It’s Not About Data Privacy

AI in Dentistry Symposium 2025,
College of Dental Surgeons of Alberta (CDSA)

Dr. Bailey Kacsmar

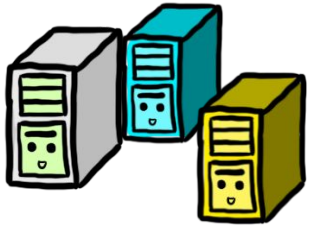
NSERC CRSNG

PUPS Research Lab

UNIVERSITY OF ALBERTA

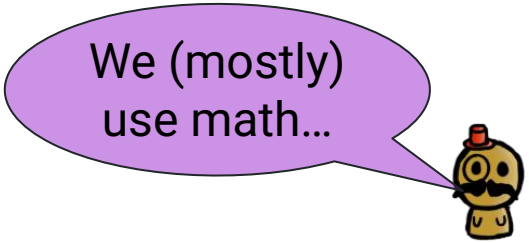
amii

Privacy, Data, and AI?



Someone wants to complete a task

But there are privacy implications and risk from that task



Researchers develop technical solutions

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

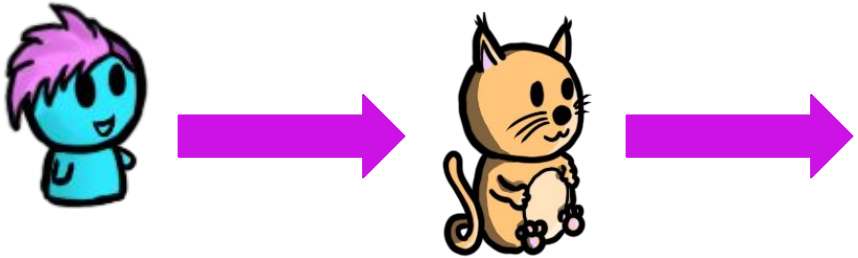
Data Generation and Collection

A Snapshot from a Day in the Life of Alice 

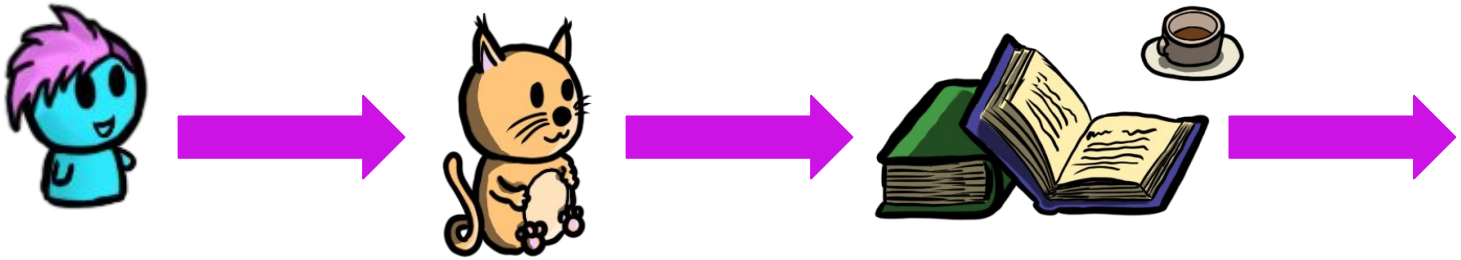
Alice, Bob, Carol, Eve…sourced from “Images by Simon Oya: https://simonoya.com/drawings/”

