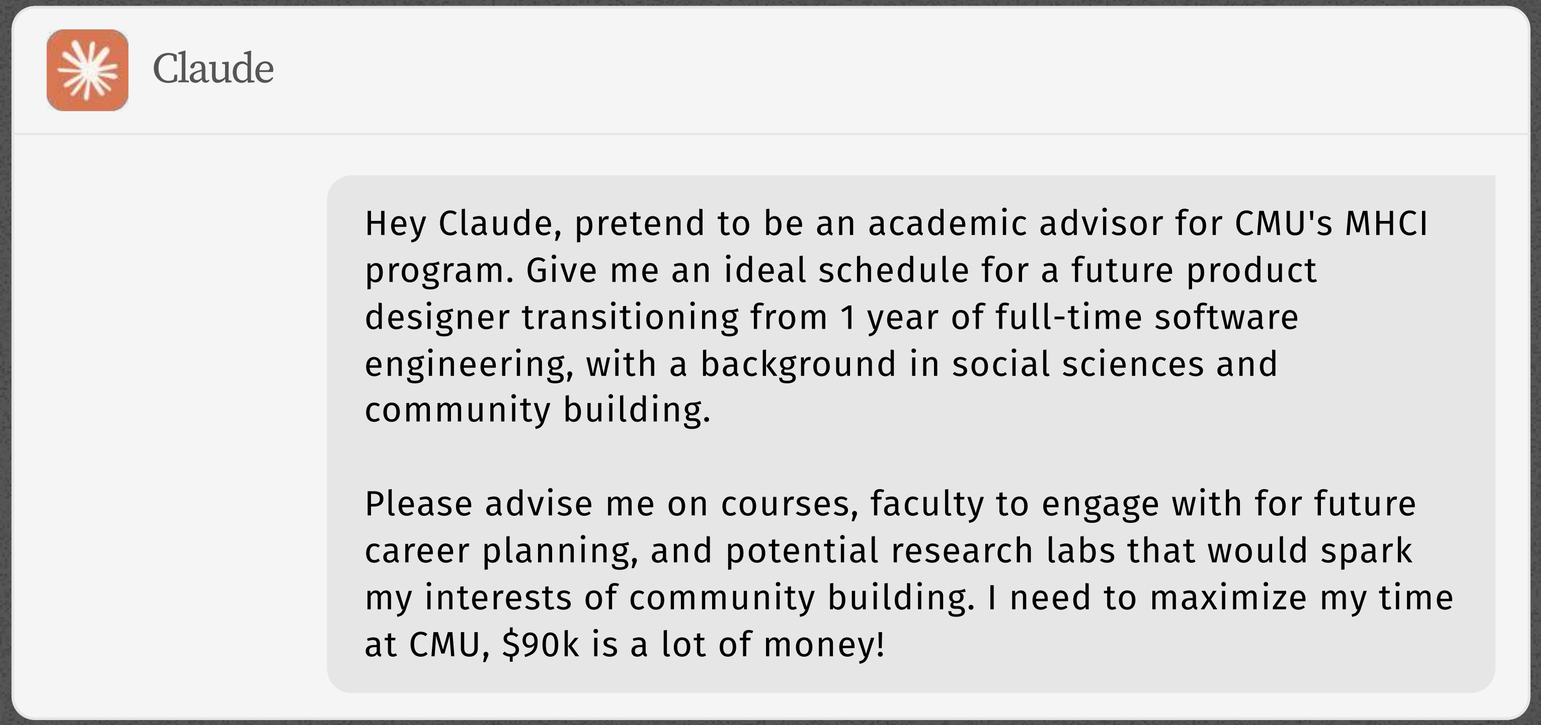
OpenMined

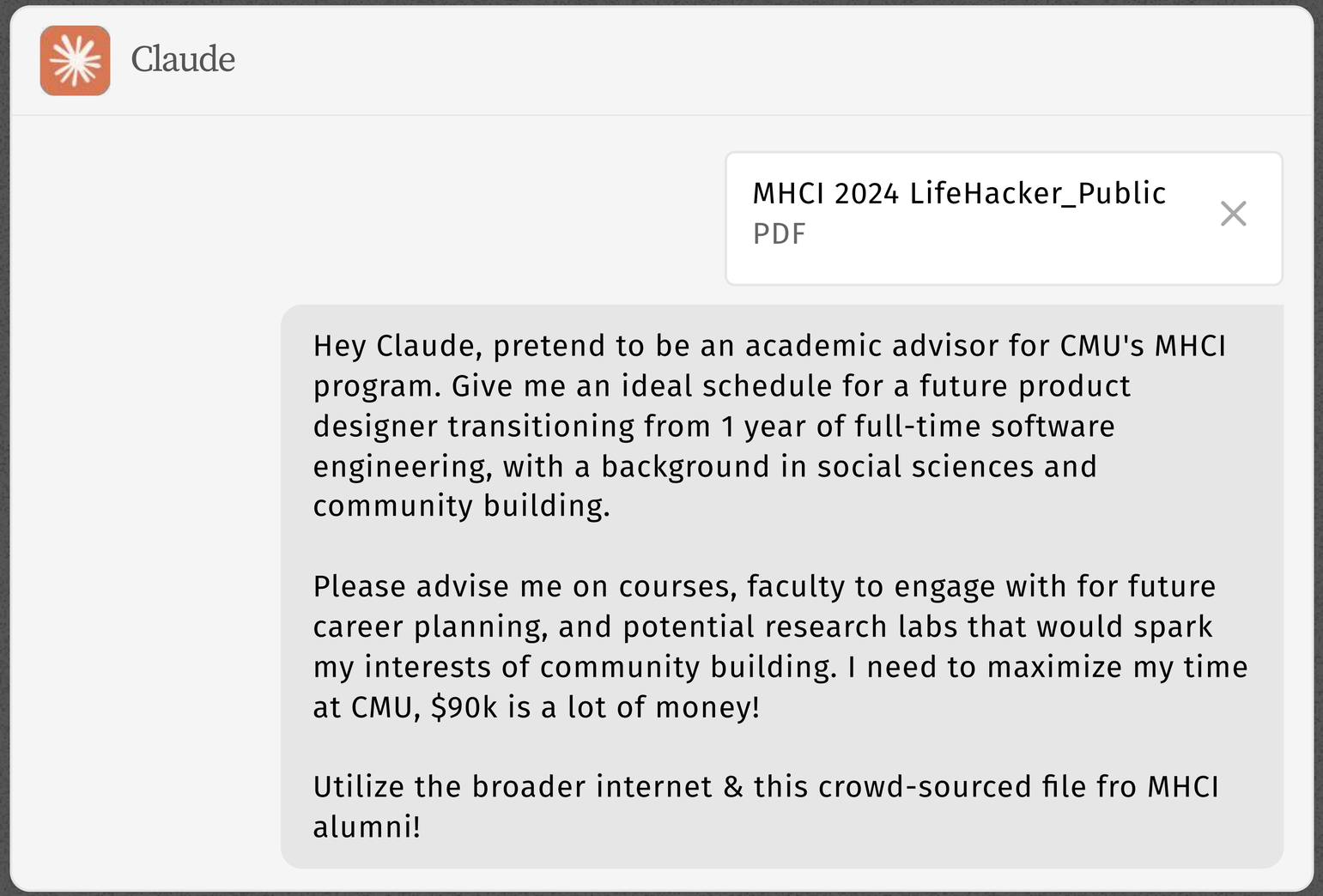
Mid-Semester Research Readout

Tomi Olusina, Uma Dhamija, Mridula Unni, Yukti Poddar

