

The Secret Life of Data

Upper Bound 2025

Bailey Kacsmar



PUPS
Research Lab



Privacy, Data Science and AI?



~~Privacy, Data Science, and AI?~~

Data Generation
and Collection



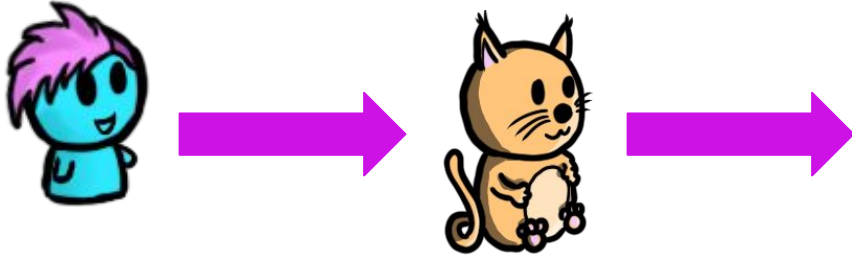
A Snapshot from a Day in the Life of Alice



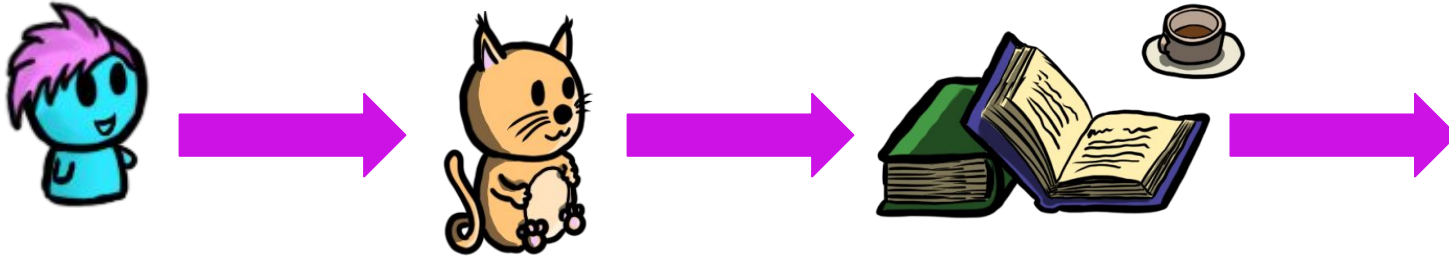
Alice's Day



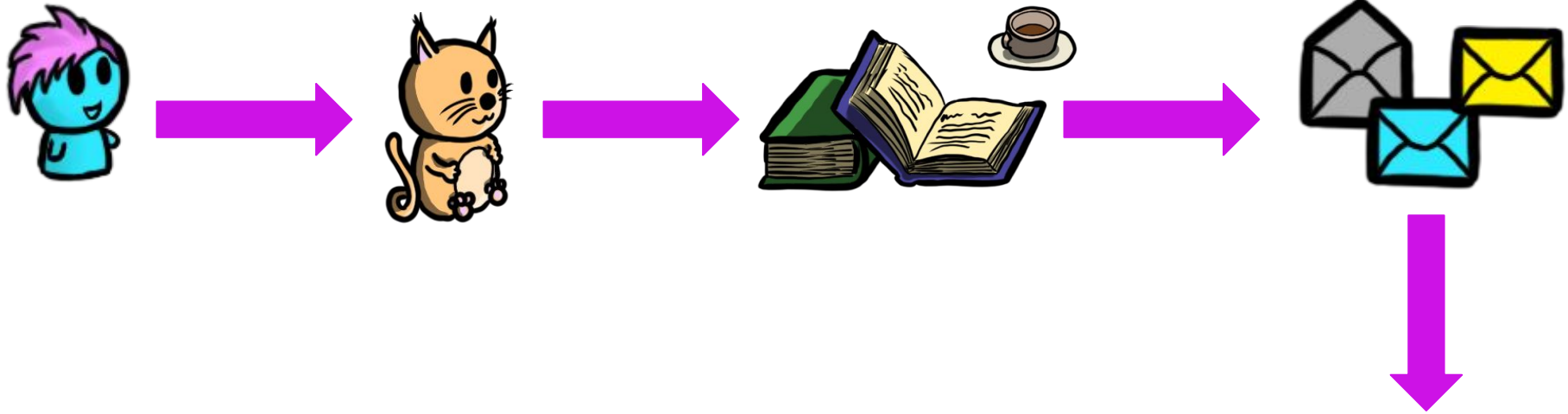
