

Towards Ment'Al'ist: Augmenting Human Development Practitioners through Behavioral Sensing

Riku Arakawa (Ph.D. student, CMU HCII)

<https://rikky0611.github.io/>

Human Development



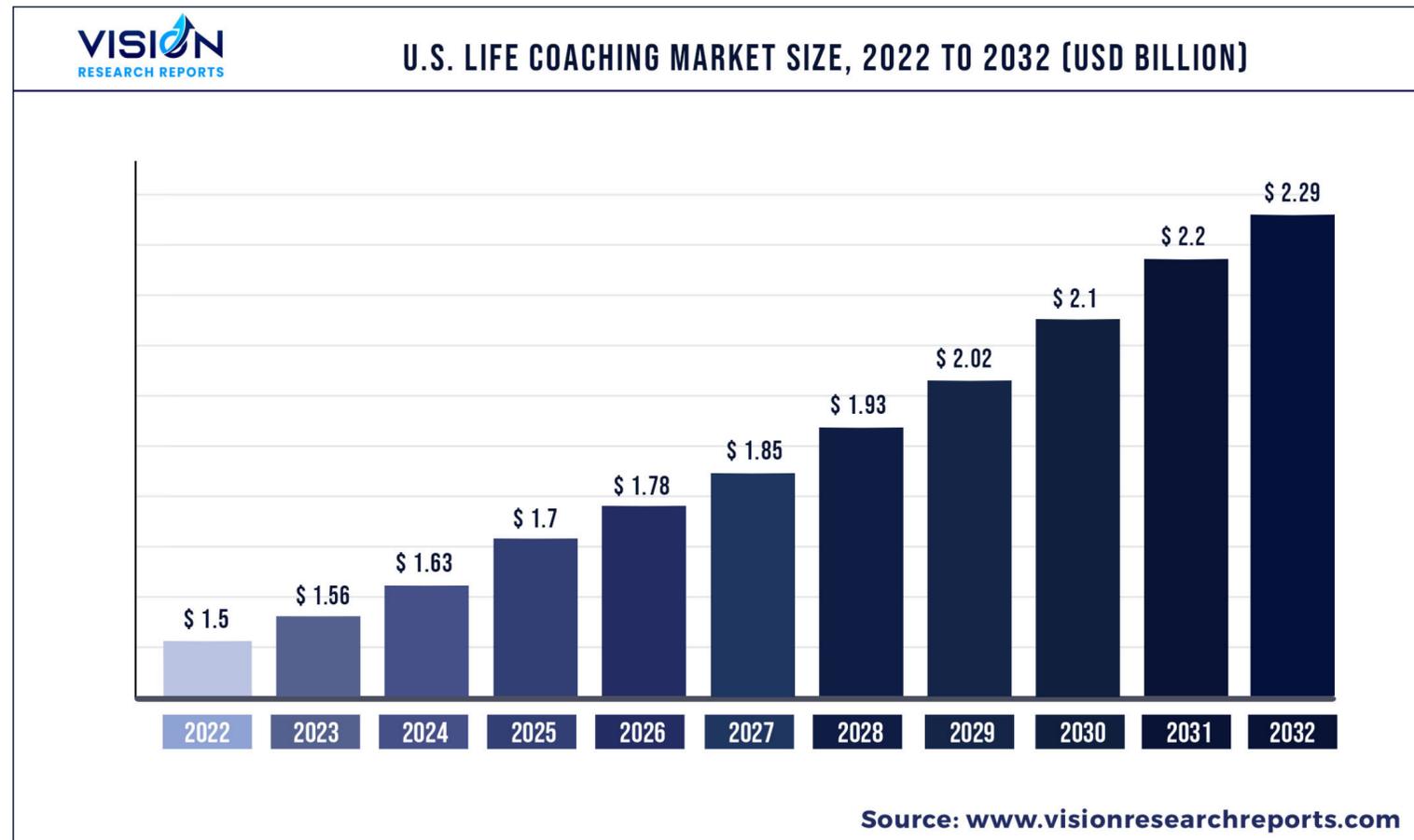
Process of improving well-being and capabilities

Reflection through Dialogue with Practitioners



coaching, counseling, therapy, etc.

Lack of Practitioners



40% reporting inadequate experience

Increasing demands and training barriers

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.

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

1on1
session



- LLM-based chatbots are effective for
 - generating action plans and answering immediate questions
 - acknowledging users' day-to-day actions

- Human coaches are critical for
 - posing reflective, critical questions
 - setting appropriately challenging goals

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.

