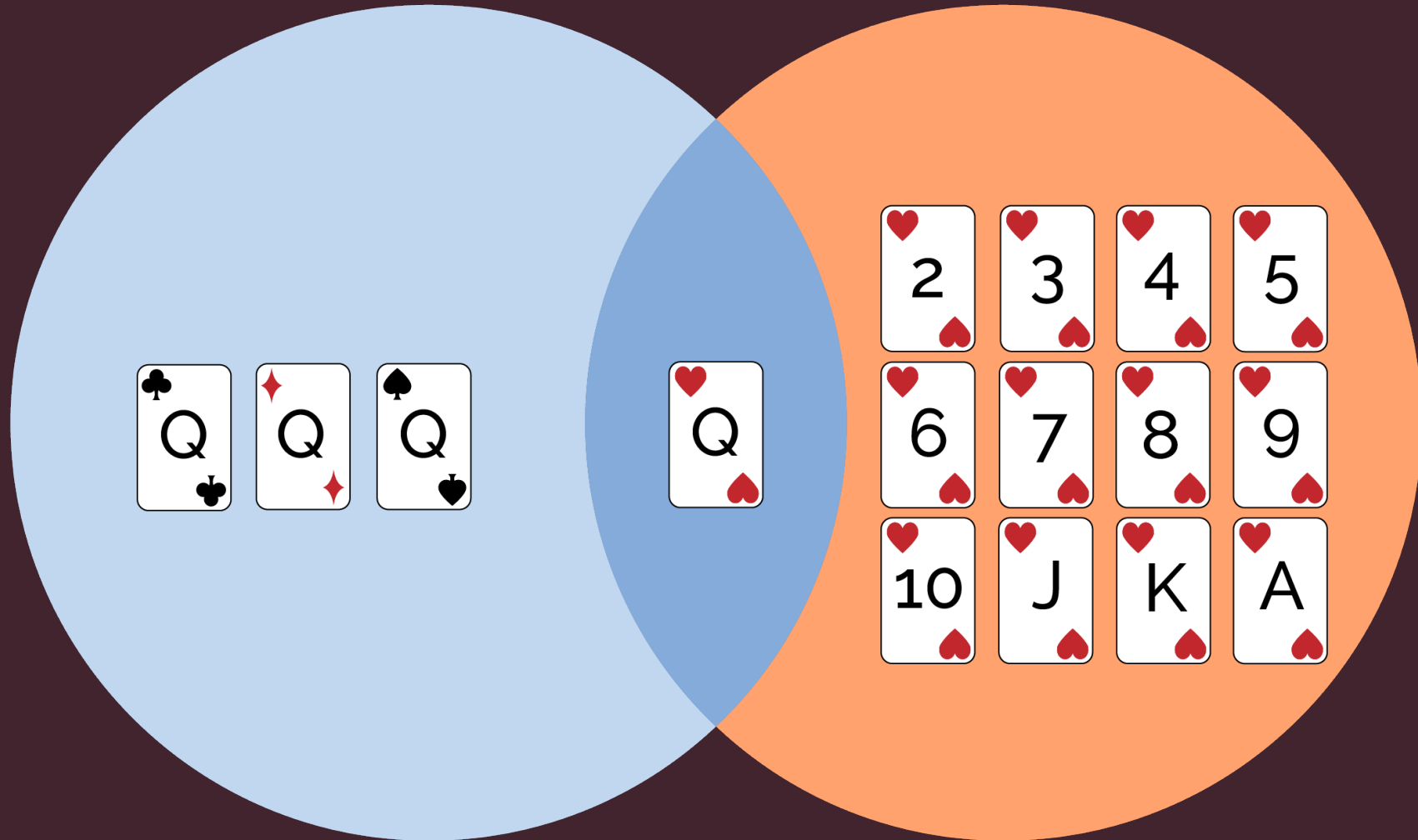
 **GENSIM**
topic modelling for humans

 **scikit
learn**

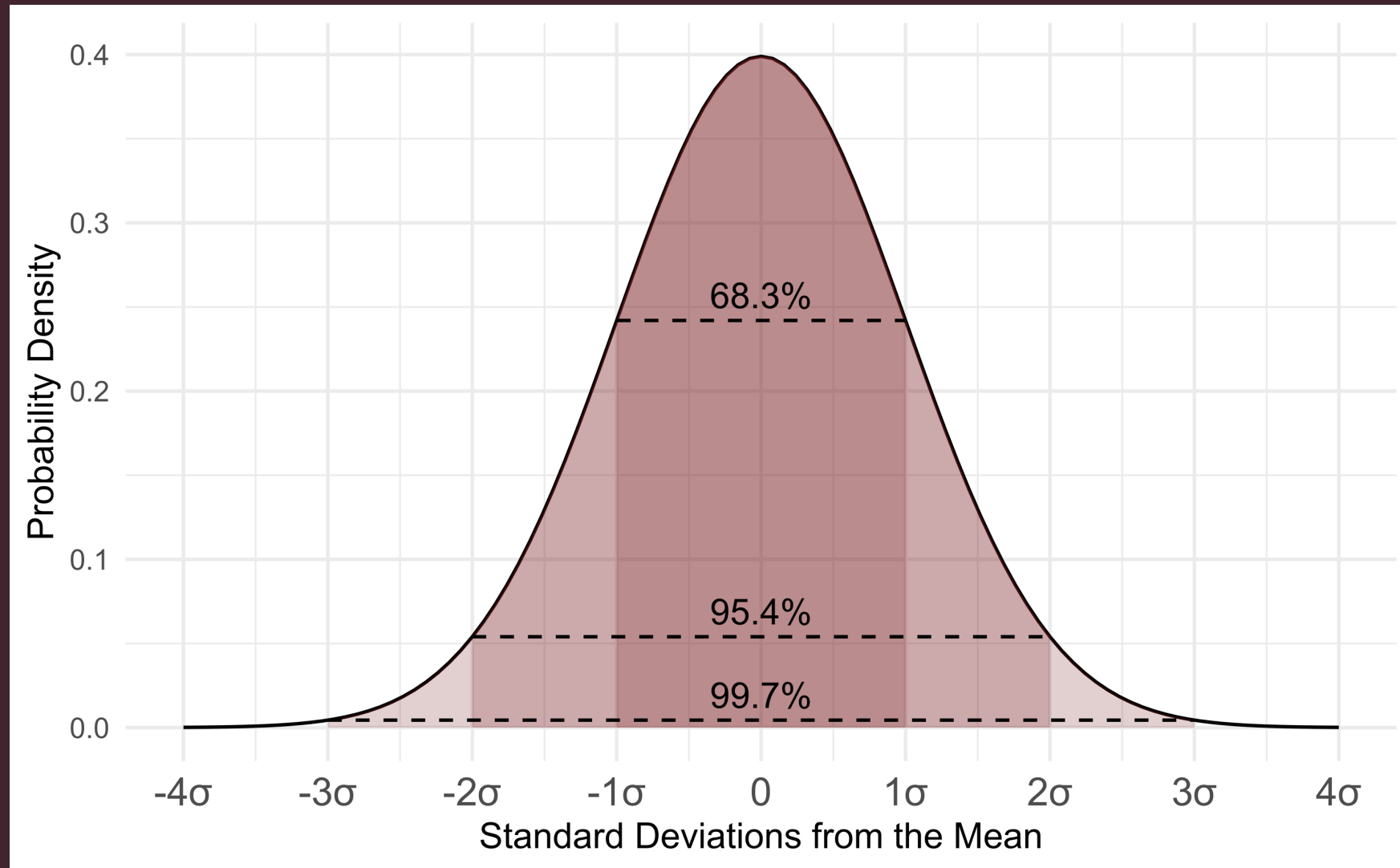
 **NumPy**

 **TensorFlow**

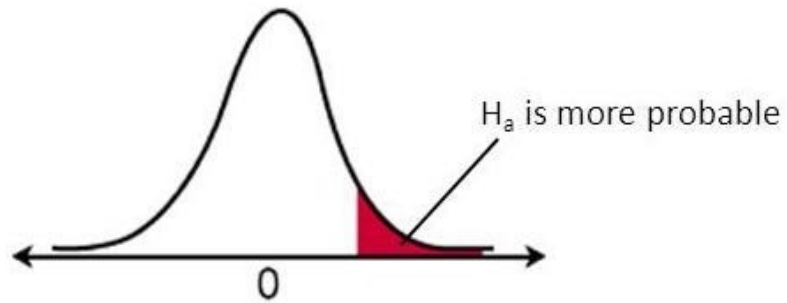
Probability



Statistics

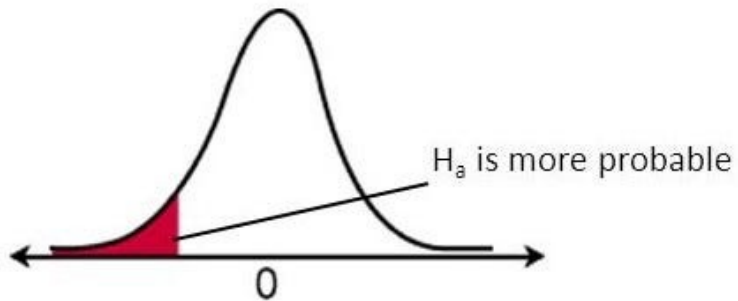


Hypothesis Testing



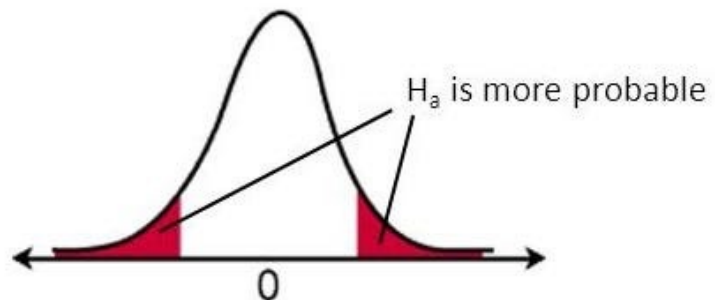
