

UBICOMP 2026
MENTAL HEALTH WORKSHOP

Beyond Prediction

Building Reliable and Responsible AI
for Digital Mental Health

Uichin Lee
KAIST

11 October 2026
Shanghai, China



Computing trends

1970s

Functional

Mainframe



1990s

Usable

Personal computers



2010s–

Ubiquitous

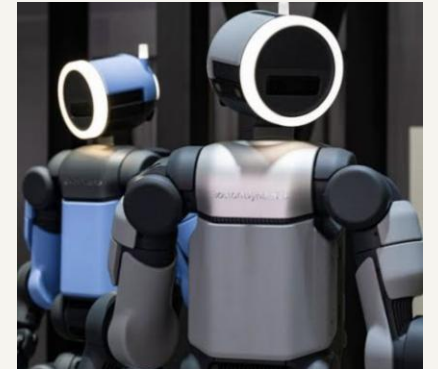
Mobile, wearable, IoT



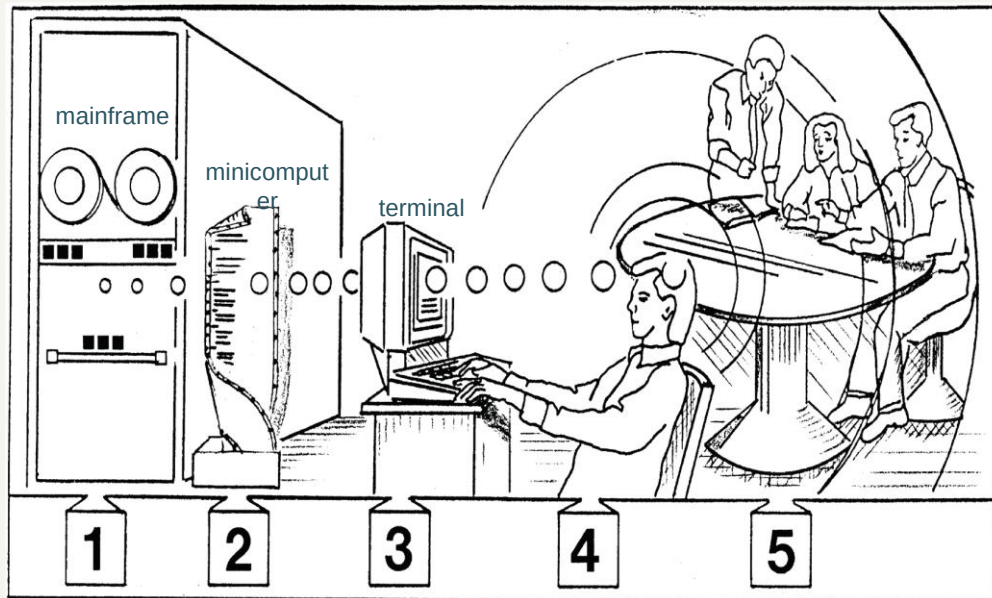
2022–

Symbiotic

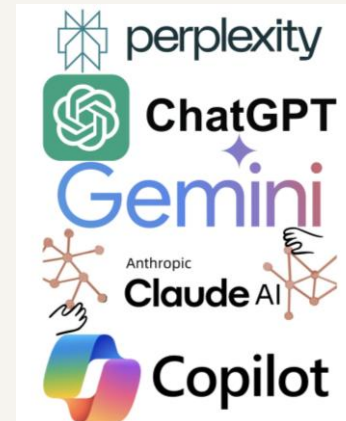
GenAI and agents



From tools to partners



Interface as Hardware Software Terminal Dialogue Social & Ubiquitous



Large Language
Models (LLMs)

Tools

People specify a task.
Systems execute it.



Assistants

Systems **infer context**
and **suggest actions**.



Partners

People **negotiate goals**,
interpret evidence and
retain **meaningful control**.

(Man-Computer Symbiosis
J.C.R. Licklider 1960)

Human-centered AI for mental healthcare

Mobile phones, wearables and IoT extend observation into everyday life.

Observation

Behavior, physiology,
social interaction
and daily context.



Analyzing

Repeated observations
reveal patterns and
changes over time.



Support

Feedback, coaching
and just-in-time adaptive
Interventions (JITIs).

Human-centered AI for mental healthcare

Mobile phones, wearables and IoT extend observation into everyday life.

Observation

Behavior, physiology, social interaction and daily context.



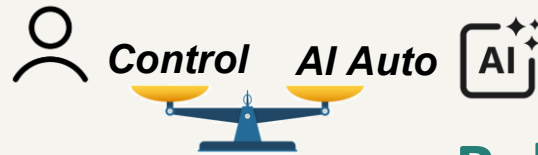
Analyzing

Repeated observations reveal patterns and changes over time.



Support

Feedback, coaching and just-in-time adaptive Interventions (JITIs).



Human control

People define goals and inspect evidence.
They can question an inference, change preferences and seek human expert support.

Reliable automation

AI reduces avoidable effort and supports demanding tasks.
It communicates its capabilities, limitations and uncertainty.

Digital mental health foundations: 2007–2020

Infrastructure, everyday symptom models and intervention studies

2007–2014

Data & platforms

MyExperience
StudentLife (2014)

State & future risk

StudentLife
Behavior & mood

Intervention evidence

MONARCA (2012)
Self-monitoring

2015–2017

Data & platforms

AWARE
mCerebrum / MD2K

State & future risk

cStress (2015)
CrossCheck (2016)

Intervention evidence

MONARCA I
Woebot RCT

2018–2019

Data & platforms

RADAR-base
mindLAMP

State & future risk

Depression
Dynamics (2018)

Intervention evidence

MONARCA II
Adaptive trials

2020

Data & platforms

CARP / CAMS

State & future risk

COVID-19
Anxiety / mood

Intervention evidence

IntelliCare RCT
CRM app pilot

Sensing created new opportunities for observation and support

Reliable and responsible AI: 2021–2026

Longitudinal evidence, transfer across settings and user agency

2021–2023

Data & platforms

GLOBEM / RADAR-MDD
K-EmoPhone

State & future risk

Multisite evaluation
Bias in personal informatics

Intervention evidence

AirHeart RCT
Customization limits

2024

Data & platforms

College Experience
PriviAware

State & future risk

Stress Pipeline
MoodCapture

Intervention evidence

Minder RCT
Student support

2025

Data & platforms

DiversityOne
DataSentry

State & future risk

Mood trajectories
Early-warning limits

Intervention evidence

Therabot RCT
Stress JITAI trials

2026

Data & platforms

EmoPhone
EnPULSE

State & future risk

CrossShift
Dual-indicator stress

Intervention evidence

CRM RCT
DYMOND field probe

The research agenda now includes reliability in use and human agency

Three perspectives on digital mental-health AI

01

Model-centric AI

Prediction

How can models predict mental states better?

02

Data-centric AI

Characterization

How are data generated?
What do they represent?
How does their meaning vary?

03

User-centric AI

Agency

Can people understand, influence and benefit from AI?

