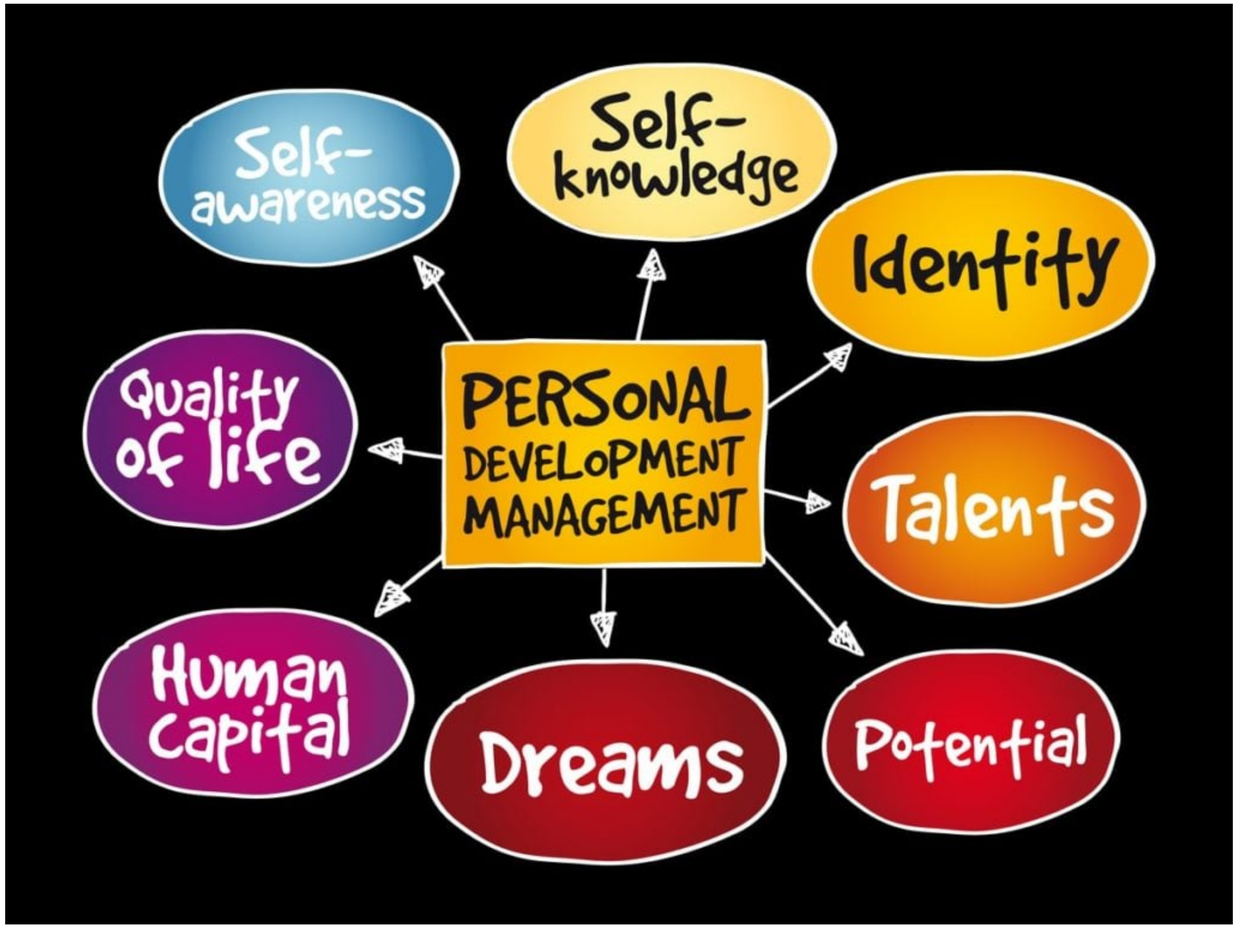
Human-Computer Interaction Institute

# **Towards Ment'AI'ist:**
# Augmenting Human Development Practitioners through Behavioral Sensing

Riku Arakawa (Ph.D. student, CMU HCII)
https://rikky0611.github.io/

# Human Development



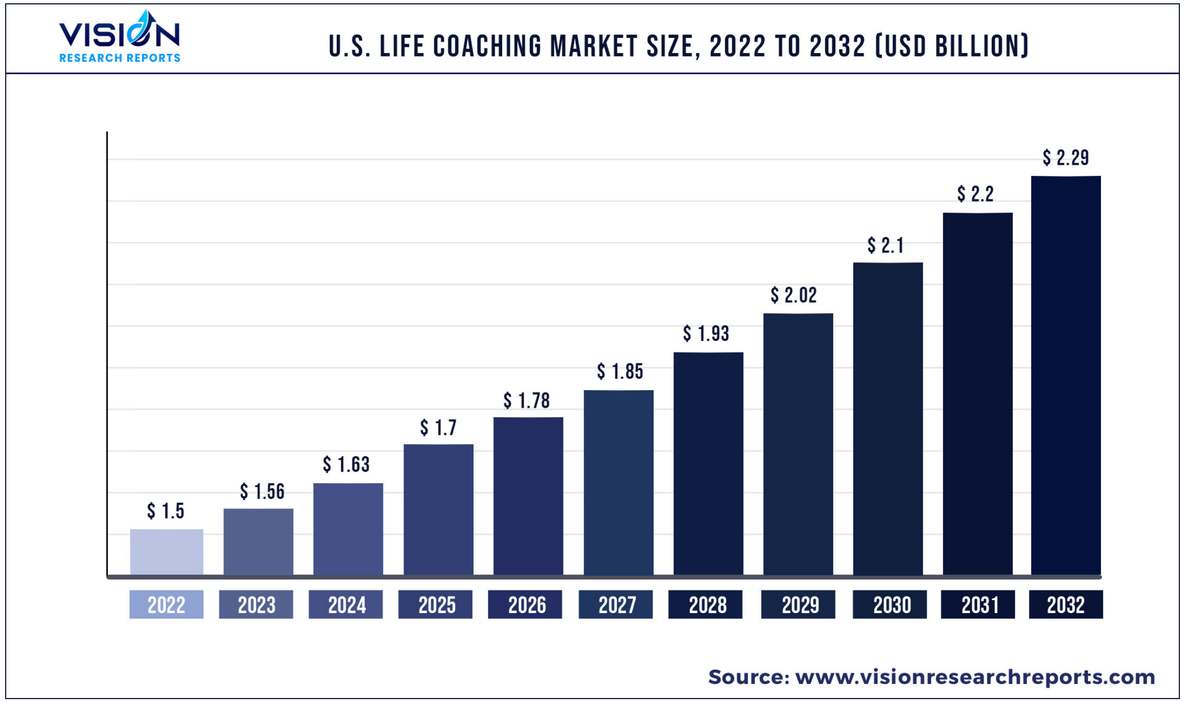

Process of improving well-being and capabilities

https://www.aworldfullofrae.com/why-personal-development-coaching/

# Reflection through Dialogue with Practitioners

coaching, counseling, therapy, etc.

# Lack of Practitioners




40% reporting inadequate experience

## Increasing demands and training barriers

https://www.visionresearchreports.com/us-life-coaching-market/39949
https://www.ddi.com/about/media/coaching-research-2024

# Motivation and Research Questions

**Support human development practitioners with AI technology**

- How can we blend human and AI in human development scenes?
- What is the unique strength of human practitioners and AI during face-to-face conversation?
- What is needed to accelerate human-AI interaction in this domain?

# Motivation and Research Questions

**Support human development practitioners with AI technology**

- How can we blend human and AI in human development scenes?
- What is the unique strength of human practitioners and AI during face-to-face conversation?
- What is needed to accelerate human-AI interaction in this domain?

# **Coaching Copilot:**
# Blended Form of an LLM-Powered Chatbot and a Human Coach to Effectively Support Self-Reflection for Leadership Growth

Riku Arakawa, Hiromu Yakura (equal contribution)
ACM Conversational User Interface (CUI) 2024

# Field: Executive Coaching

- The coach attempts to develop the client's leadership qualities.
- The dialogue often involves goal setting, action planning, and reflection.

Robert G. Hamlin, et al. 2008. The emergent ‘coaching industry’: a wake-up call for HRD professionals. Human Resource Development International.

# Research Question

Can LLM-Based chatbot replace human coach?

# Interviews with Coaches

- 8 coaches from a coaching company, June 2023
- Can we use well-prompted ChatGPT (*GPT-4) as a coach?

# Results from the Interviews

**It is not ready to use in the session without a supervision**

## LLM Limitation (as of the study period)

It requires many skills to deal with highly human context. Just asking template questions would never facilitate their reflection.

Coach

## Value of Human Presence

One of the reasons coaching is effective is the sense of being invested time by other people

# Blended Approach: Use Chatbot as a Complement


1. Set a goal by the next session with a human coach
2. Chatbot is prompted to facilitate reflection for the goal

# User Study

- 10 coach-client pairs, 2 weeks
- (Metrics related to self-reflection)
- Exit interviews with coach and client individually

# Summary of Findings

- Both the coach and client appreciate chatbot as a 24/7 system
  - **Reduced coach's burden to some extent**
- The client kept engaged over two weeks
  - Asking about action plans. Enjoying acknowledgement.