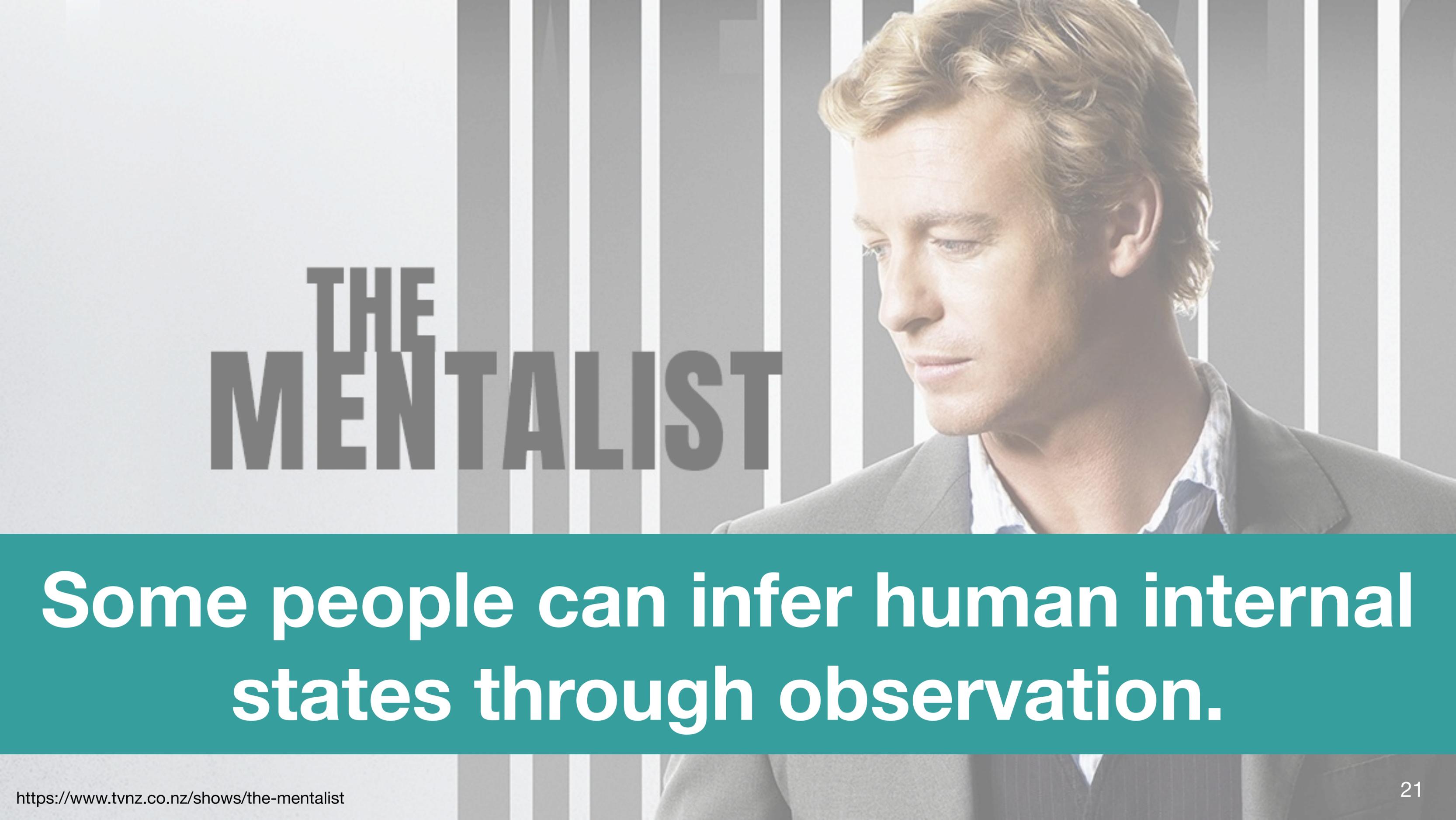
Behavior analysis is a critical skill

Nonverbal Behavior and Internal States

Internal states
ex. impatient, nervous



**“spontaneous & unregulated
expression”**



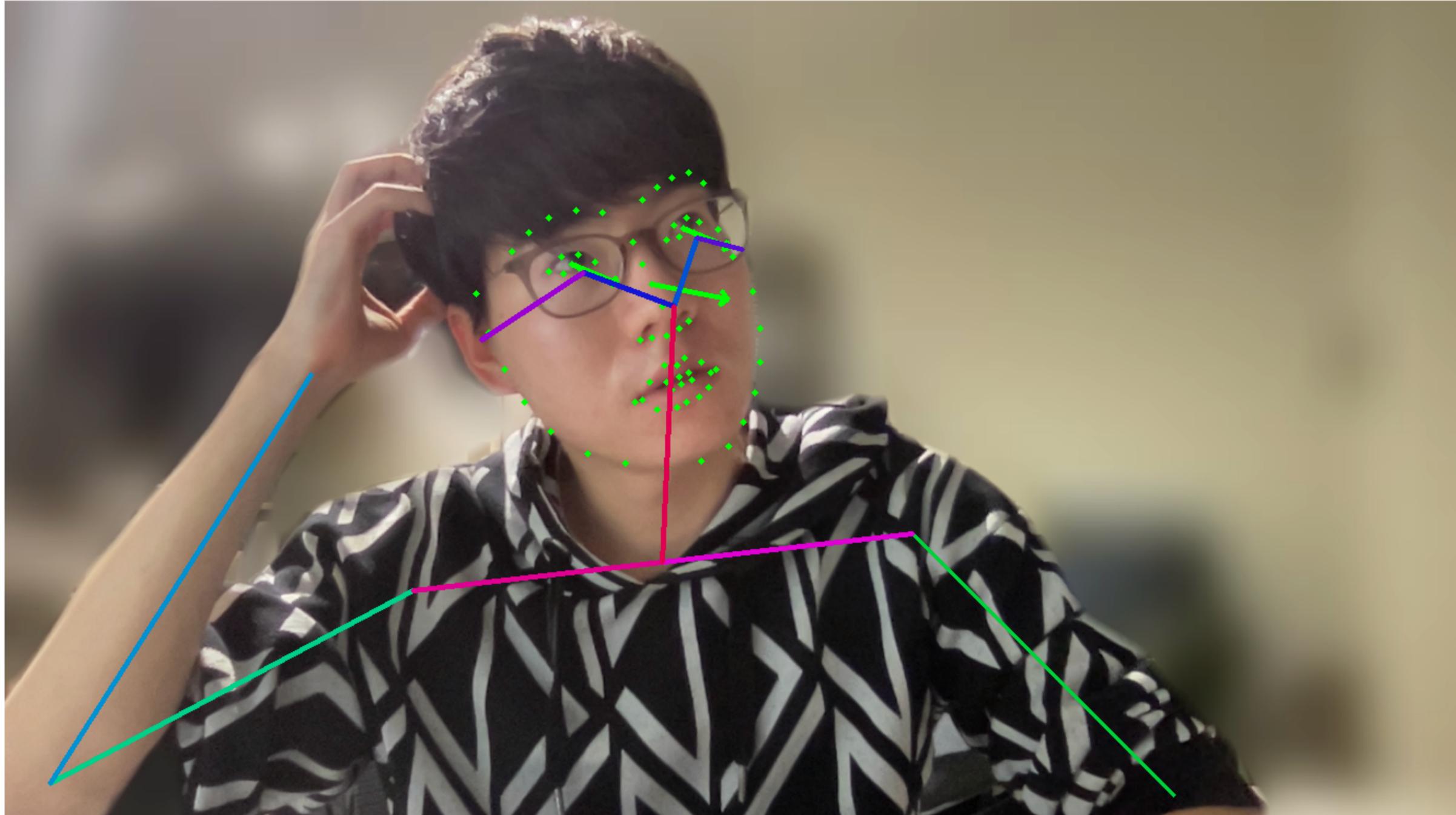
THE MENTALIST

Some people can infer human internal states through observation.

