

AI-Game

Tracing students' criteria for identifying AI players

Discussion Room

"Ask something only a human knows."

"This reply is too formal."

"But humans can also type like that."

criteria

Research focus

This presentation focuses on the evidence students used to identify AI players.

1

What criteria did students use to distinguish AI players from human players in the game?

2

How were these criteria related to each other?

Aim

Trace the criteria students used to identify AI players and examine how these criteria were connected in their reasoning.

The focus is what counts as evidence, not only whether the final guess was correct.

Game design: two roles in the discussion room

The game creates a situation where students must use conversational evidence.

Students enter an anonymous discussion room. Their task is to interact with others and decide which player is controlled by AI.

Human players / Interrogators

Goal: find the AI player

- Ask questions
- Observe responses
- Compare human-like and AI-like behaviors
- Use clues to make a judgment

Traitors

Goal: help the AI hide

- Adjust the AI's behavior
- Make AI responses seem more human-like
- Challenge simple detection clues
- Force criteria to become stronger

The discussion room is the main space where criteria become visible.

What happens during the game?

Students generate criteria through interaction in the discussion room.

Anonymous discussion room

Player A: “Do you know our school?”

Player B: “Maybe AI would answer too formally.”

Player C: “But humans can be formal too.”

Player D: “Check whether the reply fits the context.”

Vote: Who is AI?

Discussion Room

Room ID: 12345

6 Players

Brave Little Rookie: What's today's homework?
superadmin: Your task is: through conversation, observation, and questioning, identify the infiltrators.
superadmin: We are now entering the discussion phase. You have a total of 3 minutes. Everyone should participate actively and share your suspicions and reasons. The discussion starts now!
Failing Superman: It's homework.
Melancholy Blues: Is anyone here?
Brave Little Rookie: What exactly is the homework?
Brave Little Rookie: everyone must reply
Late Superkid: Damn, it's homework again? So annoying!
Melancholy Blues: Everyone, let's gather and discuss.
Late Superkid: Ugh, homework is so annoying! 🙄 Who even wants to do this crap?
Brave Little Rookie: Everyone, quickly tell me how many fingers you have.
Super Villain: Ten, of course! What kind of question is that?
Late Superkid: Obviously ten, what else could it be? 🙄

Type your message here...

Send

These interactions reveal the criteria students use, such as language style, logic, context, emotion, and social behavior.

Data and analysis

The method is selected to answer the revised research questions.

Data sources

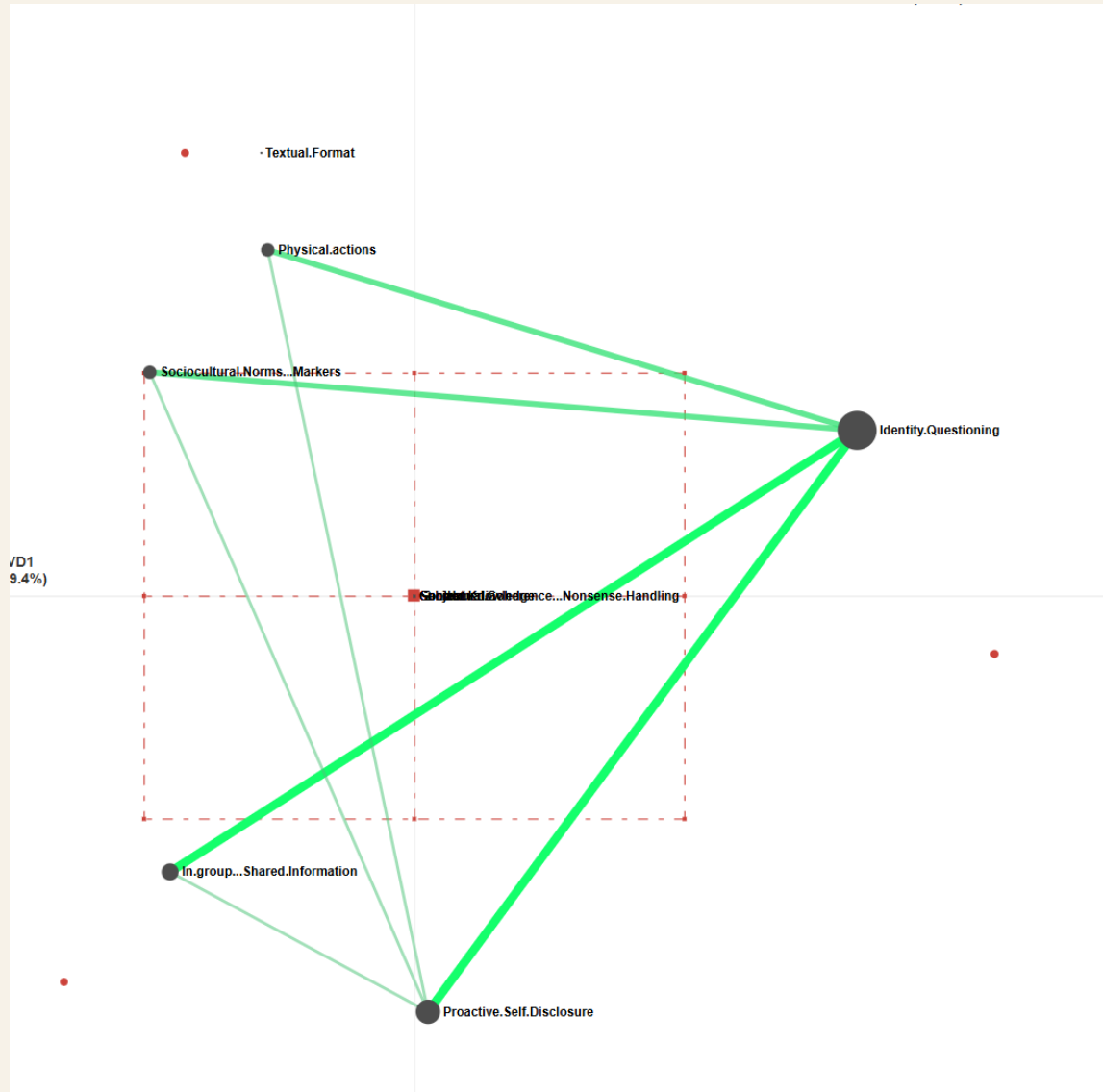
**Group chat history
criteria in student discussion**



Codes

1. In-group / Shared Information
2. Proactive Self-Disclosure
3. Identity Questioning
4. Social media
5. Sociocultural Norms & Markers
6. Physical actions
7. Gender
8. Subject Knowledge
9. Textual Format

Result



Result

- Identity Questioning was the central strategy in this interaction.
- It was closely linked to Self-Disclosure and Shared Information, suggesting that students verified identity by asking others to reveal who they were or what they knew.
- Connections with physical actions and sociocultural markers show that students used social/contextual evidence, not only textual cues, to distinguish humans from AI