Accepted for publication, to appear in:
Discourse & Communication

DOI: 10.1177/17504813241267069

Can AI explain AI? Interactive co-construction of explanations among human and artificial agents

Nils Klowait[1,2*], Maria Erofeeva[3*], Michael Lenke[1,2], Ilona Horwath[1,2], and Hendrik Buschmeier[1,4]

Abstract

This study investigates the potential of using advanced conversational artificial intelligence (AI) to help people understand complex AI systems. In line with conversation-analytic research, we view the participatory role of AI as dynamically unfolding in a situation rather than being predetermined by its architecture. To study user sensemaking of intransparent AI systems, we set up a naturalistic encounter between human participants and two AI systems developed in-house: a reinforcement learning simulation and a GPT-4-based explainer chatbot. Our results reveal that an explainer-AI only truly functions as such when participants actively engage with it as a co-constructive agent. Both the interface's spatial configuration and the asynchronous temporal nature of the explainer AI – combined with the users' presuppositions about its role – contribute to the decision whether to treat the AI as a dialogical co-participant in the interaction. Participants establish evidentiality conventions and sensemaking procedures that may diverge from a system's intended design or function.

Keywords

ChatGPT, co-constructed explainability, ethnomethodological conversation analysis, human-computer interaction, multimodality

Introduction

Artificial intelligence (AI) is increasingly used in sensitive environments. It decides whether we are likely to commit a crime, whether we should receive the requested bank loan, and what career paths we should pursue (Elliott 2022). There is thus a need to

[*]These authors contributed equally. [1]SFB/Transregio 318 Constructing Explainability, Germany [2]Paderborn University, Germany [3]Free University of Brussels, Belgium [4]Bielefeld University, Germany

Corresponding author:
Nils Klowait, SFB/Transregio 318 Constructing Explainability, Mersinweg 7, Paderborn 33100, Germany.
Email: nils.klowait@uni-paderborn.de

be able to understand how AI makes its decisions. Moreover, there are calls to move beyond one-sided explainability, where what needs to be explained is predetermined. Co-constructed explainability, the notion that explanations need to be able to emerge from a more equally-footed, mutually-oriented interaction between participants and AI, is a new paradigm that aims to more equitably address issues surrounding the spread of pervasive and opaque AI (Rohlfing et al. 2021).

With the release of the 'ChatGPT' conversational agent, the question emerged whether the matter of co-constructed explanations had thereby been solved. With GPT-4, the next version of models underlying ChatGPT, it became possible to also assign a stable role to the conversational agent, with the resulting agent displaying seemingly humanlike competency on a range of cognitive tasks (Han et al. 2024).

We thus had the real possibility of assigning a flexible, knowledgeable, and remarkably competent artificial intelligence the role of being an explainer AI. This explainer AI could then receive information about another AI system that is to be explained. The explainer AI could thus mediate between an opaque AI system – by having access to its source code and description – and a lay user. The lay user would be able to formulate novel questions and articulate possibly unexpected explainability needs, whereas the explainer AI could attend to the success of the explanation by monitoring the follow-up responses. Participants could respecify their questions by giving feedback on prior AI-generated responses and co-construct an explanatory situation across multiple turns.

In traditional rule-based conversational agents, the resultant 'dialogue' is substantially limited as topics and conversational projects are mostly predetermined (Moore et al. 2017). If rule-based conversational agents can't be more than a simulacrum of a conversation (Button 1990), would a system like ChatGPT be capable of engaging an interaction partner on human terms? However, mutual exchanges of words do not become "conversations" because of an underlying principle or an inherent feature of a system, but as a local accomplishment. Much like somebody might be mistaken as a robot, so can even simple systems – such as Weizenbaum's ELIZA (Bassett 2019) – be taken as full-fledged interactants by human interlocutors.

This paper is positioned within the tradition of conversation-analytic human-computer interaction (for a review, see Reeves and Porcheron 2023), which commonly traces its genealogy from Suchman's distinction between plans and situated actions (Suchman 2007), where any planned-out interactional scenario inscribed into a system needs to be investigated as it becomes oriented-to by in-situ participants. Thus, instead of situating the question of ChatGPT's interactivity on the level of its technical capabilities, we aim to investigate its potential to help co-produce new conversational complexities by setting up a naturalistic encounter between human participants and a conversational AI.