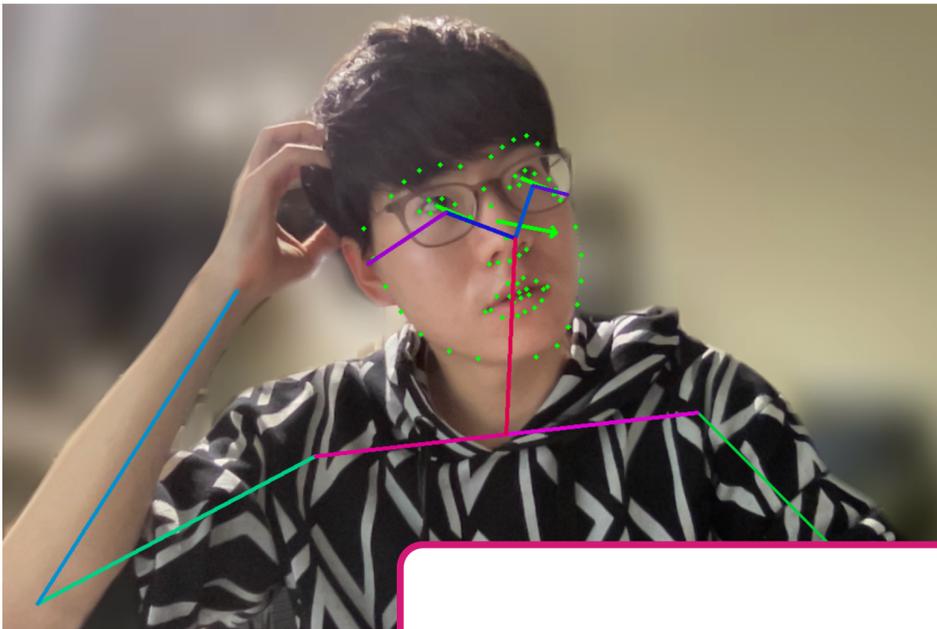
Research Question

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

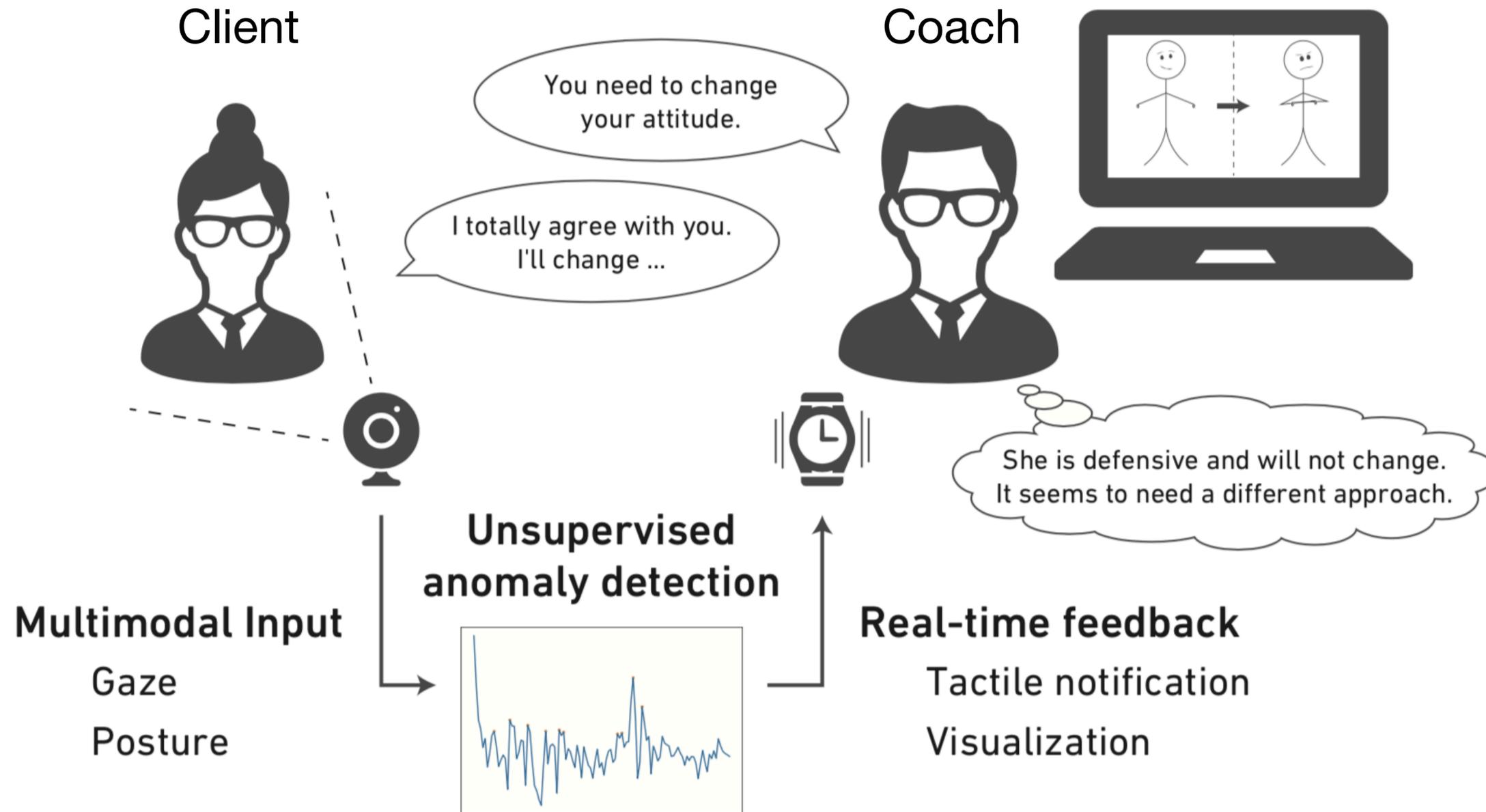
Pros: Stable performance

Cons: Not good at dealing with context



Observation for AI, judgment for human

Proposed Framework: REsCUE



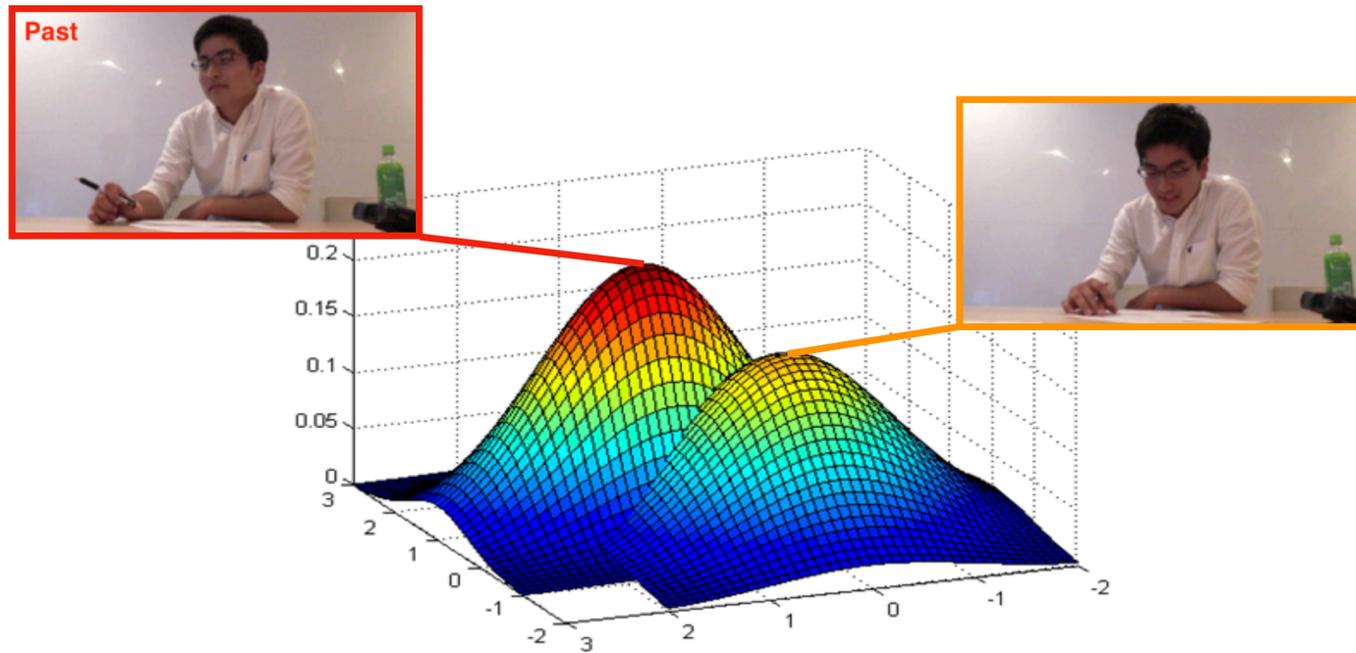
**We show a simplified example
to illustrate **REsCUE**'s assist for coaches.**



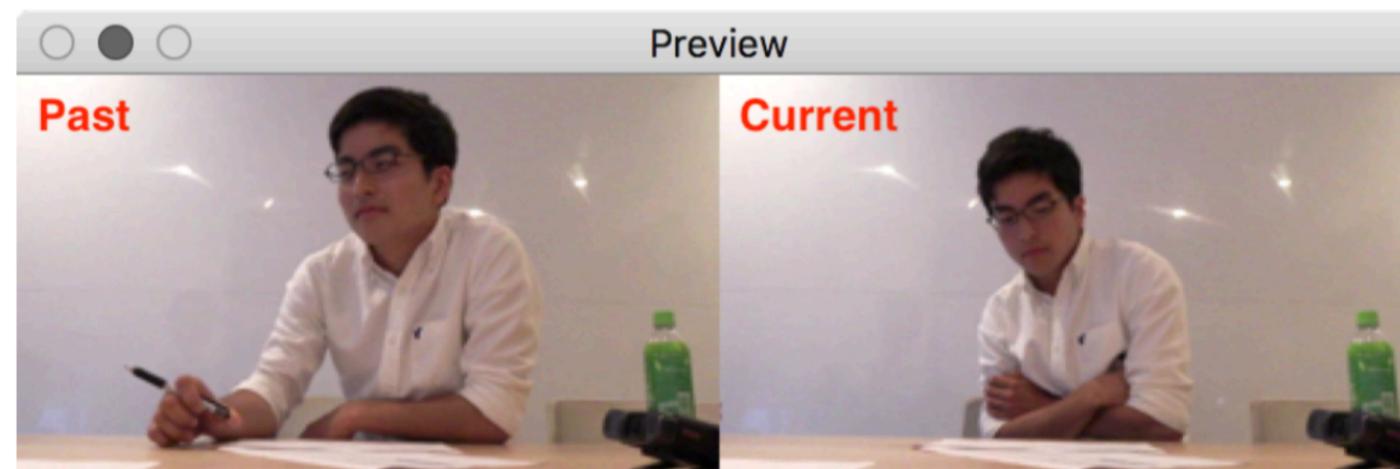
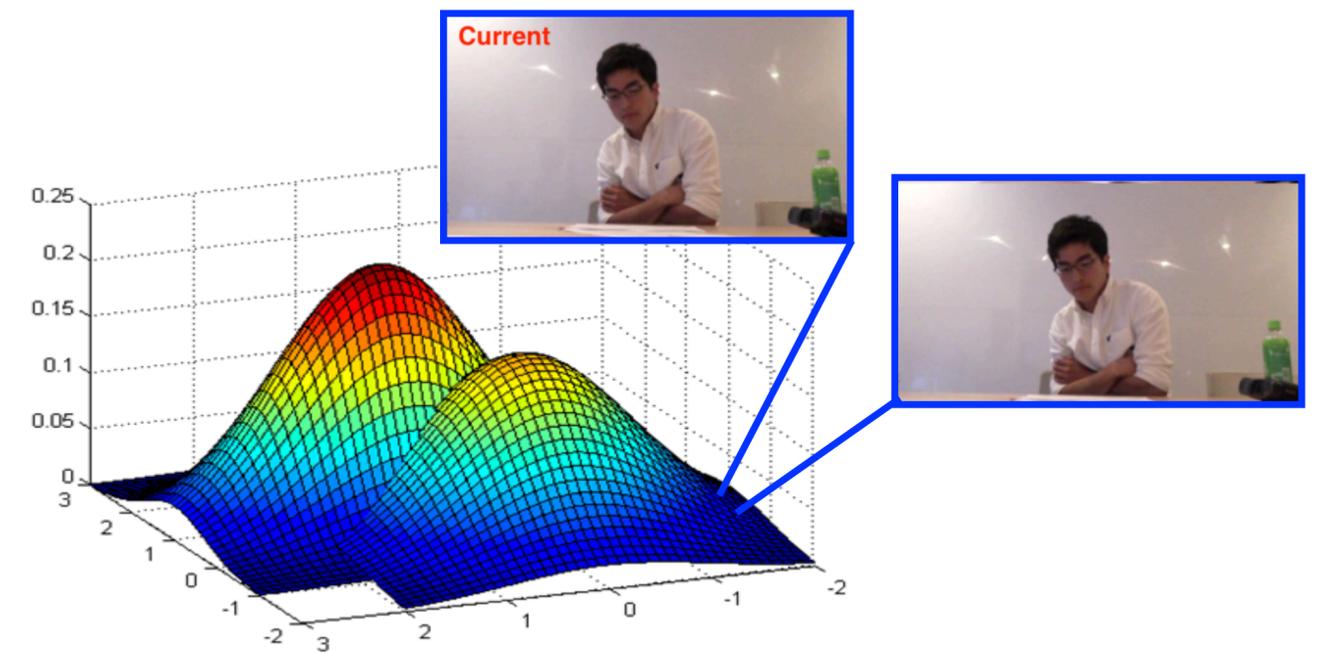
Coach: Do you agree with your decision?

