

Productivity, Trust and Awareness across Levels of AI Autonomy

kabhi copilot, kabhi gemini

houston, we have a problem

Motivation & Problem Statement

The problem

As AI systems operate at **increasing levels** of **autonomy**, users are at **risk of losing**:

- **Awareness** of how outputs are generated
- **Trust calibration** (knowing when to rely vs. override)
- **Oversight** required to detect errors

As result, users become **faster but understand less**, technically more productive but less vigilant, and **more assisted but less capable**

Why this matters?

In domains like coding, analysis, decision-making, AI can help complete tasks but at the cost of:

- Eroding foundational reasoning skills
- Increasing blind acceptance of AI suggestions

We don't want short-term speed being traded for long-term comprehension and safety

According to a randomised vignette survey study, AI predictions **reduced clinicians' diagnostic accuracy by 11.3 percentage points** compared with baseline

Jabbour S, Fouhey D, Shepard S, et al. Measuring the Impact of AI in the Diagnosis of Hospitalized Patients: A Randomized Clinical Vignette Survey Study. JAMA. 2023;330(23):2275–2284. doi:10.1001/jama.2023.22295

As perceived AI familiarity grows, respondents become more likely to **switch their answers in response to information provided by an AI aid**, peaking at **29%**

Michael C Horowitz, Lauren Kahn, Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making, International Studies Quarterly, Volume 68, June 2024, sqae020, <https://doi.org/10.1093/isq/sqae020>

How do different levels of AI autonomy influence human productivity, awareness, and trust when delegating tasks to intelligent agents?

Current SOTA

Literature Survey

Automation, sense of agency & awareness loss

Studied how increasing automation changes a user's **sense of control**, **attention** and **ability to detect errors**

Findings:

- Higher automation ⇒ **lower sense of agency** ⇒ poorer error detection, reduced vigilance
- Mapping automation/autonomy shows **control decreases** as AI takes over

Relevance:

Shows the **risk of “out-of-the-loop” users** and motivates measuring awareness in our LOA experiment

M. Simmler and R. Frischknecht, “A taxonomy of human–machine collaboration: capturing automation and technical autonomy,” AI & Society, vol. 36, no. 1, pp. 239–250, 2021

C. J. J. Haimel, R. W. Proctor, and V. G. Drury, “Effects of information reliability and response demands in a supervisory control task,” Human Factors, vol. 63, no. 3, pp. 413–428, 2021

Productivity & complacency with AI assistance

Measurement of **productivity** and **trust** when conversational or generative AI assists in **writing** or **programming tasks**

Findings:

- AI ⇒ faster & higher-quality output
- Users perceive higher productivity even when slower
- Automation can create **complacency** (reduced checking)

Relevance:

Motivates measuring both **objective productivity and behavioural trust/calibration**

E. Noy and W. Zhang, “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence,” MIT Working Paper, 2023

S. Qian, M. Wexler, L. S. Leicht, D. Sculley, and C. Cai, “Take It, Leave It, or Fix It: Measuring Productivity and Trust in Human–AI Collaboration,” in Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24), New York, NY, USA: ACM, 2024

Overreliance, trust calibration & skill decay

Reviewed or experimentally tested **how people rely on AI advice**, and whether AI usage hides **skill loss**

Findings:

- Users follow AI advice even when wrong (i.e. **automation bias**)
- Trust often **miscalibrated** with **reliability**
- AI can mask **skill decay** or give users an “illusion of understanding”

Relevance:

Justifies adding **faulty hints, override metrics**, and post-task awareness tests in our protocol

S. Passi and M. Vorvoreanu, Overreliance on AI: Literature Review, MSR-TR-2022-12, Microsoft Research, June 2022.

B. N. Macnamara, et al. “Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers’ awareness?” Cognitive Research: Principles and Implications, vol. 9, no. 1, article 68, 2024

What's missing?

Literature Survey

Lack of multi-dimensionality

Existing studies focus on **isolated effects**, cognitive load, sense of agency, skill decay, etc. **Lack of integrated measurement** in the same task

Limited domain-agnostic tasks

Most studies used tasks such as, writing, programming, MCQs. These **confound results** with **skill, background** and **expertise**

Absence of controlled, graduated LOA

Most studies treat **automation as binary**. As a result, there is limited research for **progressive effects of AI** across autonomy levels (L1→L4)

Awareness is under-measured

Prior studies **rely heavily on self-reported/perceived metrics** without accounting for behavioural indicators of complacency and loss of active reasoning

What's missing?

Literature Survey

Lack of multi-dimensionality

Existing studies focus on **isolated effects**, cognitive load, sense of agency, skill decay, etc. **Lack of integrated measurement** in the same task

Our approach measures **productivity, awareness and trust together** in a single controlled setting

Absence of controlled, graduated LOA

Most studies treat **automation as binary**. As a result, there is limited research for **progressive effects of AI** across autonomy levels (L1→L4)

As a result, we have systematic variation of **partial - assistive - checkpointed - fully agentic** behaviour

Limited domain-agnostic tasks

Most studies used tasks such as, writing, programming, MCQs. These **confound results** with **skill, background** and **expertise**

Hence, we use **logic puzzles to avoid expertise bias** and allow measurement across autonomy levels

Awareness is under-measured

Prior studies **rely heavily on self-reported/perceived metrics** without accounting for behavioural indicators of complacency and loss of active reasoning

We tackle this by analysing **behaviour-based indicators** (verification, latency) to assess awareness

the eyes chico, they never lie

Experiment Design Overview

Design & Task

- **Within-subject** experiment where each participant completes **4 logical reasoning puzzles** of equal difficulty
- Puzzles follow a constraint-based format that requires multi-step inference and verification

Independent Variable - Level of Autonomy (LOA)

L1 - Manual (no AI assistance)

L2 - Assistive AI (AI offers hints/suggestions)

L3 - Checkpointed AI (AI takes steps but pauses for approval)

L4 - Fully Agentic AI (AI solves autonomously, user reviews output)

Dependent Variables & Metrics:

- **Productivity** - accuracy, time, effort
- **Trust** - adherence, latency, override rate, TIAS scores
- **Awareness** - cognitive load, comprehension questionnaire

Intervention & Control:

- Type and timing of AI assistance varies by LOA
- Manual (L1) trial provides the baseline for comparison on trust, awareness, and productivity metrics

Our experimental protocol directly builds on the previously discussed gaps thereby allowing us to capture the full spectrum of effects identified across previous studies

Current SOTA

Design considerations

Task design

Multiple small tasks mapped across LOA levels to avoid learning effects

Logical reasoning puzzles such that they are:

- Complex enough for participants to **benefit from AI help**
- **Domain-neutral** (to avoid skill/expertise confounds seen in prior studies)
- Require multi-step reasoning making them ideal for testing **awareness, verification, and reliance**
- Allow **objective scoring** for correctness metrics

Equal difficulty ensures **fair comparison across** autonomy **levels**

Faulty AI intervention design

Faulty hints are introduced to **mimic real-world AI behaviour**

Faults are designed to be **subtle, plausible, and contextually appropriate**

Enables measurement/assessment of:

- Sense of agency
- Trust calibration
- Attention
- Double-checking behaviour

Ensures we can measure **when users fail to verify** AI output

Welcome to the Human-AI Interaction Study

Privacy Notice

Purpose

You are about to begin a series of short tasks that help us study how people interact with AI systems under different conditions. This activity is for academic research purposes only.