- **Chatbot could not pose critical questions for deep reflection**

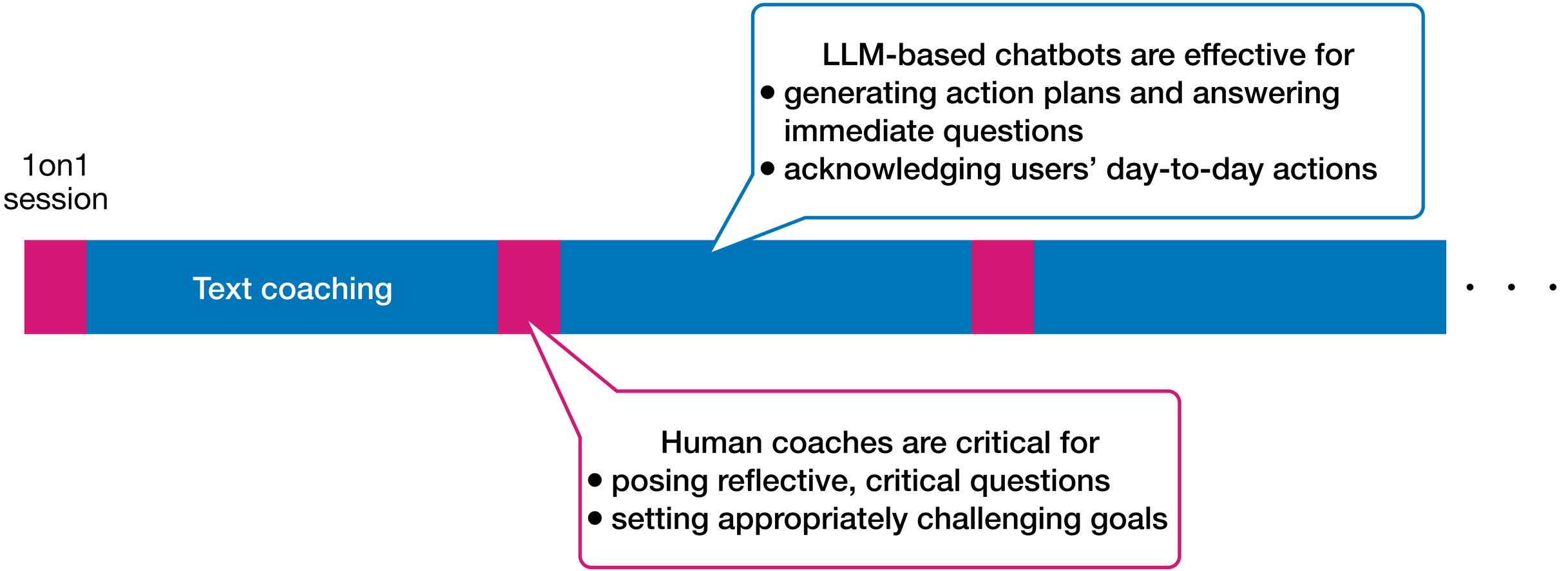
# Reconfirmed Benefits of Face-to-Face Session

Coach

Questions like **identifying behaviors the client unconsciously avoids** would deepen the conversation but were not observed.

It requires many skills to deal with highly human context, not only listening to their words but also **monitoring their behavior**.

# Summary




**This is still challenging and requires skills!**

# Motivation and Research Questions

**Support human development practitioners with AI technology**

- How can we blend human and AI in human development scenes?
- What is the unique strength of human practitioners and AI during face-to-face conversation?
- What is needed to accelerate human-AI interaction in this domain?

# **REsCUE**:
# A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection

Riku Arakawa, Hiromu Yakura (equal contribution)
ACM CHI 2019

# Coach’s Role during 1on1 Session

**Coaches**

- ask questions to induce deep reflection
- notice a discrepancy between the verbal response and the actual thoughts using nonverbal cues.

**Behavior analysis is a critical skill**

Goman 2008. The Nonverbal Advantage: Secrets and Science of Body Language at Work.

# Nonverbal Behavior and Internal States

**Internal states**
ex. impatient, nervous

**“spontaneous & unregulated expression”**

Darwin, 1872. The Expression of the Emotions in Man and Animals.
DePaulo, 1992. Nonverbal behavior and self-presentation.


# Some people can infer human internal states through observation.

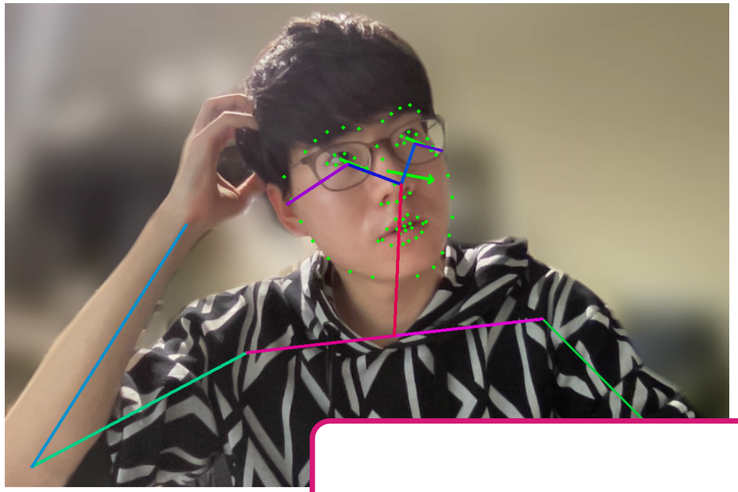
https://www.tvnz.co.nz/shows/the-mentalist

# Research Question

Can AI have this “mentalist” skill?

# Behavior Sensing Technology

# Should AI “Recognize” Behavioral Cues?



Estimate → Upset → Share

Is it correct?

Is it beneficial?

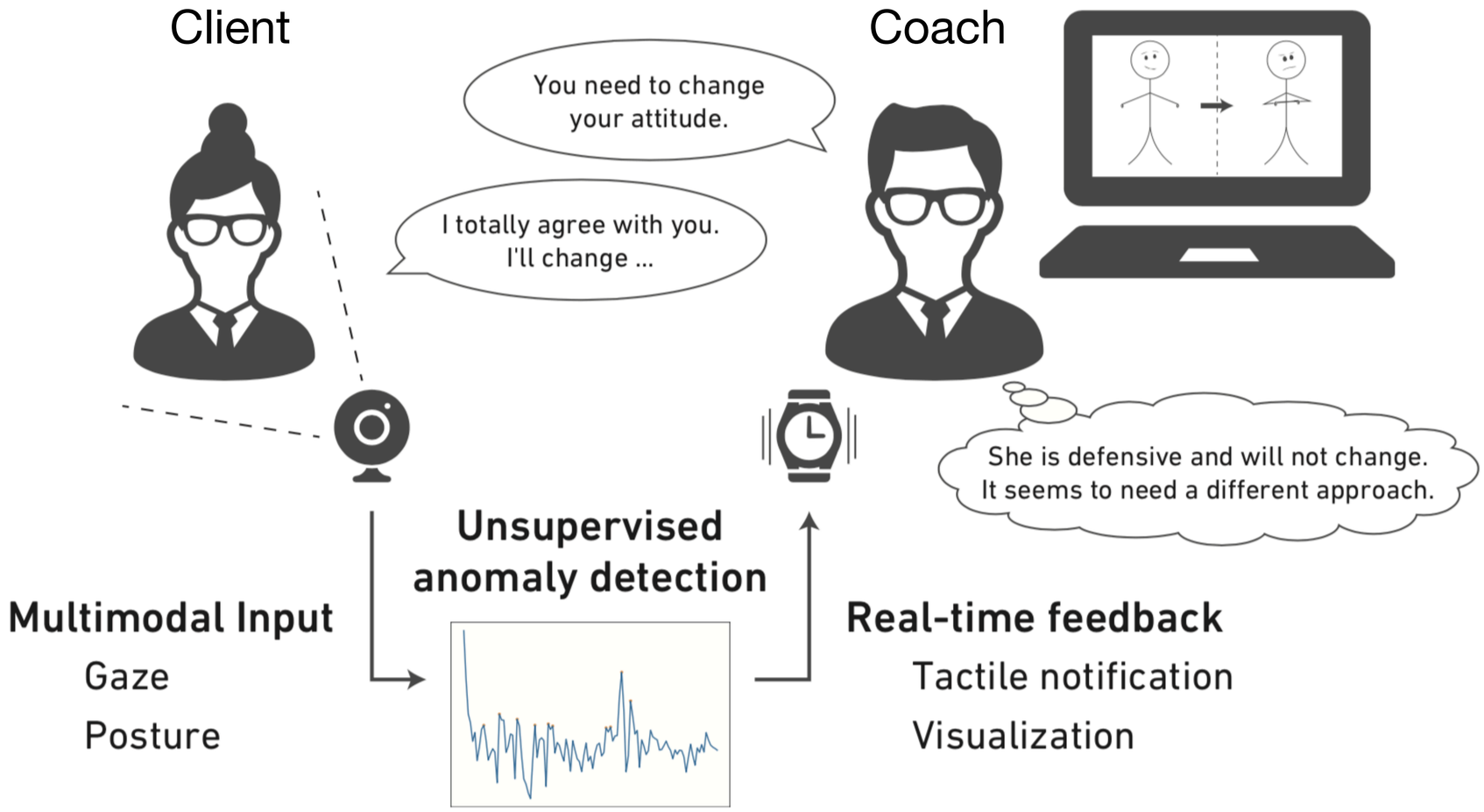
# What's the Strength of AI an Human?