Model-centric AI

Prediction

How can models predict mental states better?

Prediction targets and time horizons

Phone, wearable and IoT signals become features for a defined prediction target.

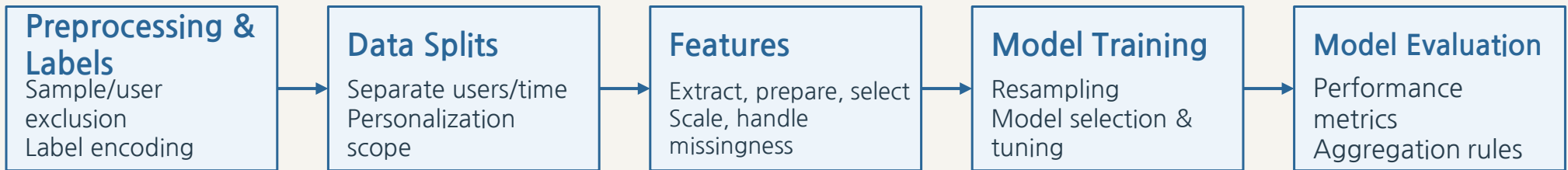
Now	Across weeks	Before an event	At intervention
Momentary affect	Symptoms over time	Relapse or deterioration	Receptivity
EMA mood, stress or arousal	Repeated symptom questionnaires	Clinical or adjudicated event	Availability and prompt response
What is the state now?	How does it change across weeks? (e.g., depression dynamics)	Can we identify risk before an event? (e.g., bipolar, depression, schizophrenia-spectrum)	When is support welcome?

Reproducibility is a pipeline property

Results depend on choices throughout the pipeline

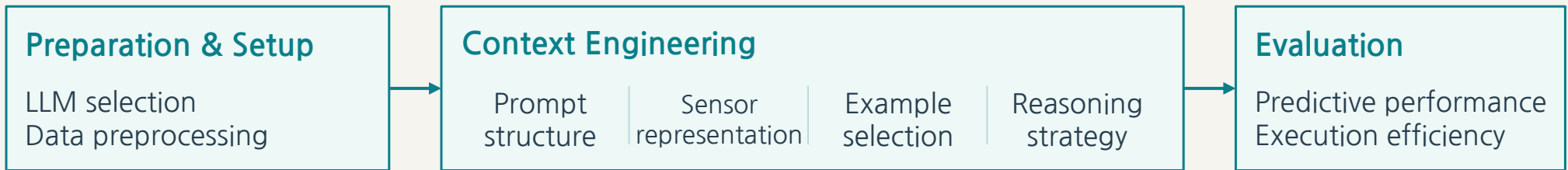
ML

Feature-based learning



LLM

Language-model approaches



Gap

Arises when records omit the design choices and execution settings needed to reproduce results

The complete record makes a result reproducible

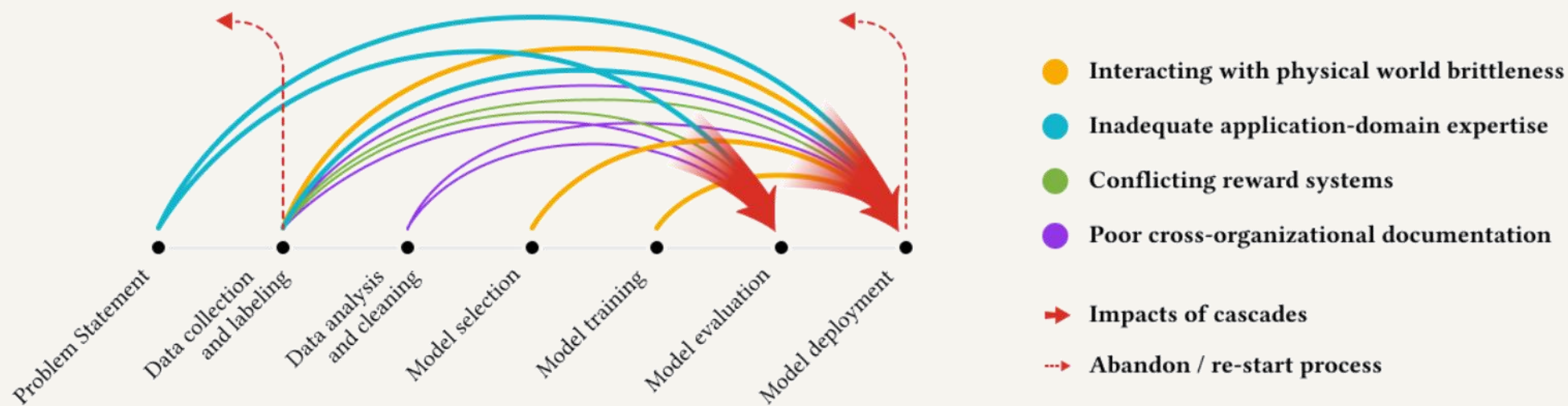
Data-centric AI

Characterization

How are data generated, represented and interpreted?

Data cascades and the limits of model improvement

Missing context and weak measurement can persist through model optimization



Data Cascades: compounding events causing negative, downstream effects from data issues, resulting in technical debt over time

Data cascades begin with measurement

Missing context and weak measurement can persist through model optimization

Technical failure

Permission, battery or upload problem

How to handle?

Diagnose and restore collection

Non-wear / non-use

Device absent from daily activity

How to handle?

Check observation opportunities

User choice

A person pauses or limits sharing

How to handle?