Alice's Day



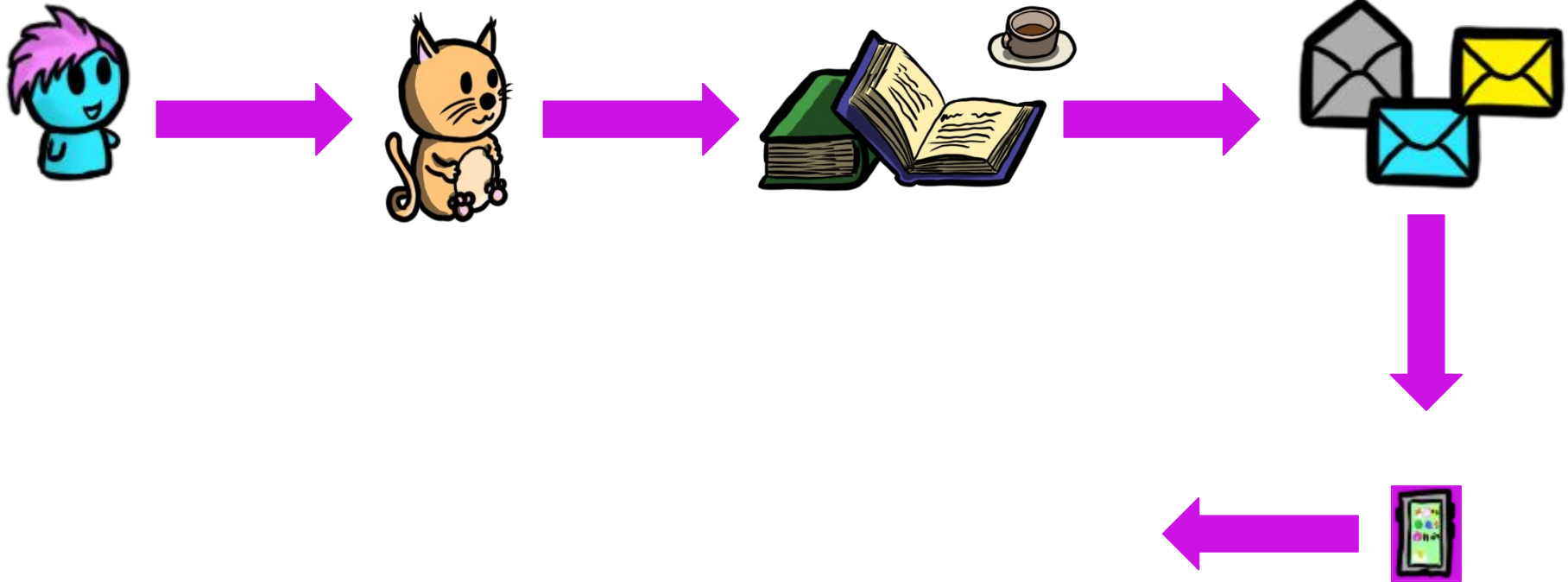
Alice's Day



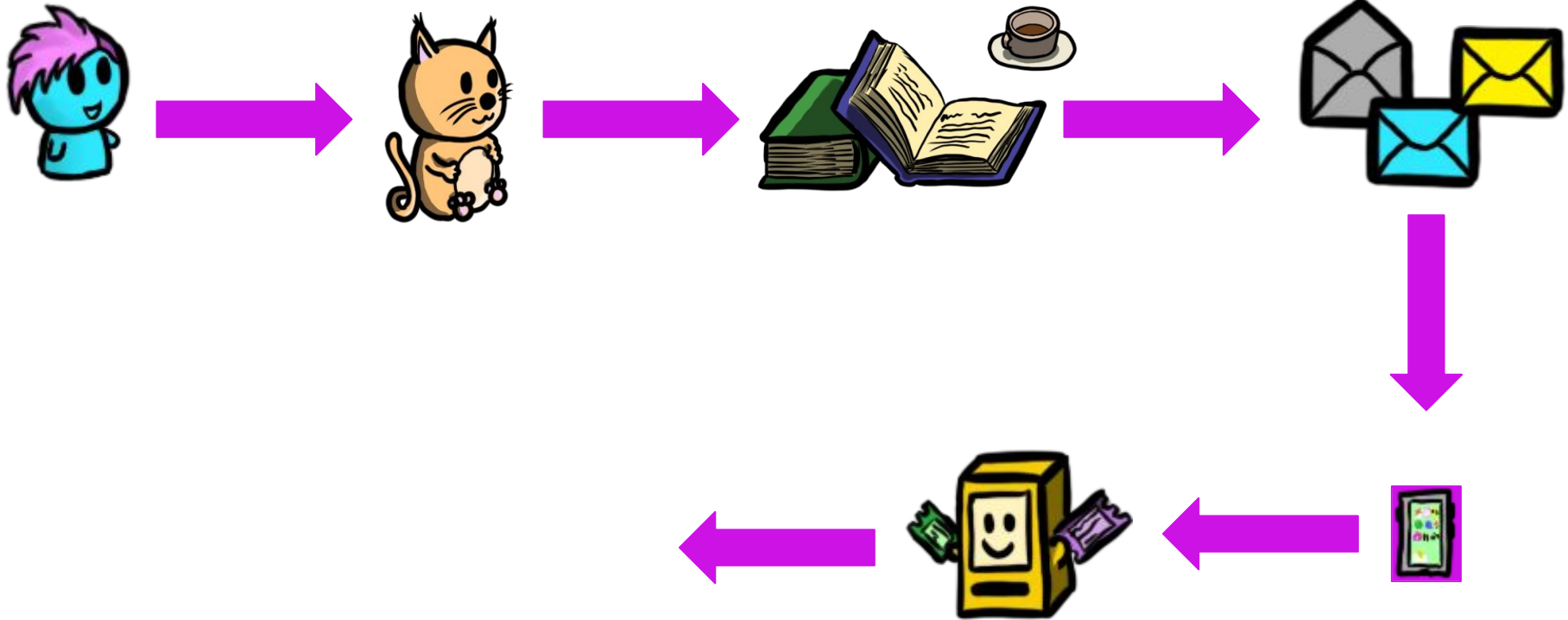
Alice's Day



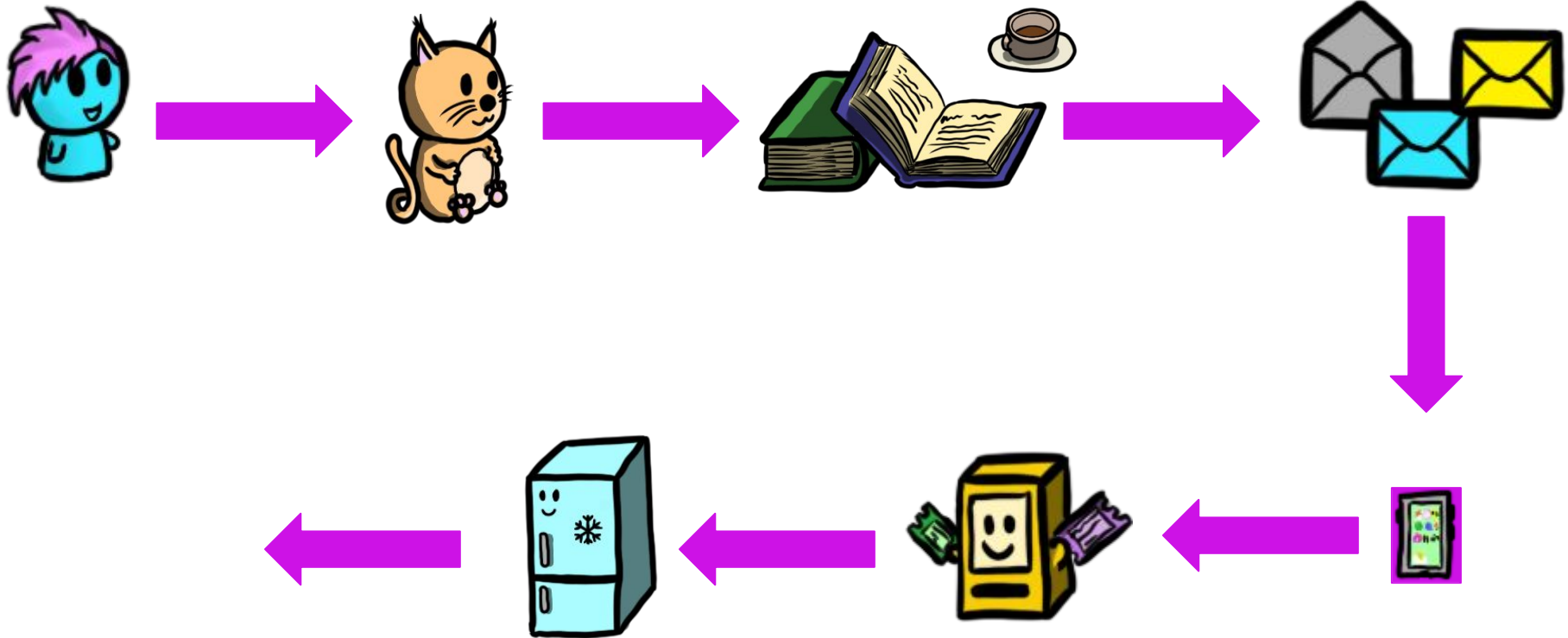
Alice's Day