Imagine Planning Your First Semester at CMU With AI



What Changes When Knowledge Has a Verified Source?

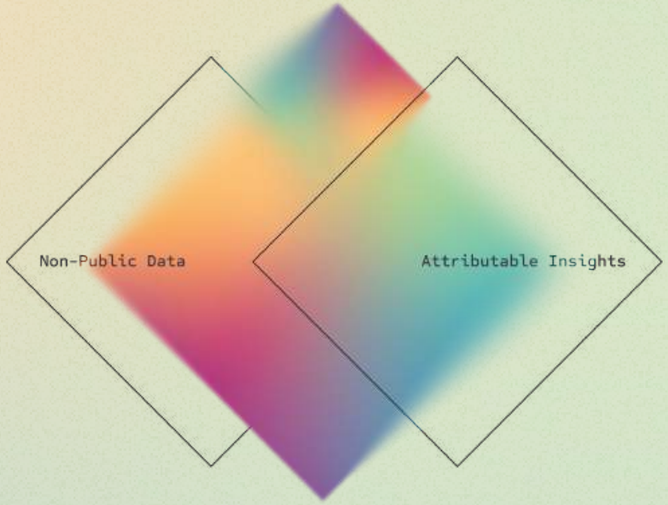


Agenda

The Mission

What is OpenMined?