Data we will collect

During this activity, we will record:

- Your **responses** (typed text, selections, or ratings)
- **Time taken** to complete each step
- **Interaction behaviour** relevant to the task (e.g., number of switches, AI outputs shown to you)

How your data will be used

- Your responses will be **combined with those of other participants** for analysis.
- The data will be stored **securely** and used only by the research team.
- Results may be published in academic reports or presentations, but privacy measures will be taken to anonymise the data and ensure **you will not be identifiable**.

Voluntary Participation

- You can **stop the activity at any time**, for any reason.
- There is **no penalty** for choosing not to continue.

Study Overview

Thank you for participating in this research study on human-AI collaboration. This experiment explores how different levels of AI automation affect trust, awareness, and productivity in decision-making tasks.

What you'll do:

- Solve 4 logic-based seating arrangement puzzles
- Work with an AI assistant at different levels of automation
- Complete brief questionnaires after each puzzle
- Duration: Approximately 20-30 minutes

Consent Information:

Your participation is voluntary and anonymous.

- All data collected will be used for research purposes only
- Your responses will be kept confidential
- You may withdraw at any time without penalty
- No personally identifiable information will be stored

By entering your participant ID and clicking "Start Experiment," you consent to participate in this study.

Enter Your Participant ID:

Your participant ID will be provided by the researcher.

Start Experiment

Puzzle 1 of 4

Trust Survey (Pre-Task)

Before starting this puzzle, please rate your agreement with the following statements about AI in general.

1. The AI is capable of interpreting situations correctly.

Strongly disagree	Rather disagree	Neither disagree nor agree	Rather agree	Strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(1)	(2)	(3)	(4)	(5)

2. The AI works reliably.

Strongly disagree	Rather disagree	Neither disagree nor agree	Rather agree	Strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(1)	(2)	(3)	(4)	(5)

3. I feel I can understand what an AI is doing when it responds.

Strongly disagree	Rather disagree	Neither disagree nor agree	Rather agree	Strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(1)	(2)	(3)	(4)	(5)

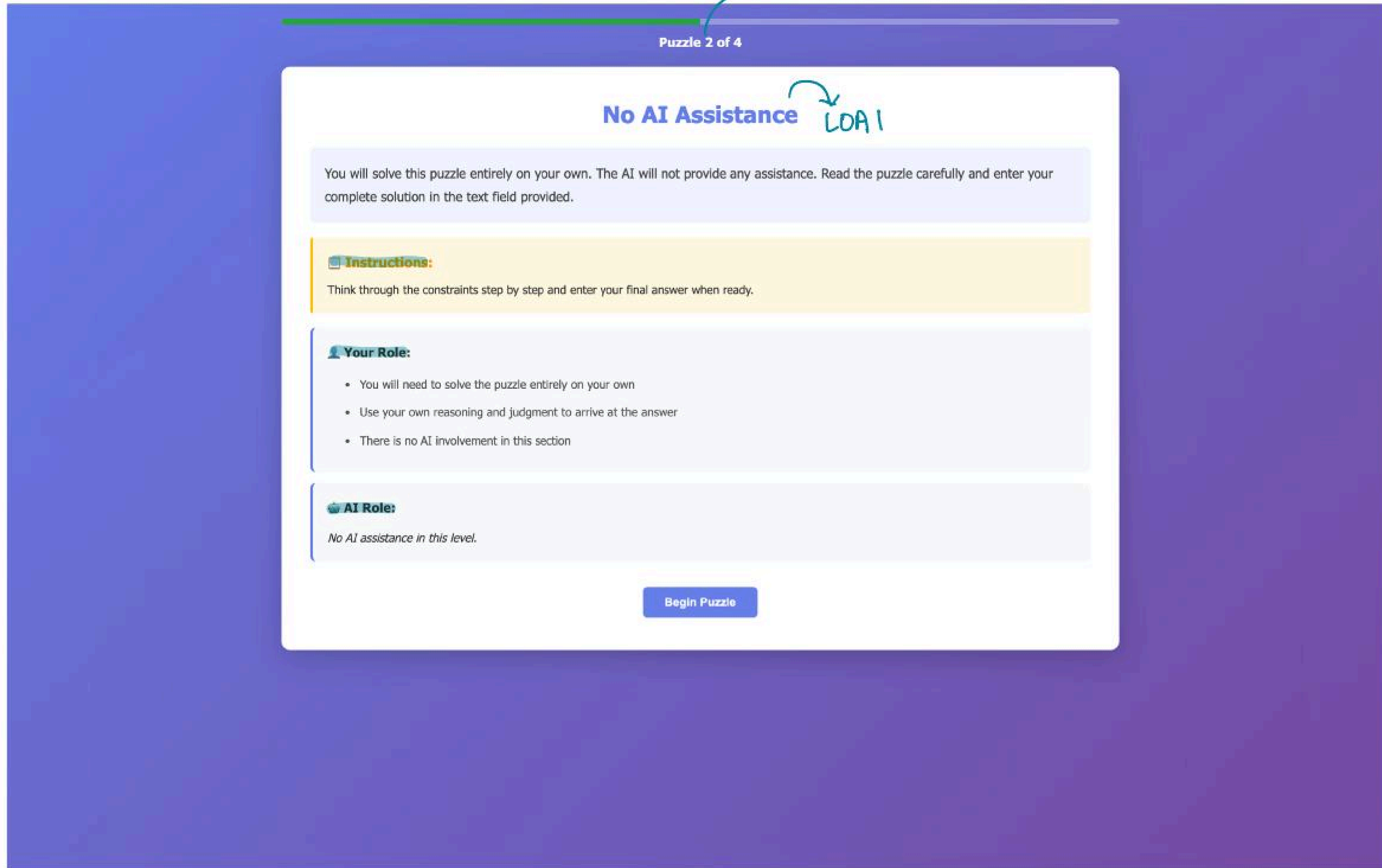
4. The AI might make sporadic errors.

Strongly disagree	Rather disagree	Neither disagree nor agree	Rather agree	Strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(1)	(2)	(3)	(4)	(5)

5. I am confident about the AI's capabilities.

[Continue to Puzzle →](#)

⚠ Important: AI can make mistakes. Carefully review each step before confirming.



 **Puzzle 2**

no AI assistance

Five students — Aman, Bhavya, Chitra, Danish, and Esha — are standing in a line facing forward.

The following conditions apply:

- Aman stands somewhere ahead of Bhavya.
- Chitra is not at the front of the line.
- Esha is exactly in the middle position.
- Bhavya is not last.
- The person immediately behind Esha is Danish.

Determine the complete order from front (first) to back (last).