Respect and record the choice

Life or health change

Routine, capacity or context changes

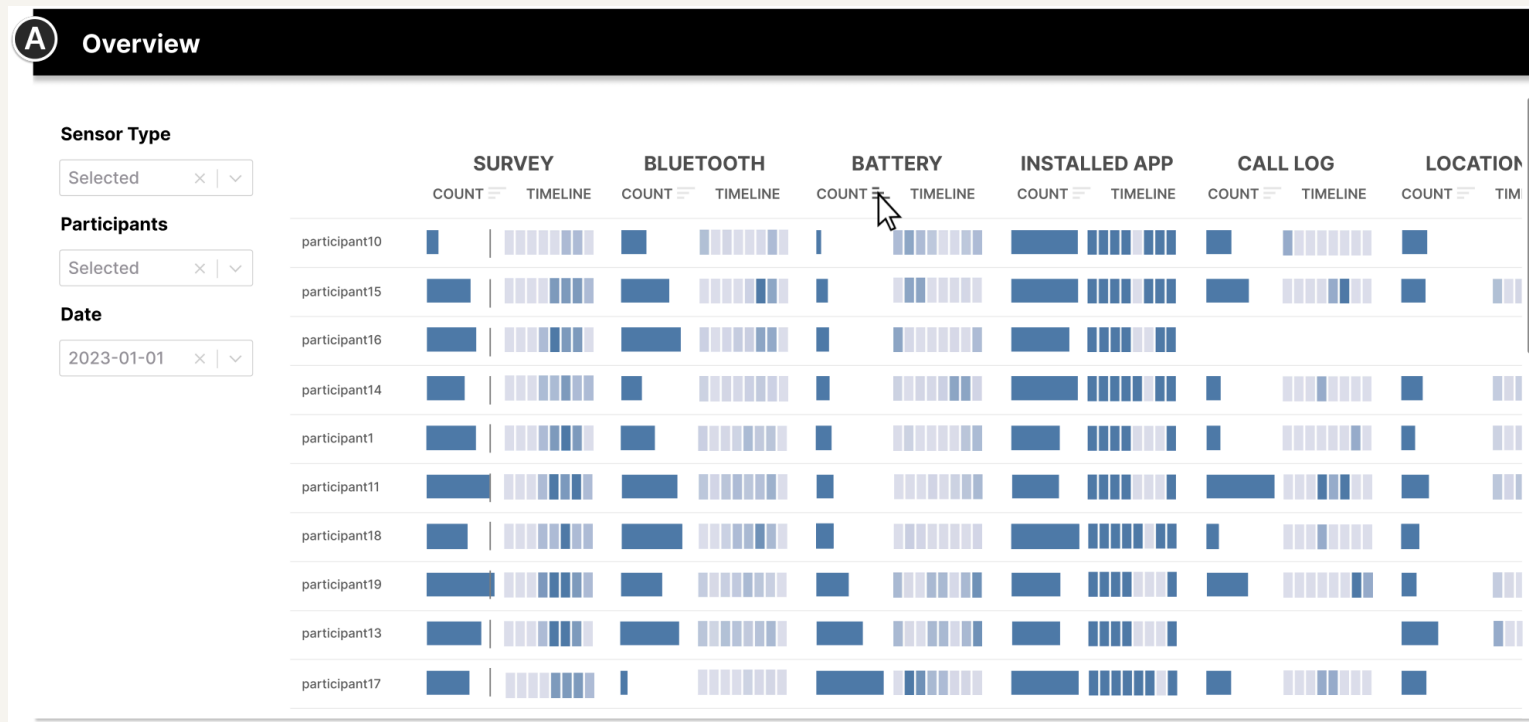
How to handle?

Interpret with personal context

Retention, availability, signal validity and usable labels differ

Quality data collection: DataSentry

Researchers need support to detect gaps, diagnose causes and act



Generalizability: datasets and study contexts

Different populations, contexts and times create different transfer tests

StudentLife

48 10 weeks

people

Phone + EMA

Daily behavior and symptoms

CrossCheck

61 12 months

people

Phone + EMA

Symptoms and relapse

RADAR-MDD

623 Up to 24 months

people

Phone, Fitbit, PHQ-8

Long-term MDD monitoring

GLOBEM

497 10 weeks / year

people

Phone + wearable

Cross-study transfer

CES

215 ~4 years / cohort

people

Phone + weekly EMA

Transitions and pandemic

K-EmoPhone

77 7 days

people

Phone + wearable + EMA

Momentary affect

EmoPhone

297 ~4 weeks / wave

people

Phone + Fitbit + EMA

User, time and wave transfer

DiversityOne

782 4 weeks / 8 countries

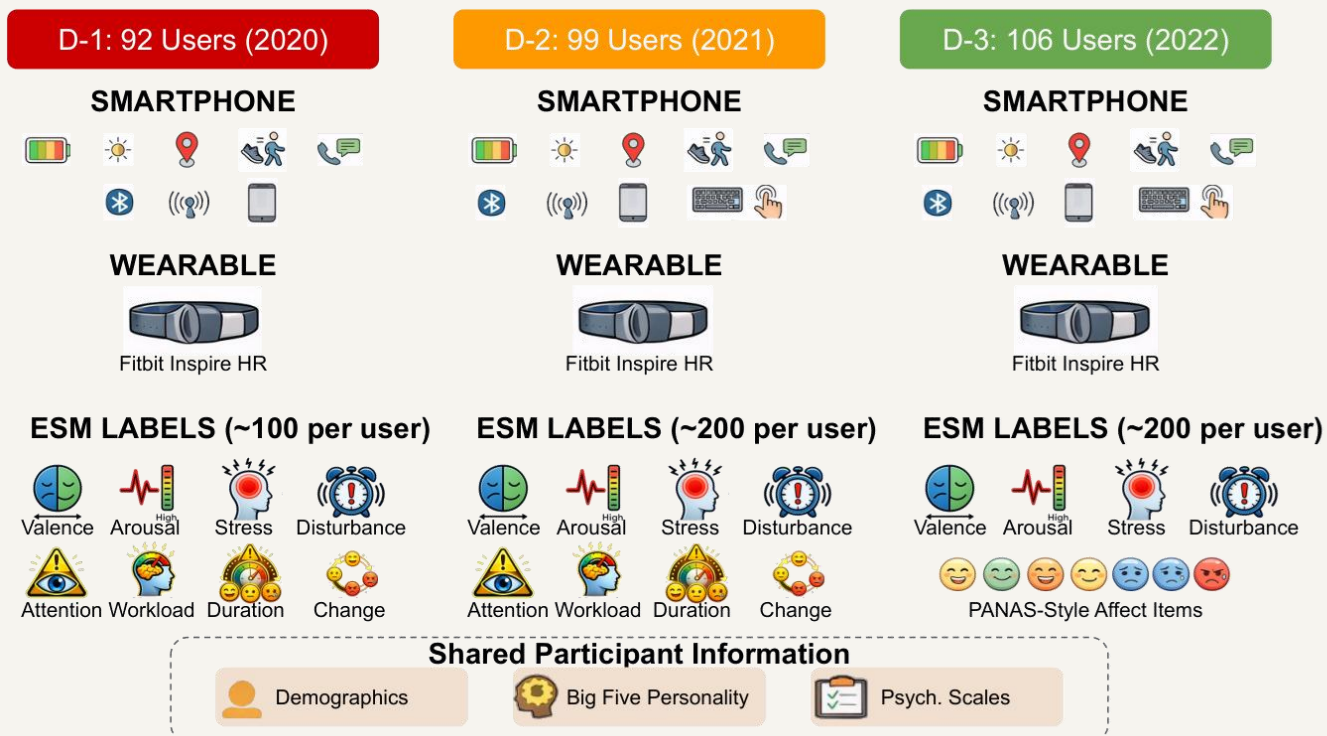
people

Phone + EMA

Cross-country transfer

EmoPhone: generalization across collection waves

297 participants across three annual waves



Within user

Future observations from the same person

Across users

An unseen person in a known wave

Across waves

A new wave with changed conditions

Generalizability: interpersonal differences

Example: Predicting stress (Y) from physical activity (X)

1 Covariate Shift

The distribution of activity $P(X)$ differs across users



User A



User B

$P_A(X: \text{activity}) \neq P_B(X: \text{activity})$

2 Label Shift

The distribution of stress $P(Y)$ differs across users



User A



User B

$P_A(Y: \text{stress}) \neq P_B(Y: \text{stress})$

3 Conditional Shift

Activity differs even at similar stress levels



User A



User B

$P_A(X: \text{activity} \mid Y: \text{stress} = \text{high}) \neq$
 $P_B(X: \text{activity} \mid Y: \text{stress} = \text{high})$