A non-profit open-source community developing the decentralized networks of the future.



Privacy First

Federated learning & differential privacy ensure raw data never leaves its source.

Decentralized Network

SyftBox nodes connect data owners, researchers, and consumers in a permissioned mesh

Consumer Empowerment

Individuals contribute data to AI training while retaining ownership and earning value

99.99% of the world's data remains locked away

A hospital has patient data.
A bank has fraud patterns.
A research lab has rare disease records.

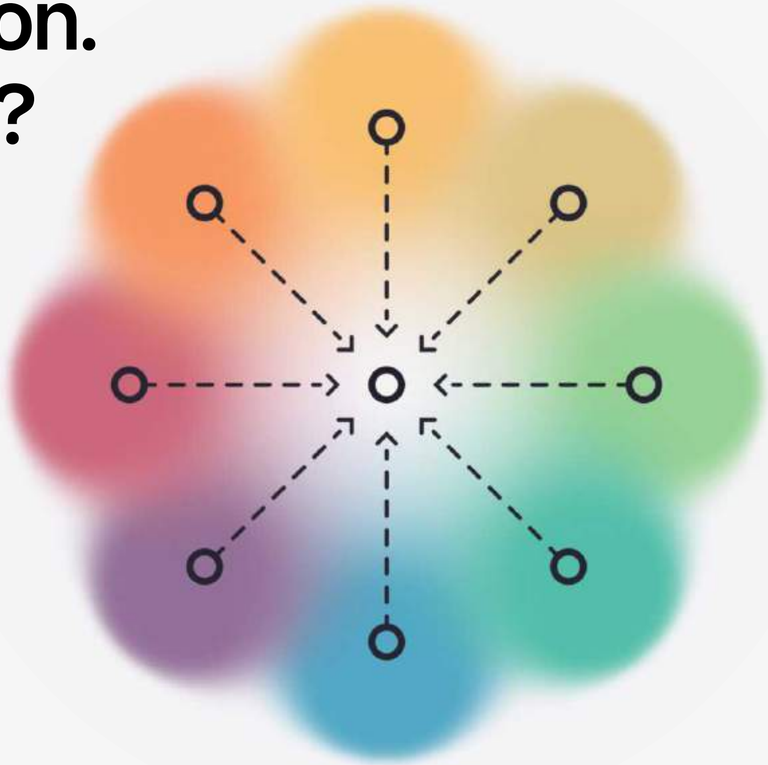
Each holds something valuable.
None can share it because of regulatory and proprietary restrictions.