Right-tail test

$$H_a: \mu > \text{value}$$



Left-tail test

$$H_a: \mu < \text{value}$$



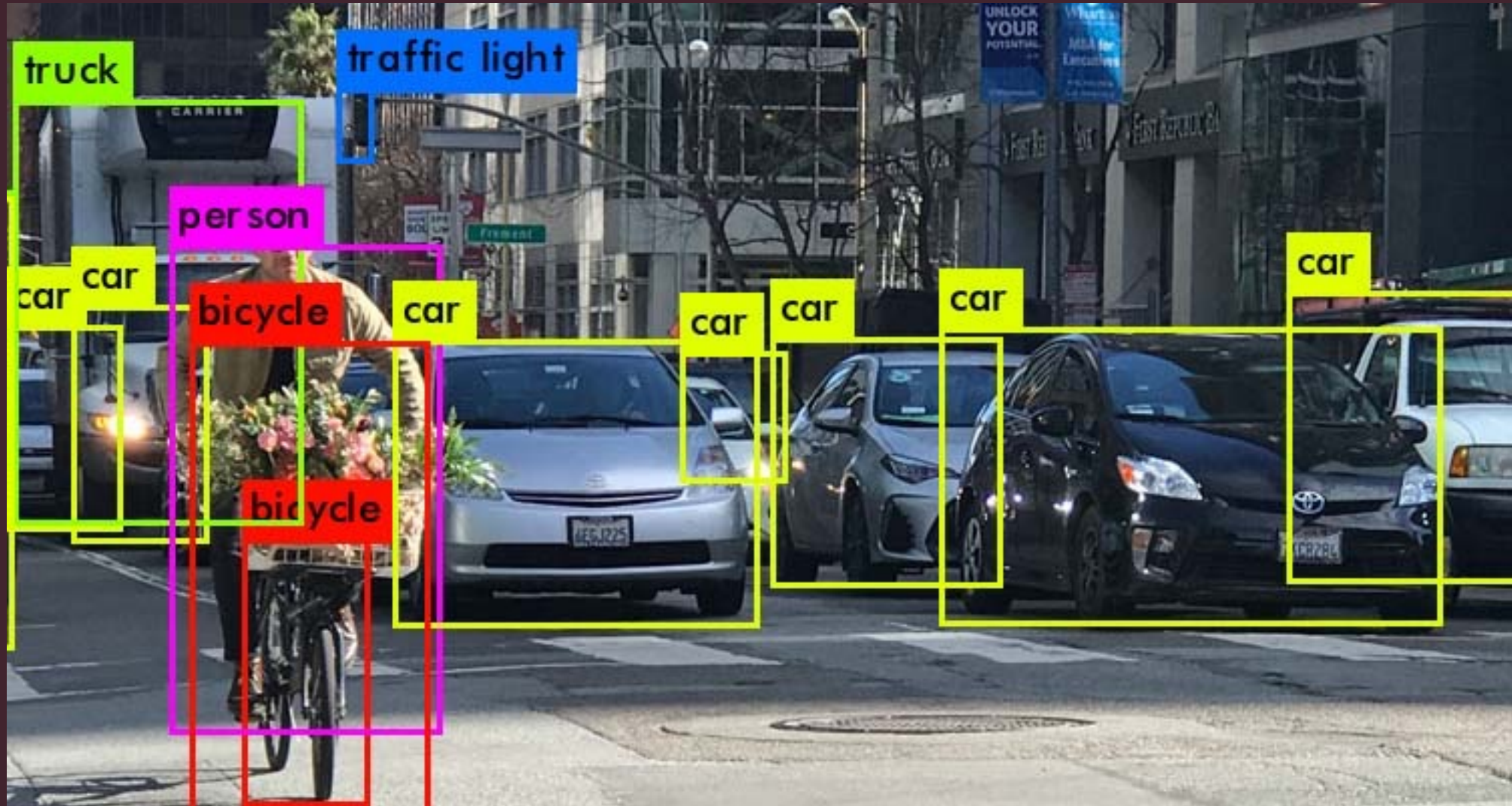
Two-tail test

$$H_a: \mu \neq \text{value}$$

Recruitment and Data Collection

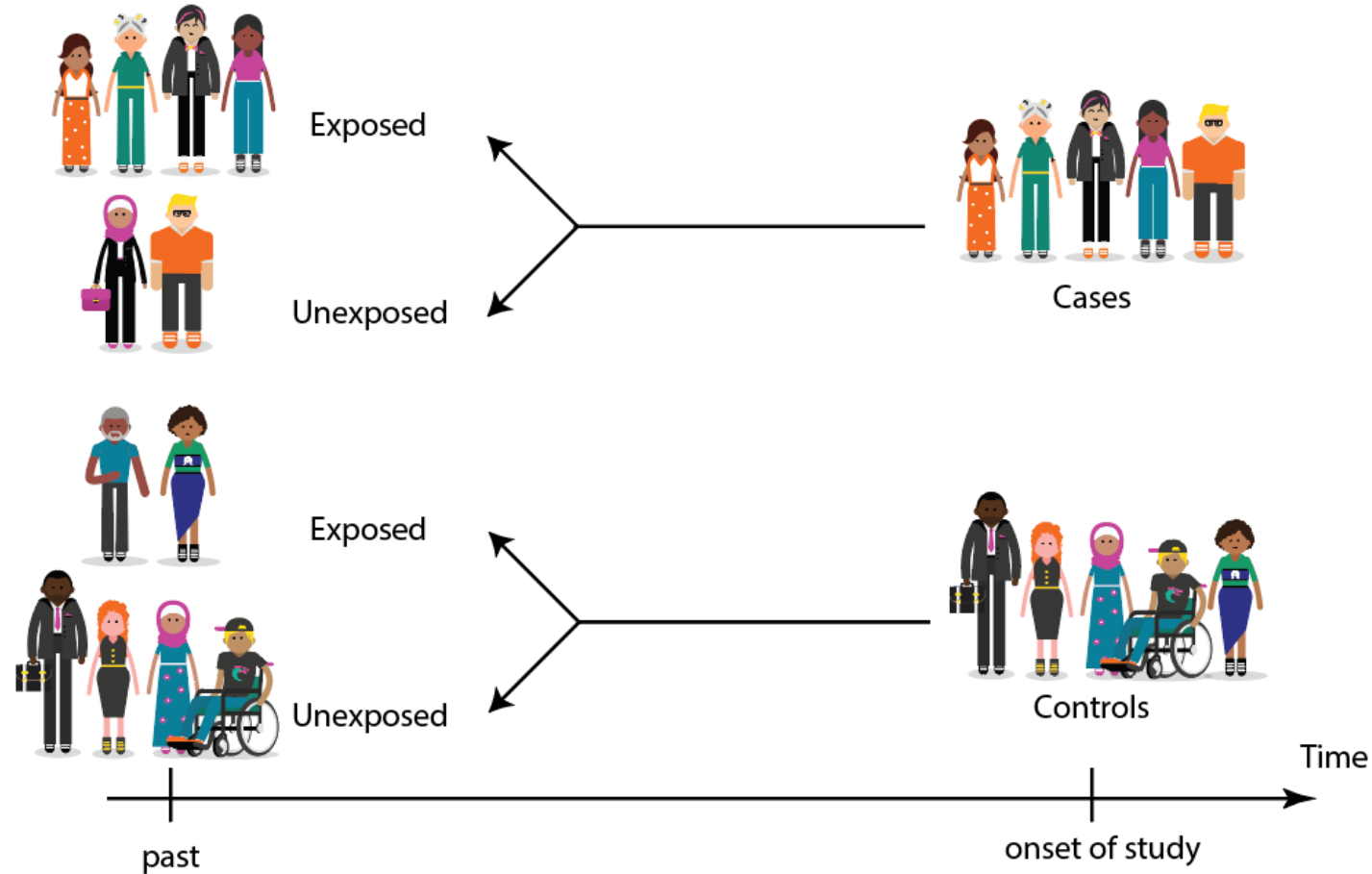


Data Labeling



Study Design

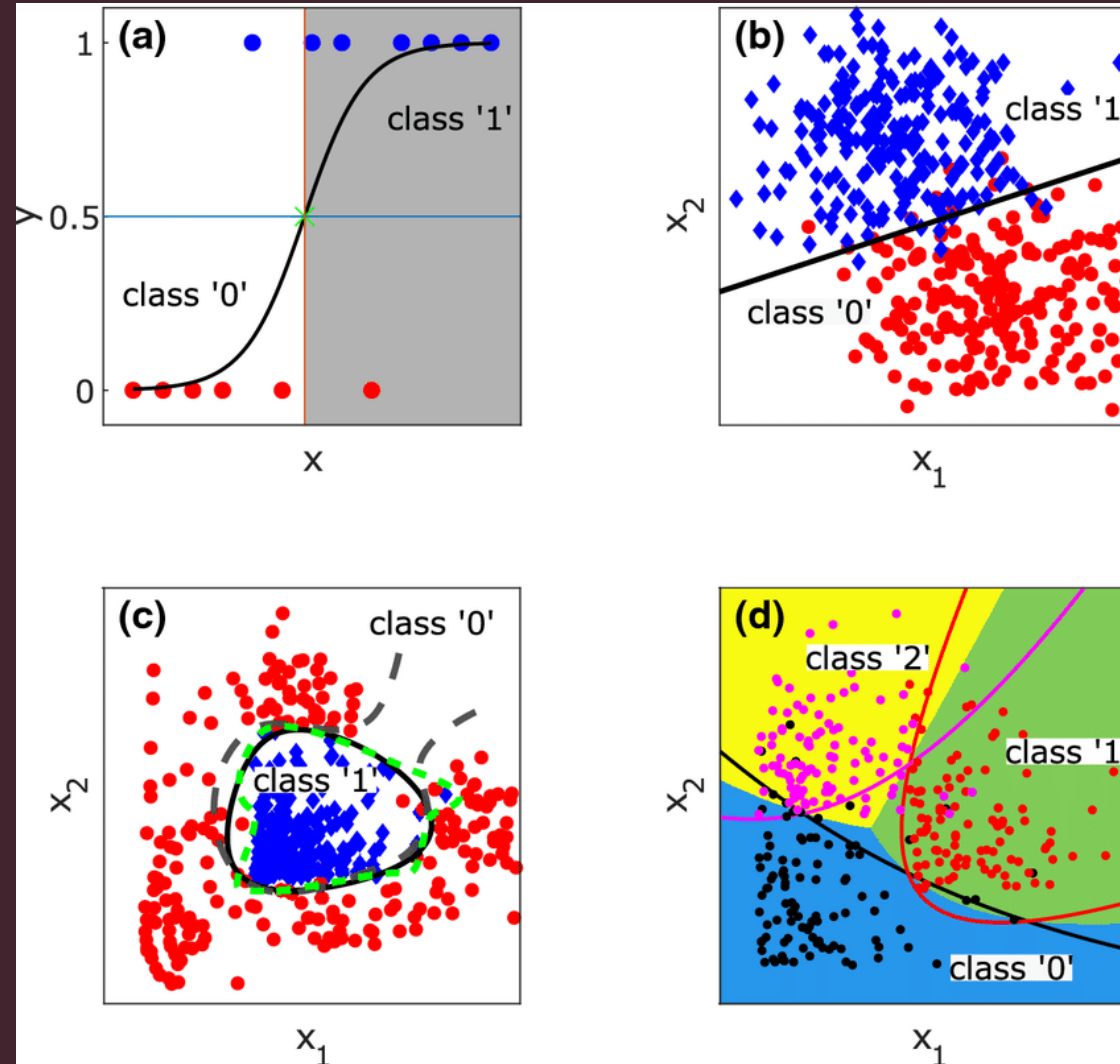
Case-Control Study