4 Concept Shift

Stress differs even at similar activity levels



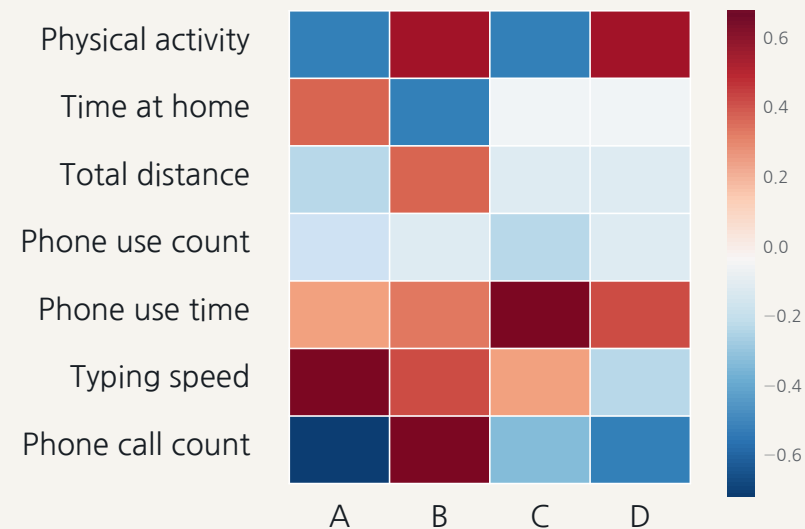
User A



User B

$P_A(Y: \text{stress} \mid X: \text{activity} = \text{high}) \neq$
 $P_B(Y: \text{stress} \mid X: \text{activity} = \text{high})$

Correlations Between Stress and Features



CrossShift: personal mappings matter for transfer

Differences remain after normalization and correlate with performance decline (stress prediction)

8

datasets

297 in-house + 399 public

Analysis samples across eight datasets

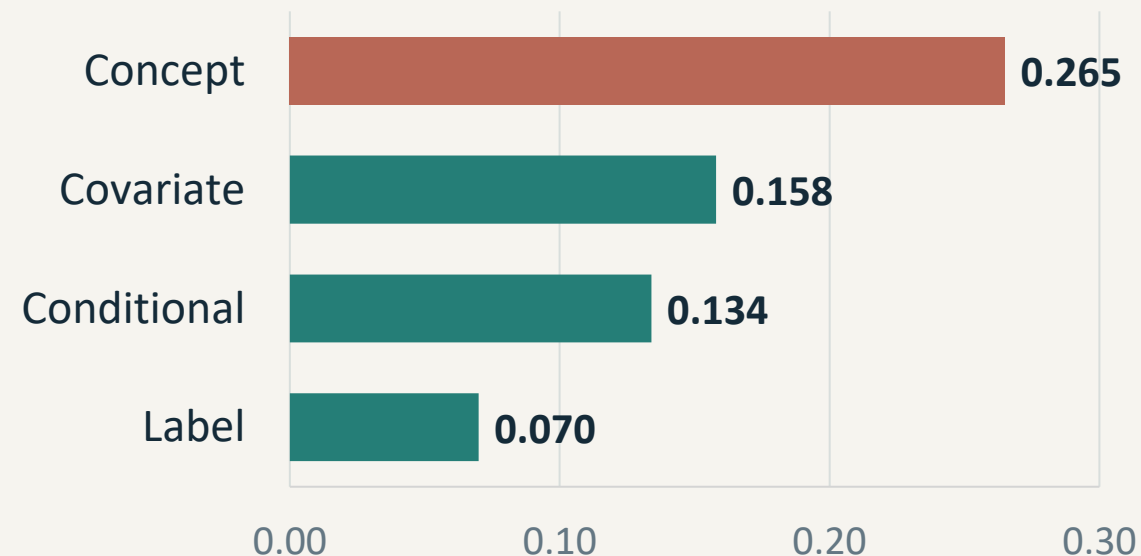
696

participants

Within-user and cross-user models

Regression of decline on four shifts (from within-user to cross-user prediction)