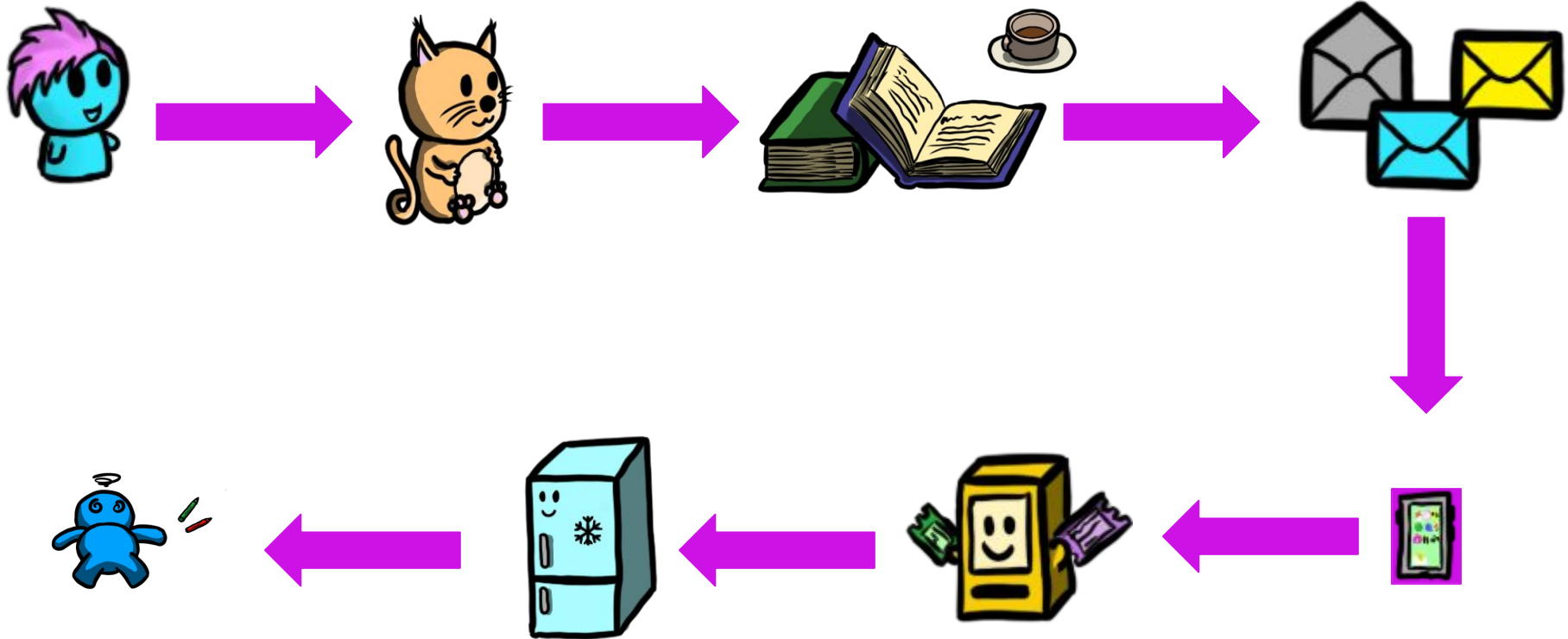
Alice's Day



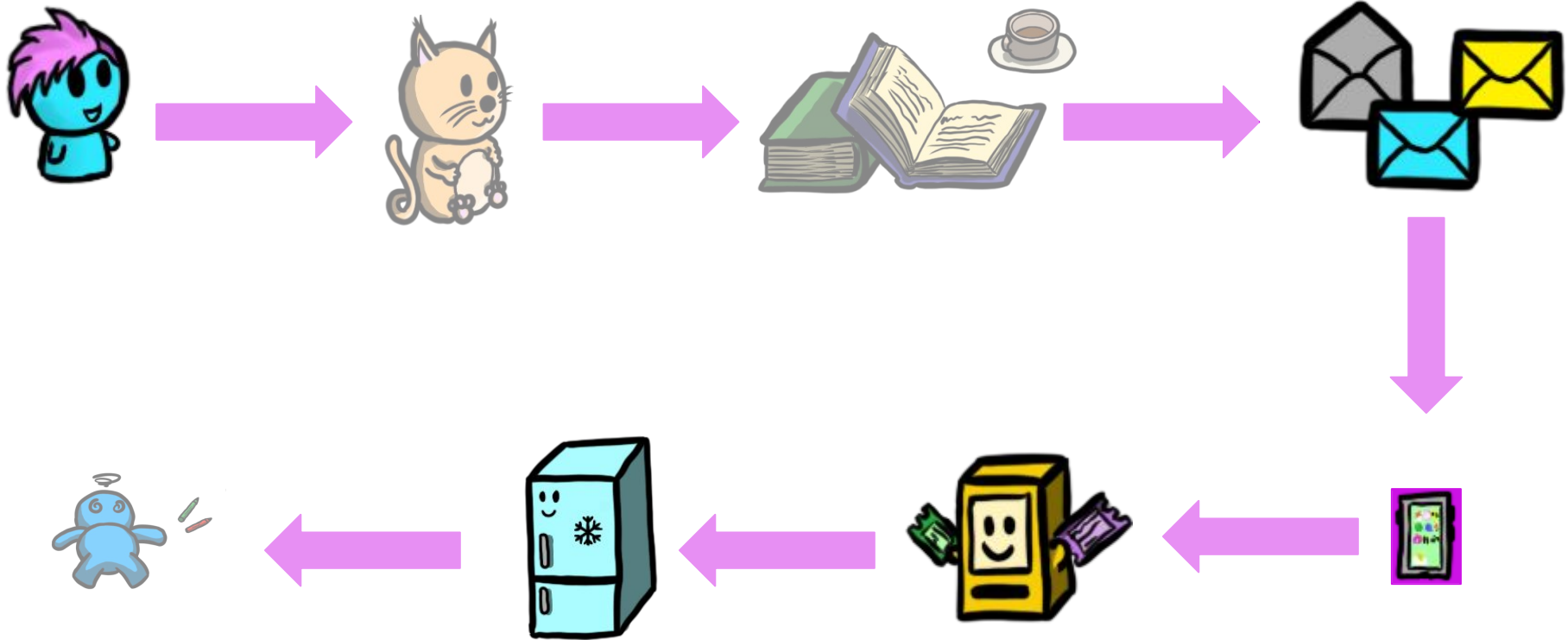
Alice's Day



Alice's Day



Data Happened



Another look at the
snapshot view of
Alice's life...