Alice’s Day

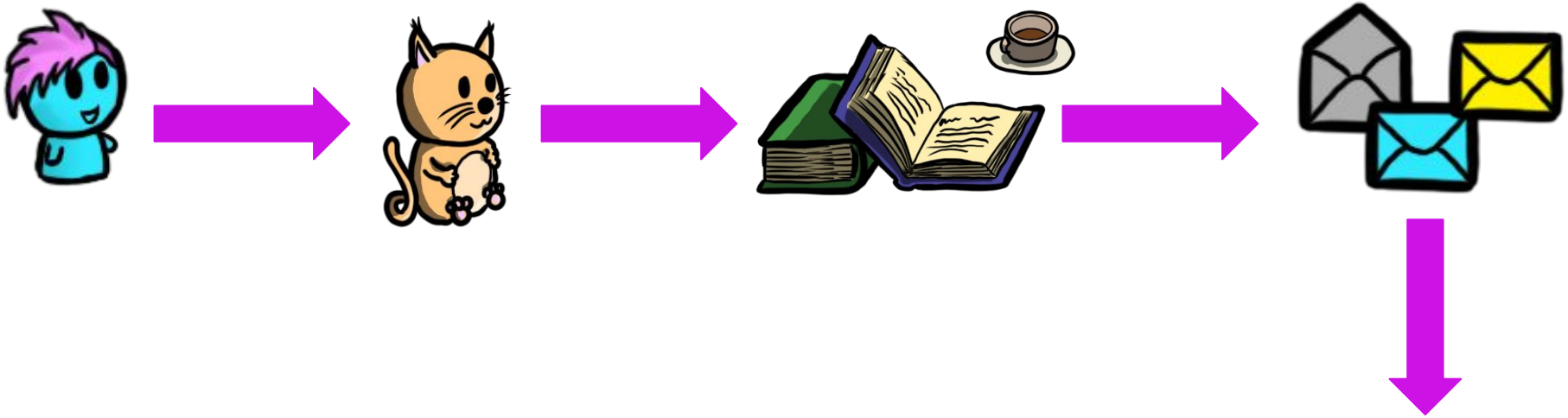
Alice’s Day



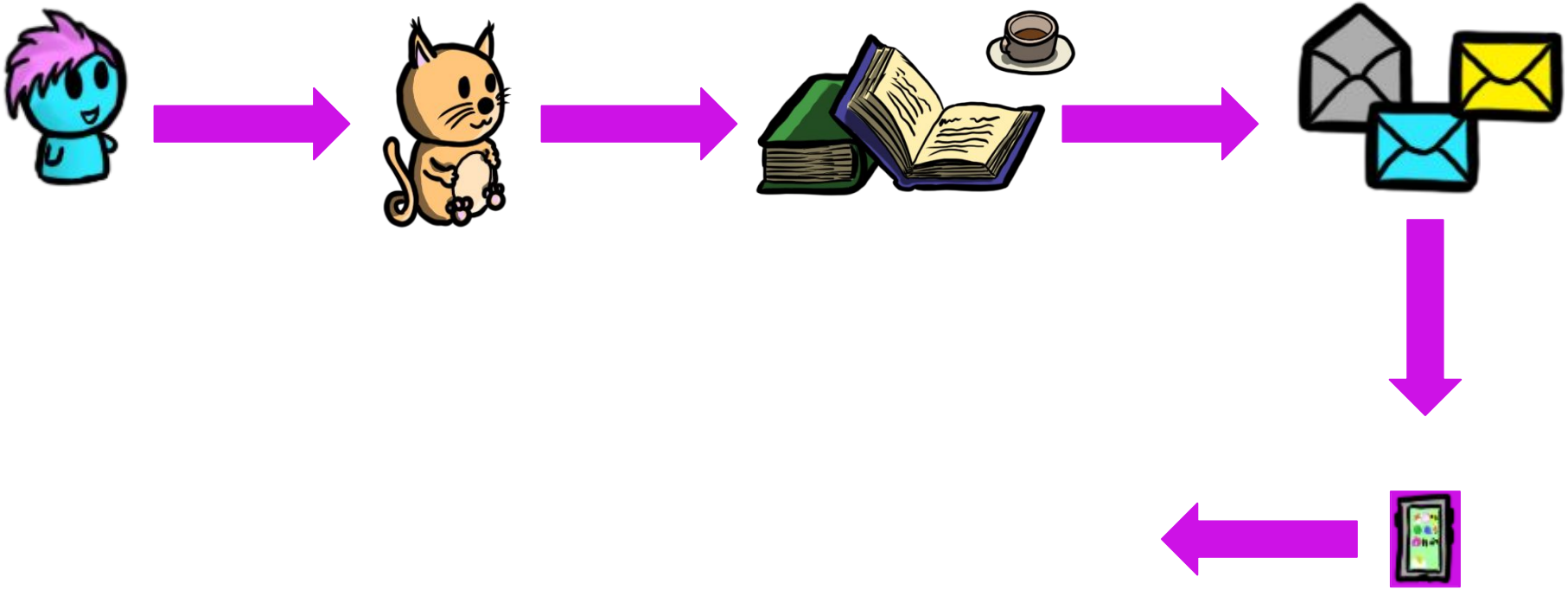
Alice’s Day



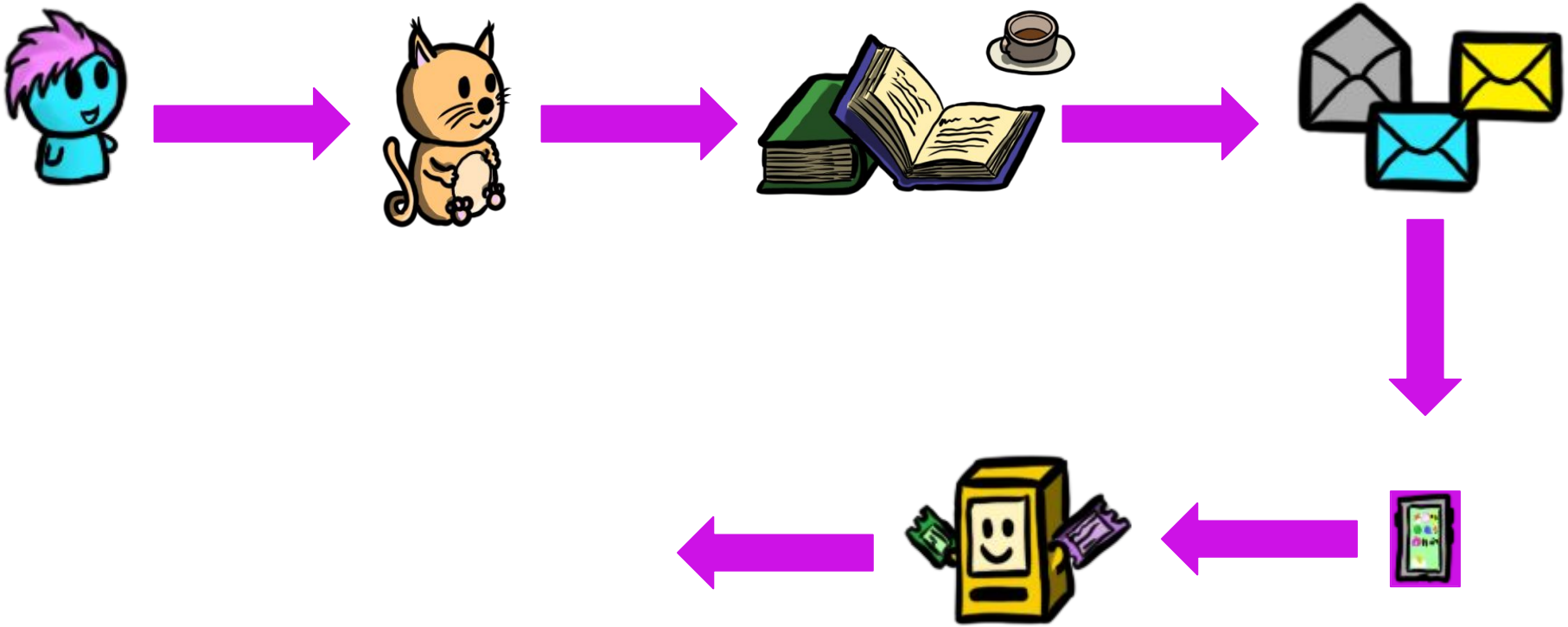
Alice's Day



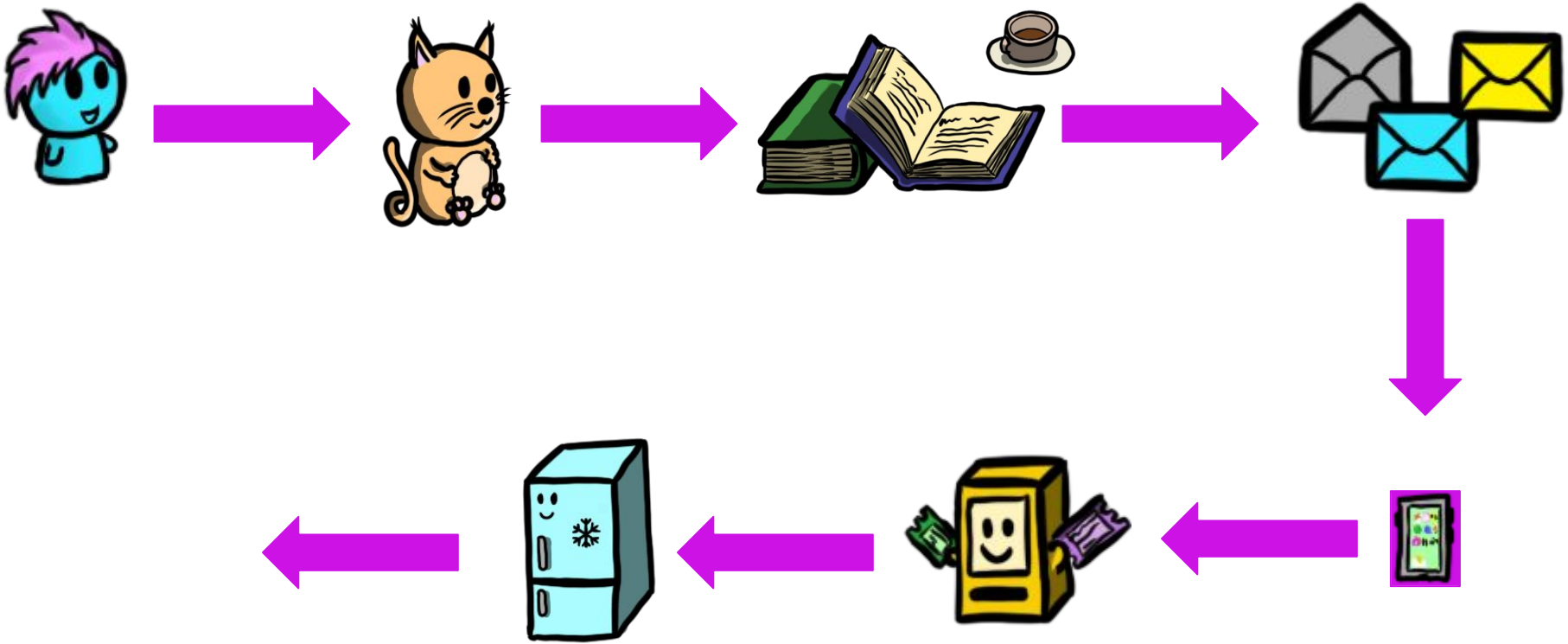
Alice's Day



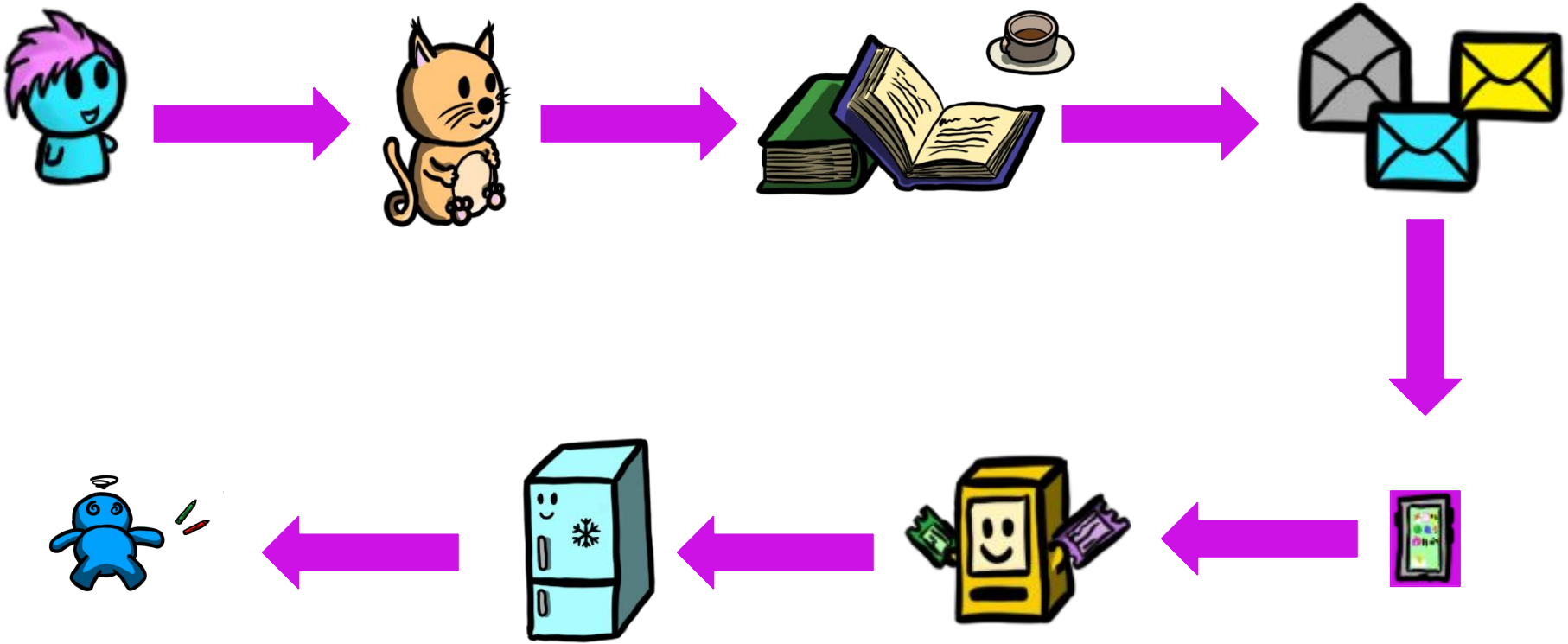
Alice's Day



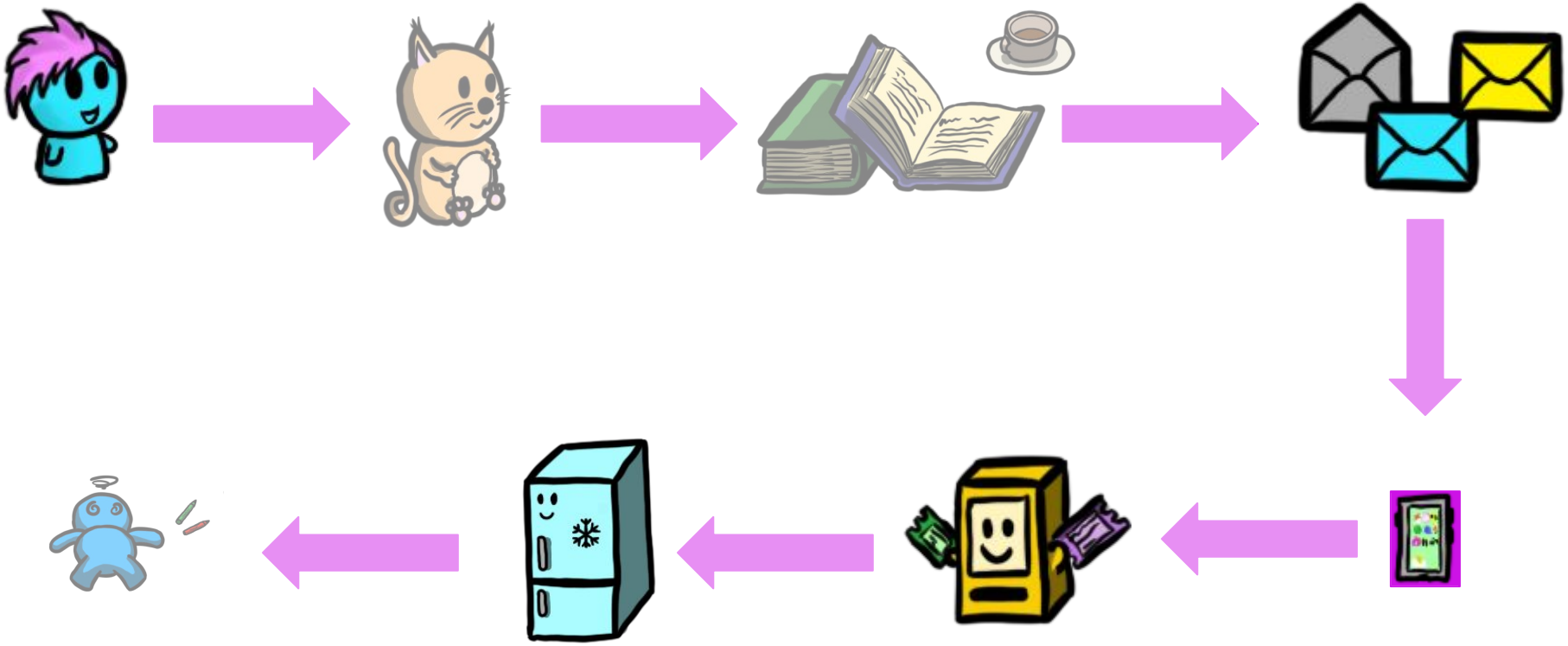
Alice's Day