Visualize Anomaly Detection with GMM

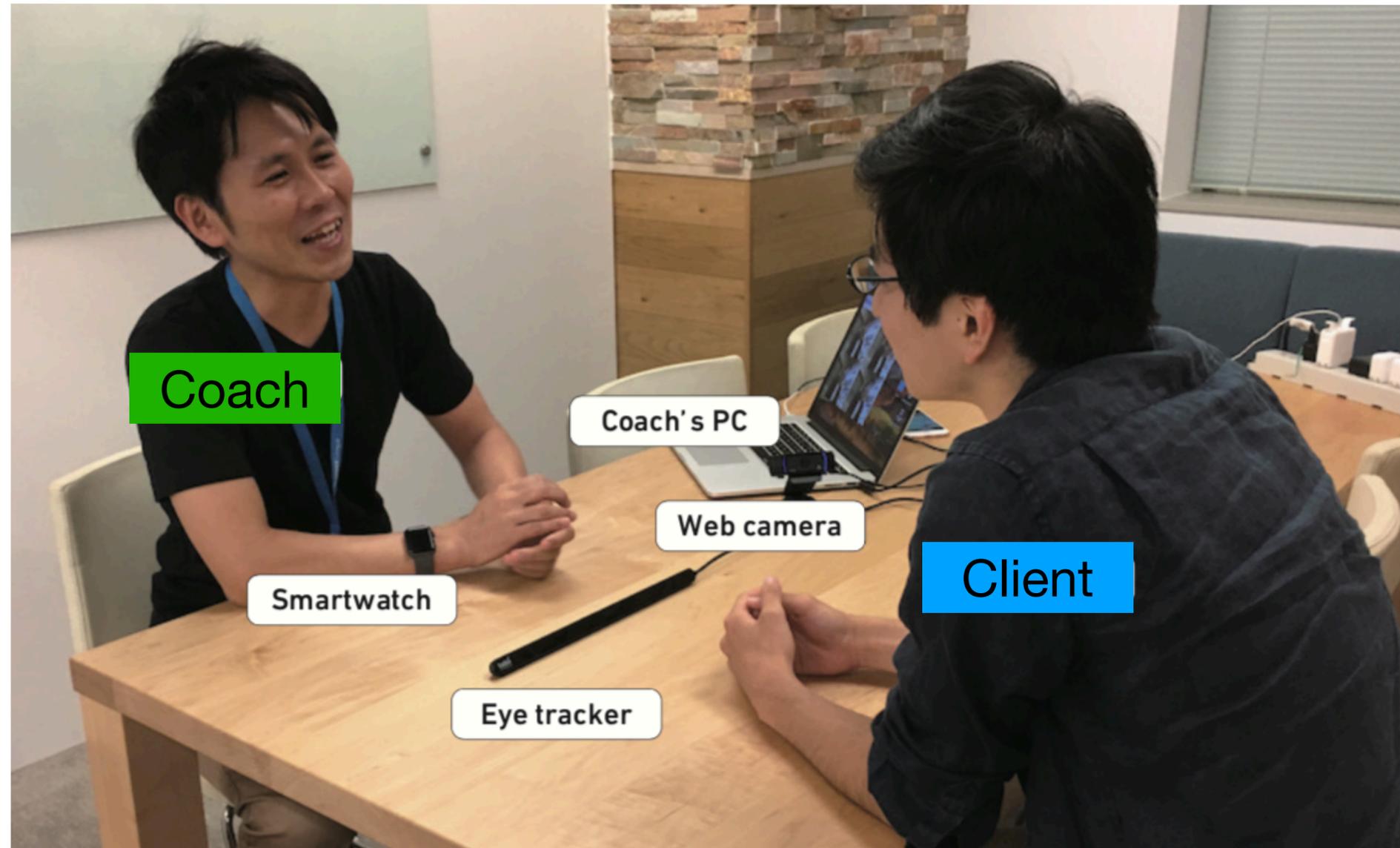
Past Frames



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.

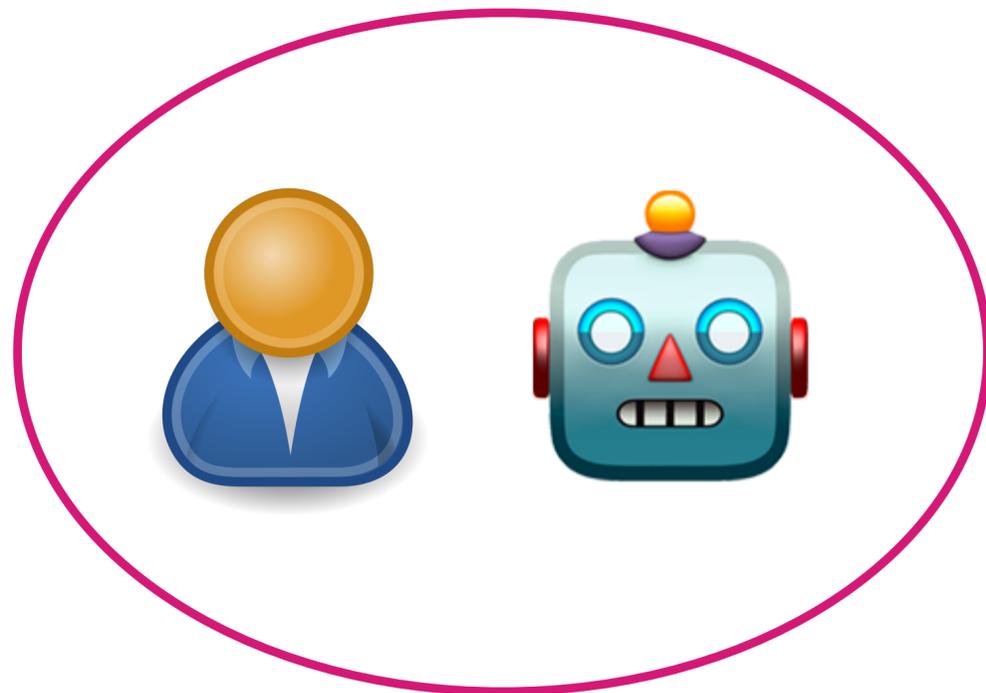
Feedback on REsCUE

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.