Concept shift shows the largest mean coefficient

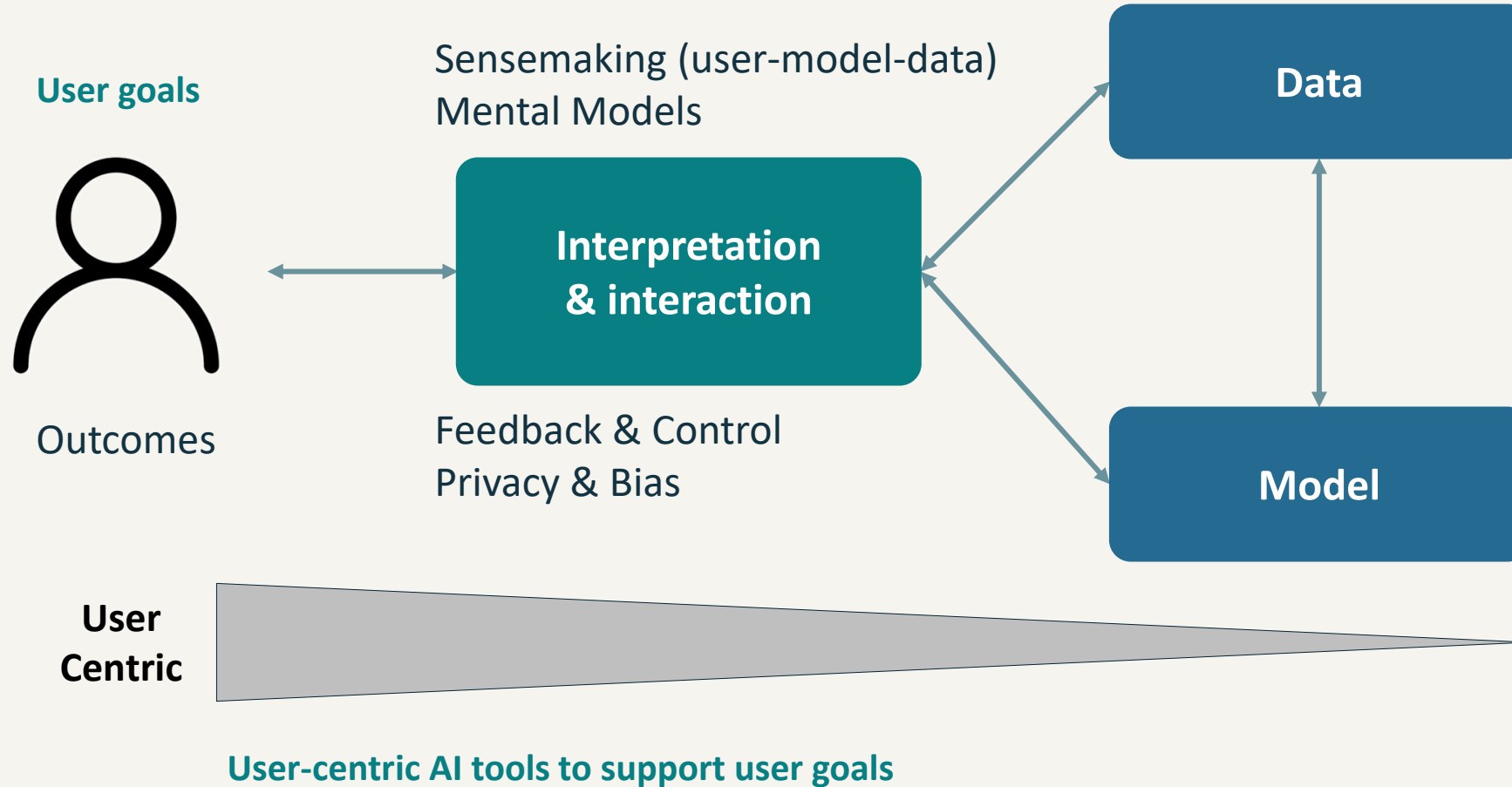


Mean regression coefficient magnitude

User-centric AI

Agency: can people understand, influence, and benefit from AI?

User-centric AI: People, interfaces, data, and models



Explainable AI, causability and sensemaking

Explainable AI

Why did the model produce this prediction?

Inspect influential inputs, model behavior and uncertainty.

Causability

Does the explanation help me understand and evaluate the result?

Judge whether the explanation is understandable and useful.

Personal sensemaking

What does the pattern mean in my life?

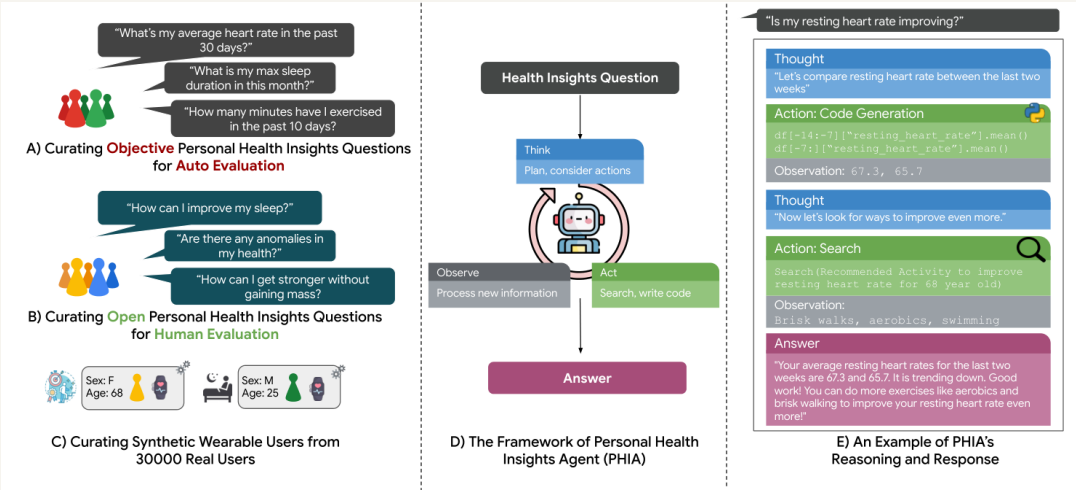
Compare experiences, add context and reflect on alternatives.

Understanding the model and understanding oneself are related goals

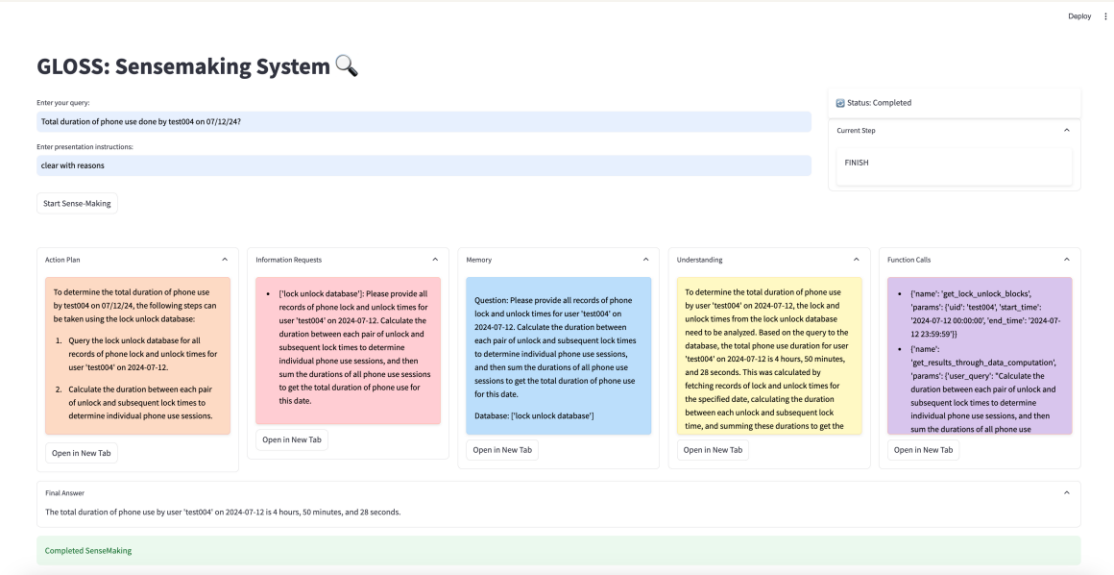
Conversational access to sensing data

People can ask questions that go beyond a fixed prediction dashboard.