Misleading Expectation



**Ticket Seller
Accessed via
Browser**



Enhanced ad privacy in Chrome

We're launching new privacy features that give you more choice over the ads you see.

Chrome notes topics of interest based on your recent browsing history. Also, sites you visit can determine what you like. Later, sites can ask for this information to show you personalized ads. You can choose which topics and sites are used to show you ads.



To measure the performance of an ad, limited types of data are shared between sites, such as the time of day an ad was shown to you.

More about ads in Chrome



You can make changes in Chrome settings

Settings

Got it

Source article from EFF: <https://www.eff.org/deeplinks/2024/07/why-privacy-badger-opts-you-out-googles-privacy-sandbox>

The Unexpected and Unauthorized

Consider mobile apps, beyond permissions,

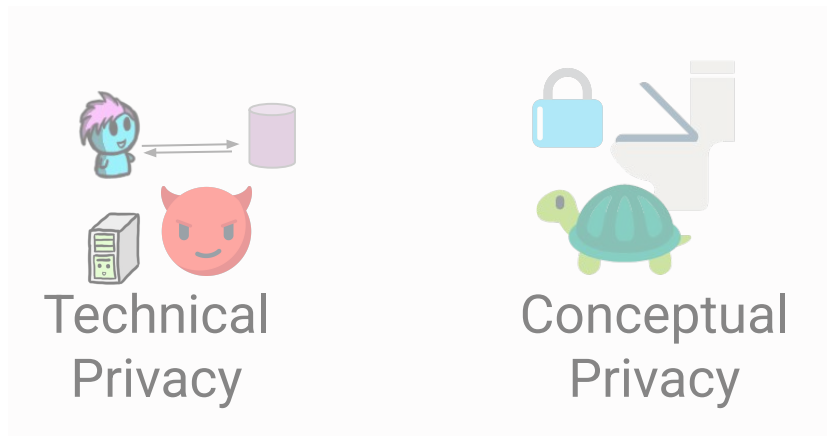
- Side channels - accessing without permissions via other means
- Covert channels - via a set of colluding apps

The consequences are you are sharing more than you think!

Wait...isn't this illegal?

J. Reardon et al. "50 ways to leak your data: An exploration of apps' circumvention of the android permissions system." In 28th USENIX security symposium 2019.

Legal Privacy is Complicated and Vague



“Trusted-third parties”, “Partners”

PetSmart's [privacy_policy](#) states: "We may share the information we collect with companies that provide support services to us."

Data Brokers Also Have Alice's Data

- Purpose is the **collection of data** about people
- Source data includes both **public information and sold information**
- Is an organization that **has your data, without you every interacting with them**
- Depending on jurisdiction...**regulated by very different laws** than other organizations



CFPB Quietly Kills Rule to Shield Americans From Data Brokers

Russell Vought, acting director of the Consumer Financial Protection Bureau, has canceled plans to more tightly regulate the sale of Americans' sensitive personal data.

<https://www.wired.com/story/cfpb-quietly-kills-rule-to-shield-americans-from-data-brokers/>


The Unaware - AI've Got Your Records

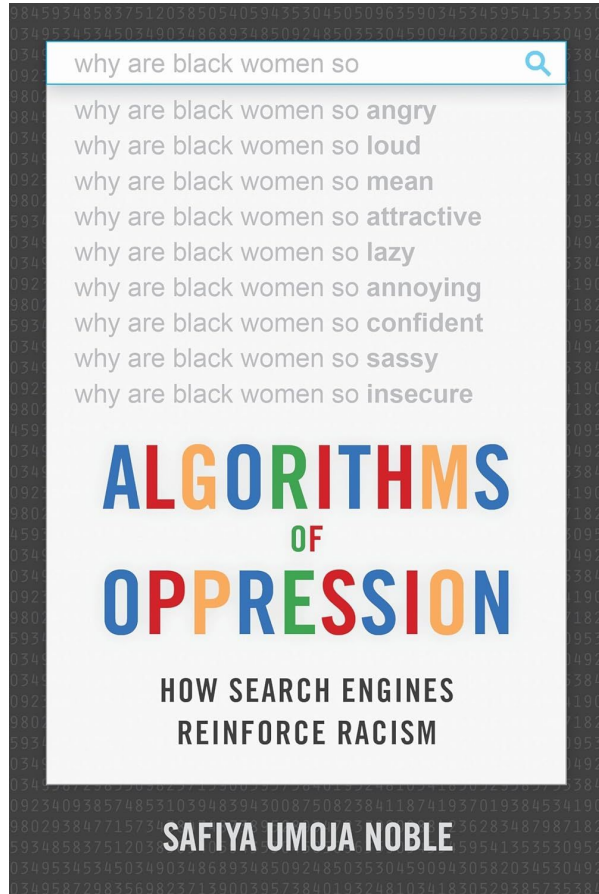
Technology

Revealed: How the UK tech secretary uses ChatGPT for policy advice

New Scientist has used freedom of information laws to obtain the ChatGPT records of Peter Kyle, the UK's technology secretary, in what is believed to be a world-first use of such legislation