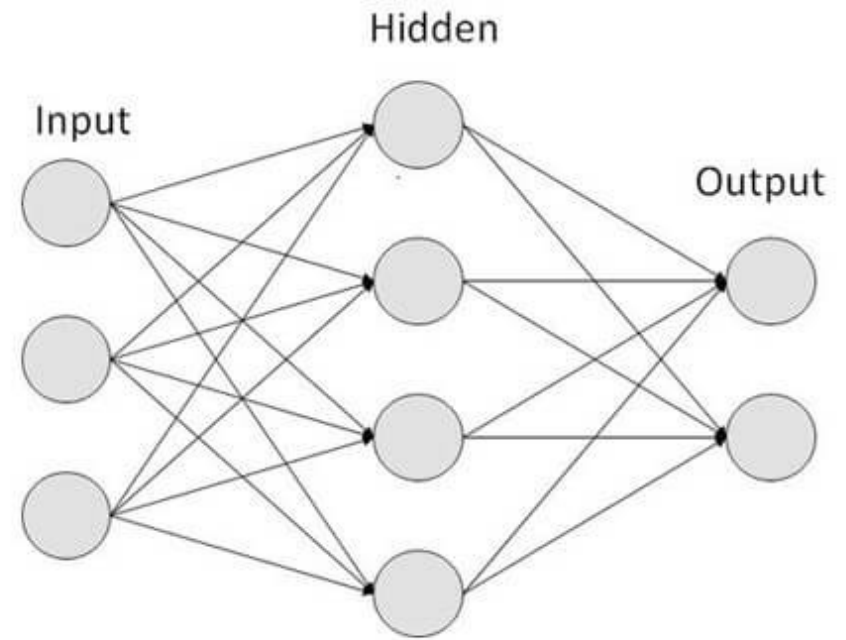
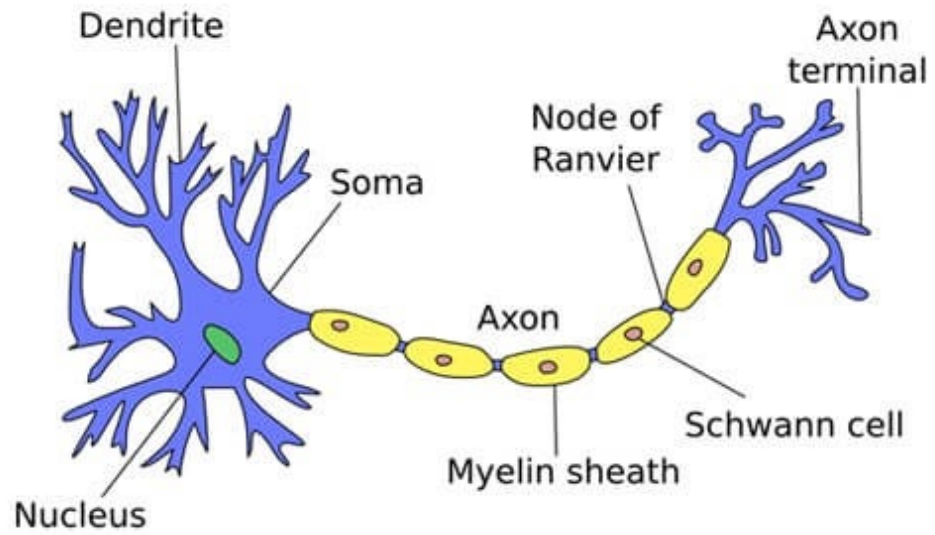
Protection of Human Subjects



Machine Learning



Deep Learning



HOW TO WRITE A RESEARCH PAPER



What is a Research Paper?

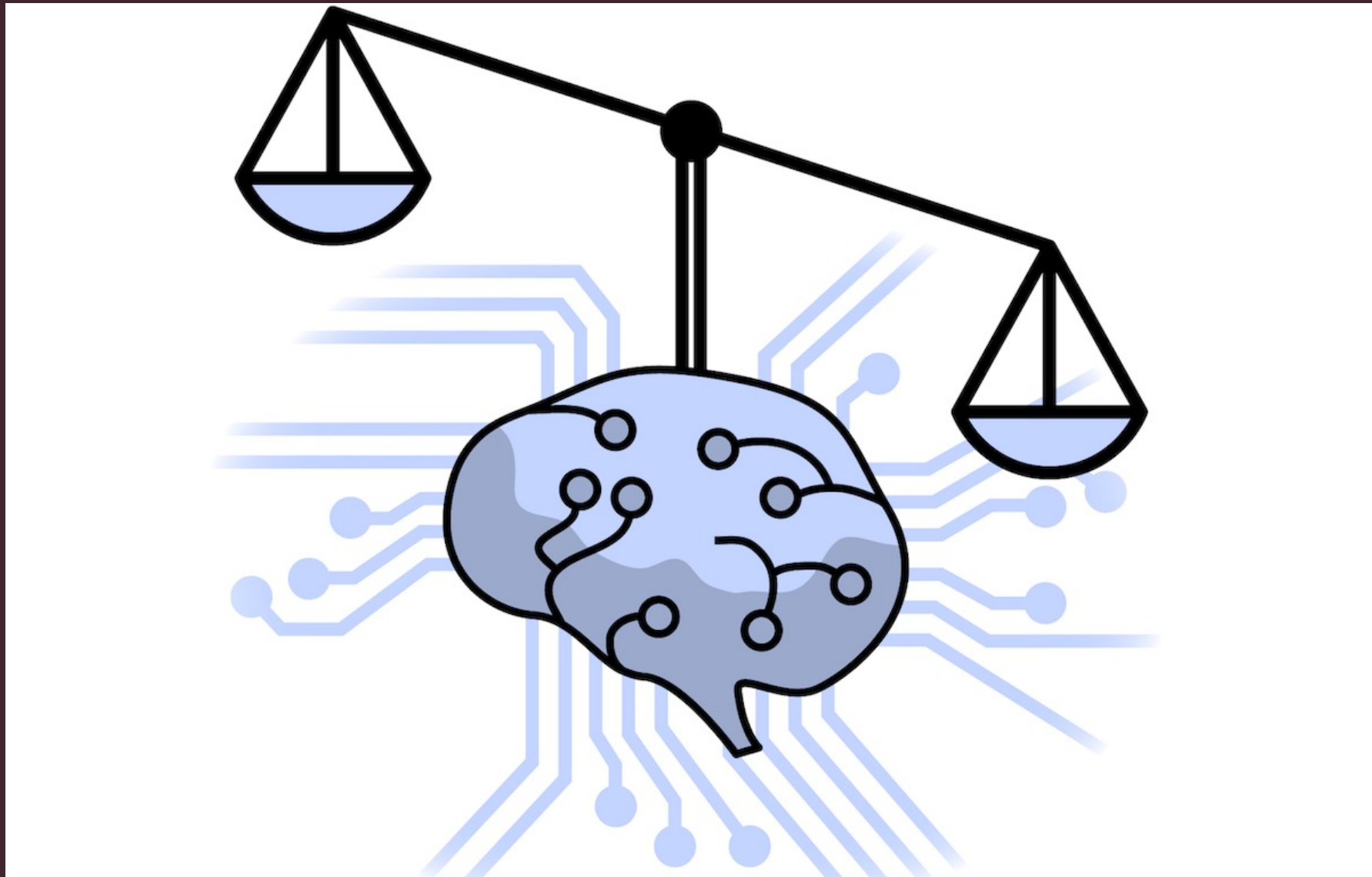
A research paper is a piece of academic writing, consisting of analysis, interpretation, and argument-based research. In comparison to essays, research papers are far more detailed. They don't only focus on your writing skills, but your research skills as well, hence the name 'research paper'. When writing a research paper, you don't just share your opinion on the matter, but also include a lot of factual information including any relevant sources.