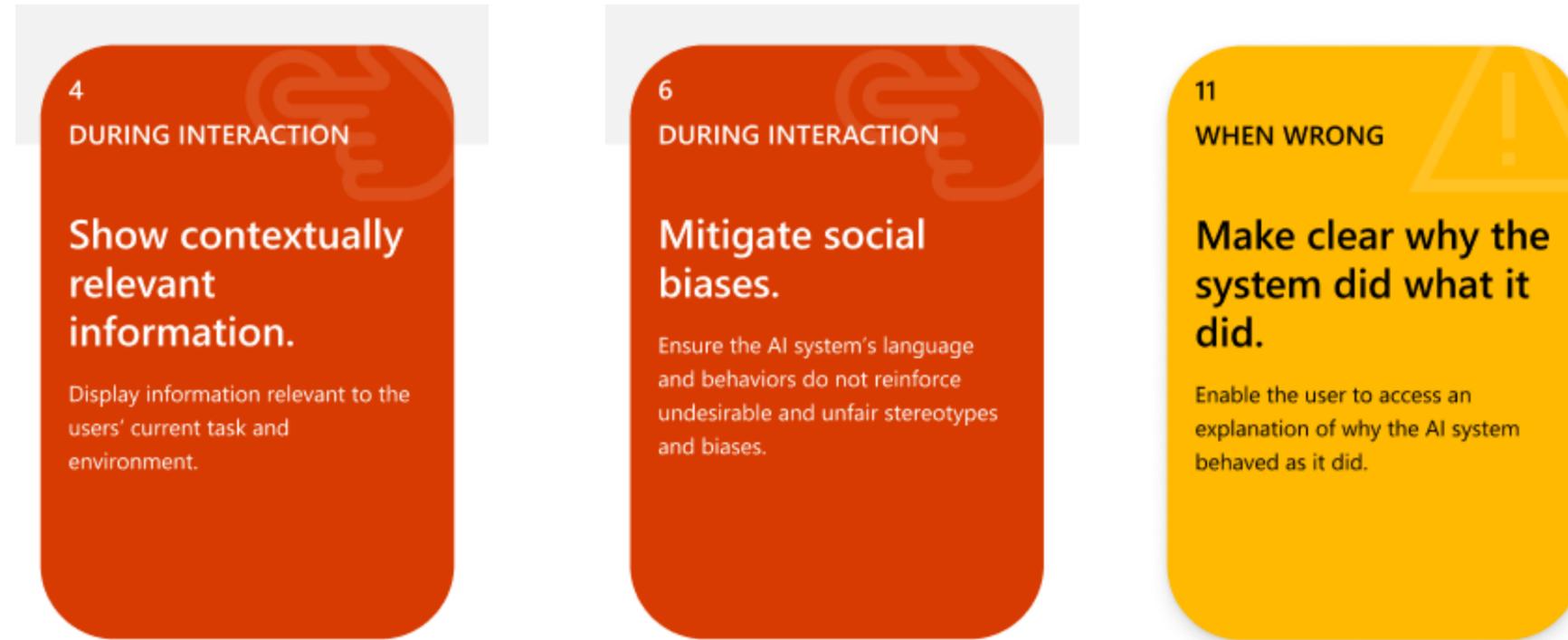
Human-AI Interaction



Guidelines for Human-AI Interaction

<p>1 INITIALLY</p> <p>Make clear what the system can do.</p> <p>Help the users understand what the AI system is capable of doing.</p>	<p>2 INITIALLY</p> <p>Make clear how well the system can do what it can do.</p> <p>Help the user understand how often the AI system may make mistakes.</p>	<p>3 DURING INTERACTION</p> <p>Time services based on context.</p> <p>Time when to act or interrupt based on the user's current task and environment.</p>	<p>4 DURING INTERACTION</p> <p>Show contextually relevant information.</p> <p>Display information relevant to the users' current task and environment.</p>	<p>5 DURING INTERACTION</p> <p>Match relevant social norms.</p> <p>Ensure the experience is delivered in a way that users would expect, given their social and cultural context.</p>	<p>6 DURING INTERACTION</p> <p>Mitigate social biases.</p> <p>Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.</p>	
<p> INITIALLY</p>		<p> DURING INTERACTION</p>				
<p>7 WHEN WRONG</p> <p>Support efficient invocation.</p> <p>Make it easy to invoke or request the AI system's services when needed.</p>	<p>8 WHEN WRONG</p> <p>Support efficient dismissal.</p> <p>Make it easy to dismiss or ignore undesired system services.</p>	<p>9 WHEN WRONG</p> <p>Support efficient correction.</p> <p>Make it easy to edit, refine, or recover when the AI system is wrong.</p>	<p>10 WHEN WRONG</p> <p>Scope services when in doubt.</p> <p>Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.</p>	<p>11 WHEN WRONG</p> <p>Make clear why the system did what it did.</p> <p>Enable the user to access an explanation of why the AI system behaved as it did.</p>		
<p> WHEN WRONG</p>						
<p>12 OVER TIME</p> <p>Remember recent interactions.</p> <p>Maintain short-term memory and allow the user to make efficient references to that memory.</p>	<p>13 OVER TIME</p> <p>Learn from user behavior.</p> <p>Personalize the user's experience by learning from their actions over time.</p>	<p>14 OVER TIME</p> <p>Update and adapt cautiously.</p> <p>Limit disruptive changes when updating and adapting the AI system's behaviors.</p>	<p>15 OVER TIME</p> <p>Encourage granular feedback.</p> <p>Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.</p>	<p>16 OVER TIME</p> <p>Convey the consequences of user actions.</p> <p>Immediately update or convey how user actions will impact future behaviors of the AI system.</p>	<p>17 OVER TIME</p> <p>Provide global controls.</p> <p>Allow the user to globally customize what the AI system monitors and how it behaves.</p>	<p>18 OVER TIME</p> <p>Notify users about changes.</p> <p>Inform the user when the AI system adds or updates its capabilities.</p>
<p> OVER TIME</p>						

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 🏆

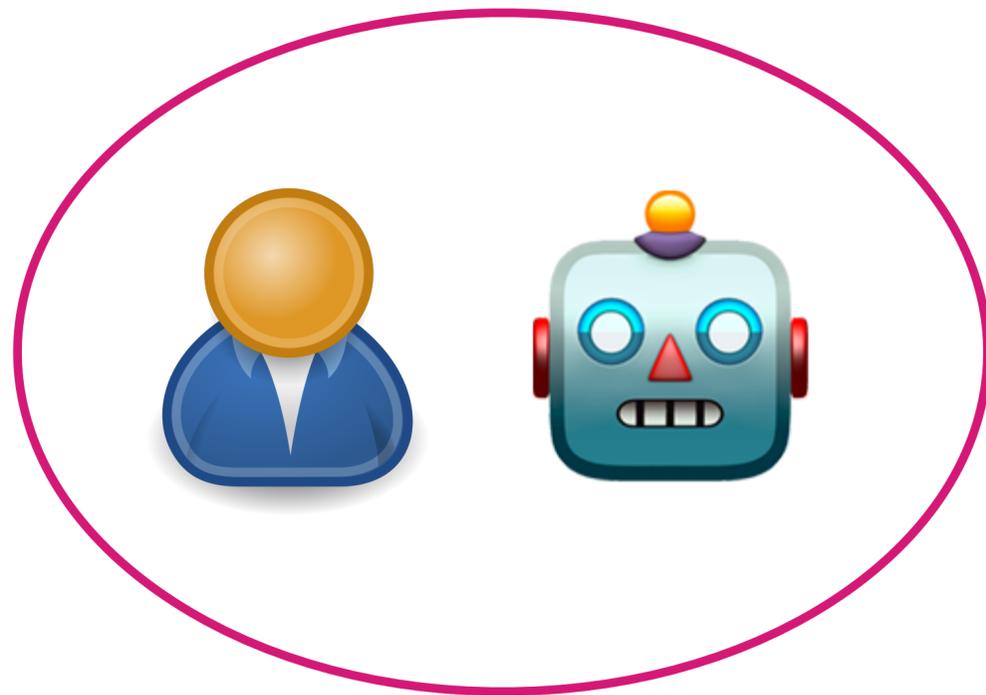


SaaS deployment (> 1M recordings)
CHI 2025, case study

Practitioners want to **customize the detection algorithm**

- ignore “touching glasses”
- capture “struggling with question-answering”

Human-AI Interaction



Guidelines for Human-AI Interaction

1 INITIALLY

Make clear what the system can do.

Help the users understand what the AI system is capable of doing.

2 INITIALLY

Make clear how well the system can do what it can do.

Help the user understand how often the AI system may make mistakes.

3 DURING INTERACTION

Time services based on context.

Time when to act or interrupt based on the user's current task and environment.

4 DURING INTERACTION

Show contextually relevant information.

Display information relevant to the users' current task and environment.

5 DURING INTERACTION

Match relevant social norms.

Ensure the experience is delivered in a way that users would expect, given their social and cultural context.

6 DURING INTERACTION

Mitigate social biases.

Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.

7 WHEN WRONG

Support efficient invocation.

Make it easy to invoke or request the AI system's services when needed.

8 WHEN WRONG

Support efficient dismissal.

Make it easy to dismiss or ignore undesired system services.

9 WHEN WRONG

Support efficient correction.

Make it easy to edit, refine, or recover when the AI system is wrong.

10 WHEN WRONG

Scope services when in doubt.

Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.

11 WHEN WRONG

Make clear why the system did what it did.

Enable the user to access an explanation of why the AI system behaved as it did.

12 OVER TIME

Remember recent interactions.

Maintain short-term memory and allow the user to make efficient references to that memory.

13 OVER TIME

Learn from user behavior.