PHIA



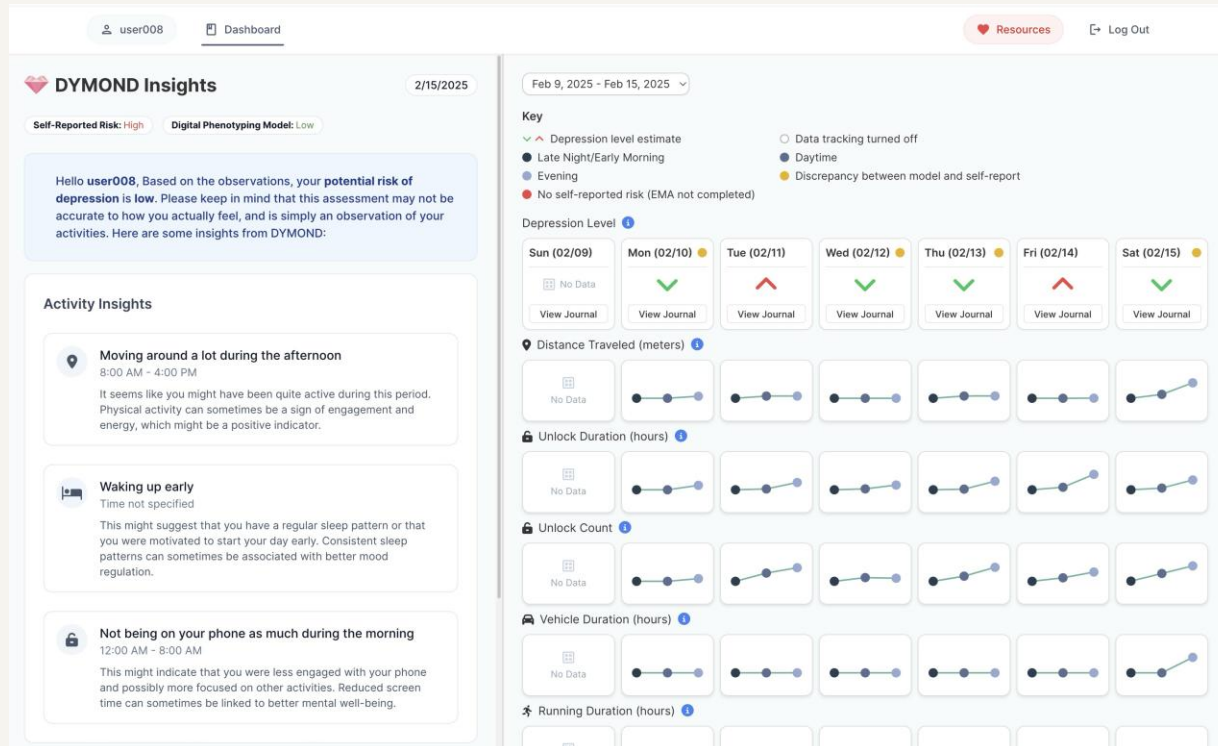
GLOSS



Wearable health questions
Code and web search tools

Multimodal data inquiry
Visible analysis steps with multi-agent processing

Collaborative reflection with DYMOND



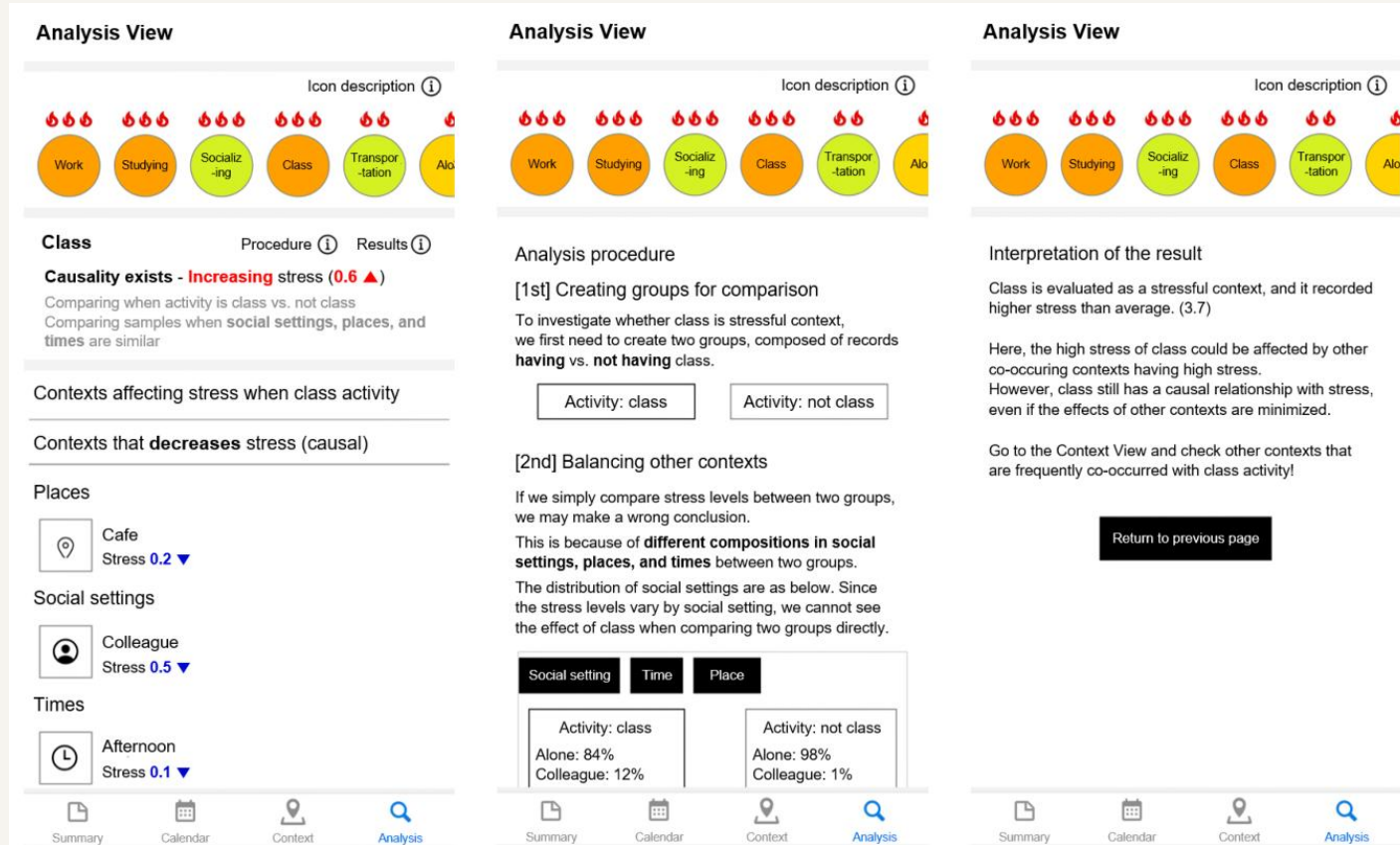
22 people
6-week technology probe

Data
Does this trace represent my day?

Model
Does the mapping fit my experience?

Output
Why does the estimate disagree with me?

Contextual sensemaking with DeepStress



Question

Which contexts might influence my stress?