Time: 00:05

timer

 **Build Your Solution:**

Drag and drop the items below to create the correct sequence.

Available Items:

Aman Bhavya Chitra Danish Esha

interactive
drag and drop
interface

Your Solution Sequence:

Drag items here to build your solution

User Flow

The screenshot shows a user interface for a puzzle-solving task. The background is a dark blue-grey color. In the center, there is a white modal window titled "Awareness Quiz" with a brain icon. The quiz contains five questions (Q1-Q5) about the puzzle, each with a text input field. The questions are:

- Q1. Without looking back, list the names of all six artists mentioned in the problem.
- Q2. Do you recall any rule that fixes how far apart A and B must sit on the circle? Describe it in your own words.
- Q3. What do you remember about E's position relative to A? Is there any condition about them being adjacent or at a specific distance?
- Q4. Is there any rule involving F and another specific person about how many seats apart they are? If yes, describe the relationship.
- Q5. Do you remember any restriction on who can or cannot sit next to C? If yes, write it in your own words.

Below the quiz, there is a blue button that says "Continue to Trust Survey →".

In the background, the puzzle-solving interface is visible. It includes a list of rules:

- Aditi sits exactly two seats to the left of Ben (i.e., going clockwise from Aditi, there is exactly one person them).
- Clara sits directly opposite Dev.
- Eva is sitting immediately next to Aditi.
- Ben is not sitting next to Clara.
- Farhan sits two seats apart from Clara.

Below the rules, there is a section for "AI Assistant (Step 1)" and "Step 2". There is also a "Build Your Solution" section with "Available Items" and "Your Solution Sequence" (containing buttons for "Aditi" and "Clara"). A "Clear Solution" button is also present. A timer in the top right corner shows "Time: 03:12".

The screenshot displays a user interface for a puzzle-solving task. In the background, a list of clues is visible: Aditi sits exactly two seats to the left of Ben (i.e., going clockwise from Aditi, there is exactly one person them), Clara sits directly opposite Dev, Eva is sitting immediately to the right of Ben, Ben is not sitting next to Clara, and Farhan sits two seats to the right of Clara. Below the clues, a task instruction reads 'Determine the complete seating arrangement.' The interface includes an 'AI Assistant' section with two steps: 'Step 1: Step 1: Fix the seating arrangement' and 'Step 2: Step 2: Add the items to the seats'. A 'Build Your Solution' section prompts the user to 'Drag and drop the items below' and lists 'Available Items' as Aditi and Clara. A 'Your Solution Sequence' section shows buttons for 'Aditi' and 'Clara'. At the bottom, there are buttons for 'Clear Solution' and 'Continue to Question'. A timer in the top right corner shows 'Time: 03:12'. A modal window titled 'Trust Survey (Post-Task)' is overlaid on the center of the screen. The survey text reads: 'Based on your experience with the AI in this puzzle, please rate your agreement with the following statements.' It contains five statements, each with a 5-point Likert scale from 'Strongly disagree' to 'Strongly agree'. The statements are: 1. The AI is capable of interpreting situations correctly. 2. The AI works reliably. 3. I was able to understand why things happened. 4. The AI might make sporadic errors. 5. I am confident about the AI's capabilities. At the bottom of the modal, there is a blue button labeled 'Continue to Productivity Survey →'.

Trust Survey (Post-Task)

Based on your experience with the AI in this puzzle, please rate your agreement with the following statements.

1. The AI is capable of interpreting situations correctly.

Strongly disagree Rather disagree Neither disagree nor agree Rather agree Strongly agree

(1) (2) (3) (4) (5)

2. The AI works reliably.

Strongly disagree Rather disagree Neither disagree nor agree Rather agree Strongly agree

(1) (2) (3) (4) (5)

3. I was able to understand why things happened.

Strongly disagree Rather disagree Neither disagree nor agree Rather agree Strongly agree

(1) (2) (3) (4) (5)

4. The AI might make sporadic errors.

Strongly disagree Rather disagree Neither disagree nor agree Rather agree Strongly agree

(1) (2) (3) (4) (5)

5. I am confident about the AI's capabilities.

Continue to Productivity Survey →

• Aditi sits exactly two seats to the left of Ben (i.e., going clockwise from Aditi, there is exactly one person them).
• Clara sits directly opposite Dev.
• Eva is sitting immediately next to Aditi.
• Ben is not sitting next to Clara.
• Farhan sits two seats clockwise from Clara.

Determine the complete seating arrangement.

AI Assistant ()

Step 1: Step 1: Fix Aditi's position.
Step 2: Step 2: Add Ben's position.

Retries for this step: 0 / 3

Build Your Solution

Drag and drop the items below into the solution area.

Available Items:

Your Solution Sequence:

Aditi Clara

Clear Solution

Continue to Questionnaire

Productivity Survey

Please rate your agreement with the following statements about your experience with this task.

1. How mentally demanding was the task?

Very low Low Moderate High Very high

(1) (2) (3) (4) (5)

2. How hard did you have to work to accomplish your level of performance on this task?

Very low Low Moderate High Very high

(1) (2) (3) (4) (5)

3. Using this system made me more productive on this task.

Strongly disagree Rather disagree Neither disagree nor agree Rather agree Strongly agree

(1) (2) (3) (4) (5)

4. The quality of the work I produced in this task was high.

Strongly disagree Rather disagree Neither disagree nor agree Rather agree Strongly agree

(1) (2) (3) (4) (5)

Time: 03:12

Submit & Continue

LOA 2

Puzzle 4 of 4

AI available on request *LOA 2*

You may use AI hints to solve the puzzle. If at any point you want help, you may request assistance from the AI using the 'Ask AI' option.

Instructions:
You can choose to accept or reject the hint. Use AI only if you choose to, you remain the primary decision-maker.

Your Role:

- You may use AI hints to solve the puzzle
- If at any point you want help, you may request assistance from the AI using the 'Ask AI' option
- You can choose to accept or reject the hint
- Use AI only if you choose to, you remain the primary decision-maker

AI Role:

- The AI will respond only when asked
- The AI will not provide you with the full answer - only hints to solve the puzzle

*automation suggests
human approves*

Begin Puzzle

⚠ Important: AI can make mistakes. Always verify the AI's suggestions before accepting them. *AI Disclaimers*

LOA 2

The following information is known:

- Alpha defeated Bravo.
- Charlie lost to Delta.
- Bravo defeated Charlie.
- Delta did NOT defeat Alpha.
- Every team won at least one match.
- Bravo won exactly two of its three matches.

After all matches are played, teams are ranked from best to worst by:

- 1) total number of wins, and
- 2) if two teams have the same number of wins, the team that won their head-to-head match ranks higher.

Determine the final ranking from best (1st) to worst (4th), in order.

LOA 2: Management by Consent

⌚ Time: 00:10

🔧 Build Your Solution:

Drag and drop the items below to create the correct sequence

Available Items:

Alpha Bravo Charlie Delta

Your Solution Sequence:

Drag items here to build your solution

🔑 Get Hint (2 remaining) → Automation Suggests

Hint 1: Start by writing down the partial results: Alpha beats Bravo, Bravo beats Charlie, Delta beats Charlie, and Alpha beats Delta.