The Solution

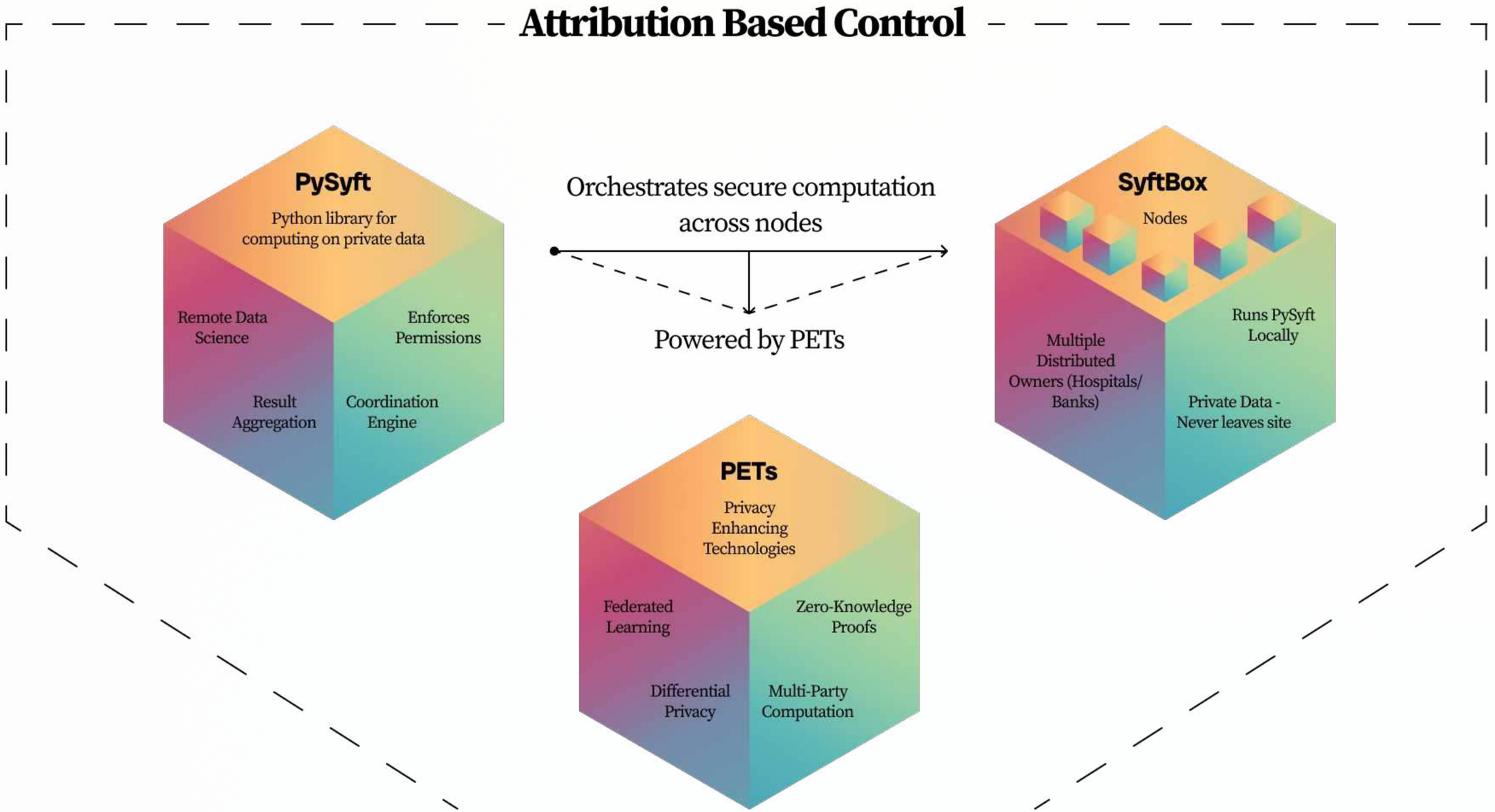
Attribution Based Control (ABC)

It's technically powerful.
But technology alone does not drive adoption.
The question is: Where does it matter most?

- Sensitive data is siloed
- Consumer trust in AI is fragile
- AI and privacy literacy are limited



Tech Stack


Attribution Based Control
PySyft
Python library for computing on private data
Remote Data Science
Enforces Permissions
Result Aggregation
Coordination Engine
Orchestrates secure computation across nodes
Powered by PETs
SyftBox
Nodes
Runs PySyft Locally
Multiple Distributed Owners (Hospitals/ Banks)
Private Data - Never leaves site
PETs
Privacy Enhancing Technologies
Federated Learning
Zero-Knowledge Proofs
Differential Privacy
Multi-Party Computation