Personalize the user's experience by learning from their actions over time.

14 OVER TIME

Update and adapt cautiously.

Limit disruptive changes when updating and adapting the AI system's behaviors.

15 OVER TIME

Encourage granular feedback.

Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.

16 OVER TIME

Convey the consequences of user actions.

Immediately update or convey how user actions will impact future behaviors of the AI system.

17 OVER TIME

Provide global controls.

Allow the user to globally customize what the AI system monitors and how it behaves.

18 OVER TIME

Notify users about changes.

Inform the user when the AI system adds or updates its capabilities.

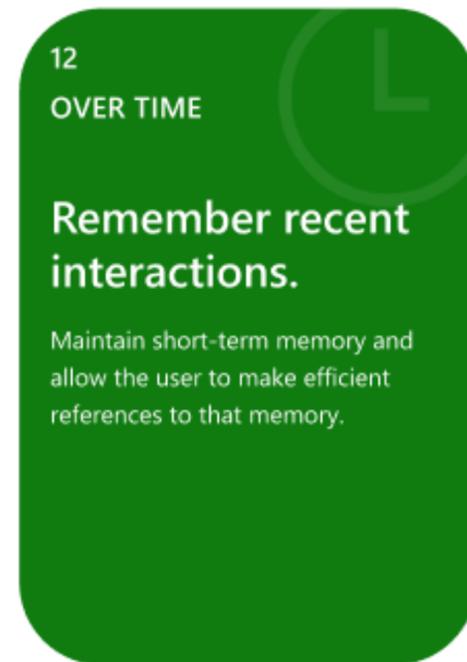
INITIALLY | **DURING INTERACTION** | **WHEN WRONG** | **OVER TIME**

Human-AI Interaction “Over Time”

12
OVER TIME

Remember recent interactions.

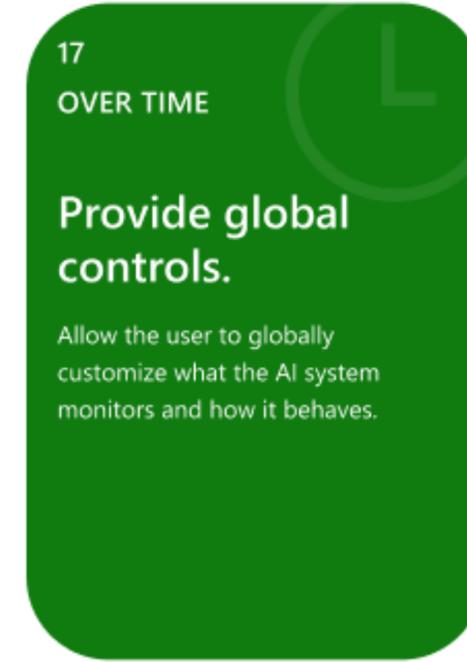
Maintain short-term memory and allow the user to make efficient references to that memory.



17
OVER TIME

Provide global controls.

Allow the user to globally customize what the AI system monitors and how it behaves.



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

×

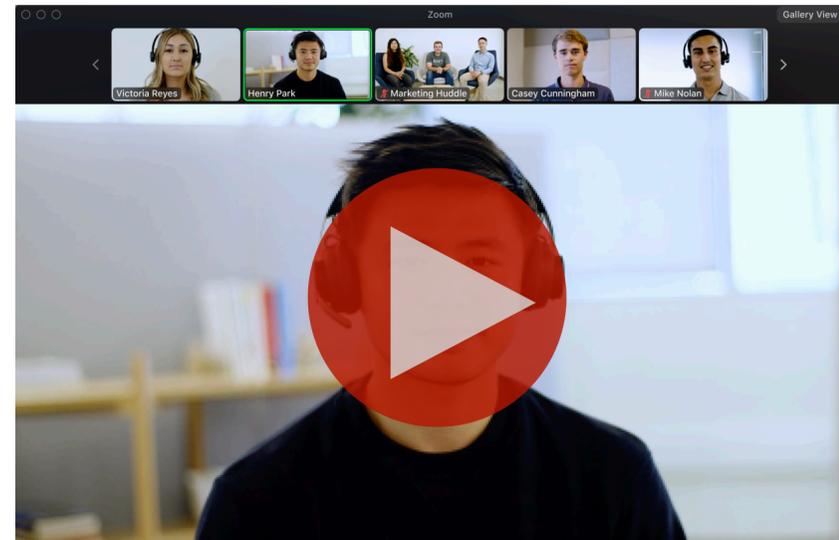
Paralinguistic



Query

×

Non-verbal



Result

key scenes



00:00:31 — 00:00:43

Example: “Struggles with Question-Answering”

Is questioned?