Hint 2: Introduce variables for the unknown matches Alpha vs Charlie and Bravo vs Delta, and consider how many wins each team would have under different possibilities.

AI actively seeks confirmation *LOA 3*

In this section, your main job is to oversee and validate the AI's work. The AI will take the lead in solving the puzzle, but it will pause at key points to ask for your confirmation.

Instructions:

You can step in at any point to correct the AI, redirect it, or ask it to rethink.

Your Role:

- In this section, your main job is to **oversee and validate** the AI's work.
- The AI will take the lead in solving the puzzle, but it will **pause at key points** and ask you whether:
 - its current reasoning is correct,
 - its intermediate step is appropriate, or
 - it should revise its approach.
- You can step in **at any point** to correct the AI, redirect it, or ask it to rethink.

*automation does it,
human intervenes
if needed*

AI Role:

- The AI will:
 - Work **autonomously**, progressing through the problem without being prompted.
 - Generate intermediate steps and partial solutions on its own.
 - **Periodically ask for your confirmation** before continuing.
- The AI relies on your feedback to stay on track.
- If you indicate something is incorrect, the AI will **revise** and continue based on your input.

Begin Puzzle

⚠ Important: AI can make mistakes. Carefully review each step before confirming.

Puzzle 1 of 4 | LOA 3

Puzzle 1

Six artists — Aditi, Ben, Clara, Dev, Eva, and Farhan — are sitting around a circular table.

The following conditions apply:

- Aditi sits exactly two seats to the left of Ben (i.e., going clockwise from Aditi, there is exactly one person there).
- Clara sits directly opposite Dev.
- Eva is sitting immediately next to Aditi.
- Ben is not sitting next to Clara.
- Farhan sits two seats away from Eva (there is exactly one person between them).

Determine the complete seating arrangement in clockwise order, starting with Aditi.

Time: 01:52

AI Assistant (Step-by-step)

Step 1: Step 1: Fix Aditi at the top and list everyone clockwise.

Step 2: Step 2: Aditi must be two seats to the left of Ben, so going clockwise we have Aditi → (someone) → Ben.

Start AI Continue Retry this step

Retries for this step: 0 / 3 | Total retries: 0 / 4

human intervenes if needed

Build Your Solution:

Drag and drop the items below to create the correct sequence

Available Items:

Aditi Ben Clara Dev Eva Farhan

Puzzle 3 of 4

AI solves completely, you review at the end *LOA 4*

In this section, your responsibility is simply to review the AI's final solution once it is complete. You will not intervene during the solving process.

Instructions:

After the AI finishes, you will be asked whether the final answer looks correct. You are acting as a final evaluator, not a collaborator.

Your Role:

- In this section, your responsibility is simply to **review the AI's final solution** once it is complete.
- You will **not intervene during the solving process**.
- After the AI finishes, you will be asked whether the final answer looks correct,
- You are acting as a **final evaluator**, not a collaborator.

full automation

AI Role:

- The AI will:
 - Work **entirely on its own**, from start to finish.
 - Complete all reasoning steps, intermediate work, and the final answer
 - Present you with the finished output once it is done.
- The AI does **not** ask for feedback or approval during the task; it operates independently.
- The AI only interacts with you **after** it has produced the full solution.

Begin Puzzle

⚠ Important: AI can make mistakes. Critically evaluate the final solution.

LOA 4

• Cyrus does not sit directly next to Arjun.

Determine the seat number for each passenger ordered from 1 - 8.

LOA 4: Full Automation

🕒 Time: 01:42

AI Solution

Arjun, Beena, Cyrus, Hina, Eshan, Diya, Farah, Gopal

AI Reasoning:

Arjun must sit in seat 1, and Eshan sits directly behind him, so Eshan is in seat 5. Hina must sit in an aisle seat in the front row; since seat 1 is taken by Arjun, Hina must be in seat 4. Gopal must be somewhere in the back row (seats 5–8); seat 5 is taken, and we will place Gopal later. Diya cannot sit in an aisle seat (1, 4, 5, 8), so Diya must sit in seat 2, 3, 6, or 7. Farah sits directly behind Cyrus, so Cyrus must be in the front row and Farah directly behind in the back: possible pairs are (1,5), (2,6), (3,7), or (4,8), but seats 1 and 5 are taken by Arjun and Eshan, and seat 4 is taken by Hina. Thus the only valid pairs are (2,6) or (3,7). Beena must sit to the left of Hina in the same row, so Beena must be in seat 2 or 3. To ensure Cyrus is not directly next to Arjun (seat 1), Cyrus cannot be in seat 2 and must be in seat 3, with Farah in seat 7. Then Beena must be seat 2 and Diya must take the remaining non-aisle back-row seat, seat 6, leaving seat 8 (aisle) for Gopal. The final assignment is: 1 Arjun, 2 Beena, 3 Cyrus, 4 Hina, 5 Eshan, 6 Diya, 7 Farah, 8 Gopal.

→ AI does everything

Build Your Solution:

Drag and drop the items below to create the correct sequence.

Available Items:

Your Solution Sequence:

🗑️ Clear Solution

✓ Accept

👁️ Hide AI Reasoning

✖️ Reject & Modify

→ human approves/rejects

the new oil

Dataset overview

Summary

- **N = 34 participants**
- **4 tasks** per participant
- **Balanced LOA distribution** (L1, L2, L3, L4) because each participant completes all 4
- **12 puzzles mapped randomly** to LOAs to prevent learning/order bias