Data and methods

This paper's findings are based on videographic recordings of 6 online and 23 in-person sessions involving diverse participants ($N = 77$) recruited in the context of AI literacy workshops and open calls for participation; data was collected in Q3 of 2023, across German-language workshops and English-language online sessions. In the

Figure 1. Artemis, the AI-to-be-explained, with an interface to set parameters (left) and Pythia, the explainer-AI (right). Picture desaturated. Contrast increased for legibility.

sessions, participants were exposed to an AI system consisting of two parts, an AI-to-be-explained and an explainer-AI. Participants were left to figure out the system's workings independently, within a given time limit ranging from 10 to 20 minutes. The data collection is contextualized in a longer-term project on investigating in situ 'AI explains AI' sessions. Our findings reported here focus on the initial English-language online sessions. In line with Suchman's speak-aloud protocols (Suchman 2007), *two* participants were invited per online session, with one participant controlling the interface. Participants were thus encouraged to verbalize their thoughts out loud. The three focal fragments from these sessions presented below are reflective of the broader dataset and were chosen to highlight recurring patterns. The collected video-material and ethnographic notes were analysed using a multimodal conversation-analytic framework (Goodwin 2017).

We aimed to create a configuration that would place the AI-to-be-explained and the explainer-AI next to each other, allowing participants to quickly move between both systems whilst keeping them within a shared view. To facilitate this, we created a website where we placed 'Artemis', our custom-made reinforcement learning simulation, next to 'Pythia', our GPT-4-based explainer chatbot (see Figure 1). Both the chatbot and the simulation were created by members of our team for the purpose of the present research endeavour.

On the left side of the screen, participants were able to observe the dynamic game-like AI Artemis, where hotdogs fall from the sky, and cats move horizontally. The cats 'die' when they collide with the hotdogs, illustrated by an 'explosion'. The cats can move to avoid the hotdogs. The 'score' indicates elapsed time. Once all cats 'die', a new generation of cats is produced from the most successful offspring of the previous generation. With the right settings, the cats will 'learn' how to dodge the hotdogs and thus maximize the score.

It was important to provide a way for participants to interact with the system without turning an autonomous system into a user-controlled game. For this reason, an interface

with sliders was added to give users control over key parameters of Artemis. The sliders themselves were set up in a way that encourages understanding-seeking: some parameters, such as 'Population Size', are more intuitively understandable, while others are entirely opaque (such as 'Top K'), need further context ('Hotdog Interval'), or are related to the underlying algorithm (such as 'Mutation Rate'). Some of this further context may be intuited by changing the parameters and observing the impact on the simulation – including the displayed parameters in the top half of the simulation screen – while others would require in-domain expertise to be meaningful. Artemis was thus set up with varying levels of opacities, mimicking the range of opacities present in the social world, creating possibilities for the generation of explanations whilst not explicitly constraining participants to a single path towards understanding.

On the right side of the screen, we placed the explainer-AI Pythia, a custom-made chatbot set up with knowledge about the purpose and design of Artemis. Its system prompt – i.e., its basic role or personality – is set up to assist the participant in understanding the simulation, resisting attempts to change the subject and engage in off-topic conversations. Based on the specific questions asked by users, additional information is dynamically passed to the AI (such as the source code, specific variables, or other system parameters). Thus, Pythia is positioned as being present for (and knowledgeable of) a discussion surrounding Artemis, occasionally asking follow-up questions and flexibly adapting to the unfolding interaction.

Analysis

In the following sections, we will explore three dimensions, using three fragments that are illustrative of the wider dataset. We employed multimodal conversation analysis to investigate situated sensemaking practices and participant orientations.

The transcripts augment classical Jeffersonian conventions (Schegloff 2007) – along with a Goodwinian (Goodwin 2017) approach to multimodal transcription – with a system for relating recorded mouse movement to ongoing turns. The illustrations on the right display a simplified version of the interface traced from the videodata. Dotted lines represent the direction of movement, and each lowercase 'c' represents a distinguishable stable position of the cursor. The corresponding markings within the transcript, placed above the relevant turns, indicate the position of the mouse relative to speech/silence. The lines indicate the start and end of the movement to the new mouse position. Typing is expressed with a different font with the preceding timing indicating the length of a typing turn. Italics are employed to show reading-aloud practices. We carried over the convention for transcribing overlapping talk across different modalities, with gaps in talk being attributed to a participant in cases of a turn being claimed by a nonverbal modality. For ecological context, we placed a still frame from the videorecording in the top-left of each transcript.