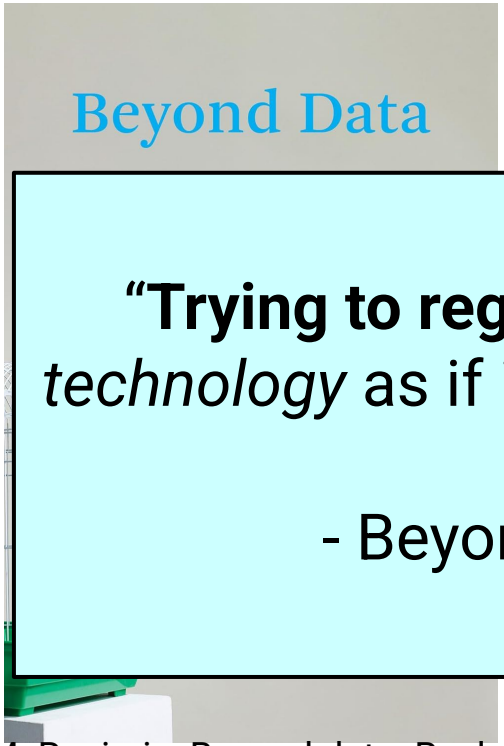
Alice's Day



Data Happened



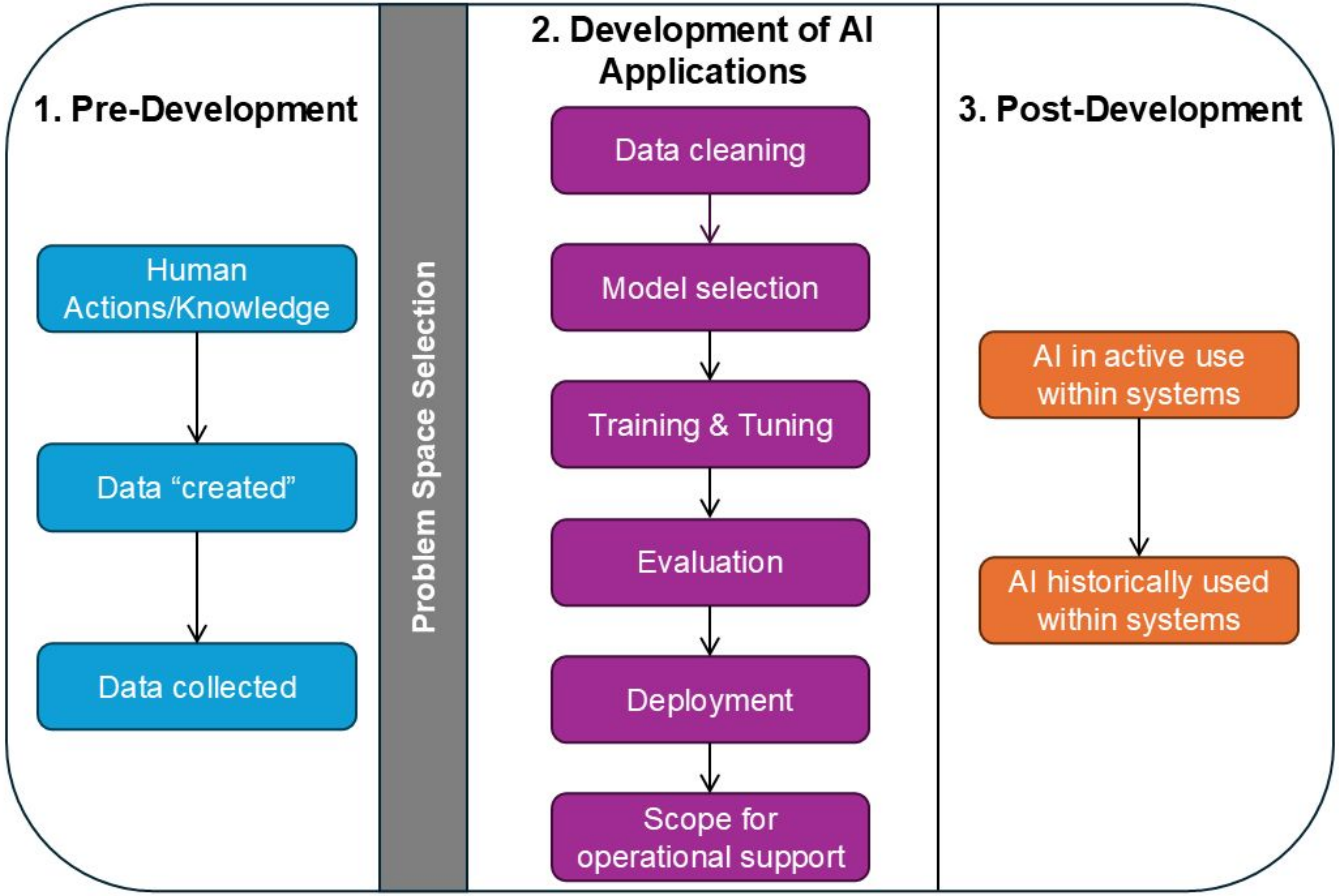
From Data Back to People

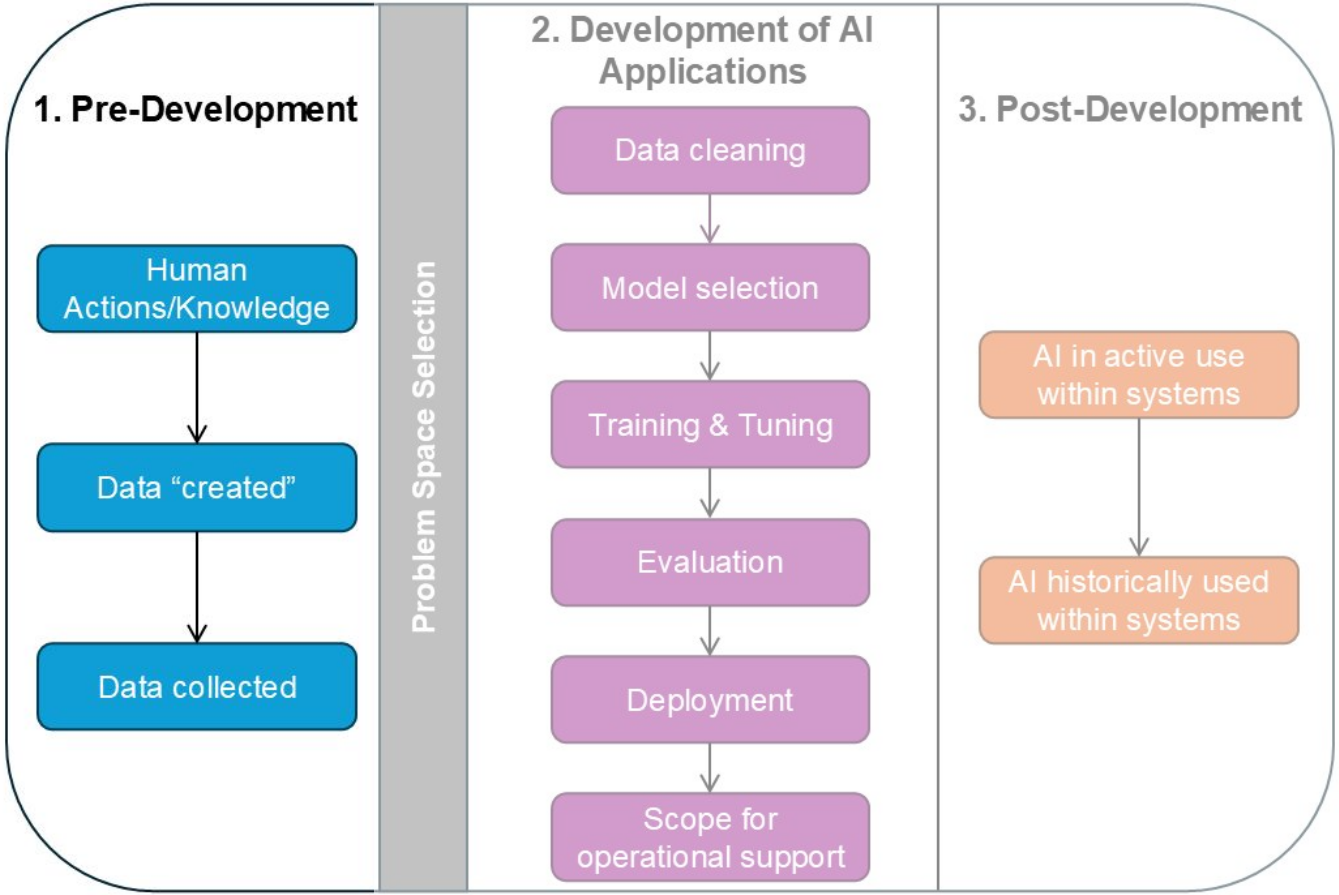


"...when everything is digital, there

"**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

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