**Human**
Pros:  Good at understanding context
Cons: Difficult to notice behavioral cues due to task cognitive load

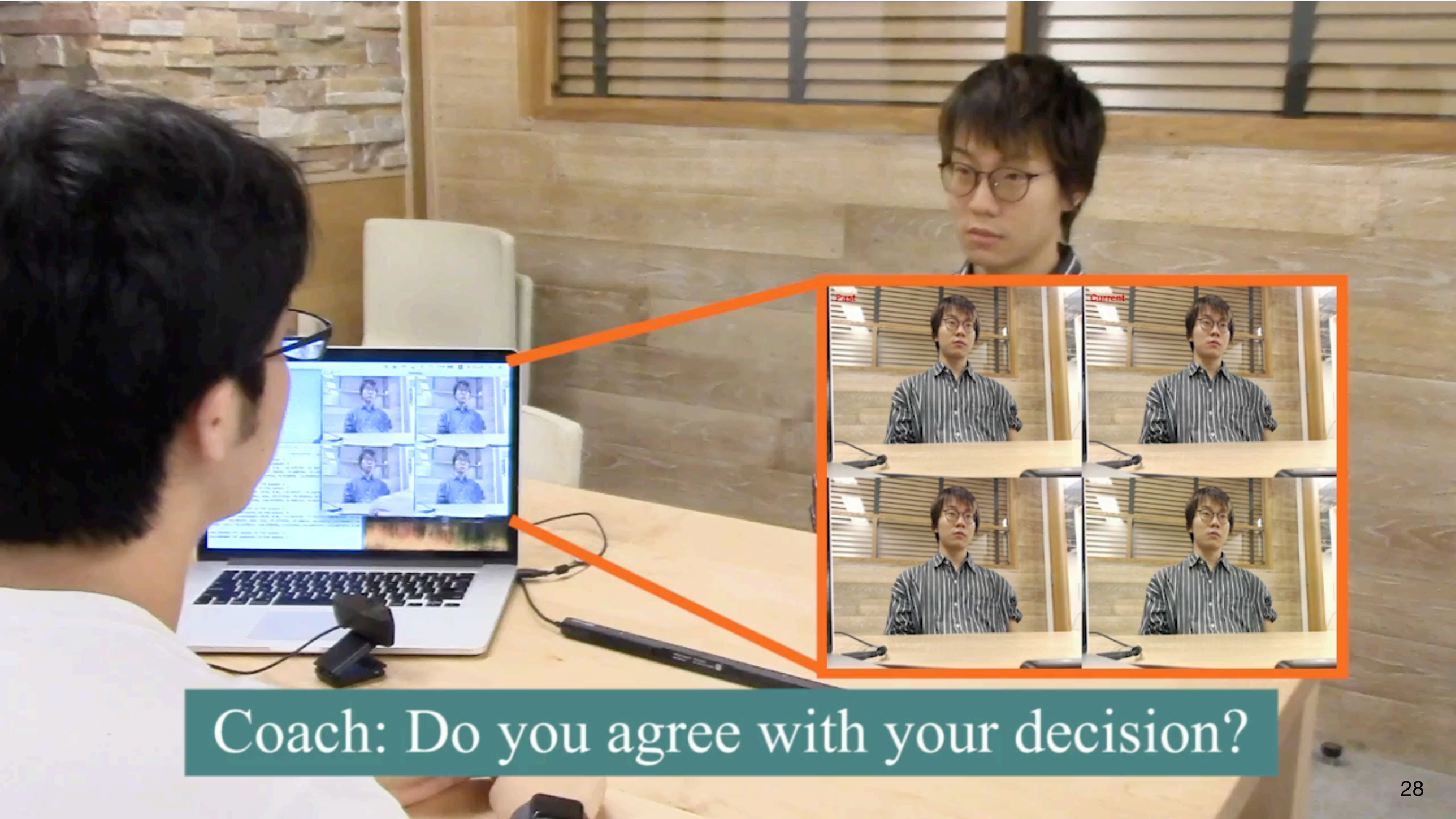
**AI**
Pros:  Stable performance
Cons: Not good at dealing with context

**Observation for AI, judgment for human**

# Proposed Framework: REsCUE

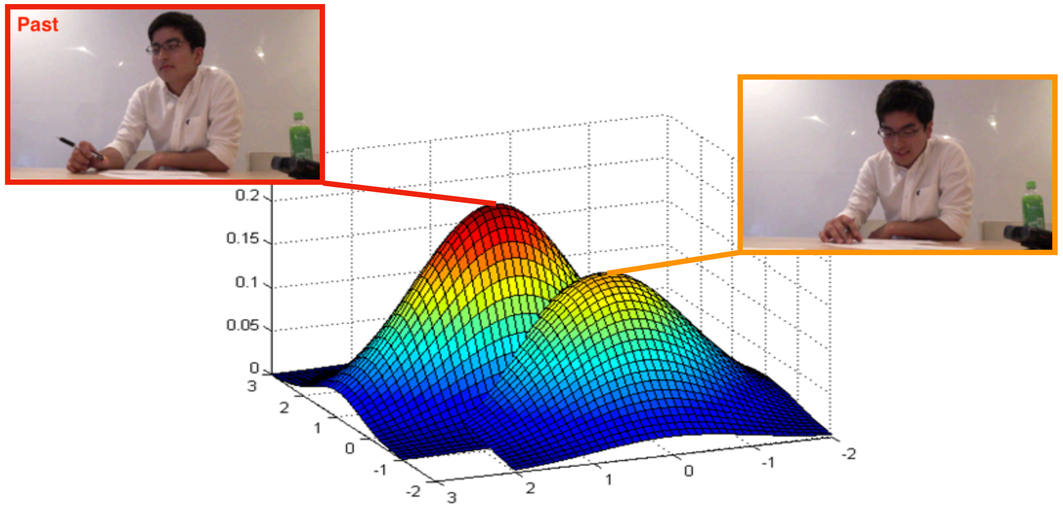


**We show a simplified example to illustrate REsCUE's assist for coaches.**

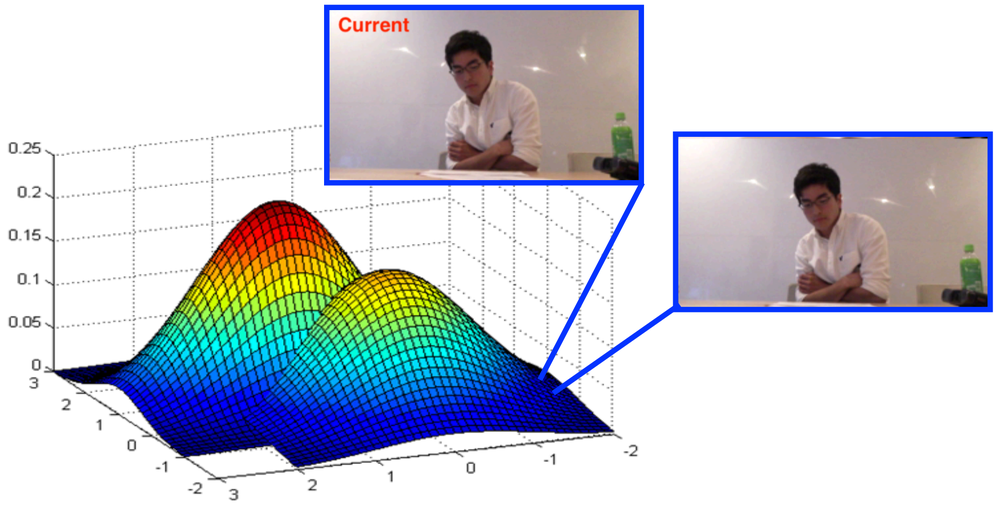

Past
Current
Coach: Do you agree with your decision?