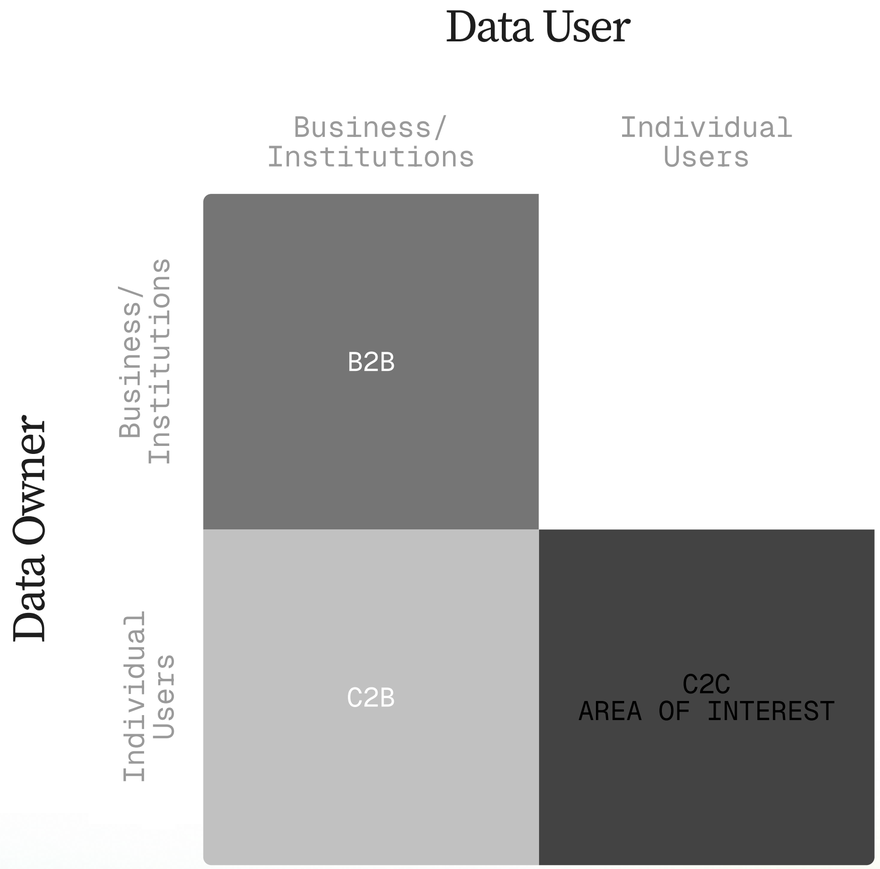
Project Brief

Who are the users?

Individual user groups in a first-mover markets that can derive value from a decentralized network.

How do we persuade them?

Create an incentive-driven product strategy that encourages value-aligned usage.



We conducted

1
PESTLE Analysis
Territory Map
Experiential Learning

2
Process Maps
Ecosystem Maps

3
SME Interviews

8
Use Case Studies

50
Primary Research Articles

73
Secondary Research Papers (ongoing)

52
Survey Responses (ongoing)

SME Interviews

OpenMined Contributor & Initiatives Lead

Osam Kyemenu-Sarsah

"Move slowly, incremental steps. Trust requires time to develop!"

- High bar for trust-earning, low bar for trust-breaking
- Contact → Education → Synthetic Data → Sensitive Data
- Christchurch call - log auditing at large tech company
- Medical Federated Learning Program (MFLP) - global medical imaging network

OpenMined Researcher & Education Lead

Irina - Madalina Bejan

"When we have a good enough story for why they should participate, the cost and risks do fade away."

- Data sharing incentives should emphasis the value of the output or benefit rather than the transparency of the process
- Reddit4Researchers - enabling researchers to access data for their studies while protecting Reddit's data from being used for commercial purposes
- Publishing/Journalism - helping media companies in Indonesia maintain control over their content against AI bots

Industry Landscape

Beekeeper AI

A collaborative platform that allows AI models to train and validate algorithms on sensitive healthcare data without exposing underlying records.

Kakao Healthcare

A digital health company that leverages AI to create secure data-sharing platforms for cross-hospital research.

ProRata.AI

Attribution technology developers that ensure content creators and publishers are credited and fairly compensated when their work is used to generate answers in AI search engines.

P
Political

- Fragmented global regulation (GDPR vs CCPA vs PIPL)
- Lack of federal US privacy law
- AI governance concerns

E
Economic

- Data sales drive significant revenue in B2B spaces
- Brokers pay cents; Users pay dollars
- Exposed private data decreases competitive edge between companies

S
Social

- Lacking of agency with corporations
- Trust driven by social proof
- Younger demographics show higher concern but also higher tech literacy