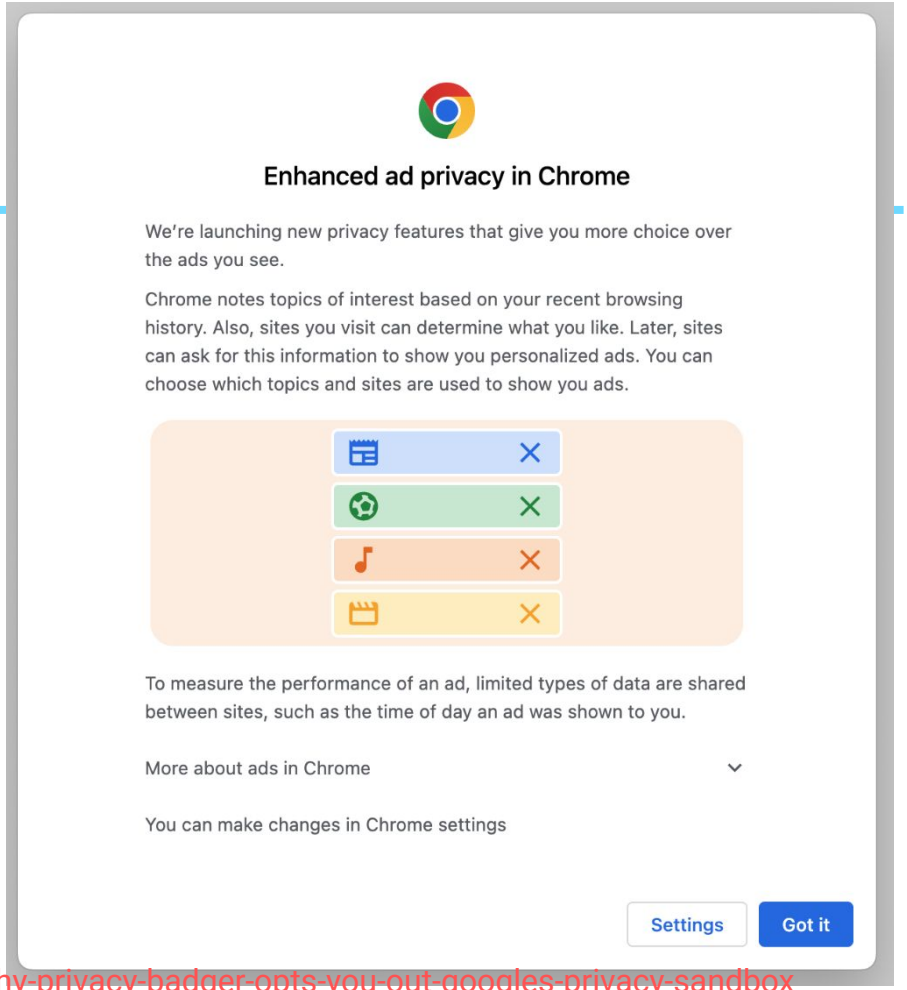

1. Pre-Development
Human Actions/Knowledge
Data "created"
Data collected
Problem Space Selection
2. Development of AI Applications
Data cleaning
Model selection
Training & Tuning
Evaluation
Deployment
Scope for operational support
3. Post-Development
AI in active use within systems
AI historically used within systems


1. Pre-Development
Human Actions/Knowledge
Data “created”
Data collected
Problem Space Selection
2. Development of AI Applications
Data cleaning
Model selection
Training & Tuning
Evaluation
Deployment
Scope for operational support
3. Post-Development
AI in active use within systems
AI historically used within systems