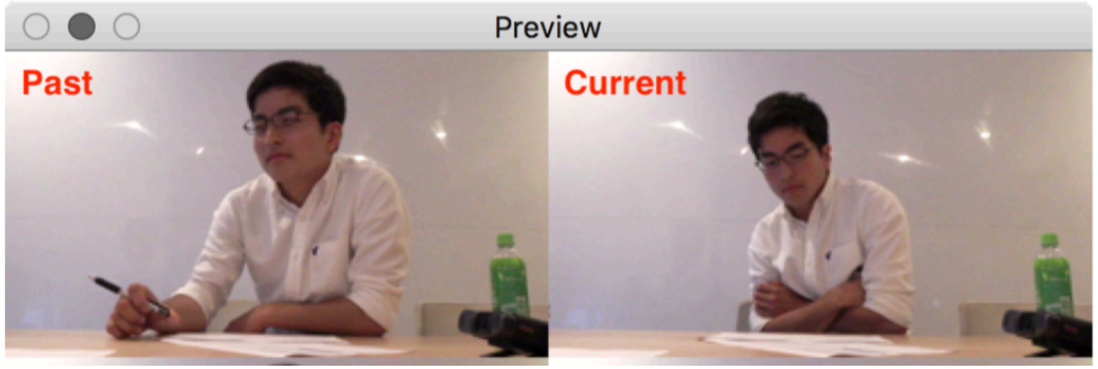
# Visualize Anomaly Detection with GMM

Past Frames

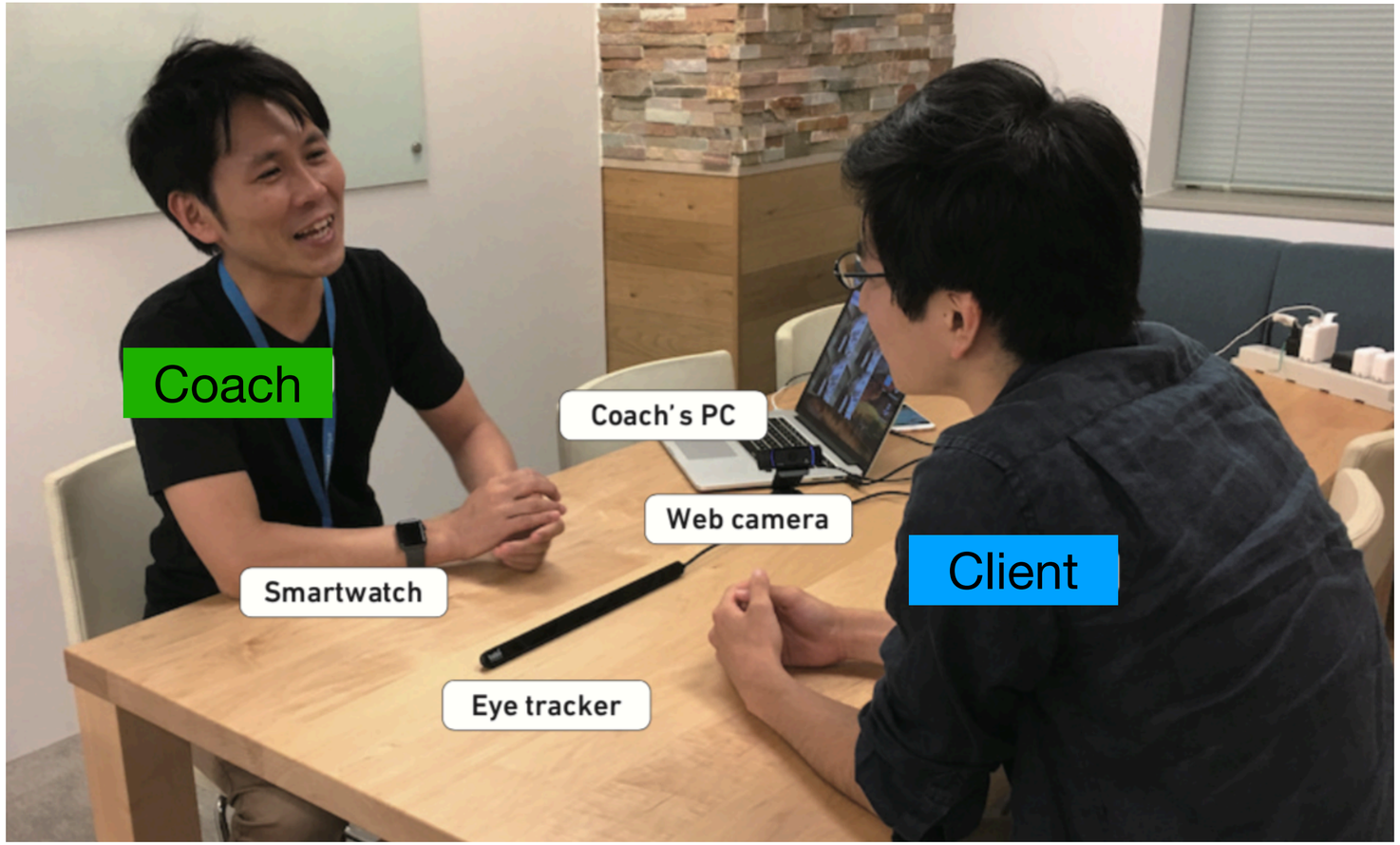



Current Frame



# User Study: Coaching with REsCUE




15 sessions with 5 coaches

# Feedback on REsCUE

Coach

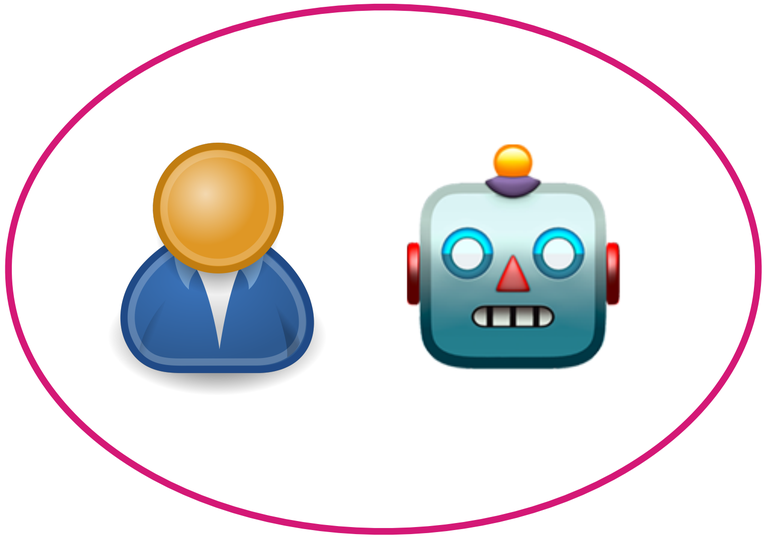
I had been convinced that the client was agreeing to my proposal, but **from the given feedback, I noticed that it didn't seem true**. So, I was able to make a decision to explain my proposal more carefully until he was satisfied.

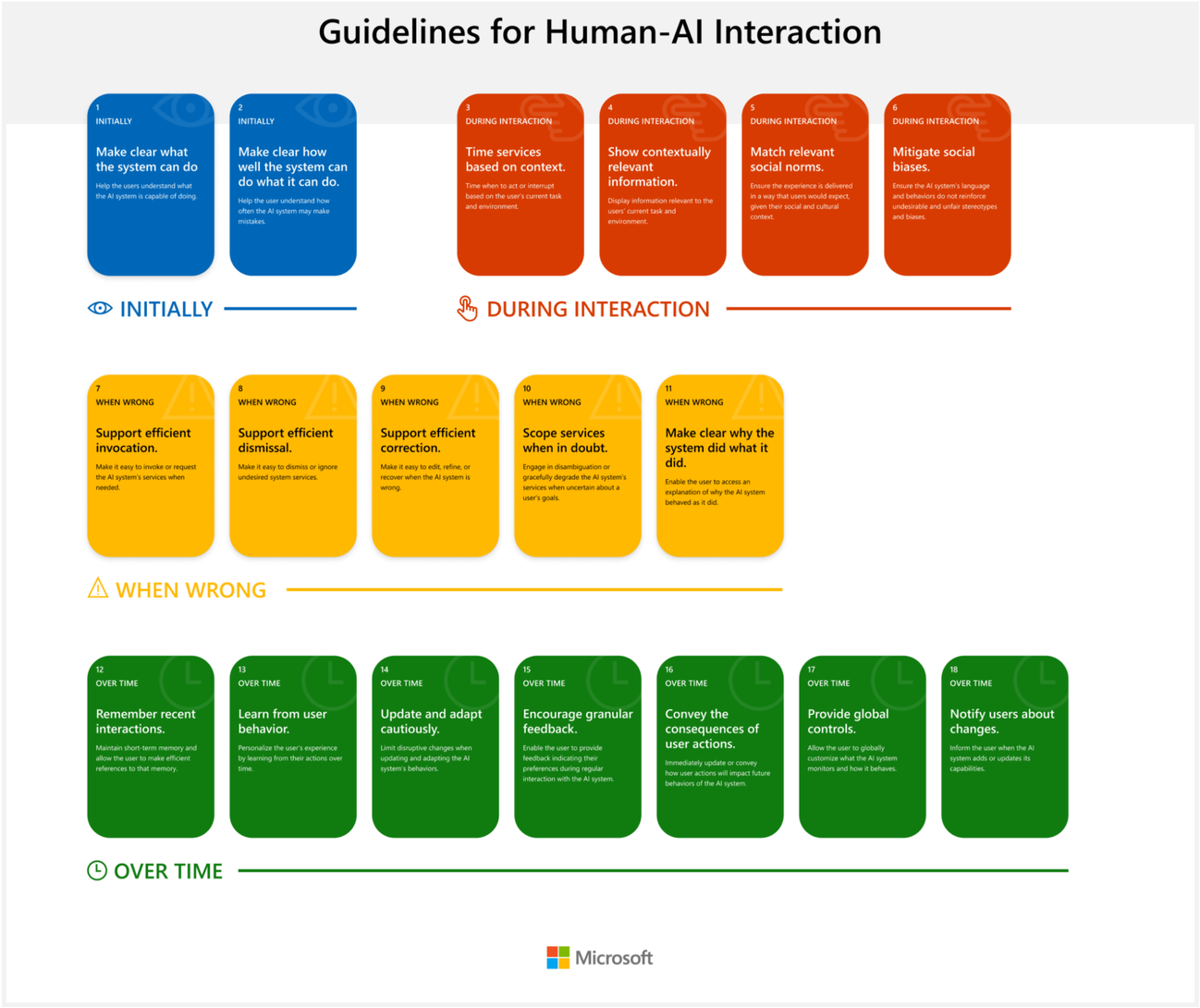
# Feedback on REsCUE

Coach

Simpler feedback such as just showing “defensive” could also be easy to understand. However, **if it contradicts my feelings, I could get confused and might ignore the feedback**. In that respect, **this system passes the initiative to me** and does not cause such confusion.