×

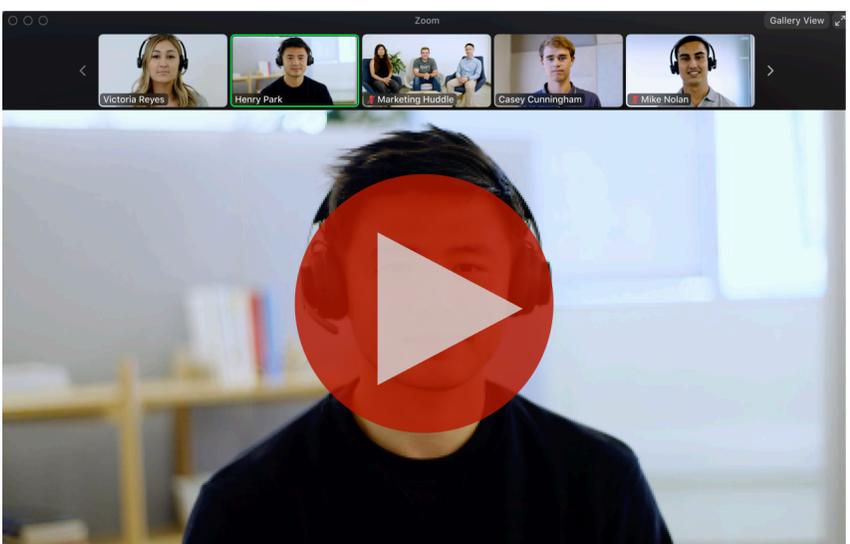
Long interval



Query

×

Non-verbal



Result

key scenes



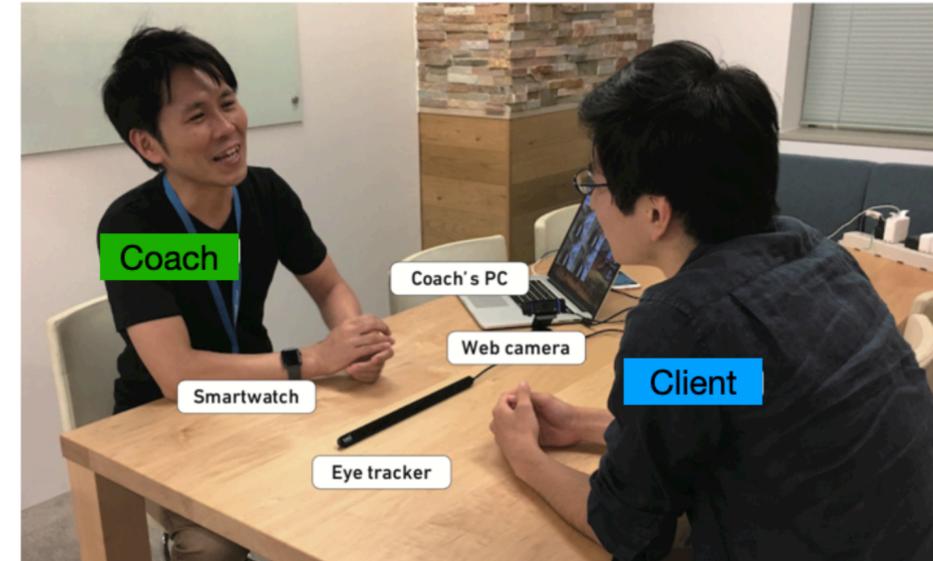
00:00:31 — 00:00:43

Application of Multimodal Scene Search

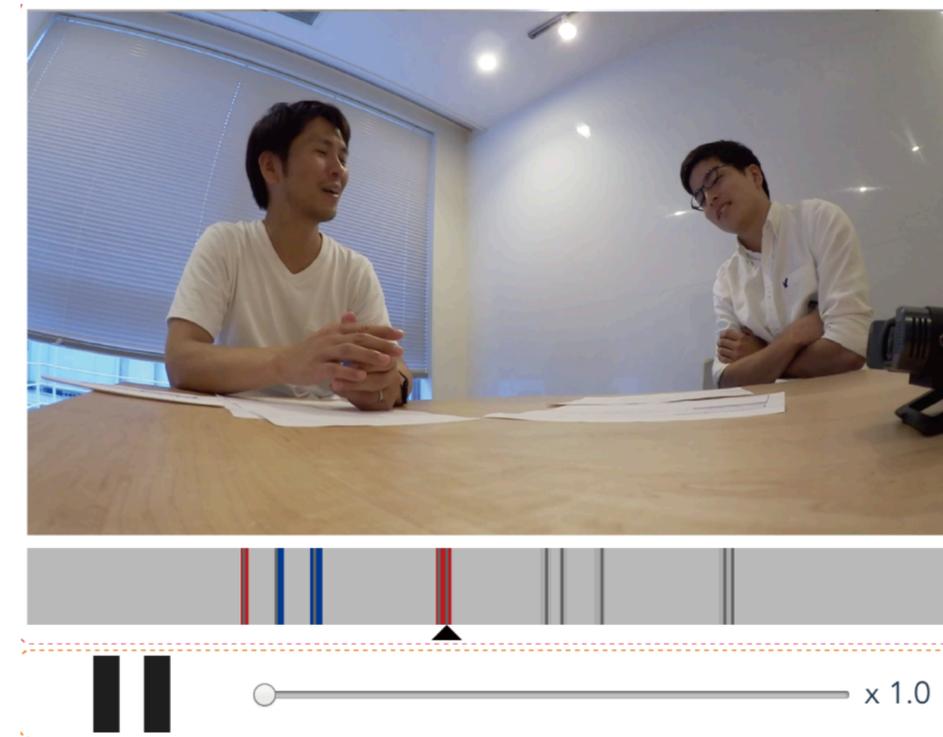
Multimodal
Query



Real-Time
Intervention



Key Frame Extraction
for Post Reflection



Input videos



#1

0:00 / 12:43

/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

Question

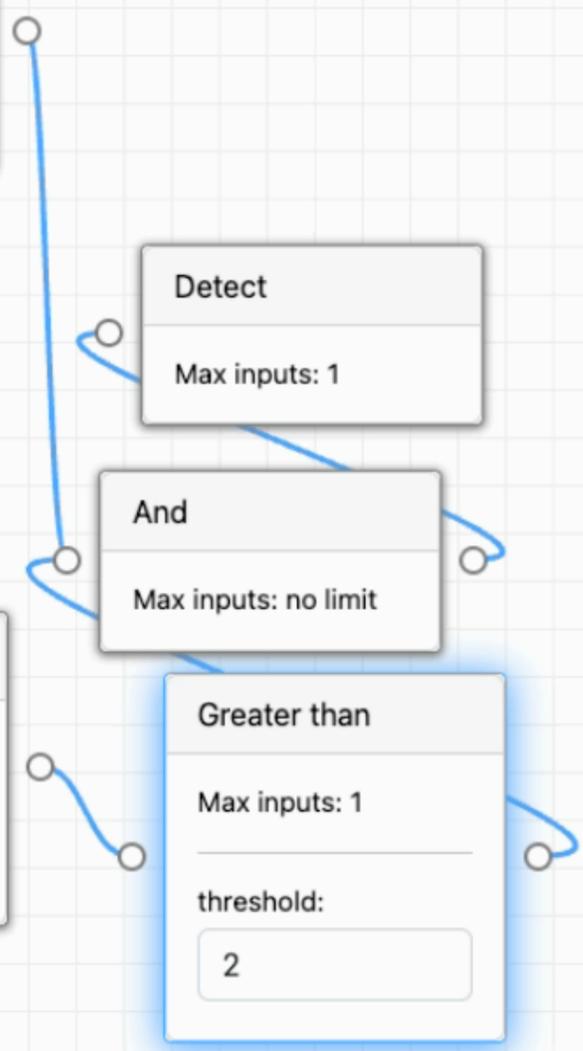
Flag specifying whether utterance is a question or not

Returns: Union[bool, NoneType]

SpeechNextIntervalLength

Length of interval between the current and next speech segments

Returns: Union[float, NoneType]



Results

#3: Tom

0:00 / 12:43

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

- / 3

#2: Tom

0:00 / 12:43

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

- / 5

#1: Tom

7:01 / 12:43

/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なつた場合にですね、弊社ですとやっぱアカウントごとの料金



Share Generate code Execute

Vite App exp1.yui

Input videos #1



/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

Buttons: +, -, [], [X]

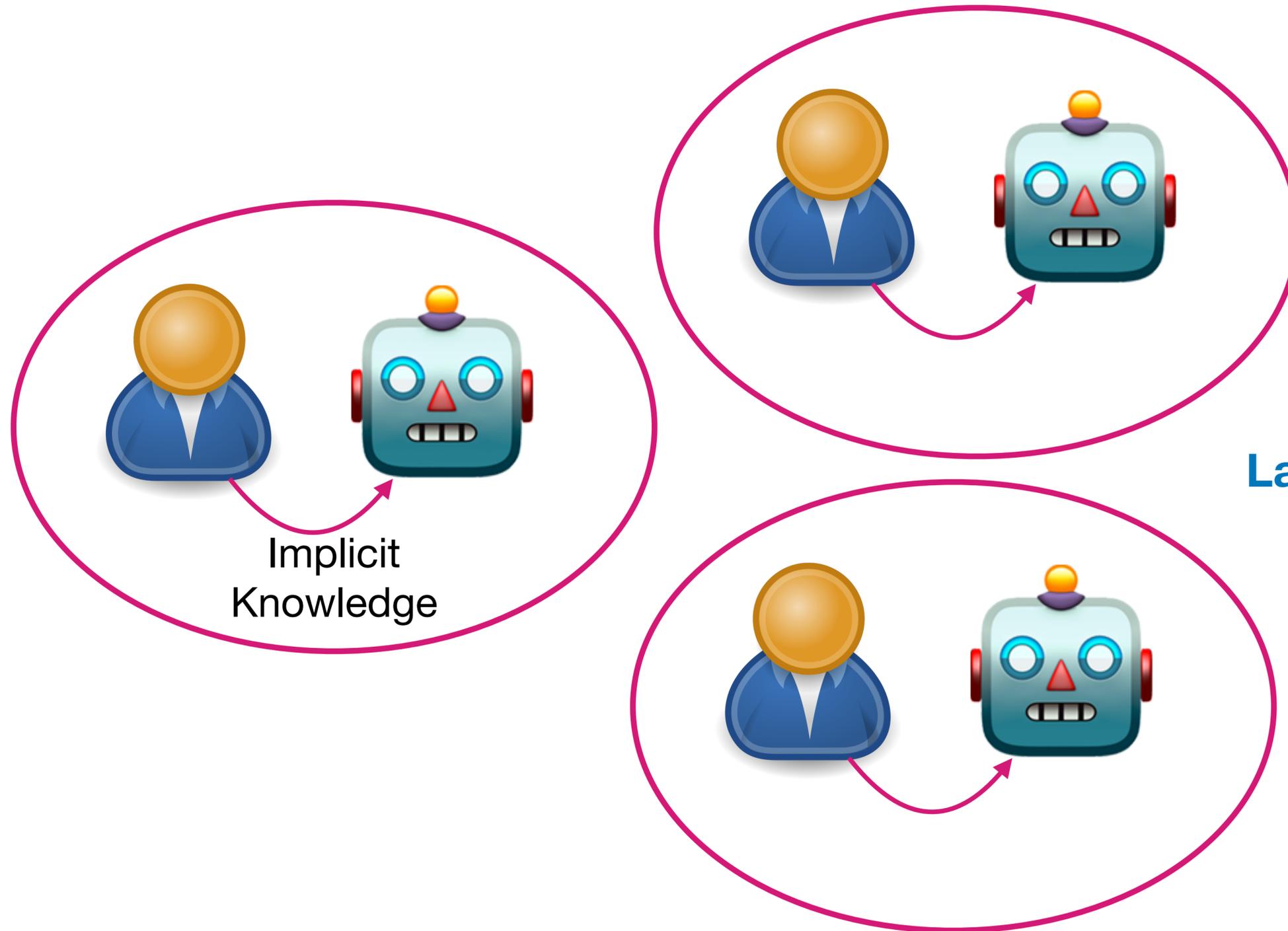
Non-verbal	Linguistic	Para-linguistic
Facial keypoints [39]	Speech text [47]	Voice activity
Body pose [40]	Filler	Speech frequency
Head pose [33]	Backchannel	Speech volume
Gaze direction [41]		Speech length
Facial expression [52]		Speech speed
		Speech interval

Nod	Question	Intonation
Blink	Sentiment	Speech overlap

Detect

Max inputs: 1

Human-AI Interaction“s” at Workspace



Lack of Knowledge Transferability

Knowledge-Share Repository

The screenshot shows a web browser window with a search bar at the top containing the text "Filter with keywords...". Below the search bar is a grid of five knowledge cards. Each card features a flowchart diagram at the top, a title, a list of keywords in blue pill-shaped buttons, a descriptive sentence, and a footer with a user initials and a timestamp.