By [Chris Stokel-Walker](#)

 13 March 2025



“While we often think of terms such as ‘**big data**’ and ‘**algorithms**’ as being **benign, neutral, or objective**, they are anything but.”

- Alg. of Opp. Safiya Umoja Noble, 2018

Data, Beyond the Abstraction

Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales

Google found the perfect way to link online ads to store purchases: credit card data

By [Mark Bergen](#) and [Jennifer Surane](#)

August 30, 2018, 3:43 PM EDT Updated on August 31, 2018, 12:40 PM EDT

[washingtonpost.com](https://www.washingtonpost.com)

Now for sale: Data on your mental health

Drew Harwell

Home Depot didn't get customer consent before sharing data with Facebook's owner, privacy watchdog finds | CBC News

Catharine Tunney · CBC News · Posted: Jan 26, 2023 9:53 AM
Updated: January 27

These retailers share customer data with Facebook's owner. Customers may not have been told | CBC News

Thomas Daigle · CBC News · Posted: Feb 07, 2023 4:00 AM EST | Last

Double-double tracking: How Tim Hortons knows where you sleep, work and vacation



James McLeod



June 15, 2020

In : Canada Privacy



0



1,169



11 min read

How did we get here?

Defining Data: Past to Present

Previously

- **Stipulations**, concessions (theologians, 'a thing given')
- **Facts** and principles

Currently:

- Data as **property**
- Data as a **raw material** or resource, something **unrefined**, **inert** and **abstract...**