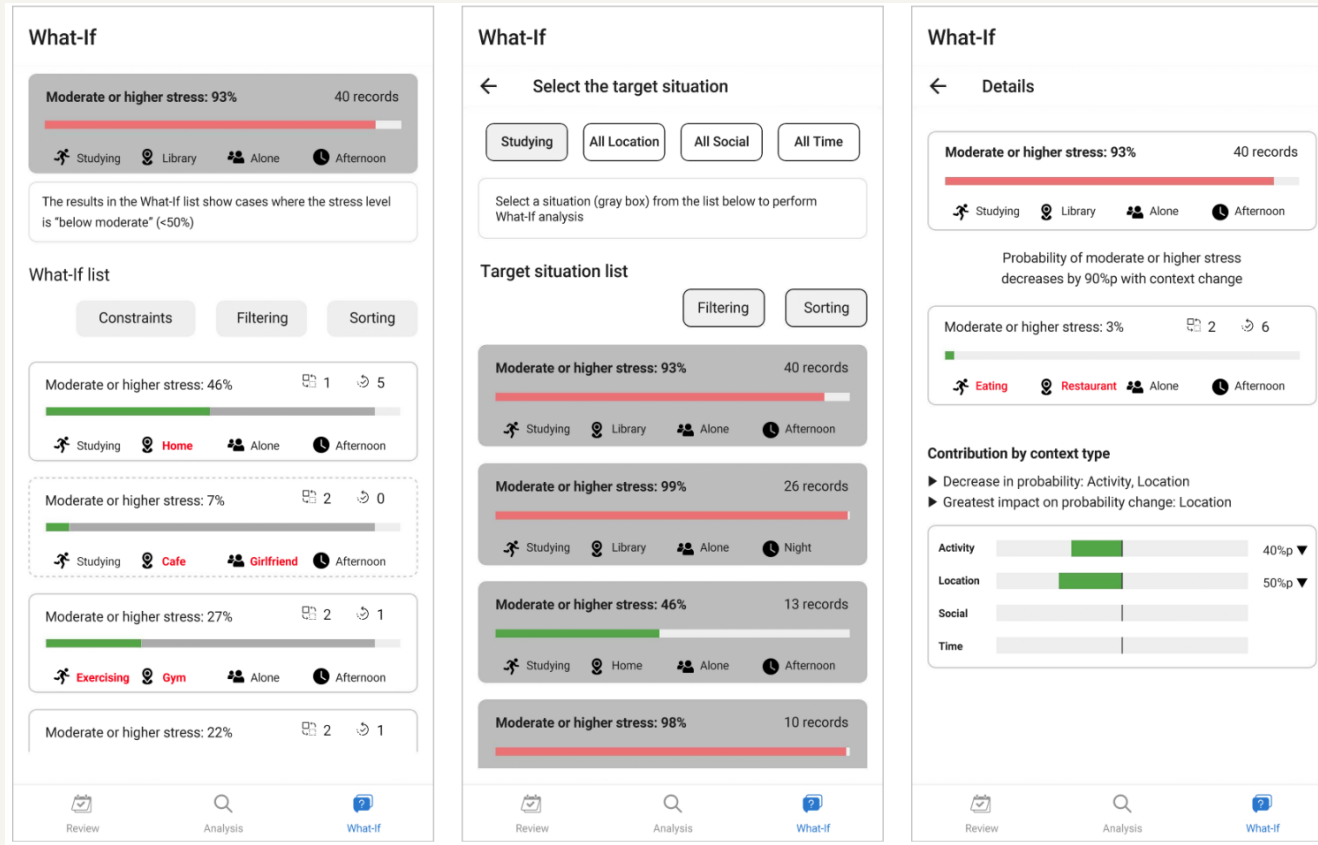
Method

Compare episodes matched on measured contexts.

Personal interpretation

Confirm, question and reinterpret patterns.

Counterfactual plans with CounterStress



Personal model

Generate alternatives with lower predicted stress.

User constraints

Lock contexts and compare feasible changes.

Everyday planning

Choose a plan using personal knowledge.

Feedback & Control

Give people clear ways to question, configure, correct, and pause AI.



Challenge an inference

Add context or dispute an estimate.



Set the error trade-off

Choose the balance between false alarms and missed detections.



Correct the learner

Show what feedback changes. Support undo and unlearning.



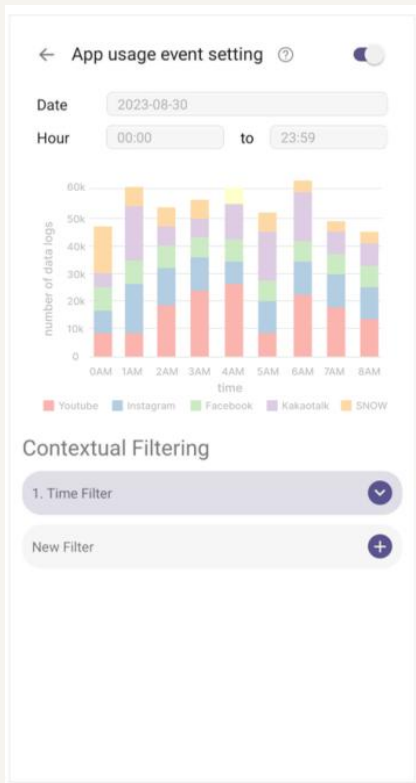
Pause and recover

Pause prompts and revise preferences.

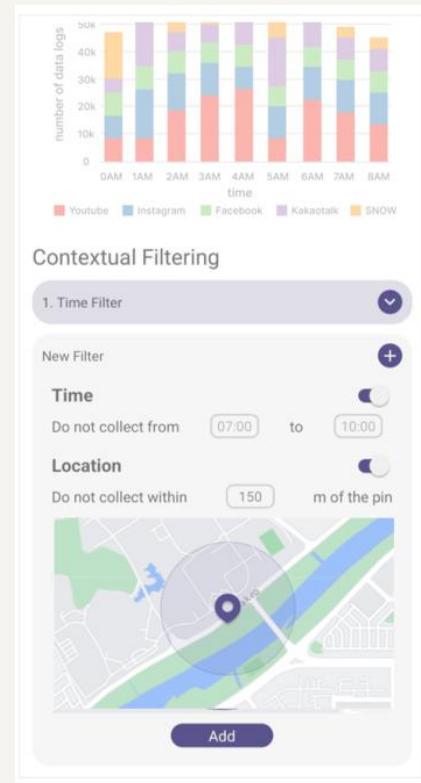
User control should change system behavior.

Visible data and granular privacy control

Raw traces



Contextual filters



PriviAware

People inspect raw sensor traces.

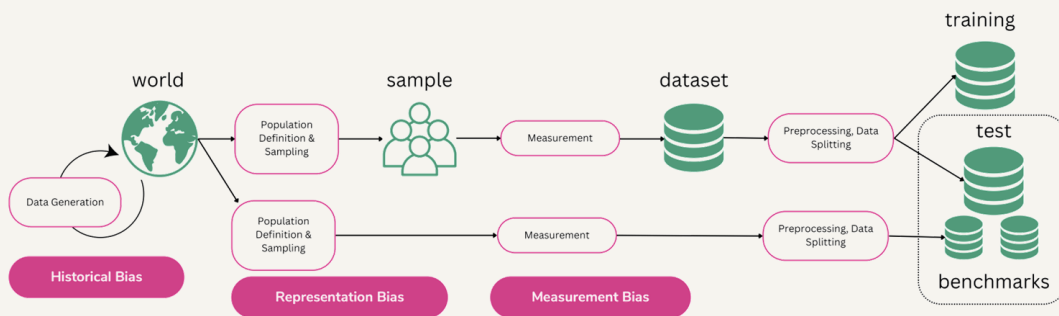
Time and location filters limit shared data.

Controls support agency while adding configuration work.

Control consequences need to be understandable.