Another look at the snapshot view of Alice’s life…

Misleading Expectation

Ticket Seller Accessed via Browser



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,

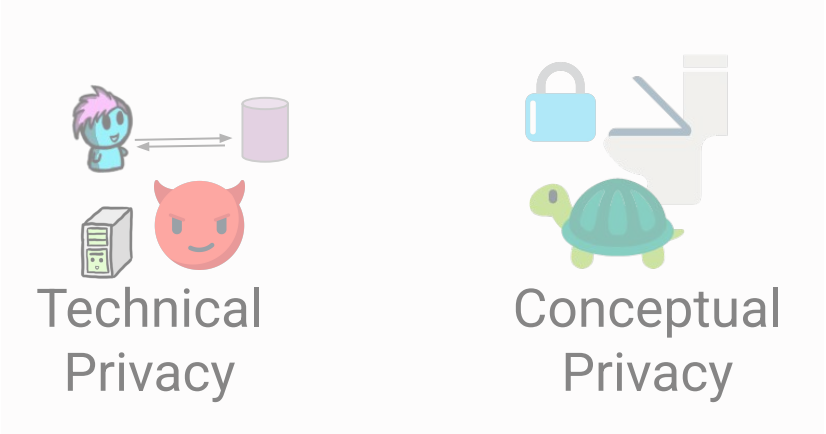
- Side channels - accessing without permissions via other means
- Covert channels - via a set of colluding apps

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 Opaque



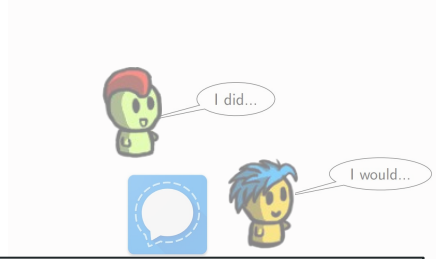
Technical Privacy

Conceptual Privacy

Bill C-27

PIPEDA

GDPR.EU



The Canadian tech company that changed its mind about using your tax return to sell stuff | CBC Radio

CBC Radio · Posted: Feb 23, 2020 4:00 AM EST | Last Updated: February 23, 2020

“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 diff**
  laws than other organizations

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

Data Generation and Collection

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

Now for sale: Data on your mental health

Drew Harwell

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

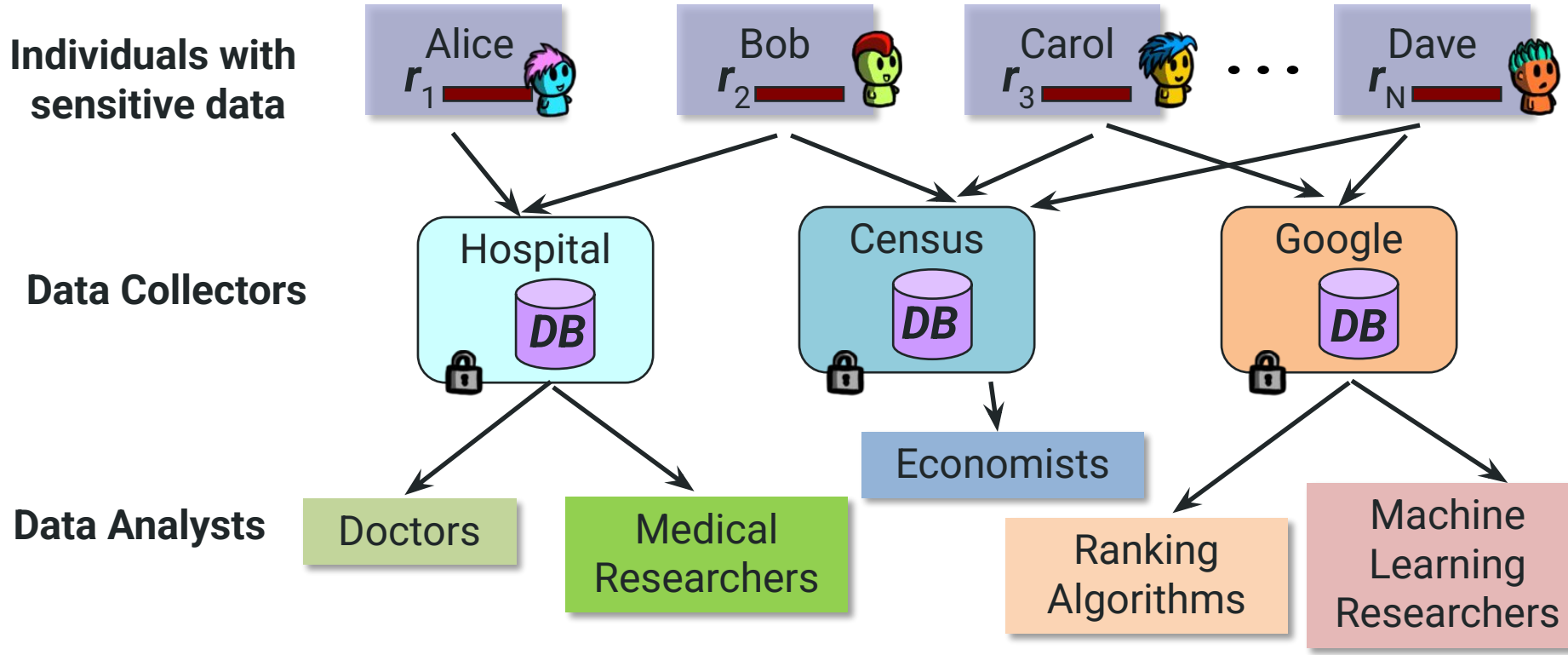
Home Depot didn't get customer consent before sharing data with Facebook's owner, privacy watchdog finds | CBC News

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

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

Where we are: Statistical Databases

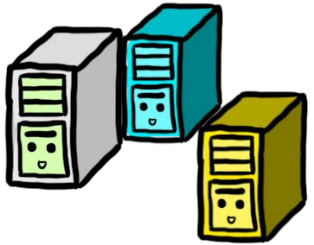


Consequences of the **Human-Data Disconnect** and AI

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 <u>not</u> be used in 'high-risk domains'**."

Misunderstandings - None Shall Pass

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

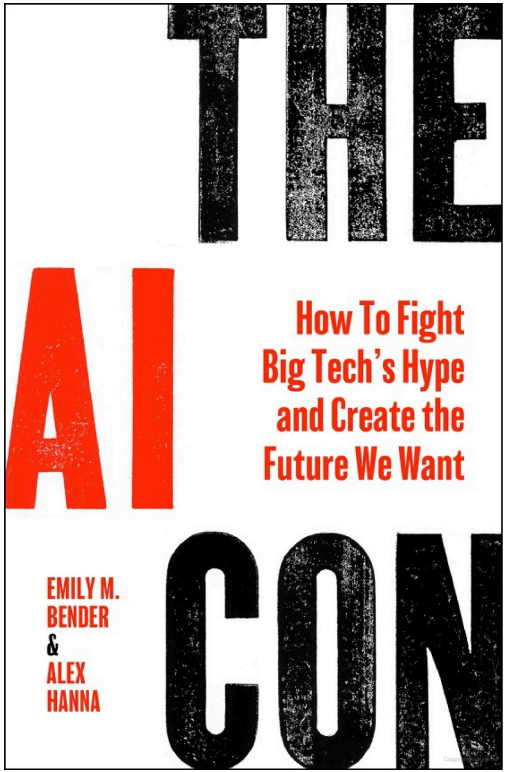
AI-generated writing is almost impossible to detect and tensions erupting at a Texas university expose the difficulties facing educators