4 STEPS FOR WRITING A RESEARCH PAPER

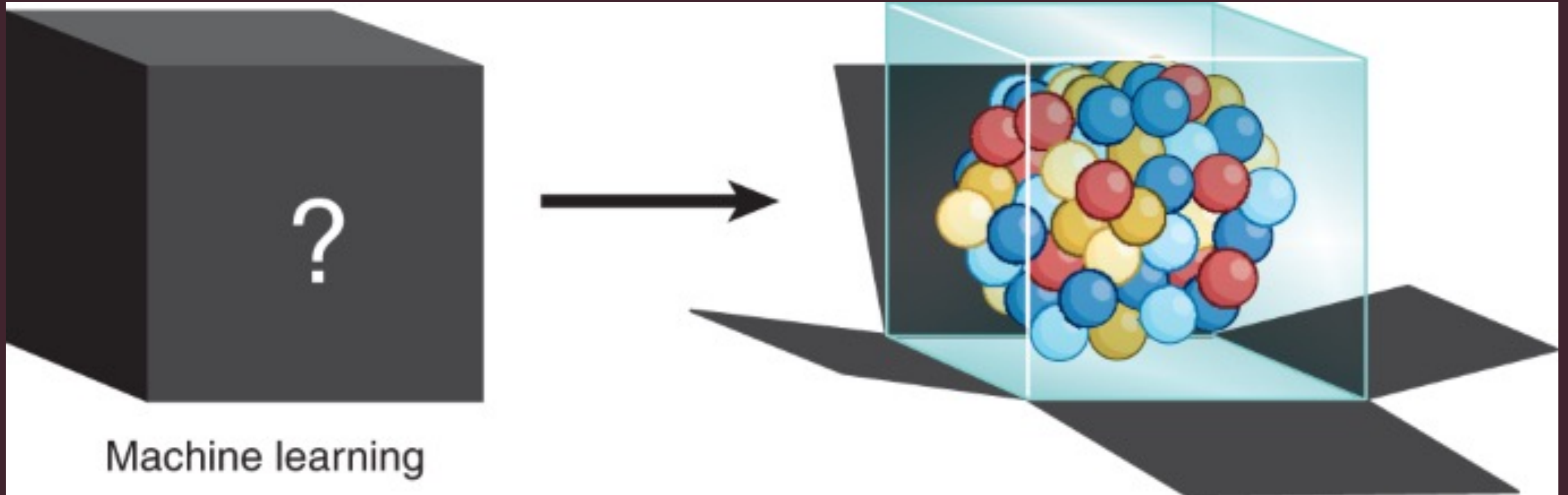
Data Privacy and Security



Bias and Fairness



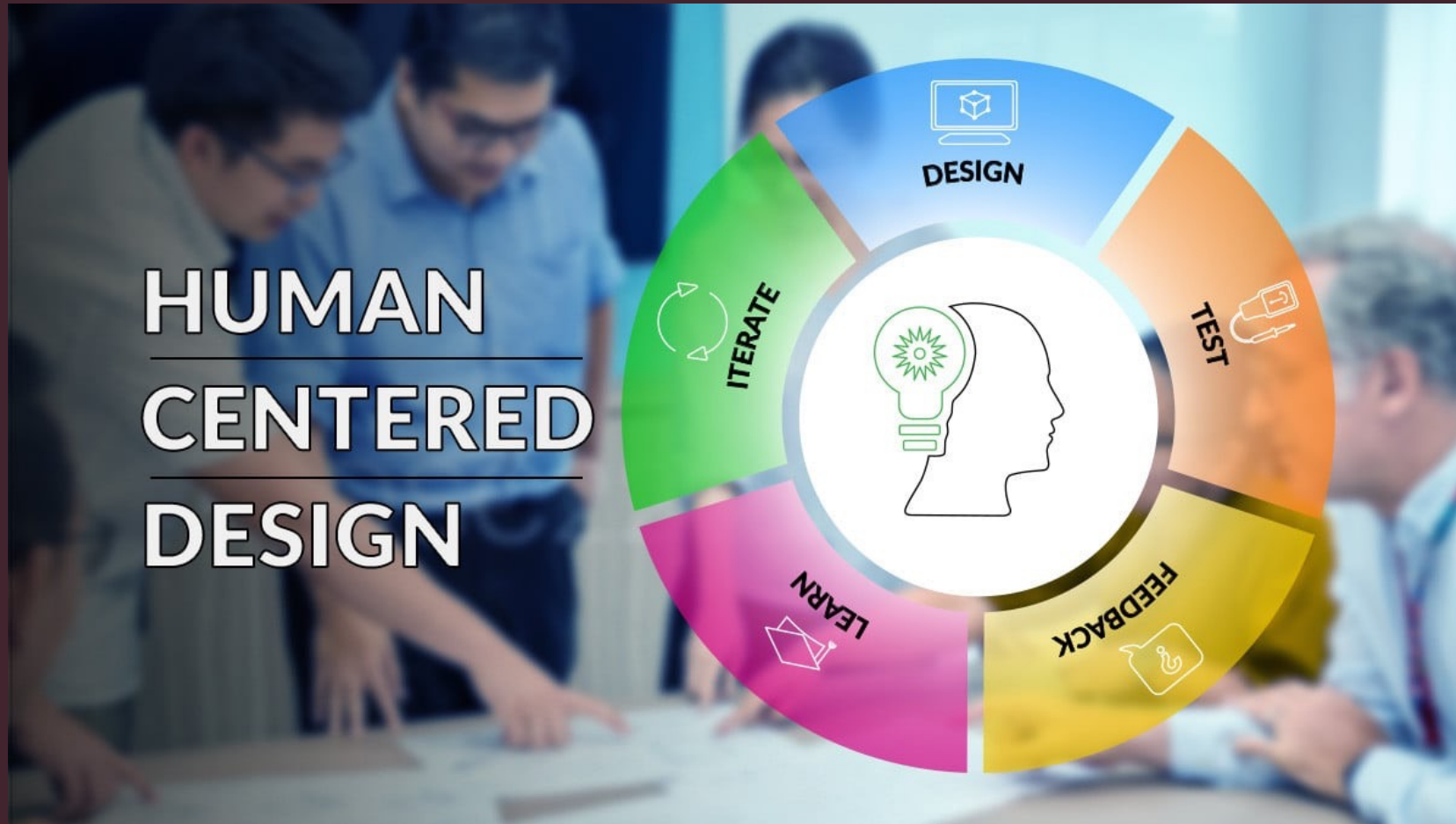
Transparency and Interpretability



Data Ethics



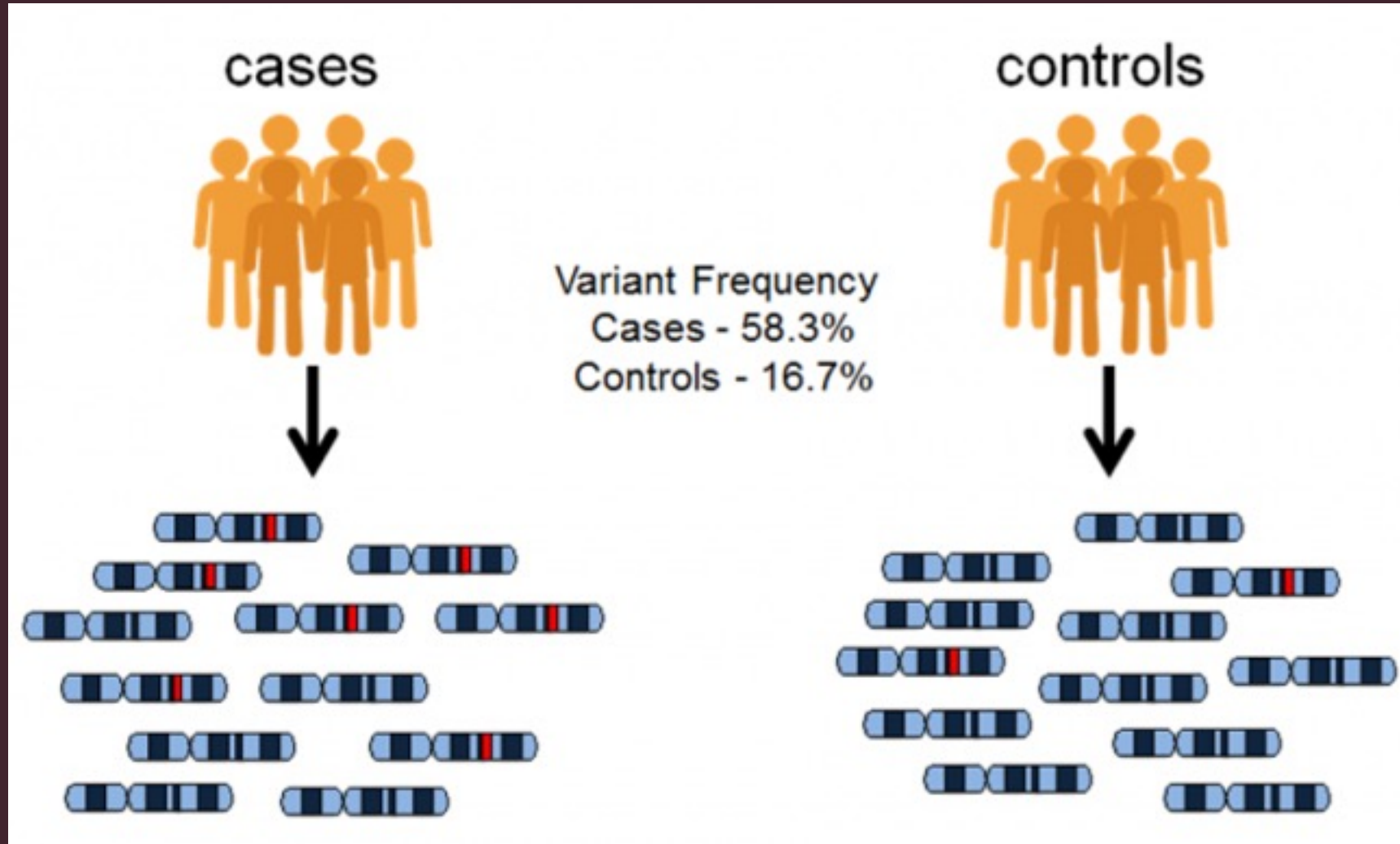
Human-Centered Design



Computational Biology



Disease Association Studies




Digital Diagnostics



Digital Therapeutics



Natural Language Processing

 M Tell me how ChatGPT works.



ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.



Social Network Analysis



Multimedia Analysis

The screenshot displays a multimedia analysis interface. At the top, there is a close button (X), a microphone icon, the word "Compare", and a menu icon (three dots). On the left, there is a stick figure icon. The main video shows a person in a dynamic pose with a yellow ball in their right hand. A green line tracks the path of the ball, and a blue arrow points to the right. A black box with the text "134°" is positioned near the person's right arm. Another black box with the text "1.772" is