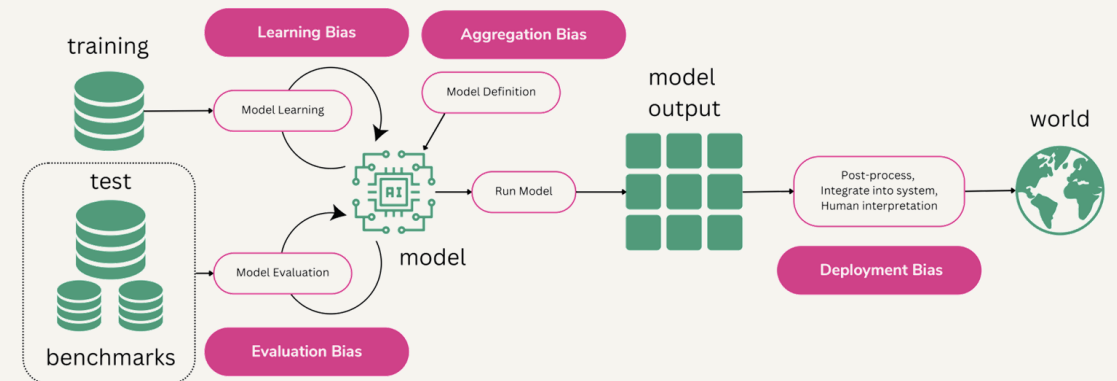
Bias across the machine-learning lifecycle

Data and measurement



Whose lives are absent or poorly measured?

Model and deployment



Whose differences disappear in a model or test?

Bias in personal informatics: subgroup and intersectional evaluation exposes hidden disparities

Next steps

- Foundation models
- Deeper understanding
- Closed-loop adaptation
- Human agency in coaching

Foundation models and context-aware systems

Foundation-model families

Wearable models

Pulse-PPG, NormWear
Across sensors and people

Time-series models

Sequence foundation models
Across contexts and time

Tabular models

TabFM, TabPFN
Small personal datasets

Context + interaction

DAIMON

Data: longitudinal study support

TRACE

Model: activity recognition

DYMOND

People: reflection and control

Efficiency, personalization and validity require empirical comparison.

Deeper understanding of data, models, and people

Data

What do signals and labels actually represent?

Measurement validity
Quality and missingness
Context and time

Models

What do models learn, and when do they fail?

Learned relationships
Uncertainty and failure modes
Personalization and transfer

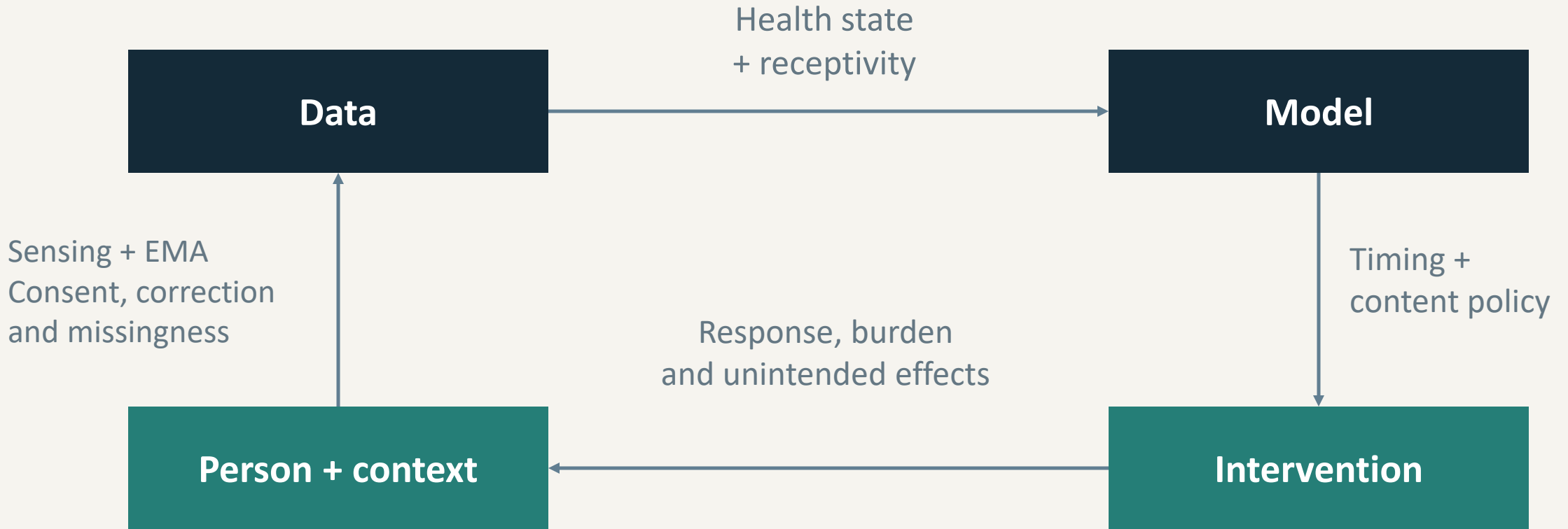
People

Why does the same behavior mean different things?

Personal baselines
Lived experience
Situational constraints

Standardized benchmarks and longitudinal studies support this agenda

Closed-loop adaptation



Actions change the future data distribution

Coaching that strengthens human agency

How can a coach strengthen independent sensemaking and coping over time?

Proactive people

Notice patterns and reflect on experience.

Practice coping skills and choose when to seek support.

Success may include needing the tool less.

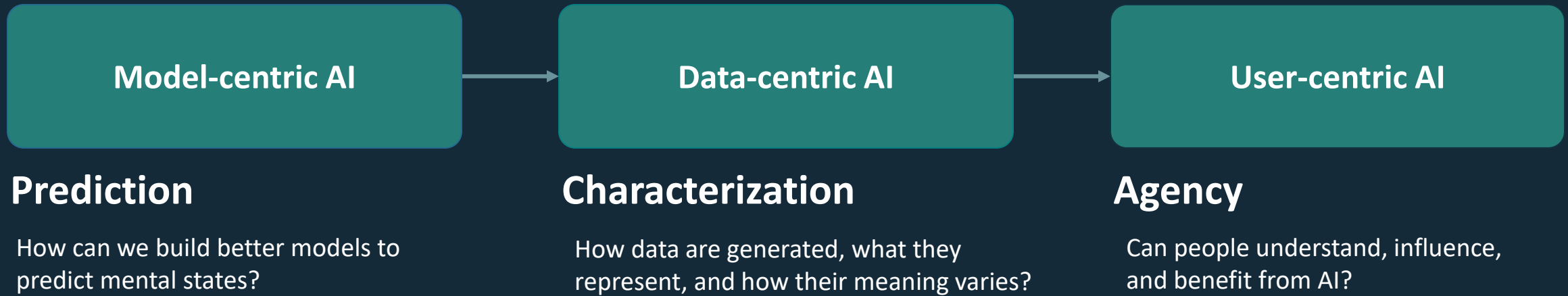
Cognitive offloading

Ready-made answers may reduce personal understanding.

Test whether assistance builds skills or increases dependence.

The effect depends on design, task and person.

Beyond prediction



Next Steps