Updated May 18, 2023

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

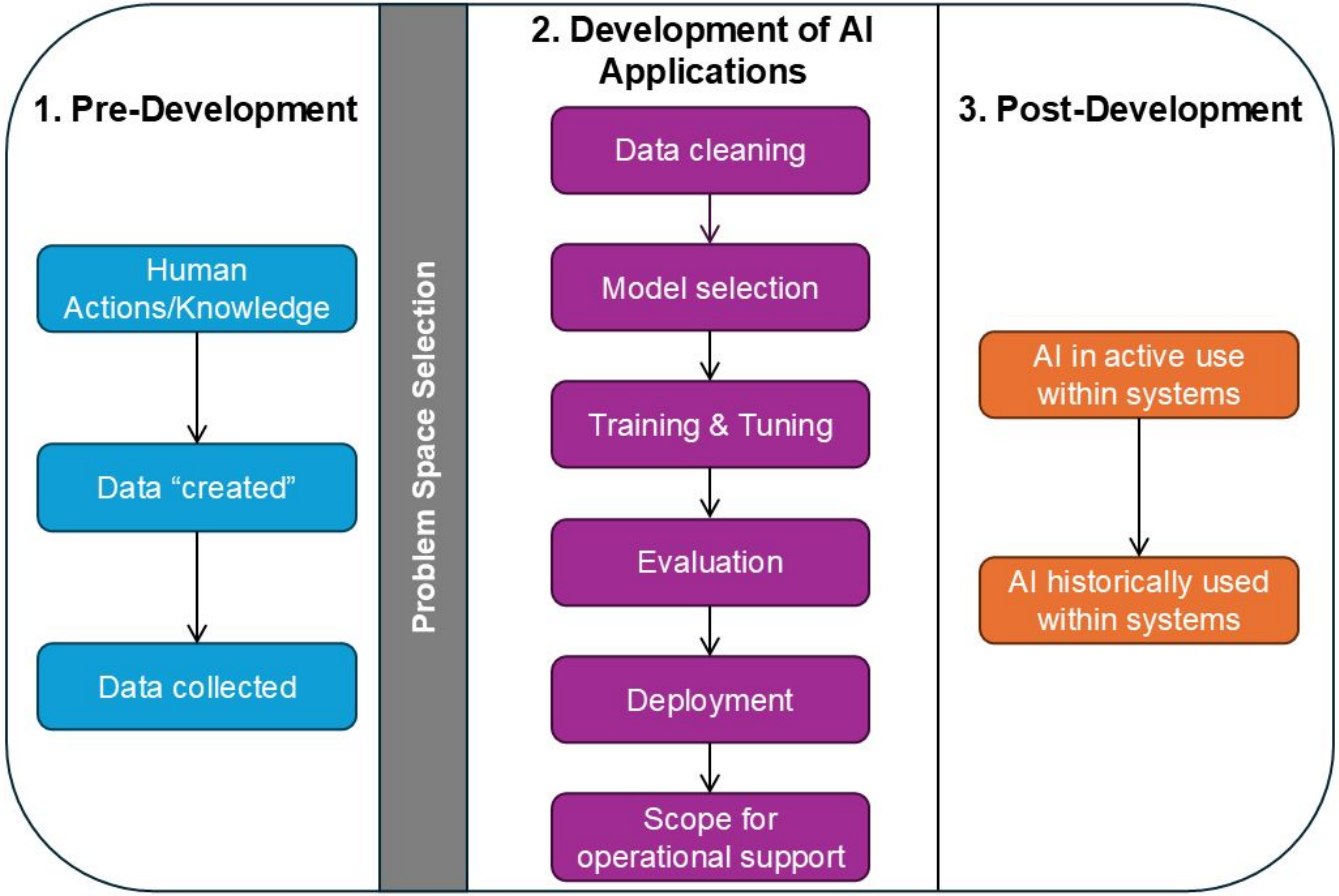
Washington Post, 2023:
www.washingtonpost.com/technology/2023/05/18/texas-professor-threatened-fail-class-chatgpt-cheating/

Oh dear… On to Overcoming the AI Hype

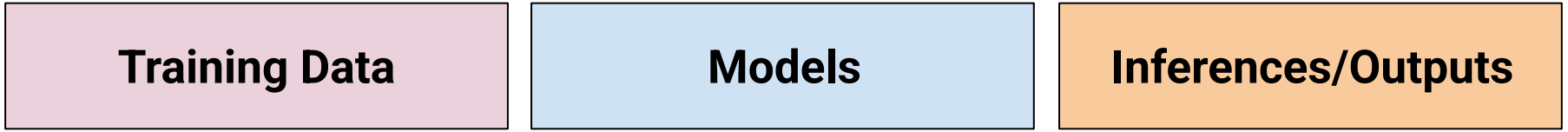


"When people can **spot AI hype**, they **make better decisions** about how and when to use automation, and they are in **a better position to advocate for policies** that constraint the use of automation by others" - The AI Con

Towards being better?


1. Pre-Development
Human Actions/Knowledge
Data "created"
Data collected
Problem Space Selection
2. Development of AI Applications
Data cleaning
Model selection
Training & Tuning
Evaluation
Deployment
Scope for operational support
3. Post-Development
AI in active use within systems
AI historically used within systems

Private Computation and Machine Learning?



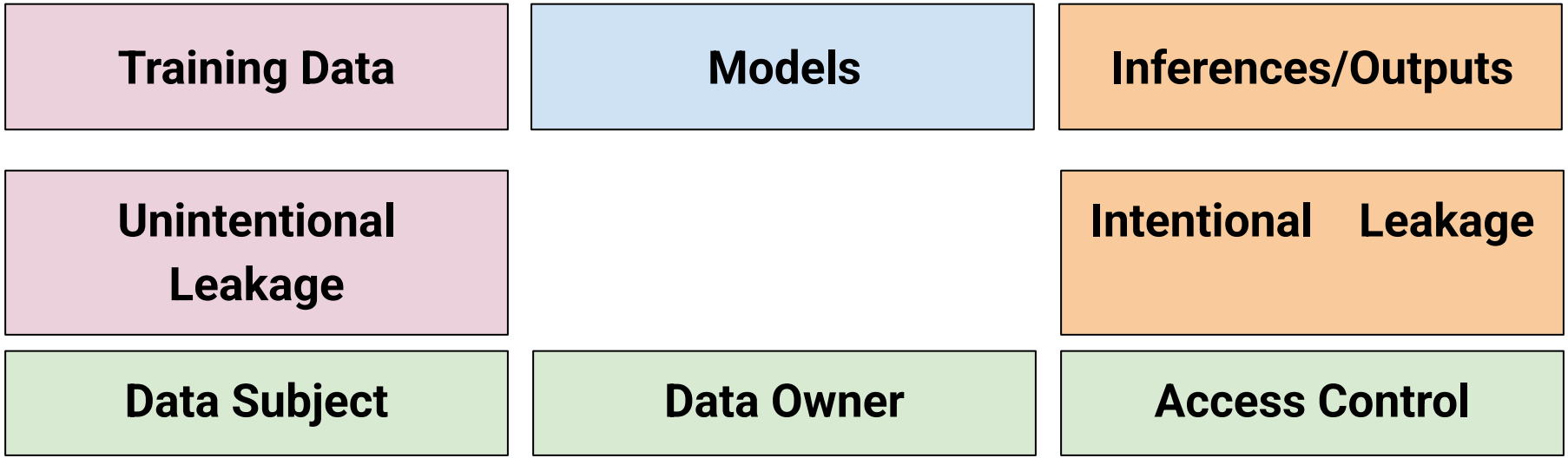
Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Private Computation and Machine Learning?