E. M. Renieris. Beyond data: Reclaiming human rights at the dawn of the metaverse. MIT Press, 2023.

Enron Corpus

Fraud investigation

The Deluge of Data, A Culture of Extractivism

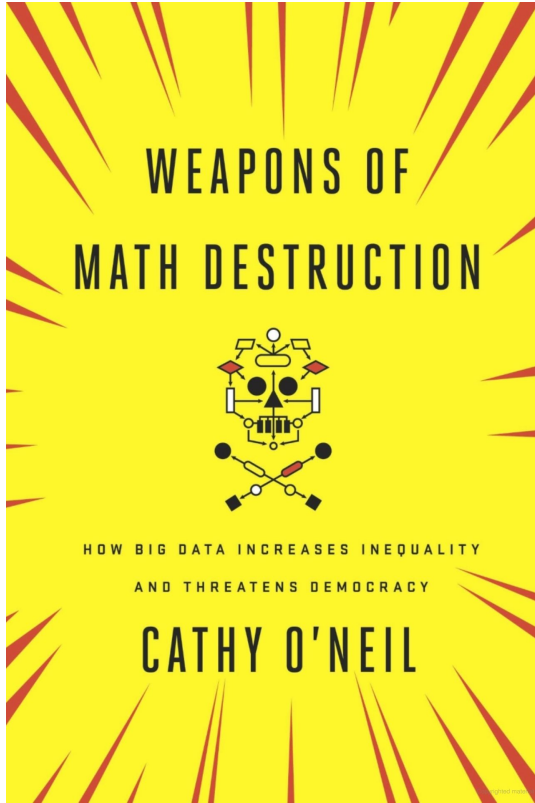
ImageNet

Scraped

MTurk Annotate

NIST Multiple
Encounter Dataset

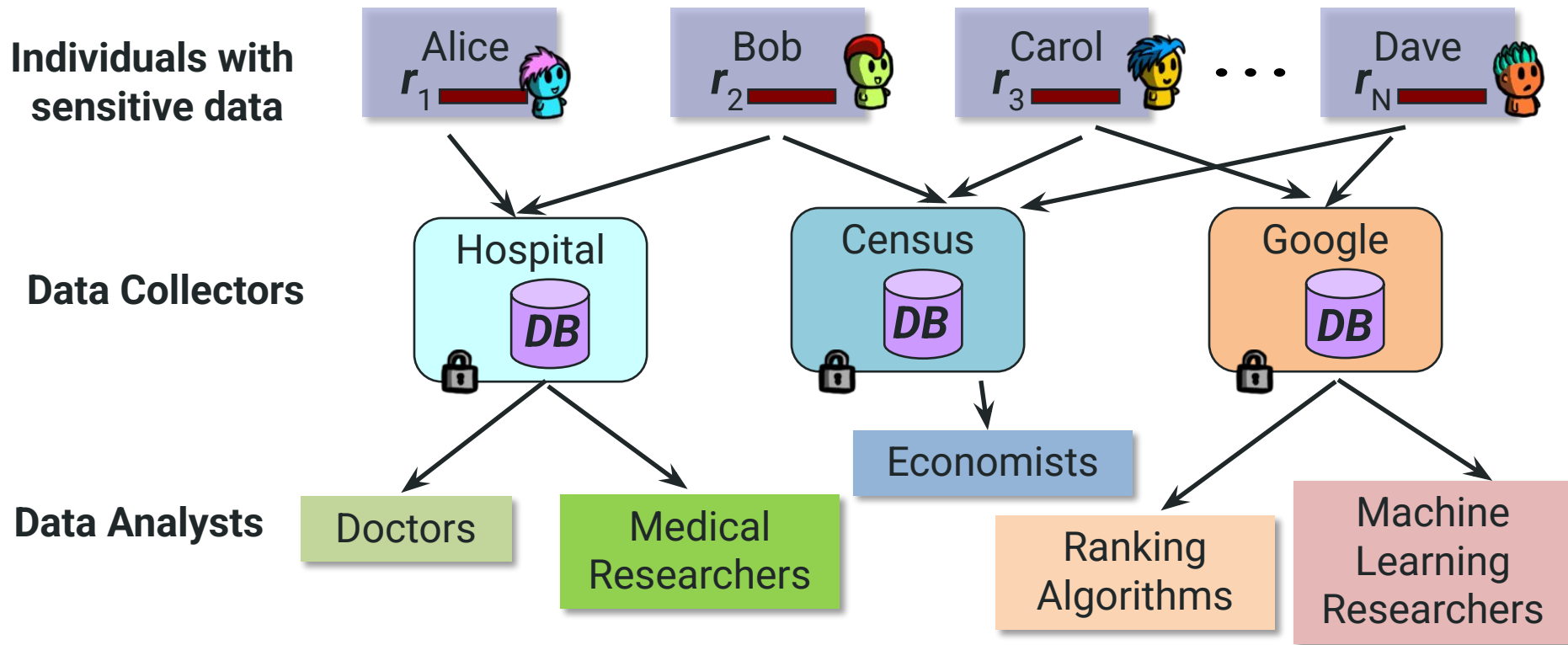
Mugshots



“They churned 24/7 through petabytes of information, much of it scraped from social media or e-commerce websites.”
- WMD, Cathy O’Neil, 2016, of the early 2000s

C. O’neil. Weapons of math destruction: How big data increases inequality and threatens democracy. Crown, 2017.

Where we are: Statistical Databases



Consequences of the Human-Data Disconnect and AI

Consequences of the Status Quo?

- Monetary
- Environmental
- Bias
- ...

“...**data centers** are the primary driver of its **electricity demand**” -
Clicking Clean, Cook et al. 2017

E. M. Bender et al. "On the dangers of stochastic parrots: Can language models be too big? 🦜." In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.

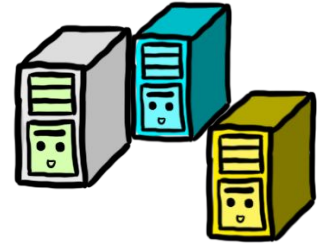
AI is Ultimately a Technology



Printing Press



**Artificial intelligence
and Machine Learning**



**Internet and
Computers**

AI is a Technology - Which has Applications

High-Risk or Otherwise Sensitive Domains

Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said

By GARANCE BURKE and HILKE SCHELLMANN

October 26, 2024

“despite OpenAI’s warnings that the tool **should not be used in ‘high-risk domains’.**”

The Automation Match of Porcha Woodruff

Eight Months Pregnant and Arrested After False Facial Recognition Match

Kashmir Hill

False positives can have catastrophic consequences for those impacted, with facial recognition frequently falsely identifying Black individuals.

The New York Times, 2023: www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html

Misunderstandings - None Shall Pass

A professor accused his class of using ChatGPT, putting diplomas in jeopardy

AI-generated writing is almost impossible to detect and tensions erupting at a Texas university expose the difficulties facing educators

Updated May 18, 2023

Not understanding what is possible and what is not has consequences for all sides of a problem.

Washington Post, 2023:

www.washingtonpost.com/technology/2023/05/18/texas-professor-threatened-fail-class-chatgpt-cheating/

Oh dear...

Oh dear... On to Overcoming the AI Hype



“When people can **spot AI hype**, they **make better decisions** about how and when to use automation, and they are in a **better position to advocate for policies** that constraint the use of automation by others” - The AI Con

Towards being
better?

The Reality

“In the case of AI, there is **no singular black box** to open, no secret to expose, **but a multitude of interlaced systems** of power. Complete transparency, then, is an impossible goal”

- The Atlas of AI, Kate Crawford

Data Norms and Consent Complications



Fig. 23. "This is me enjoying my privacy. This is the only time during the day, were I am truly alone and nothing bothers me. No man no children no dogs." By Cindy, age 54



Fig. 24. "No one come in when I am in the bathroom!" By Sydney, age 7

Consent means
choice, means
alternatives...and
also time.