positioned near the person's legs. At the bottom, there is a play button, a "1x" speed indicator, and a timeline with a yellow dot. On the right side, there is a toolbar with icons for drawing lines, circles, and arrows, and buttons for "Undo" and "Clear".

Takeaway #1: Tradeoffs Are Everywhere (There Is No “Right Answer”)

Data Sovereignty / Privacy

Model Fairness



Model Interpretability

Model Accuracy



...

What other tradeoffs have we discussed?

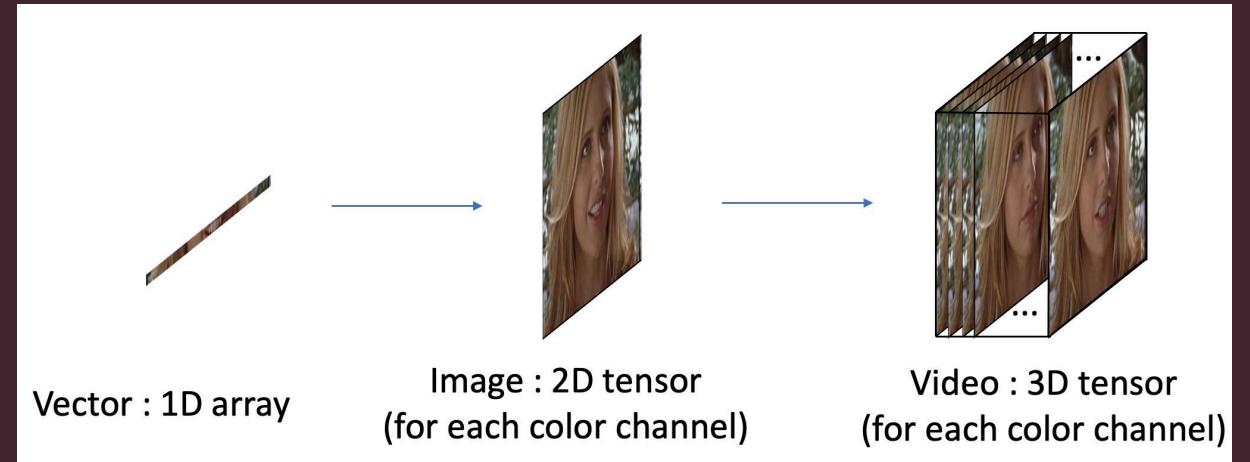
Takeaway #2: A Key Challenge is Properly Representing Data as Numbers

Tabular Data

columns = attributes for those observations

Player	Minutes	Points	Rebounds	Assists
A	41	20	6	5
B	30	29	7	6
C	22	7	7	2
D	26	3	3	9
E	20	19	8	0
F	9	6	14	14
G	14	22	8	3
I	22	36	0	9
J	34	8	1	3

Rows = observations



How might we represent the following as numbers?
Social Networks, DNA Sequences

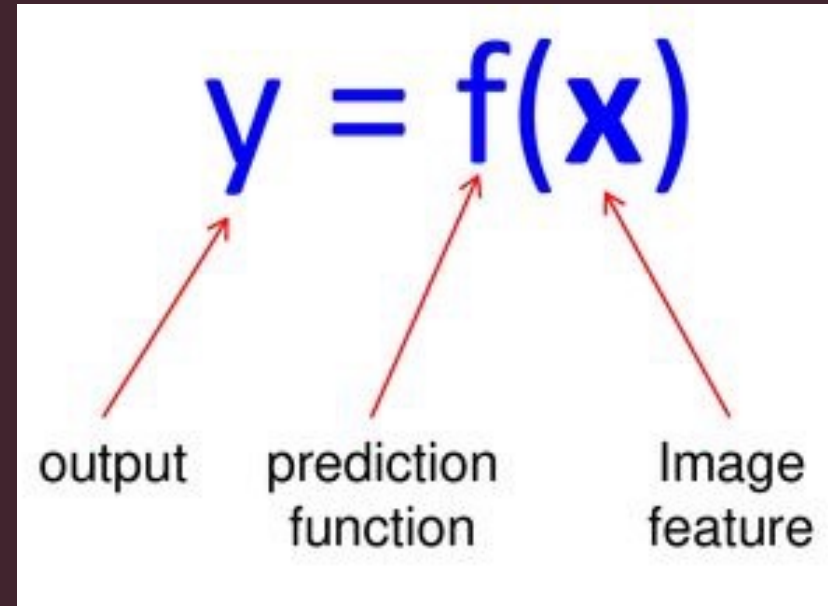
Takeaway #3: Math is Key to Data Science

The formula used by the software to determine that, is derived from [Bayes' theorem](#)