Training Data	Models	Inferences/Outputs
Unintentional Leakage		Intentional Leakage

Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Private Computation and Machine Learning?



Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

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

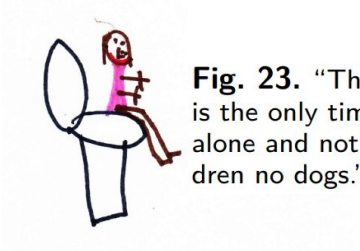
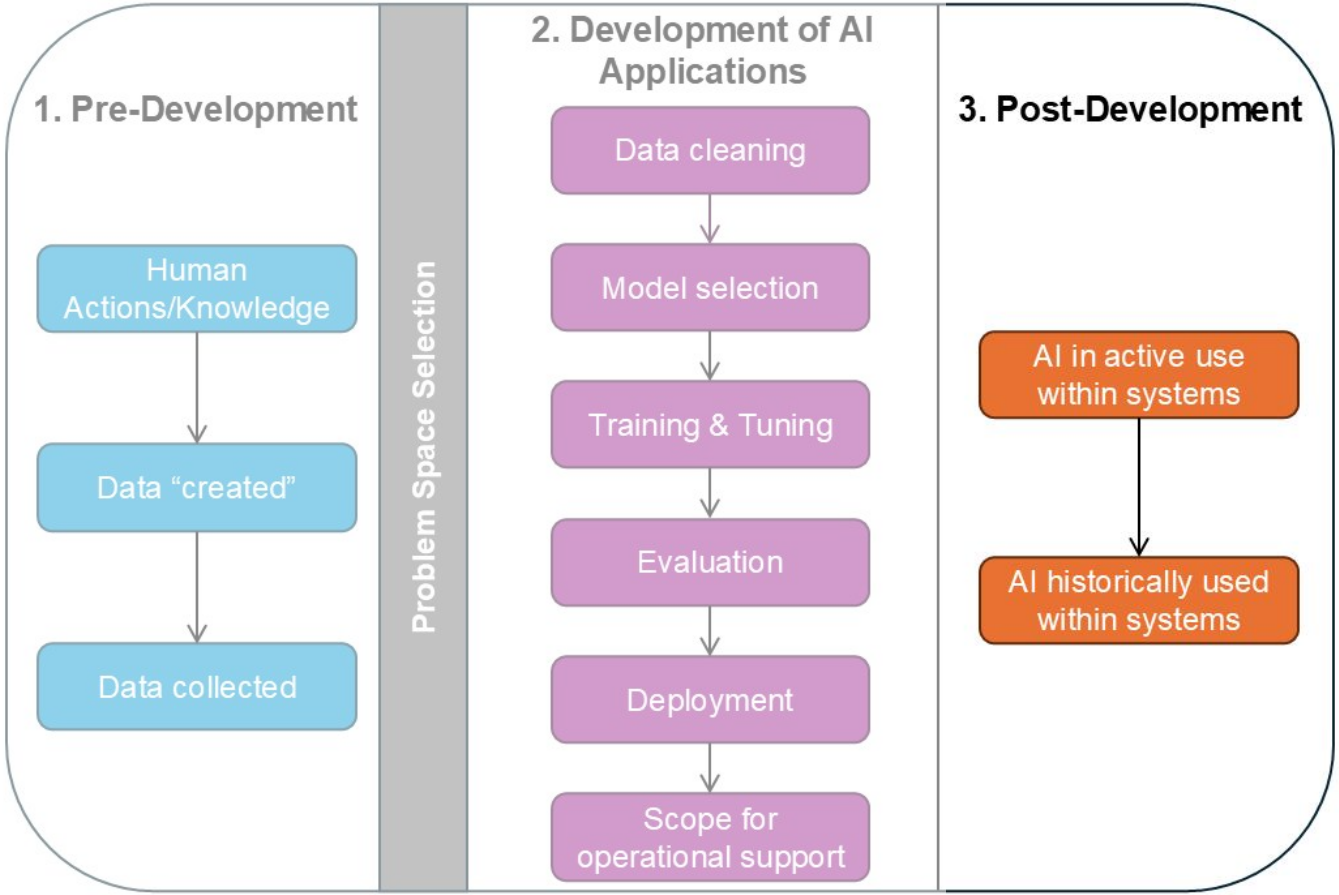


Fig. 23. "This is me enjoying my privacy. This is the only time during the day, were I am truly alone and nothing bothers me. No man no children no dogs." By Cindy, age 54

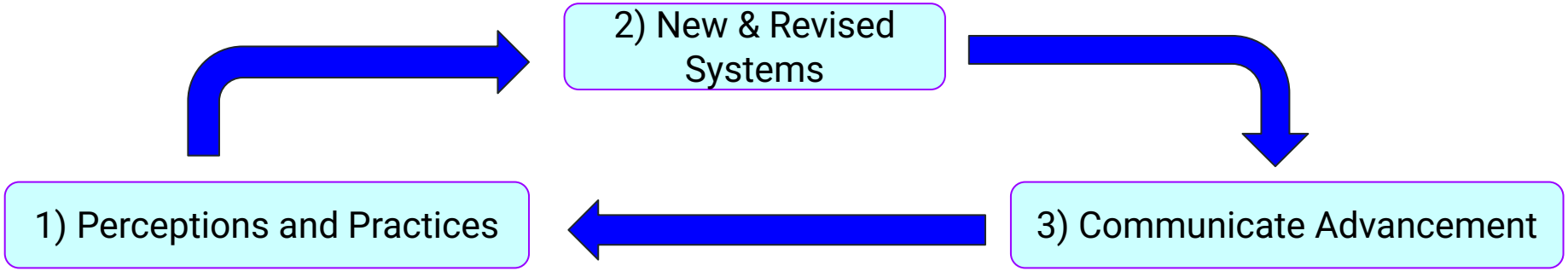
Fig. 24. "No one come in when I am in the bathroom!" By Sydney, age 7

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.


1. Pre-Development
Human Actions/Knowledge
Data “created”
Data collected
Problem Space Selection
2. Development of AI Applications
Data cleaning
Model selection
Training & Tuning
Evaluation
Deployment
Scope for operational support
3. Post-Development
AI in active use within systems
AI historically used within systems

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).

Takeaways

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