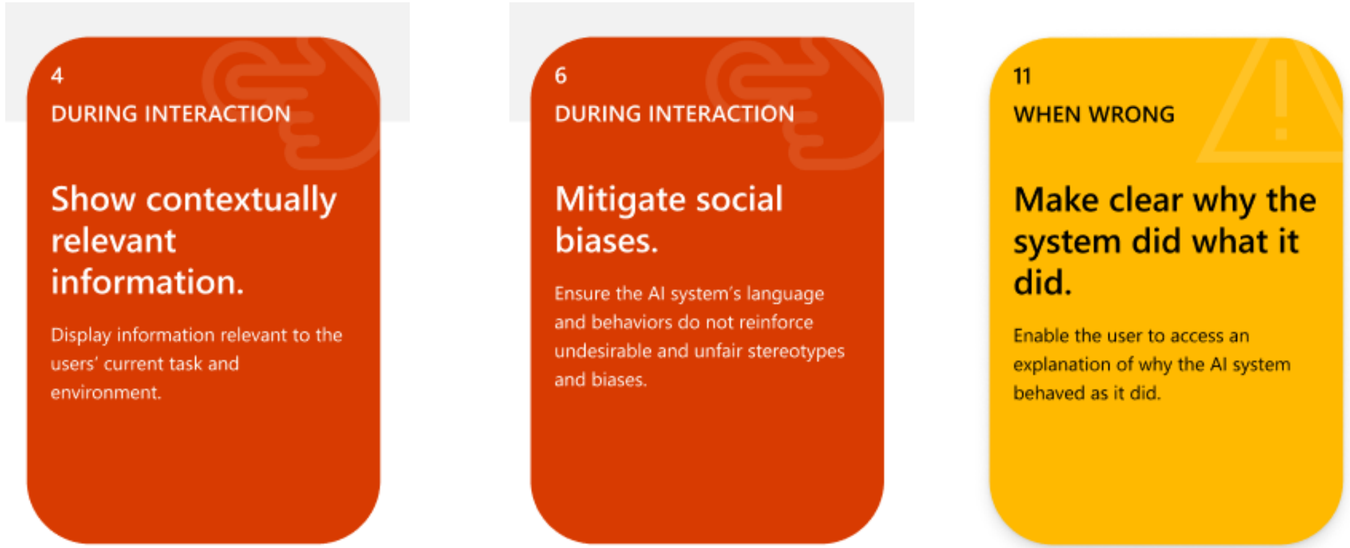
# Human-AI Interaction






Amershi et al., Guidelines for Human-AI Interaction, CHI 2019

# Human-AI Interaction




## REsCUE

- Anomaly detection relies solely on data patterns, without using heuristics.
- Judgment is made by humans.

# Applications of REsCUE

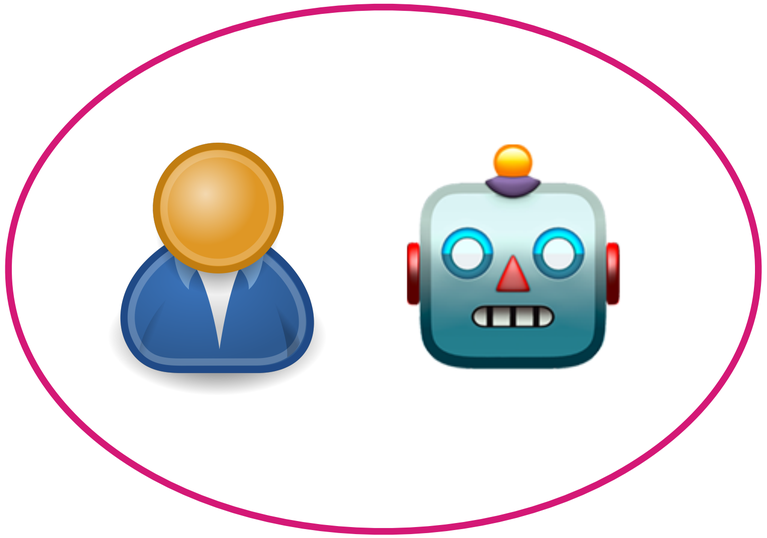

**Post reflection**
**CHI 2020**

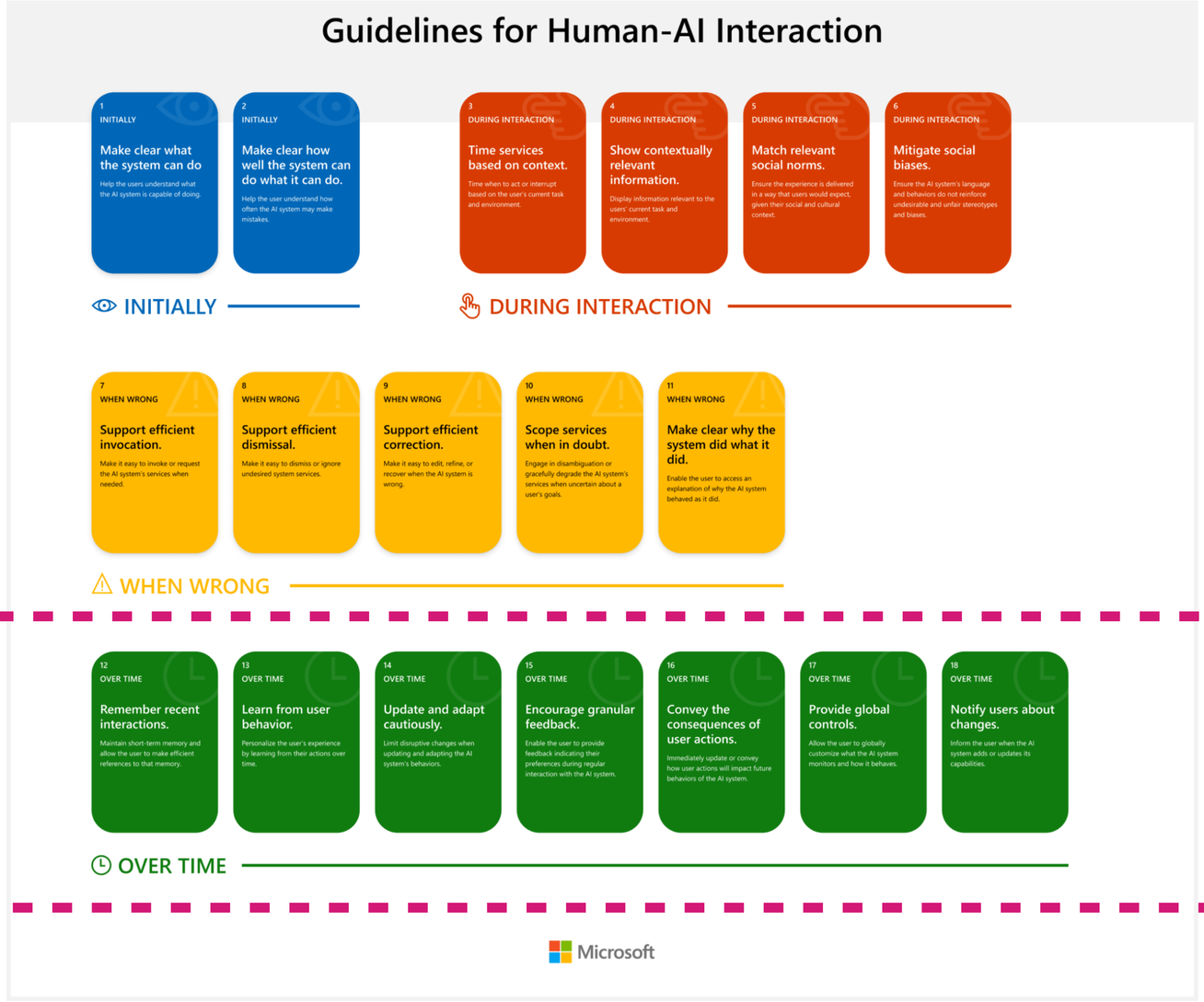
**Other domains**
**CHI 2023, case study 🥇**


**SaaS deployment (> 1M recordings)**
**CHI 2025, case study**

Practitioners want to **customize the detection algorithm**
- ignore “touching glasses”
- capture “struggling with question-answering”

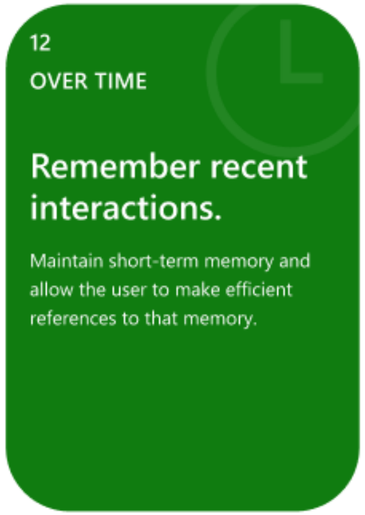
# Human-AI Interaction

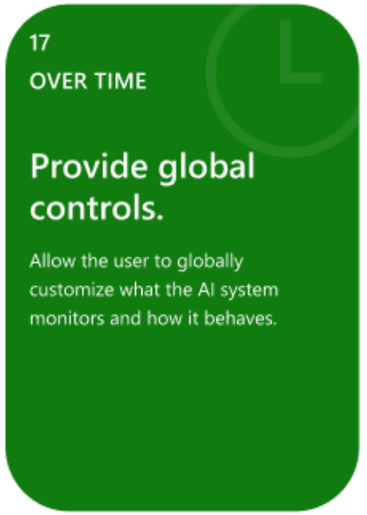





Amershi et al., Guidelines for Human-AI Interaction, CHI 2019

# Human-AI Interaction “Over Time”




## Our Case

- Practitioners want to add their own knowledge / feedback to the algorithm

# Motivation and Research Questions

**Support human development practitioners with AI technology**

- How can we blend human and AI in human development scenes?
- What is the unique strength of human practitioners and AI during face-to-face conversation?
- What is needed to accelerate human-AI interaction in this domain?