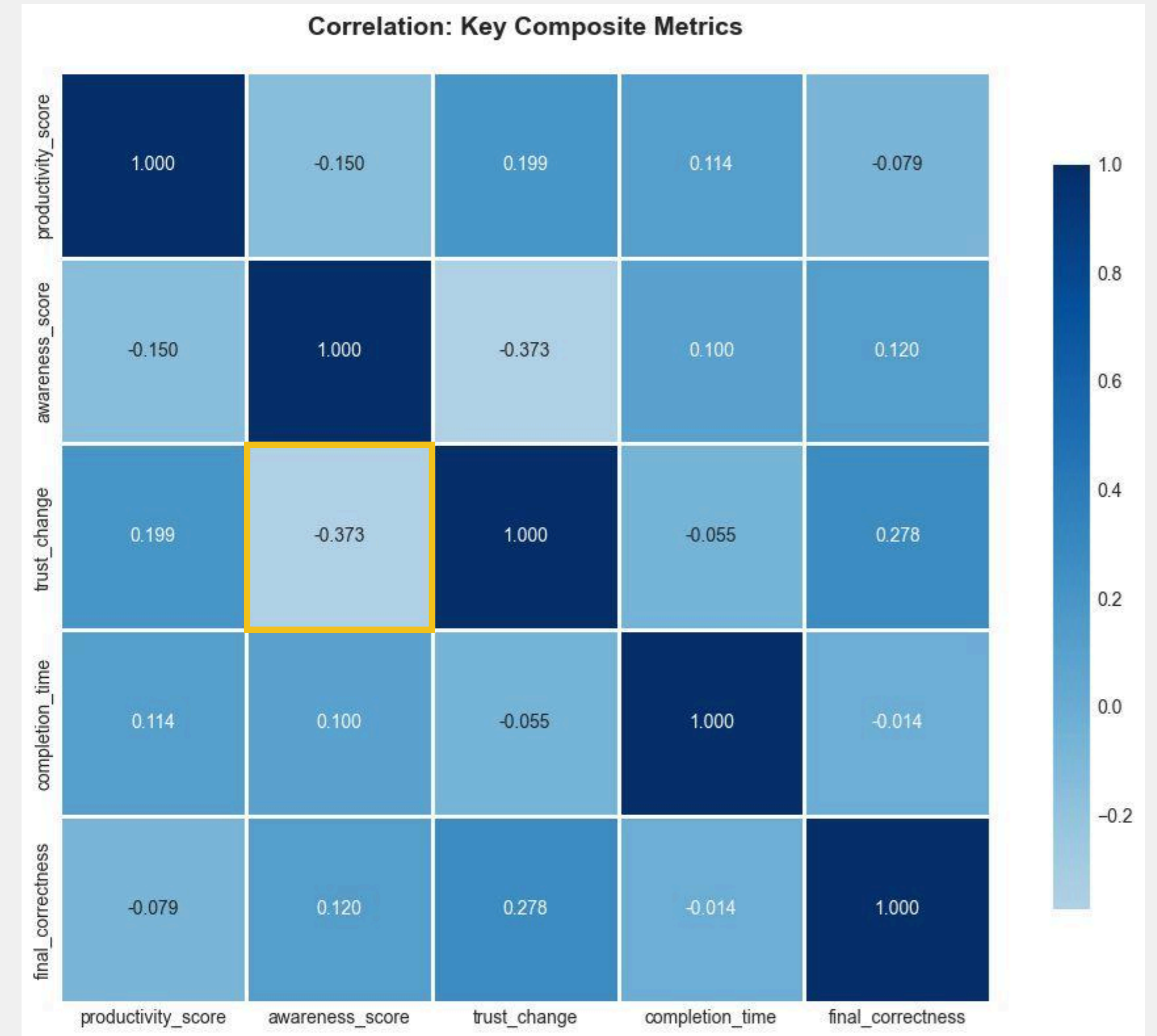
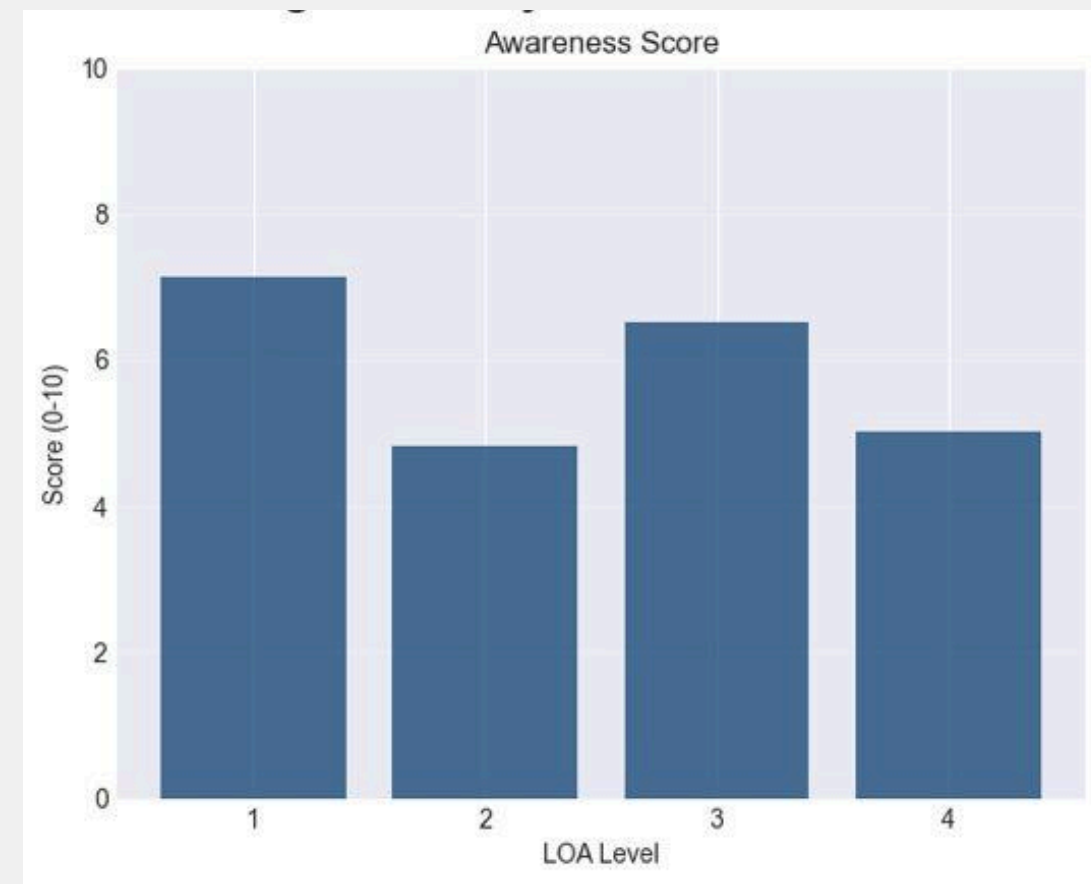
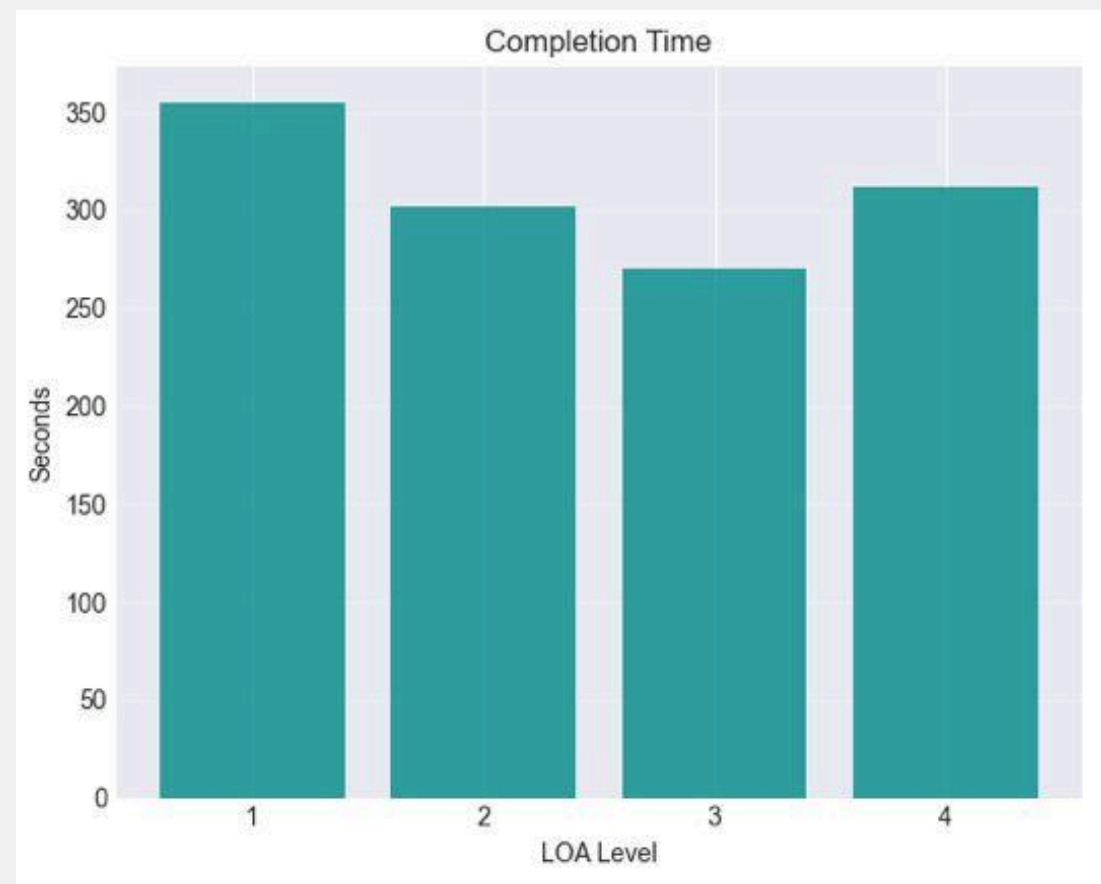
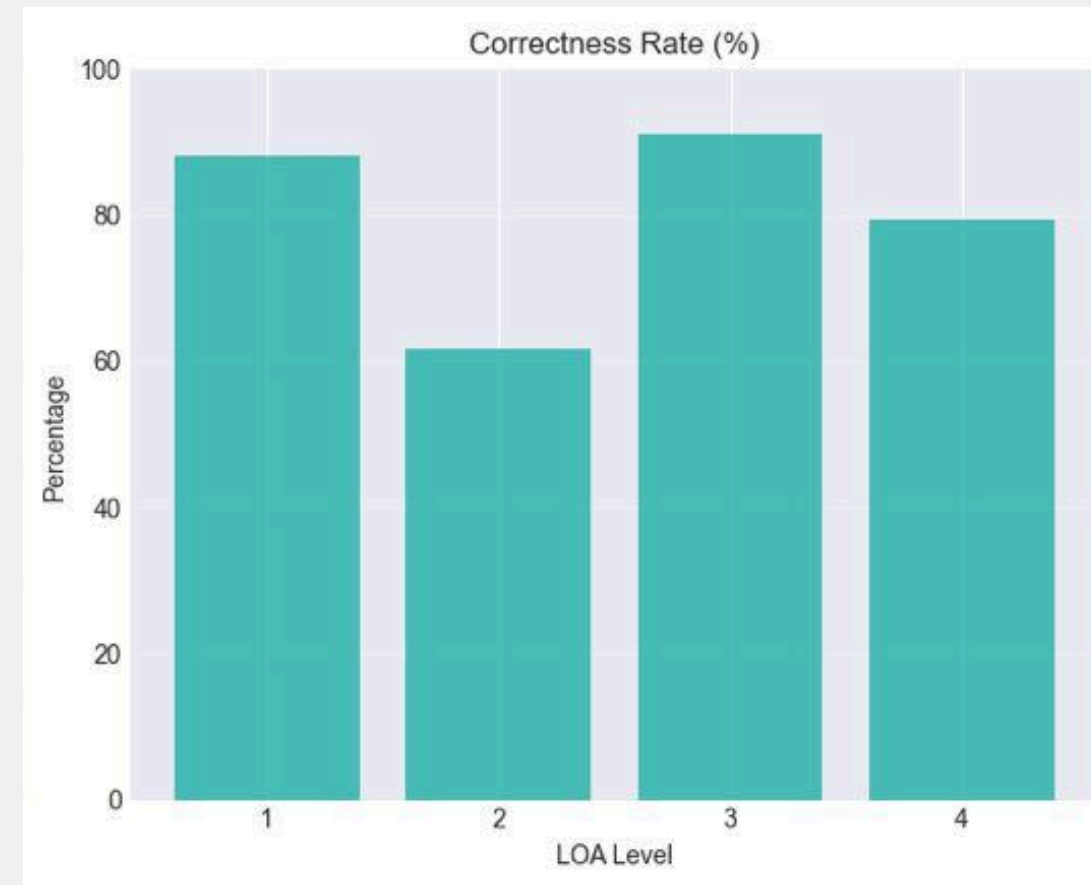
Key variables