M. Oates, et al. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration." PoPETs 2018.

From Data Back to People

Beyond Data

“...when everything is digital, there

**“Trying to regulate data as such is like trying to regulate *technology* as if it had a common definition or clear contours -
an exercise in futility”**

- Beyond Data, 2023, Elizabeth M. Renieris

We Have to Understand People



PEARL OYSTERS HAVE SOMETHING VALUABLE TO PROTECT - THE PEARL. THEY CAN DO SO BY SIMPLY 'CLOSING THE LID'. IF ONLY SAFEGUARDING THE DATA IN MY LAPTOP WERE THAT SIMPLE!

Fig. 62. "Pearl oysters have something valuable to protect - the pearl. They can do so by simply 'closing the lid.' If only safeguarding the data in my laptop were that simple!" By Sharon, age 25.



Fig. 33. "Privacy means that the thoughts in my brain are locked away. What I know does not have to go into the world, which I put an X over." By Thomas, age 19

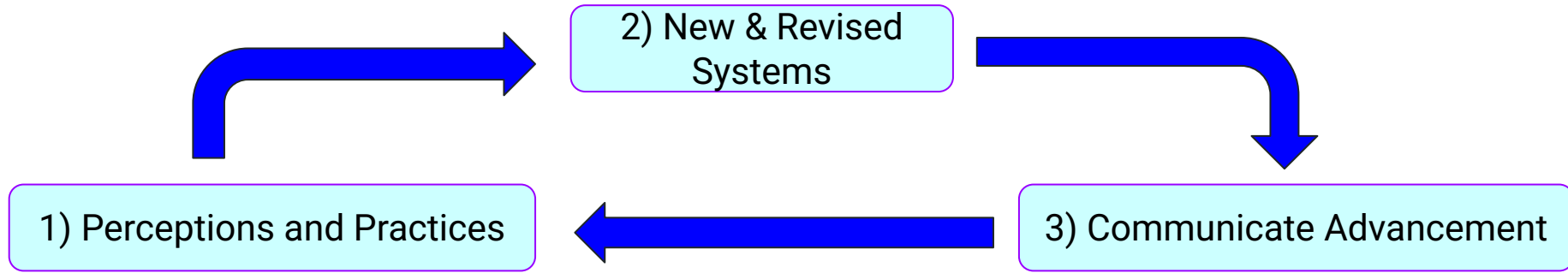


Fig. 24. "No one come in when I am in the bathroom!" By Sydney, age 7



Fig. 23. "This is me enjoying my privacy. This is the only time during the day, were I am truly alone and nothing bothers me. No man no children no dogs." By Cindy, age 54

Human-Centered AI and ML



“...that aims to make systems usable and useful by **focusing on the users, their needs and requirements**, ... counteracts possible adverse effects of use...” - ISO 9241-210:2019(E)



Perceptions and Expectations

- What do data subjects understand?
- How is a data subject's willingness to share impacted?
- How do data subjects perceive the risks?



**What they
“want”**



**What they
“need”**



**Build towards
those attributes**

Kacsmar, Duddu, Tilbury, Ur, and Kerschbaum. Comprehension from Chaos: Towards Informed Consent for Private Computation. 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS).

The Use of Health Data



P20986: “It depends. I think it can be beneficial **under certain circumstances**, but I would be hesitant having any healthcare data shared outside my practitioners. However, I recognize how it can improve goods/services, but there **has to be a lot of protection** in place **anytime data is shared**”

P94865: “**Repugnant**, especially in light of for profit health systems attempting to maximize profitability from patient interactions”



B. Kacsmar, K. Tilbury, M. Mazmudar, and F. Kerschbaum. "Caring about Sharing: User Perceptions of Multiparty Data Sharing." In 31st USENIX Security. 2022.

Bounded Impact of Private Computation

Intentions
Matter

Divulge the
Details

Regulate the
Restrictions

Consent Above
All

“At the end of the day,
they’re still like learning specific things about me” (P7)

Kacsmar, Duddu, Tilbury, Ur, and Kerschbaum. Comprehension from Chaos: Towards Informed Consent for Private Computation. 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS).

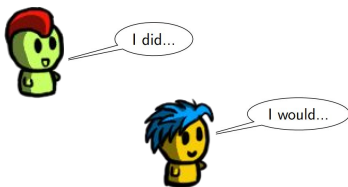
Demystifying the Life of Data in the PUPS Lab



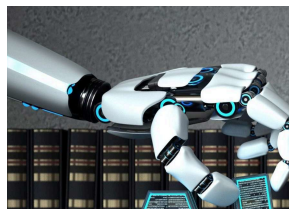
AI in Hate Speech
Moderation



Mathematically
Bound Privacy



Perceptions of ML
Design Trade-offs



Data Flows in Bionic
Limbs

Takeaways

- **Data privacy cannot alone solve the issues** of human privacy in our current society
- The meaning of data changed quickly, but now so to have **the consequences of data**
- We can make better systems, but **even good systems can cause harm** without consideration for the full picture of who they impact