# **ConverSearch:**
# Supporting Experts in Human Behavior Analysis of Conversational Videos with a Multimodal Scene Search Tool

Riku Arakawa, Kiyosu Maeda, Hiromu Yakura (equal contribution)

ACM Transactions on Interactive Intelligent Systems (TiiS) 2025

# Implicit Knowledge in Practitioners

I want the algorithm to look at “struggling question-answering”

How can we code these various knowledge in a generalizable manner?

# Multimodal Scene Search of Conversations

Linguistic

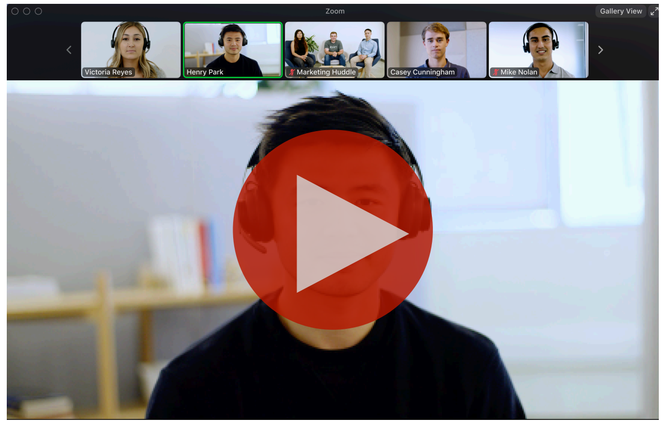
×

Paralinguistic

×

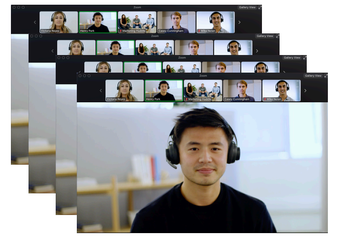
Non-verbal

Query




Result

key scenes



00:00:31 — 00:00:43

# Example: “Struggles with Question-Answering”

**Is questioned?**

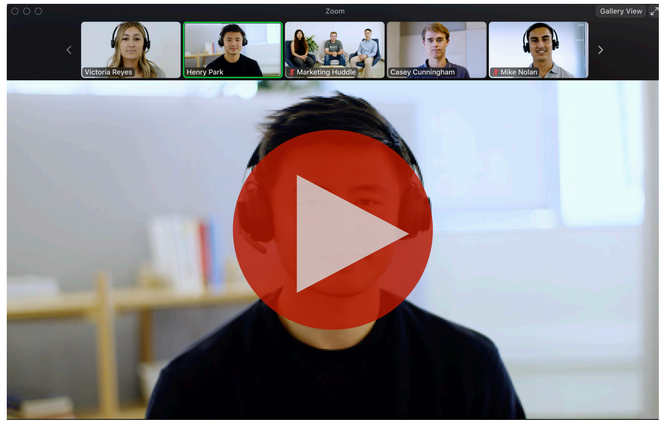
×

**Long interval**

×

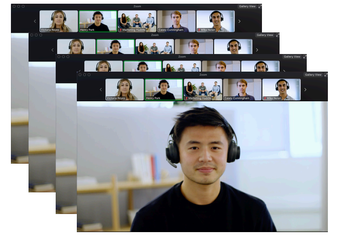
Non-verbal

Query




Result

key scenes

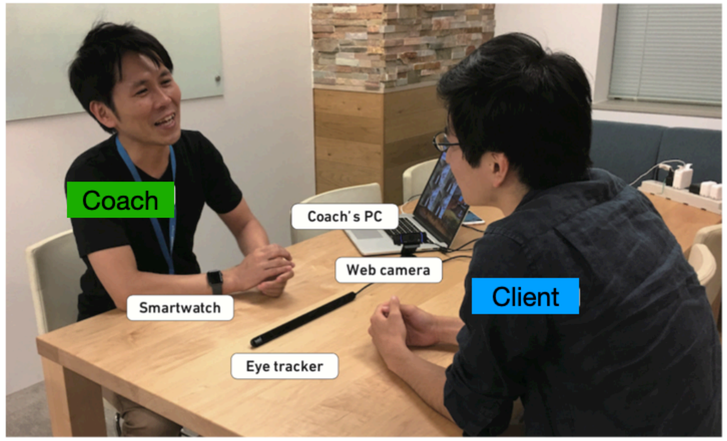


00:00:31 — 00:00:43

# Application of Multimodal Scene Search

Multimodal Query

Real-Time Intervention




Key Frame Extraction for Post Reflection

## Input videos

#1

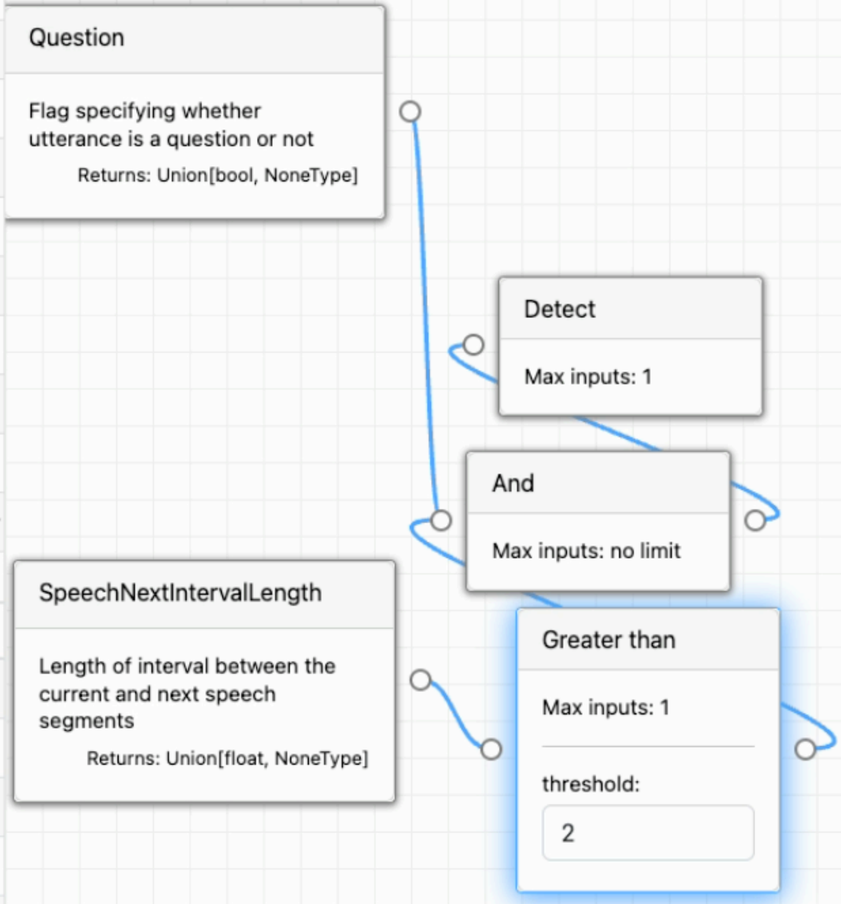

/root/datadrive/share/samples/providence/tmp/meetings/multimodal_demo_trim/anonymized_demo_video.mp4

## Features

ExpressionNeutral
ExpressionSadness
ExpressionSurprise
FillerNum
GazeX
GazeY
HeadPosePitch
HeadPoseRoll
HeadPoseYaw
Nod
NodCount
Overlap
Question
SpeechFrequencyMean
SpeechNextIntervalLength
SpeechNextIntervalSameSpeaker
SpeechPostprocessedText

## Operators

And
Or
Is none
Not
Greater than
Less than




Share | Generate code | Execute

## Results

#3: Tom


/root/datadrive/share/samples/providence/tmp/meetings/multimodal_demo_trim/anonymized_demo_video.mp4

- / 3 < >

#2: Tom


/root/datadrive/share/samples/providence/tmp/meetings/multimodal_demo_trim/anonymized_demo_video.mp4