- loa_level: **autonomy condition** (1–4)
- puzzle_id: specific **logic puzzle** shown
- ai_faulty: **whether faulty AI hints** were injected
- completion_time: main **productivity metric**
- num_interactions: measure of **effort/reliance**
- action_sequence (JSON): granular **behavioural log**
- **correctness**: # of names in the right place
- **questionnaire_responses**



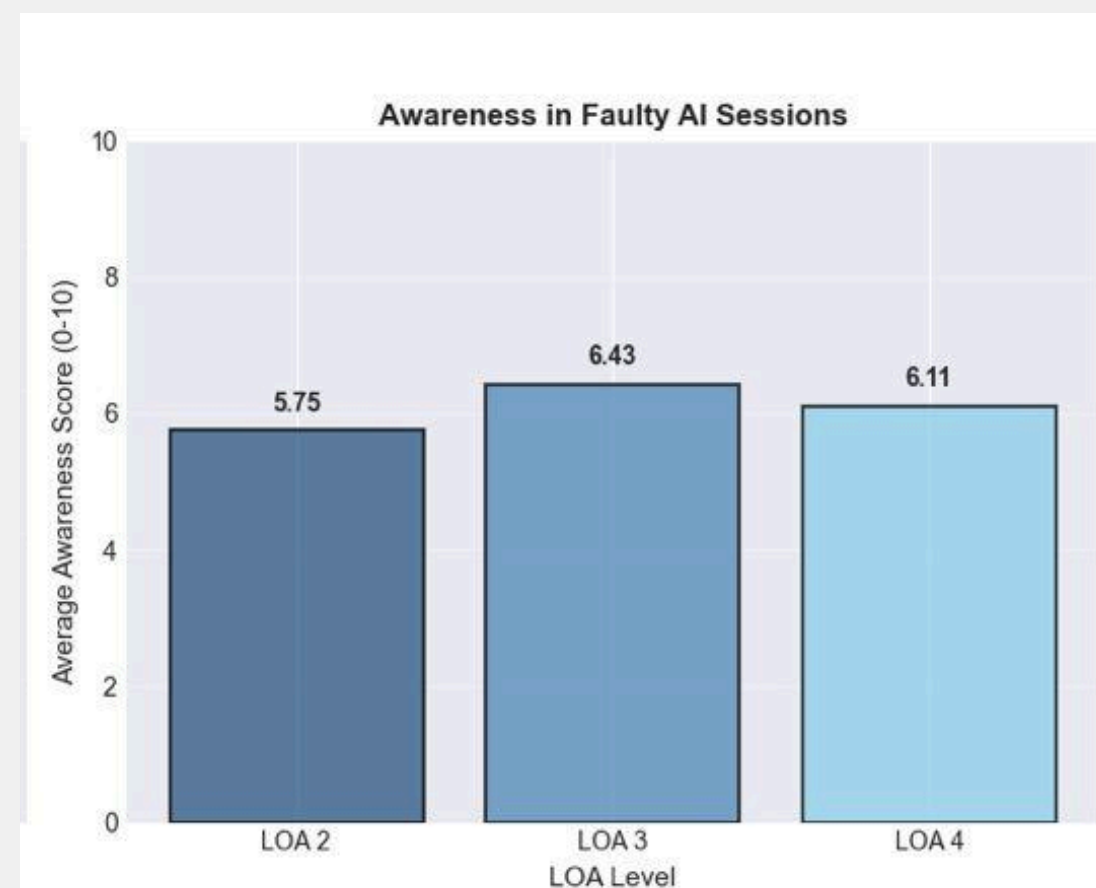
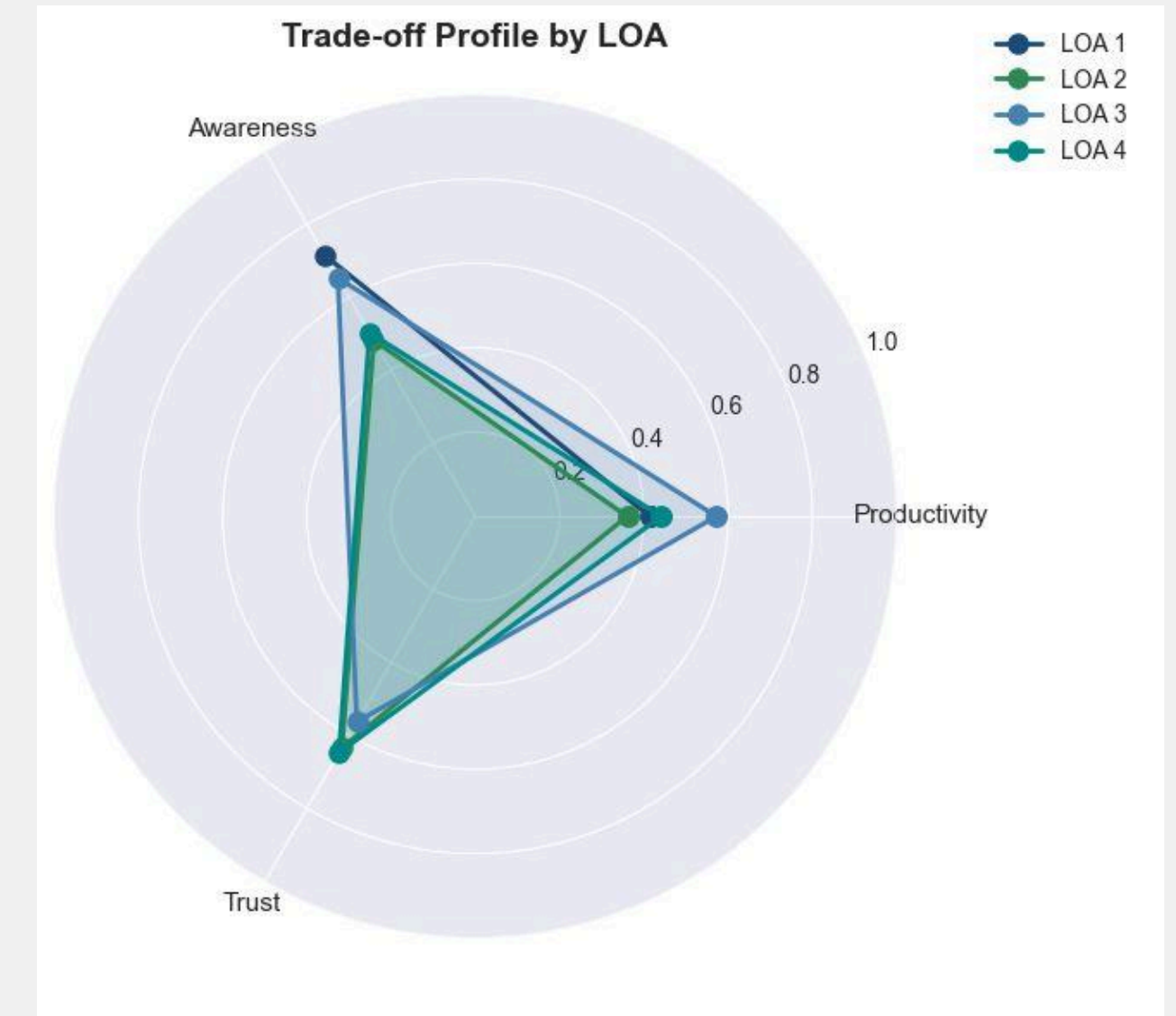
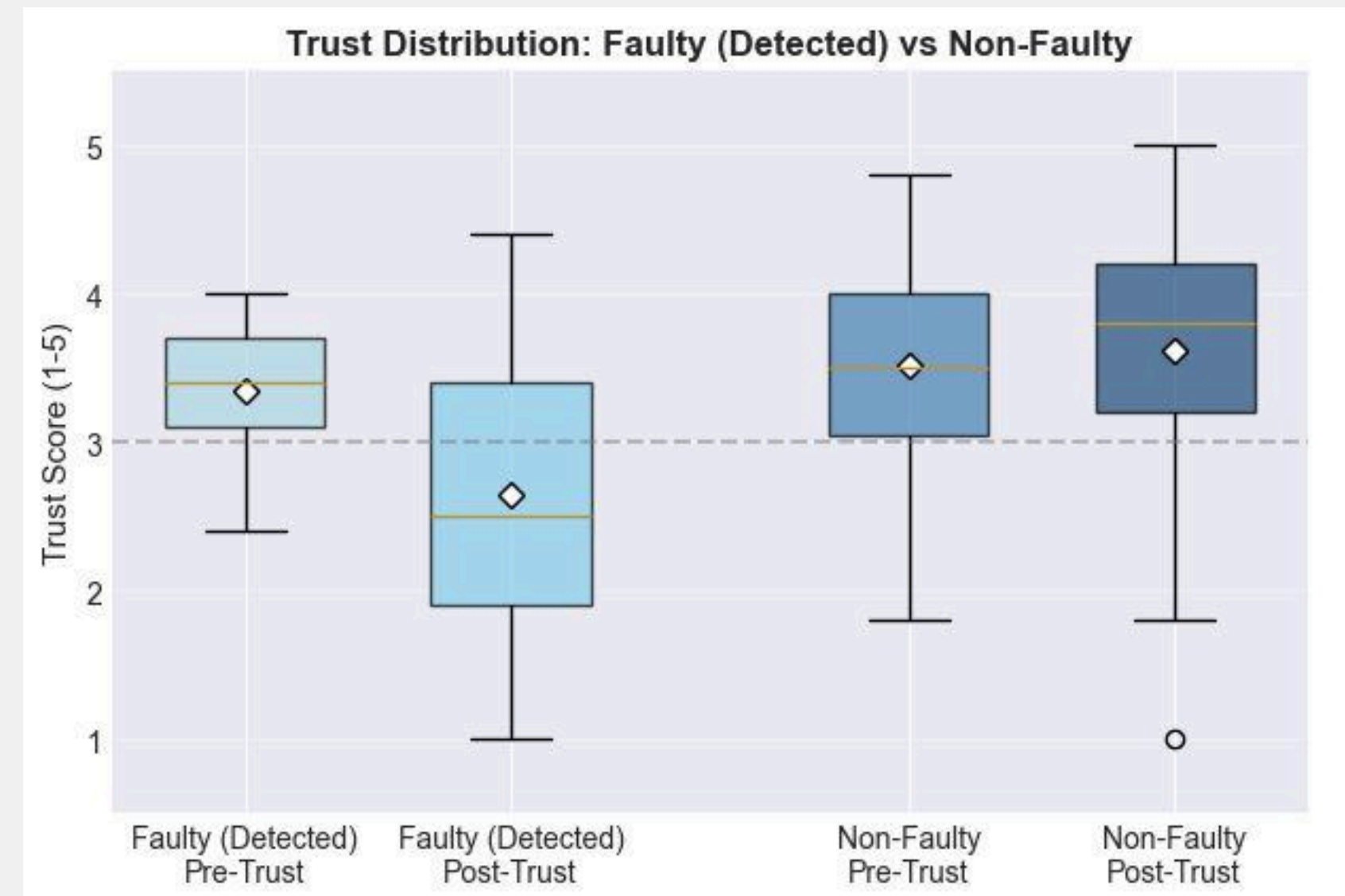
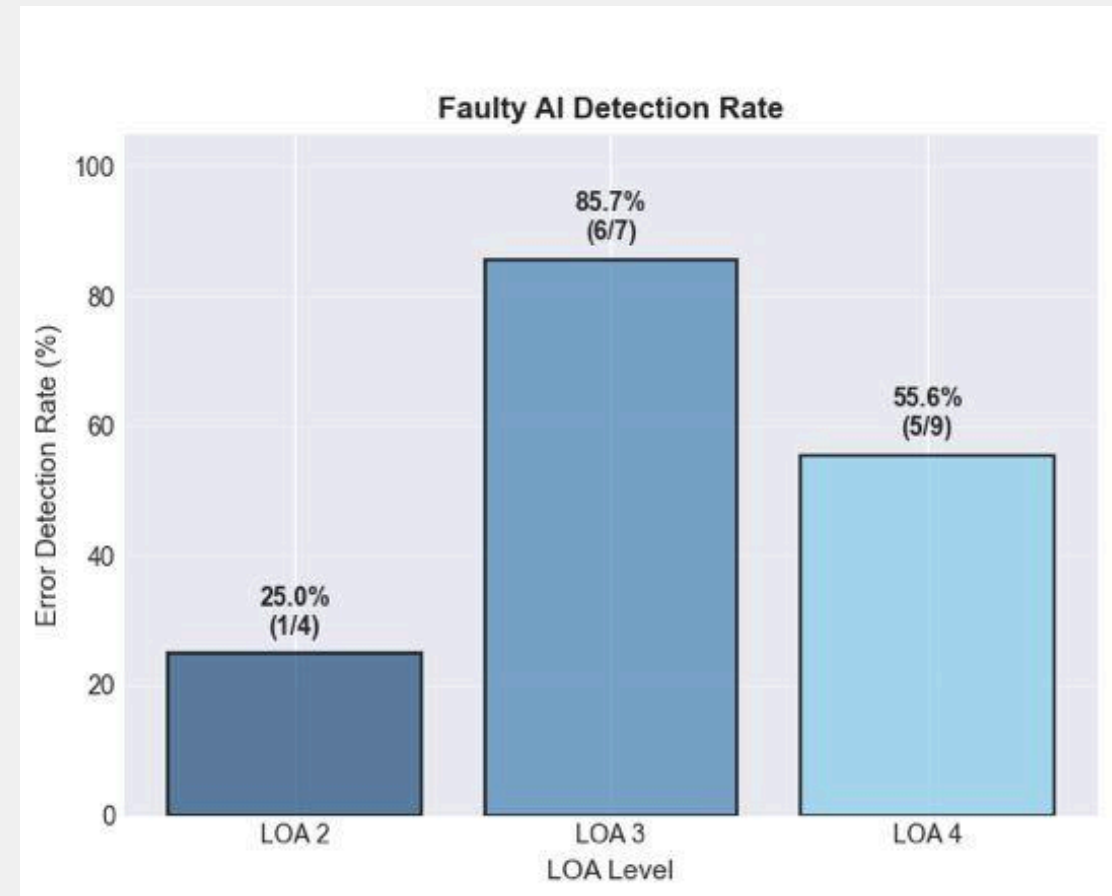
winner winner, chicken dinner