Role-attribution and role-taking

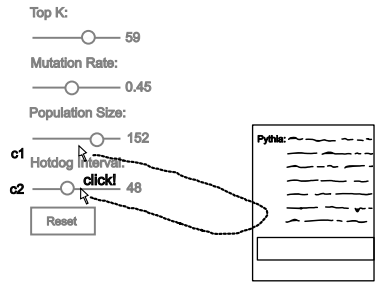
A conversational agent does not automatically create a dialogical interaction. While we, as the creators of the system, may have a clear distinction between Artemis and Pythia,

the way non-human agency was conceptualized in situ varies across participants. While Pythia is named as such, and presented as a dialogue partner in the chat, its connection to the surrounding environment can be conceived in various ways, depending on what is actively highlighted by the participants. The 'highlightable' elements include not only what is directly visible on the interface, but also what is produced as a locally known fact about the activity. More specifically, the game-like activity may be invoked as an explanation of what 'the AI' is. In this focal fragment, the participants – spouses located in the same room (the husband controls the interface) – recruit 'the AI' as an active stakeholder and potentially adversarial player in 'the game' (lines 1–2). Although the wife's question is addressed to her husband (which is evidenced by her gaze shift), he immediately readdresses it to Pythia by moving the mouse to the chatbox (c2–c3) where he starts typing the question. The question is framed by the wife as not being correct (9) revealing her presuppositions about Pythia's role in this activity: you cannot just ask it anything. Here, the activity has been conceptualised as a game where the AI plays an active role and special communication rules are in place.

Tacit knowledge of the internal workings of the AI is (de)constructed through the interaction: while the constrained communication rules have been relaxed (both participants are waiting attentively for Pythia's response and take it seriously), the AI's impartiality is questioned. Although Pythia starts its response with 'As an AI, I don't have preferences', the other part of its response gets highlighted by reading it aloud (11) which results in the conclusion that the AI 'plays for cats'. This conclusion is reached through the co-operative action of three participants where the initial questioning of the nature of an AI by the wife is supported by the husband offloading the question to Pythia, whose response is then selectively used as a resource for interpretation. Pythia's answer is taken as trustworthy, and its agentic status is enhanced by the husband's patterns of engagement: he projects an intent to involve the AI in the conversation by changing his body position from one- to two-handed typing (3–4) – attracting continuous attention from his wife after line 7 – whilst disengaging after the response to the initial question was produced (12). Through his silent participation, he takes the role of a neutral conduit and Pythia is constructed as a trustworthy participant.

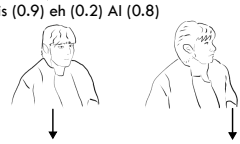
Although we 'know' that there are two AIs in this social situation, the participants do not, and there is no clear evidence that Artemis and Pythia are being distinguished. Yet, their distinctiveness – both in their spatial positioning on the screen and different rhythms – is consequential for the interaction. The division of the interface into two parts creates the possibility for a 'division of labour' between participants: typing takes considerable time, which can be spent by the second participant on observing the other part of the screen. This happens in lines 3-4, when the wife first briefly glances at Pythia and then turns back to Artemis and comments on the events in the simulation.

In sum, the nature of the AI is not given in its source code – it is defined and redefined in interaction. The outputs of the system (both Pythia and Artemis) are selectively used as a resource for sensemaking, leading to an ascription of agency ('AI plays for cats'). At the same time, this ascription is accomplished by actively delegating a response to a non-human. Thus, the AI gets constructed as a participant in interaction, dialogical or not.



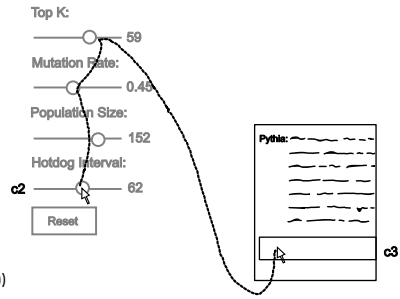
c1 → c2

1 W the question is (0.9) eh (0.2) AI (0.8)



2 does it play for ↑hotdogs or does it play for cat (3.0)

c2 → c3

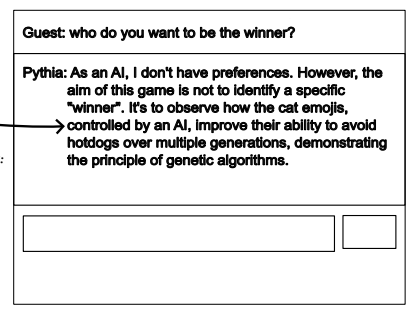


3 H (12.2) who do you want [t] ((typing with one hand))

4 W [because] now hotdogs begin to win a lot.
5 (1.6)
6 H (1.7) ((erases message))
7 W now with what you did ((H erases message))

8 H (10.2) who do you want to be the winner? ((H typing with two hands; W gazes to Pythia))=
9 W =↓no I don't think it's °correct question°
10 (5.6) ((both gaze to Pythia)) ((Pythia responds)) (7.1) ((both gaze to Pythia))

11 W >*cat emojis controlled by an AI improves*< *their ability:*



12 .hh so AI plays fo::r cats.

Transcript 1.