T
Technological

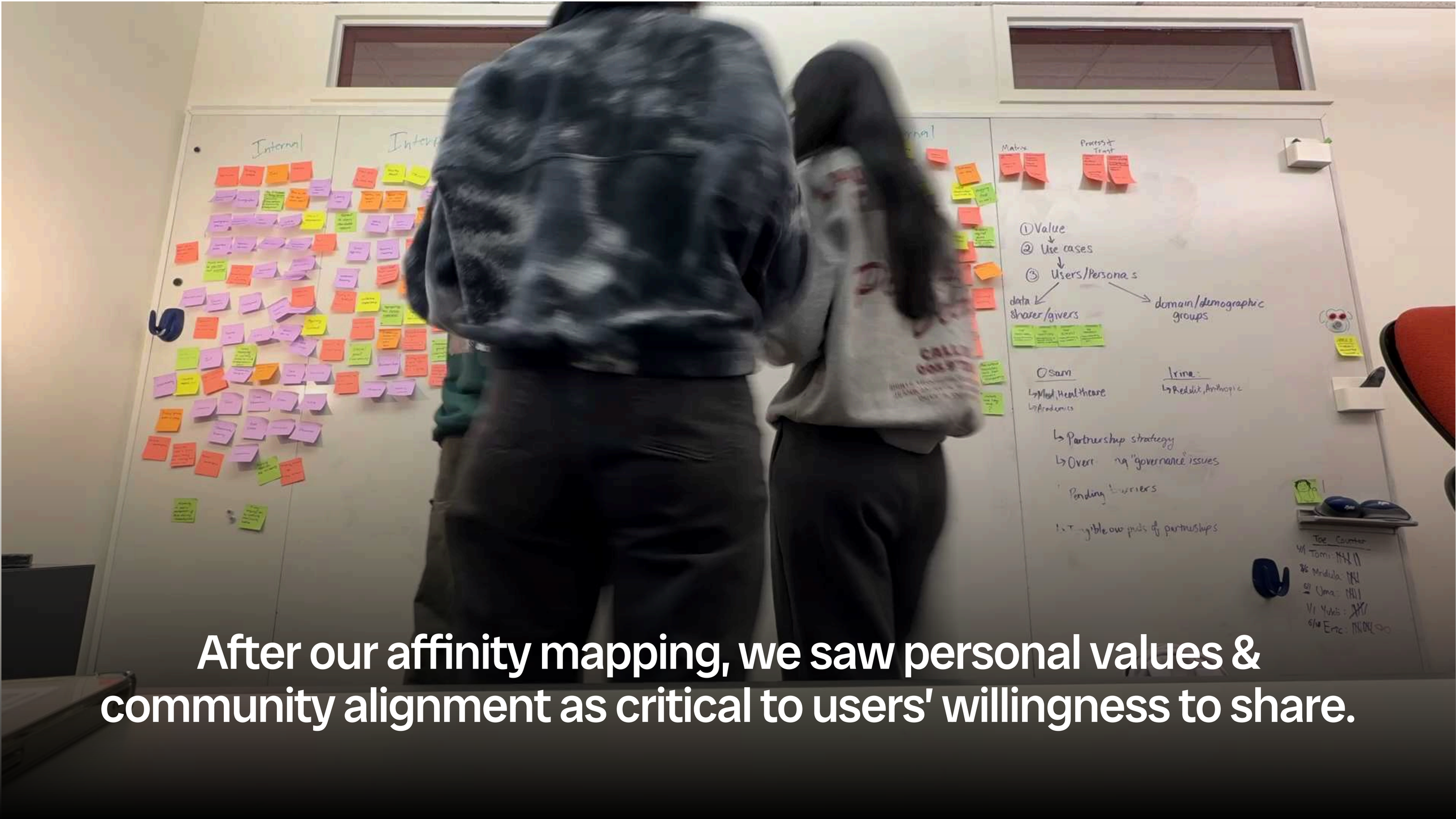
- PETs (FL, DP, MPC) are mature but complex
- Setup tax for SyftBox nodes is high
- Re-identification risk grows with AI — making true federated isolation more valuable.

L
Legal

- HIPAA gaps exploited by big tech hospital tracking
- Outdated de-identification standards
- GDPR requires specific consent and deletion rights

E
Environmental

- AI data centers have significant carbon footprints
- Decentralized computing can reduce data center energy demands!



After our affinity mapping, we saw personal values & community alignment as critical to users' willingness to share.

Current Research

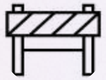
Key Insights

Negative Feedback Loop

- Unwilling ignorance in data literacy creates feelings of powerlessness and uncertainty
- Loss of agency and control drives mistrust and resignation, which results in apathy and continued ignorance
- Socio-technical trust-building and transparency are needed

Data Value Gap

- Companies generate significant revenue from user data, while users bear all the risk with zero financial gain
- Individuals value their own data more than they value other people's personal data
- Privacy paradox is the inconsistency between stated concerns and actual behavior



Trust as the Biggest Hurdle

- Trust is earned & incremental. It cannot be assumed
- Contextual transparency and data integrity inform the faith of users in a system
- Reciprocity supports the give-and-take information economy, unsustainable by trust alone

Aggregating the Human Experience

- Privacy is a negotiated social value, not a commodity to trade
- Social proof - leveraging co-creation of collective knowledge is critical
- Altruism as a driver - Public good is an accepted trade-off to personal risk

Next Steps

1

Complete Exploratory Research

- Analyze survey results
- Mapping analogous domains
- Define viable user groups and synthesize personas

2

Framing the Solution

- Narrow into a single consumer base to build around
- Recruit & interview potential users
- Identify highest-leverage intervention points in the adoption funnel

3

Generative Design

- Co-design workshop with OpenMined team and recruited participants
- Brainstorm and ideate solutions
- Begin pretotyping


OpenMined
we yearn for the mines
99.99% of the world's siloed data

QUESTIONS?

OpenMined

Mid-Semester Research Readout

Tomi Olusina, Uma Dhamija, Mridula Unni, Yukti Poddar