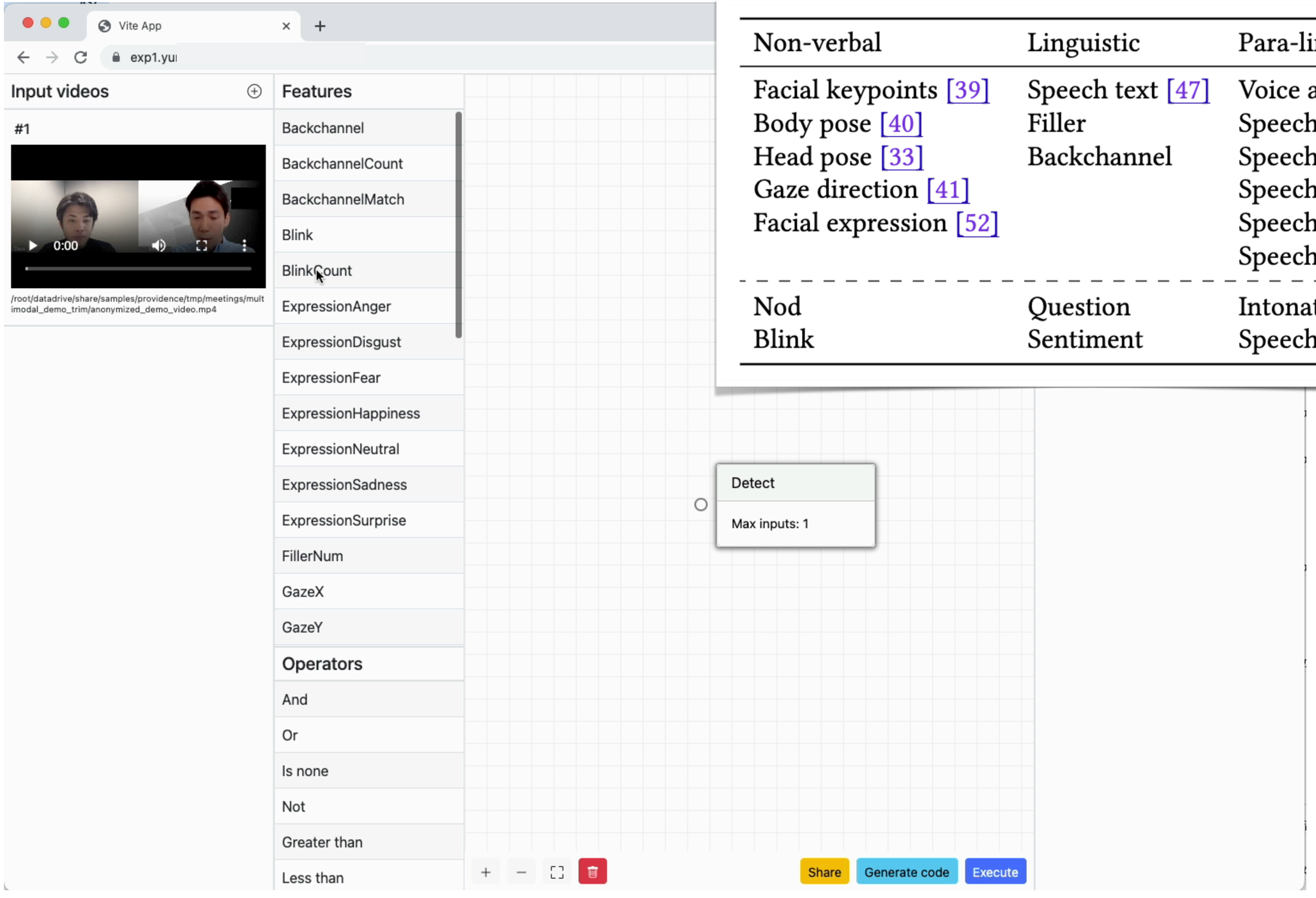
- / 5 < >

#1: Tom


/root/datadrive/share/samples/providence/tmp/meetings/multimodal_demo_trim/anonymized_demo_video.mp4