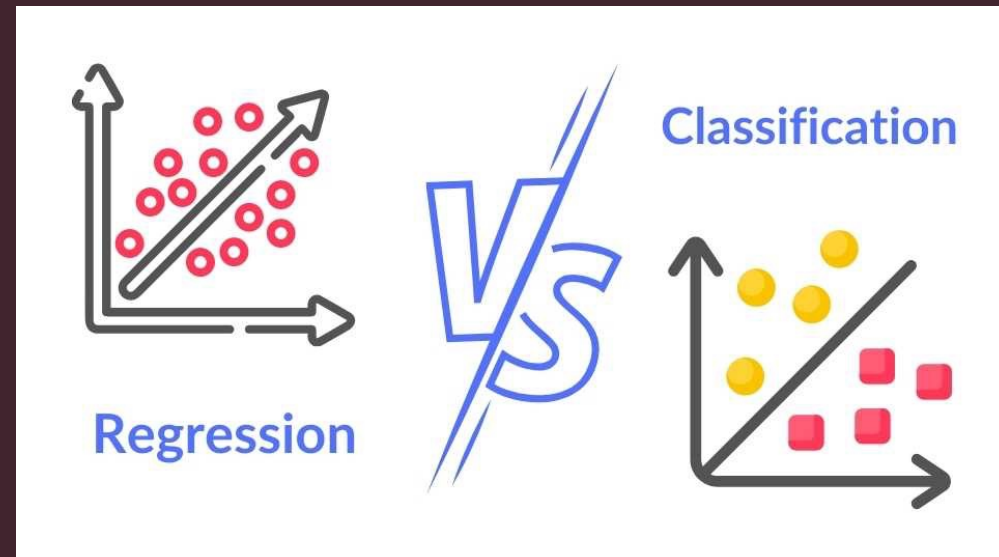
$$\Pr(S|W) = \frac{\Pr(W|S) \cdot \Pr(S)}{\Pr(W|S) \cdot \Pr(S) + \Pr(W|H) \cdot \Pr(H)}$$

where:

- $\Pr(S|W)$ is the probability that a message is a spam, knowing that the word "replica" is in it;
- $\Pr(S)$ is the overall probability that any given message is spam;
- $\Pr(W|S)$ is the probability that the word "replica" appears in spam messages;
- $\Pr(H)$ is the overall probability that any given message is not spam (is "ham");
- $\Pr(W|H)$ is the probability that the word "replica" appears in ham messages.



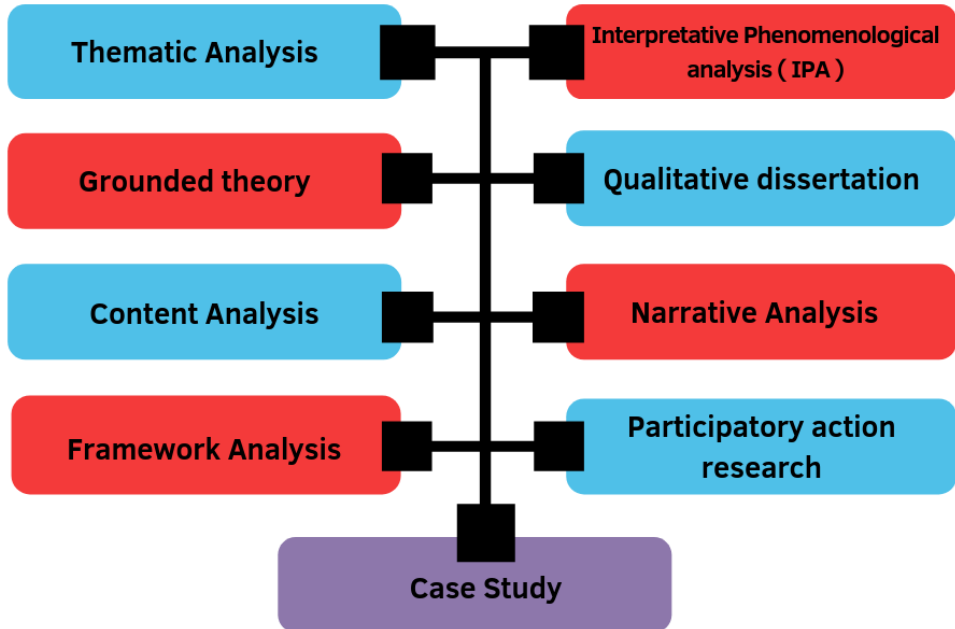
Probability Distributions				Ace Tutors
Continuous	Uniform $\mu = \frac{a+b}{2}$ $\sigma = \sqrt{\frac{(b-a)^2}{12}}$ $P(X < x) = \frac{x-a}{b-a}$	Exponential $\mu = \frac{1}{\gamma}$ $\sigma = \frac{1}{\gamma}$ $P(X < x) = 1 - e^{-\gamma x}$	Normal $z = \frac{x - \mu}{\sigma}$ $P(X < x) \Rightarrow$ Use Z-Chart	Key γ = rate parameter z = z-score p = probability of success n = # of trials N = population size K = # of success states
	Binomial $\mu = n \cdot p$ $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	Geometric $\mu = \frac{1}{p}$ $\sigma = \frac{\sqrt{1-p}}{p}$ $P(X = x) = (1-p)^{x-1} p$	Hypergeometric $\mu = n \frac{K}{N}$ $\sigma = \sqrt{n \frac{K(N-K)(N-n)}{N^2(N-1)}}$ $P(X = x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$	
	Discrete	Discrete	Discrete	



Takeaway #4: But Math Can't Answer Everything, Especially Socio-Technical Issues and Human Factors



TYPES OF DATA ANALYSIS PLANS




Career Opportunities: Data Scientist




Principal Data Scientist - Product Innovation

CVS Health
New York, NY
via Robots..Jobs
Full-time Health insurance



Data Scientist, Ad Measurements Science

Amazon
New York, NY
via ZipRecruiter
23 hours ago Full-time




Data Scientist III, User Safety

Hinge
New York, NY
via Eniac Ventures




Data Scientist, GBG Data Science

Meta
New York, NY
via LinkedIn
4 days ago \$ 134K-194K a year Full-time



Director, Data Scientist

Pfizer
New York, NY
via Pfizer Jobs
Full-time Health insurance Dental insurance
Paid time off



Data Scientist

Progyny
New York, NY



Senior Associate, Data Scientist
New York Life Insurance Co
New York, NY
via Jobs At New York Life Insurance
28 days ago Full-time



Data Scientist - Multiple Levels
Grow Therapy
New York, NY
via Greenhouse
Full-time



Data Scientist
T-Mobile
New York, NY
via T-Mobile Careers
2 days ago Full-time



Senior Data Scientist
Fidelity Investments Careers
Jersey City, NJ
via Fidelity Careers - Fidelity Investments
24 days ago Full-time Health insurance
Paid time off



Senior, Data Scientist
Walmart
Hoboken, NJ
via Walmart Careers
7 days ago Full-time and P
Health insurance Dental in



Director of Data Science (Generative AI)
New York Life Insurance Co
New York, NY



Senior Data Scientist, AI Foundations
Capital One
New York, NY
via LinkedIn
1 day ago Full-time and Part-time
Health insurance



Entry Level Data Scientist (Remote)
SynergisticIT
Anywhere
via LinkedIn
5 days ago Work from home Contractor



Principal Data Scientist, Marketing Intelligence
People Center
McLean, VA



Senior Data Scientist
Verizon
Anywhere
via Verizon Careers
6 days ago Work from home Full-time
Paid time off

Career Opportunities: ML Engineer

2 days ago Full-time and Part-time

M **Senior Machine Learning Engineer**
Mulberry Technology
New York, NY
via Wellfound

1 month ago Full-time Health insurance
Dental insurance Paid time off

J **Engineer Machine Learning Operations**
JetBlue Airways Corporation
New York, NY
via ZipRecruiter