- Card 1: Surprising speech**
 - Keywords: `speech_frequency`, `facial_expression`, `emotion`
 - Description: "It detects scenes with surprising face and high-pitch speech"
 - Footer: FY 3 days ago
- Card 2: Quick speech with fillers**
 - Keywords: `filler`, `speech_speed`
 - Description: "It detects scenes with rushed speech"
 - Footer: KM 4 days ago
- Card 3: Irritated speech**
 - Keywords: `facial_expression`, `speech_volume`, `anger`
 - Description: "It detects scenes with loud speech and anger face"
 - Footer: MK 4 days ago
- Card 4: Agreeing**
 - Keywords: `nod`, `backchannel`
 - Description: "It detects scenes a person agreeing (i.e., many backchannels or nodding)"
 - Footer: MI 4 days ago
- Card 5: Mind wandering**
 - Keywords: `gaze`, `voice_activity`
 - Description: "It detects scenes looking away without speaking"
 - Footer: KM 4 days ago

Below the first row, the top of a second row of five cards is visible, including a card titled "Backchannel".

Vite App
exp1.yumetaro.info

Guest Update

Input videos

#1



0:00

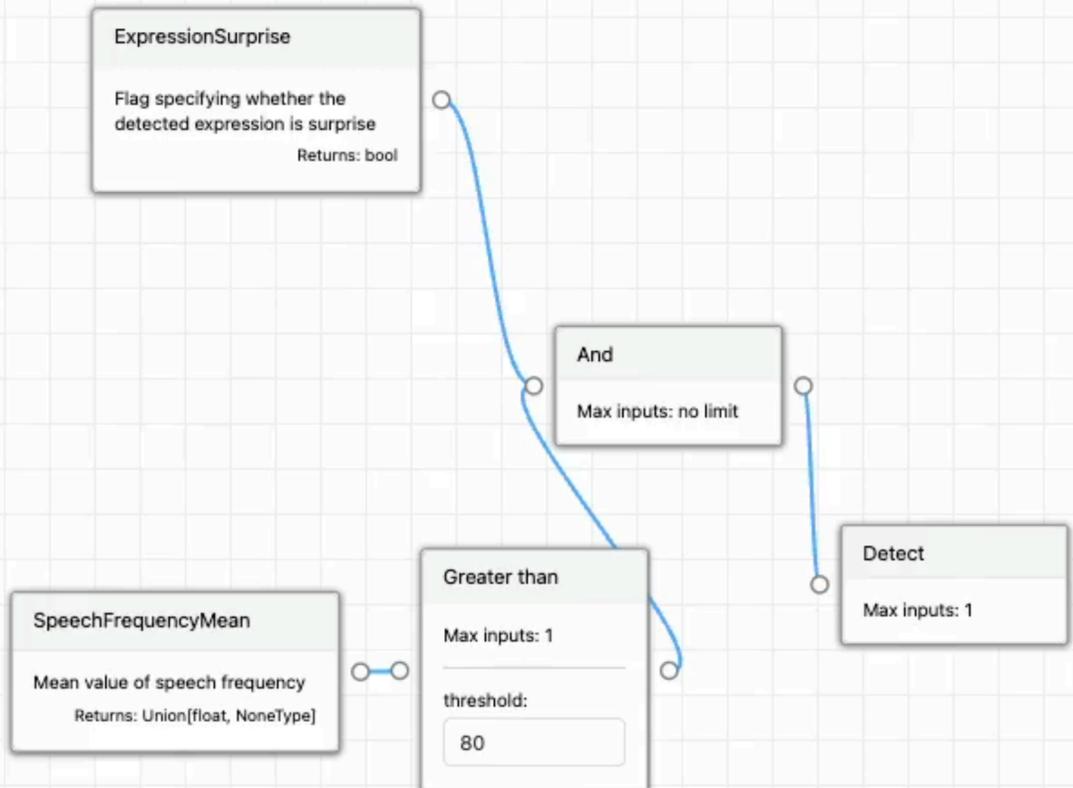
/root/datadrive/share/samples/providence/tmp/meetings/multi-modal_demo_trim/anonymized_demo_video.mp4

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



```

graph TD
    A[SpeechFrequencyMean  
Mean value of speech frequency  
Returns: Union[float, NoneType]] --> B[Greater than  
Max inputs: 1  
threshold: 80]
    B --> C[And  
Max inputs: no limit]
    D[ExpressionSurprise  
Flag specifying whether the detected expression is surprise  
Returns: bool] --> C
    C --> E[Detect  
Max inputs: 1]
    
```

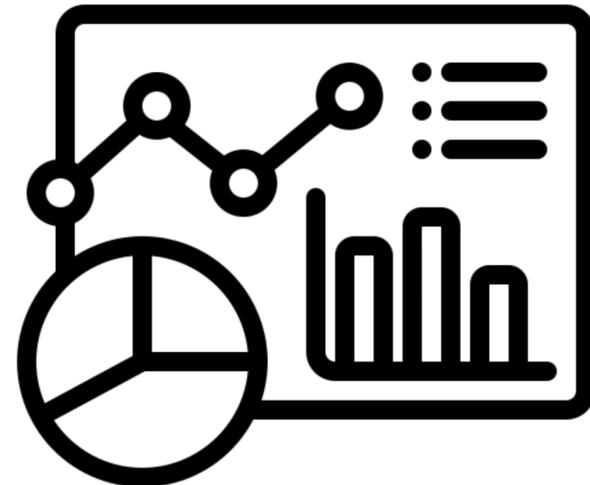
Results

+ - [] [X] Share Generate code Execute

How ConverSearch Affects Practitioners' Workflow

11 practitioners, 3 weeks

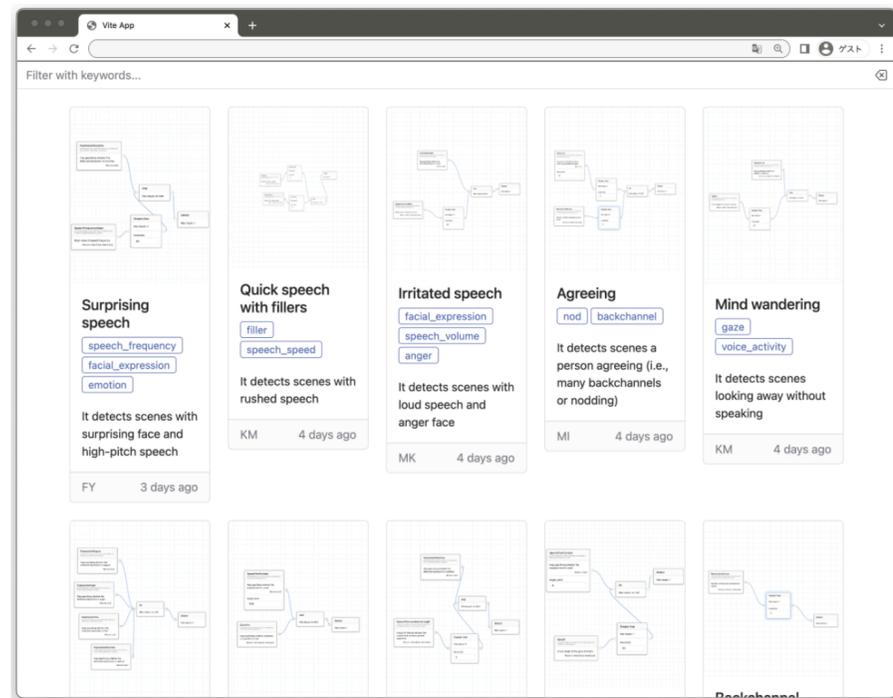
**objectively used for
multiple conversations**



**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.