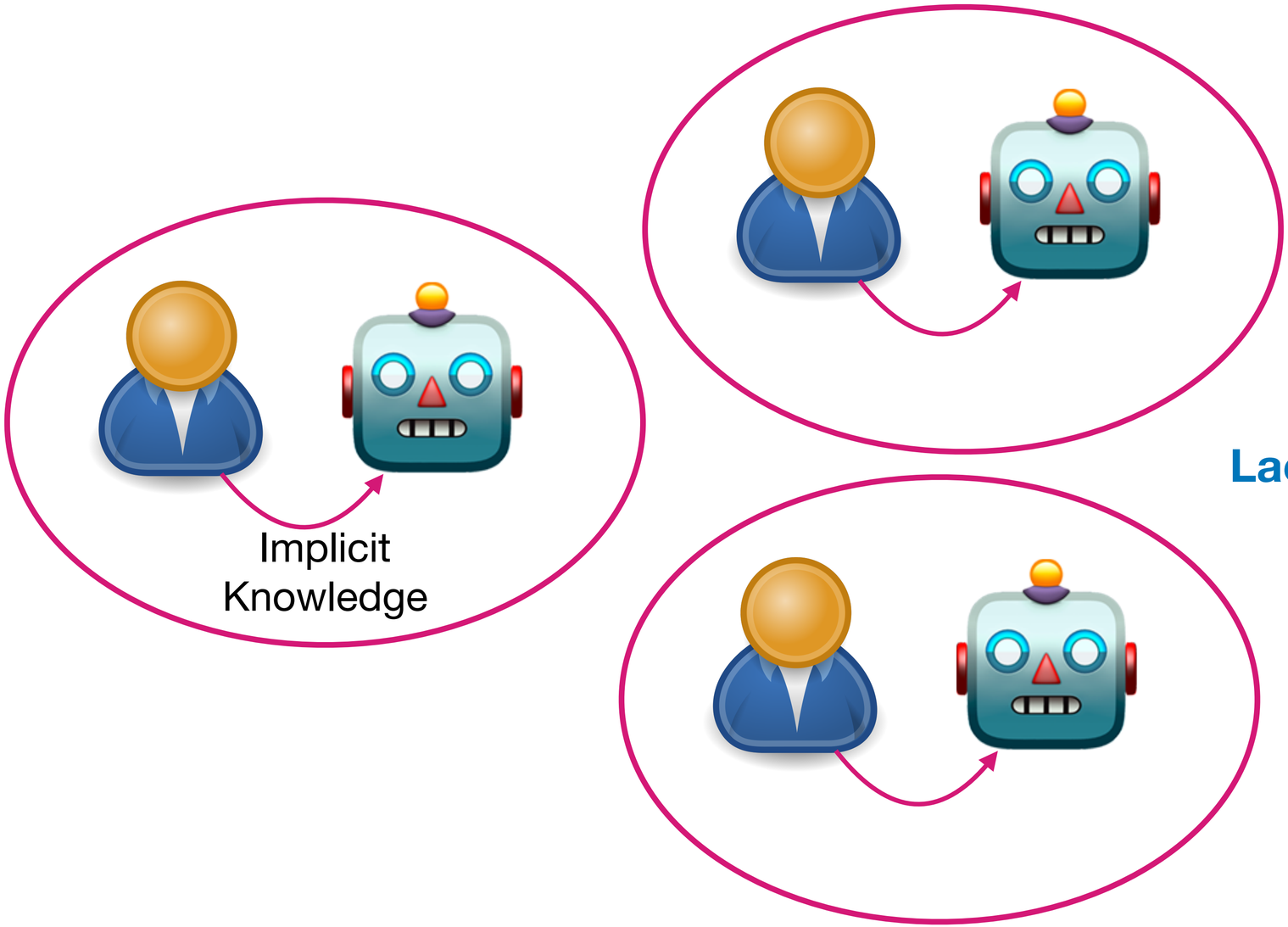
4 / 5 7:01 - 7:22 < >

ありがとうございます。具体的に仮に御社でご導入いただくと仮になった場合にですね、弊社ですとやっぱアカウントごとの料金



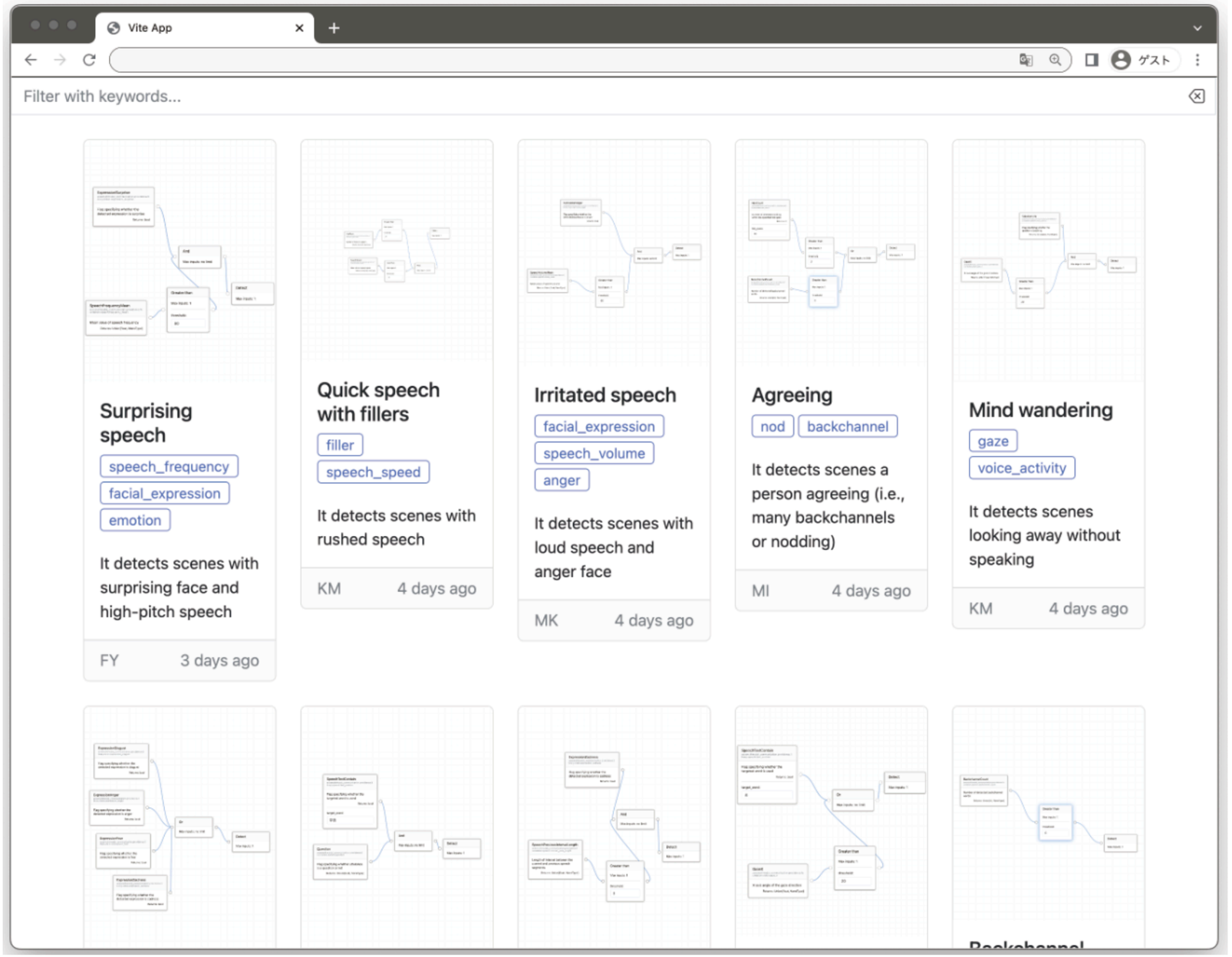

| Non-verbal | Linguistic | Para-linguistic |
|---|---|---|
| Facial keypoints [39] | Speech text [47] | Voice activity |
| Body pose [40] | Filler | Speech frequency |
| Head pose [33] | Backchannel | Speech volume |
| Gaze direction [41] | | Speech length |
| Facial expression [52] | | Speech speed |
| | | Speech interval |
| Nod | Question | Intonation |
| Blink | Sentiment | Speech overlap |

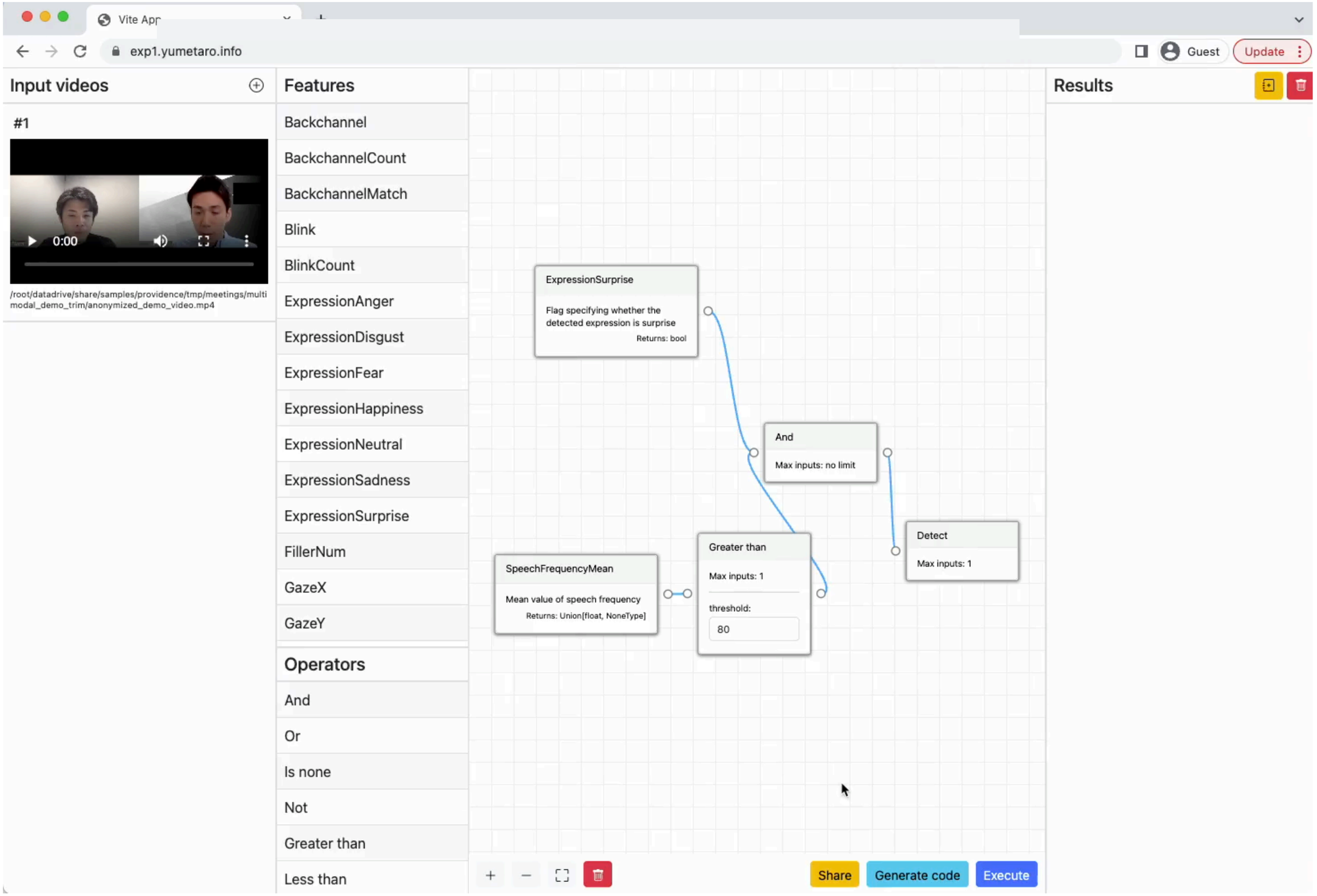
# Human-AI Interaction“s” at Workspace




**Lack of Knowledge Transferability**

# Knowledge-Share Repository



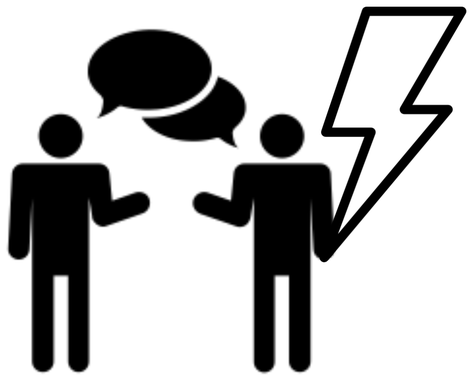

Vite App
exp1.yumetaro.info
Guest
Update
Input videos
#1
0:00
/root/datadrive/share/samples/providence/tmp/meetings/multimodal_demo_trim/anonymized_demo_video.mp4
Features
Backchannel
BackchannelCount
BackchannelMatch
Blink
BlinkCount
ExpressionAnger
ExpressionDisgust
ExpressionFear
ExpressionHappiness
ExpressionNeutral
ExpressionSadness
ExpressionSurprise
FillerNum
GazeX
GazeY
Operators
And
Or
Is none
Not
Greater than
Less than
ExpressionSurprise
Flag specifying whether the detected expression is surprise
Returns: bool
And
Max inputs: no limit
Detect
Max inputs: 1
Greater than
Max inputs: 1
threshold:
80
SpeechFrequencyMean
Mean value of speech frequency
Returns: Union[float, NoneType]
Results
Share
Generate code
Execute

# How ConverSearch Affects Practitioners' Workflow

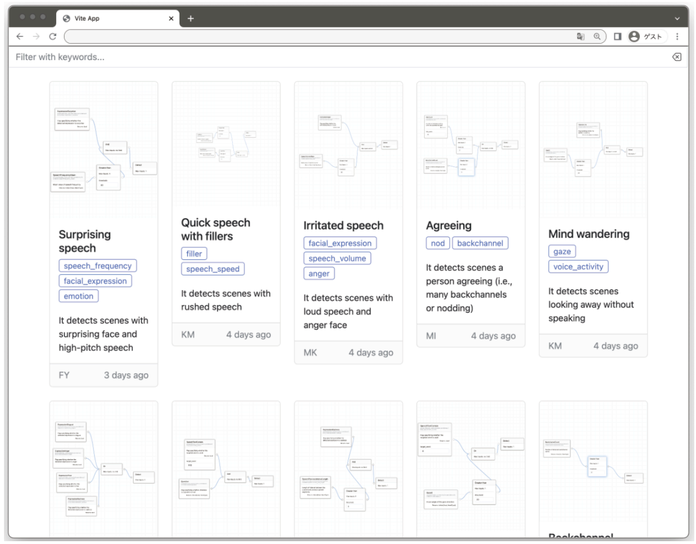
11 practitioners, 3 weeks

**objectively used for multiple conversations**

**ideate tools for developing conversation skills**



# Toward Socially Intelligent AI: Ment'Al'ist




?


Accumulation of Implicit Knowledge on Human Behavior

# Summary

## Support human development practitioners with AI technology


- How can we blend human and AI in human development scenes?
  Chatbot can be used as a supplement to face-to-face session
- What is the unique strength of human practitioners and AI?
  AI can observe behavior cues; human can make judgement
- What is needed to accelerate human-AI interaction in this domain?
  - Tool to easily code implicit knowledge on human behavior
  - Knowledge share repository for aligning knowledge among experts

# Acknowledgement


- Hiromu Yakura (Centers for Humans and Machines at the Max-Planck Institute for Human Development)
- ACES, Inc.
- Teambox, Inc.