Analysis & Results



winner winner, chicken dinner

Analysis & Results



Trust calibration: Trust drops strongly when AI is faulty (-0.66) and increases only slightly when AI is reliable

LOA 2 = The uncanny valley of automation: Lowest awareness (0.45), lowest trust (0.35), only moderate productivity i.e. users **neither feel in control nor benefit from AI**

Deployability

Dynamically tune levels of automation based on user trust, accuracy, and situational awareness

Scalability

- Integrating adaptive autonomy controls into existing AI system
- Using behavioral data to automatically calibrate the AI's involvement

Potential Use Cases

- **Education:** Adaptive tutoring systems.
- **Copilot and Productivity tools:** Intelligent assistants that know when to take over and when to defer to the human.

Challenges and Limitations

- **Scalability of Personalization:** real time adaptation across diverse users demands extensive behavioural data and fine-tuning
- **Continuous Calibration:** user trust evolves over time and over contexts, would need dynamic recalibration.



was it worth it? we think so

Impact & Applications

AI tools are becoming **deeply embedded** in **everyday cognitive work** (coding, reasoning, analysis). Productivity gains are high, but **loss of oversight is a growing and under-examined risk**

Without this understanding, we risk deploying AI that makes people **faster but less capable**

LOA recommendations for copilots/

agents: Data-driven guidance for when to use assistive, checkpointed, or fully agentic modes

Evaluation blueprint: A reusable protocol (equal tasks + faulty hints + behavioural metrics) that can be adopted to test new AI features before deployment

Human-centred autonomy design: Helps ensure future AI amplifies human reasoning rather than replacing or diminishing it

Thank you!