6 days ago Full-time Health insurance

Engineer
Capital One Financial Corp
New York, NY
via Trabajo.org

3 days ago Full-time

2σ **Software Engineer - Machine Learning: Techniques Engineering**
Two Sigma
New York, NY
via LinkedIn

16 days ago Full-time Health insurance
Dental insurance Paid time off

Associate Specialist, ML Engineer Fellowship (Hiring Our Heroes...)
Merk

Etsy **Senior ML Engineer I**
Etsy
New York, NY
via The Muse

Full-time Health insurance

H **Applied ML Engineer**
Hologram Labs
New York, NY
via Lever

Full-time

R **Machine learning Engineering Graduate**
Reesby
New York, NY
via ZipRecruiter

1 day ago Full-time and Part-time

F **Machine Learning Operations Engineer**
Fusemachines
Anywhere
via Wellfound

12 days ago \$ 100K-120K a year
Work from home Full-time

Machine Learning Engineer, Generative ML, 3+ Years of Experience
Snap Inc.
New York, NY
via LinkedIn

8 days ago Full-time Health insurance

Director of Engineering, ML Platform
MongoDB
New York, NY
via MongoDB

Full-time

Machine Learning Engineer Creator Studio
Spotify
New York, NY
via The Org

Full-time

S **Machine learning engineer**
Stripe
New York, NY
via Talent.com

Software Engineer, Machine Learning
Meta
New York, NY
via LinkedIn

3 days ago \$ 116,002-168,000 a year
Full-time

Senior ML Engineer
Sounder
New York, NY
via Wellfound

25 days ago Full-time Health insurance
Dental insurance Paid time off

H **Lead AI / ML Engineer**
Human Capital Solutions, Inc
New York, NY
via ZipRecruiter

Sr Machine Learning Engineer - ML OPS
CVS Health
New York, NY
via CVS Health

27 days ago Full-time Health insurance
Dental insurance Paid time off

AIML - ML Research Engineer, Foundation Models
Apple
New York, NY
via Careers At Apple

Full-time Health insurance Dental insurance

D **Principal Machine Learning Engineer**
Disney Entertainment & ESPN Technology
New York, NY


HELLO FRESH **Senior ML Engineer**
HelloFresh
New York, NY
via Wellfound

27 days ago Full-time Health insurance
Paid time off

2024 Intern - Machine Learning Engineer
Adobe
New York, NY
via LinkedIn


3 days ago Full-time and Internship

Career Opportunities: UI/UX Designer for Data-Centric Products



AI Human Factors Specialist / Data Scientist
National Security Agency
Fort Meade, MD
via SimplyHired


🕒 26 days ago 📁 Full-time




Data Scientist, Human Factors and User Research
Takeda Pharmaceutical Company
Lexington, MA
via Karkidi

💰 \$ 120K–190K a year 📁 Full-time

🛡️ Health insurance 🦷 Dental insurance




Human Factors Design Researcher, AI/ML
Apple
Seattle, WA




Research Scientist - Human Computer Interaction
Spotify
New York, NY
via Karkidi

💰 \$ 120K–190K a year 📁 Full-time




Human Factors Researcher, Fitbit Devices
Google
San Francisco, CA
via Karkidi

💰 \$ 120K–190K a year 📁 Full-time



Human Factors Senior Managing Scientist
Exponent Inc.
Natick, MA
via ZipRecruiter


📁 Full-time 🛡️ Health insurance





Human Factors Design Researcher, AI/ML
Apple
Seattle, WA
via Robots.Jobs

📁 Full-time


Career Opportunities: Product Manager for Data-Centric Products

 **Principal Product Manager, AI/ML**
ZoomInfo
United States
via VentureFizz
4 days ago Full-time Health insurance


 **Product Manager - Regulatory - Healthcare Data & Analytics Platform**
Oracle
Anywhere
via LinkedIn
3 days ago Work from home Full-time Health insurance Dental insurance

 **Product Manager, Data Movement & Storage, RaDS IT**
Merck
New York, NY
via CareerBuilder
3 days ago Full-time Health insurance Dental insurance Paid time off

 **Product Manager, Enterprise Data Storage**
Capital One
New York, NY
via LinkedIn
2 days ago Full-time and Part-time Health insurance

 **Specialist, Product Management (Data Science and Analytics)**
Mastercard
Arlington, VA
via Professional Diversity Network
17 days ago Full-time No degree mentioned

 **Technical Product Manager Data Pipelines**
Abbott Laboratories
Alameda, CA
via BeBee
5 days ago Full-time Health insurance


 **Technical product manager**
Verizon
Bernards, NJ
via Talent.com
6 days ago \$ 89K-165K a year Full-time and Part-time Health insurance

 **Enterprise Data & Analytics - Senior Product Manager - PVH Corp.**
Business Operations
Somerville, NJ
via SaluteMyJob
1 day ago Full-time


 **ESG Data Product Manager**
Bloomberg
New York, NY
via JobzMall
\$ 70K-137K a year Full-time No degree mentioned


 **Senior Product Manager, Preference Management**
Intuit
Mountain View, CA
NBCUniversal
New York, NY
via Talent.com
3 days ago Full-time No degree mentioned


 **Data Analytics Product Managers**
FedEx
Collierville, TN
via Chegg Internships
Full-time


 **Reference Data Product Manager**
JPMorgan Chase Bank, N.A.
New York, NY
via ZipRecruiter
3 days ago Full-time Health insurance Dental insurance

 **Product Data Managers**
Clover Health
Jersey City, NJ

 **Product Manager - Data and Monitoring (Remote)**
Paxos
Anywhere
via Built In
Work from home Full-time No degree mentioned Health insurance

 **Data Product Manager at Emsi in Moscow, ID**
EMSI
Moscow, ID
via Geebo

 **Product Manager, Security**
MongoDB
New York, NY
via MongoDB
23 days ago Full-time

 **KFC Global Sr Data Product Manager**
KFC
Anywhere

Career Opportunities: New Jobs Created by the AI Economy (Prompt Engineer, Data Annotator, etc.)

C **Artificial Intelligence/Prompt Engineering Curriculum...**
Chegg, Inc.
Providence, RI
via Monster
5 days ago Contractor
No degree mentioned

J **Gen AI Engineer**
January
Menlo Park, CA
via Lever
Full-time Health insurance Dental insurance

UR **Senior AI Prompt Engineer**
Understanding Recruitment
United States
via LinkedIn

F **Prompt Engineers**
forhyre.com
New York, NY
via LinkedIn
Full-time

Prompt Engineer
The Hartford
New York, NY
via Getwork
1 day ago Full-time No degree mentioned

Legal Prompt Engineer - LLM and GPT
CONTRACTPOD TECHNOLOGIES LT
New York, NY
via Learn4Good

U **Professional Prompt Engineer Needed - Contract to Hire**
Upwork
Anywhere
via Upwork
23 hours ago Work from home Contractor and Temp work

AI Prompt Engineer
Dropchat
Anywhere
via LinkedIn
24 days ago Work from home Full-time

U **Prompt engineering**
Upwork
Anywhere
via Upwork

Sr. Prompt Engineer
Khan Academy
Anywhere
via Fast Forward Job Board
18 days ago Work from home Full-time
Health insurance Dental insurance

U **Prompt engineer**
Upwork
United States
via Prompt Engineering Jobs
Contractor No degree mentioned

6 **Enterprise Prompt Engineer**
631 Booz Allen Hamilton_United States
Anywhere
via Workday
via Remote AI Jobs
Full-time

Prompt Engineers
Forhyre
Remote, OR
via BeBee
2 days ago Full-time

Prompt Engineer
The Hartford
Hartford, CT
via LinkedIn
1 day ago Full-time No degree mentioned