Evidence-building

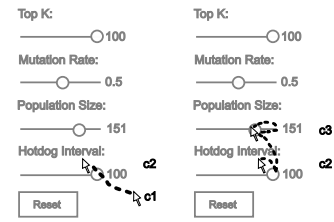
When participants encounter a hitherto unknown system, they may not necessarily orient to an underlying architecture as a focus for making the encounter explainable. A sensemaking situation can vary in terms of what is highlighted as relevant, and what kind of explanation is expected. Even though Pythia was set up with an explanatory mode, there is no guarantee that the material it generated in the chat would be taken as such. Similarly, the displayed scores, slider values, or generated events may either be partially attended to or linked in unintended ways. It may be helpful to think of novel interface interactions as a form of empirical procedure: there are items that can be manipulated, and observations that can be recorded. But how do users establish that their actions have an impact on the game? While users can seek answers from Pythia, they may come into tension with the observations of the interface, as we have seen in Fragment 1. Fragment 2 shows how participants implement previously set methodical procedures of 'evidentiality'. Specifically, the participants have established the reset button as a means of 'locking in' the values on the four manipulable sliders.

The fragment starts with the proposal of Zoe, who does not control the interface, to change the population size. The alignment of Ali is evident in that she goes with the suggestion before it is fully uttered by placing the cursor from the neutral 'home' position c1 to c2 where sliders can be manipulated. She then performs the proposed action and starts moving the mouse down from the slider in question (lines 1–2).

The mouse movements in lines 1 and 2 differ from those in line 3: while the former follow the instruction from Zoe and succeed it, the latter project a next relevant action – pressing the reset button – and precede Ali's verbal question. The reset has been established as delineating different 'runs' of the game with varying values. Before it is initiated, the game runs with previously set parameters. Therefore, there are two possible relevant actions here – to press reset to start a new run or to wait to evaluate the results of the previous one. Both possibilities are sequentially projected by Ali's verbal utterances and mouse movements, while the reset proposal is designed as preferred alternative: it comes first and is projected by the mouse movement, while the "waiting" movement c6–c7 is produced simultaneously with speech (7). Indeed, the waiting option would invalidate the instruction given in line 1.

Ten minutes earlier, Ali formulated the procedure as "When we press 'reset' and put these [the sliders] the same as quick as possible" after which they had to wait to see how their actions affected the score. A similar sequence of actions reoccurs in this fragment. After the button is pressed, Ali sequentially moves the four sliders, starting with the focal population size, and simultaneously verbalises her actions (10). The falling intonation at the end of her turn and the concomitant returning of the mouse to its home position c13 indicate that the active phase of the procedure is finished.

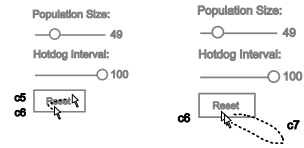
Almost immediately after the manipulations with the sliders, Ali produces the change-of-state token (Heritage 2012) 'oh' with high pitch and increased volume (11), followed by mutual confirmation of their effectiveness. Since no changes were observable at that point, we argue that the change of state referred to the procedure itself rather than its results. In ethnomethodological terms, the activity of the participants has the property of 'first time



1 Zoe well I would try to go with the population size with like maybe fifty?

2 Ali (1.4) so:: (4.8) °okay°

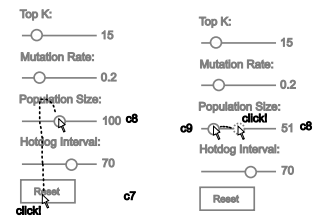
3 Ali (1.7) ((lip-smack)) and [reset?]



4 Zoe [yeh]

5 Ali (1.1)

6 Zoe uh-huh? ((nods))



7 Ali or we wait on- (0.2)

8 Zoe why [not].

9 Ali [okay].

10 Ali (6.6) <fifty:>mutation rate was (1.0) zero point

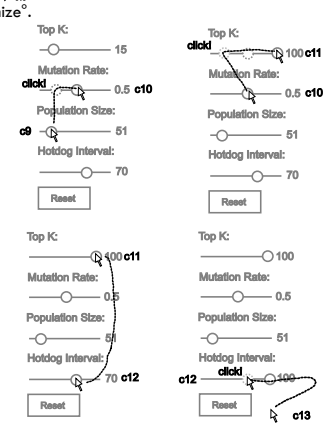
↑ ↑ <five>(0.2) top kay (.) ↑maximize and hotdog interval (0.6) °maximize°.

11 Ali (0.3) OH (.) yeah (0.4) this should

12 Zoe yeah ((nod)) (0.2) now it works

13 Zoe [yeah]

14 Ali [(there are more] hot-)



Transcript 2.

through': it reproduces it as if it were happening for the very first time (Garfinkel et al. 1981). Although the procedure has been already established, it is necessary to 'invent' it once and again. The participants do not only collect observations that they subsequently compare; they are creating the conditions of observability themselves, putting them to the test in a changing environment.

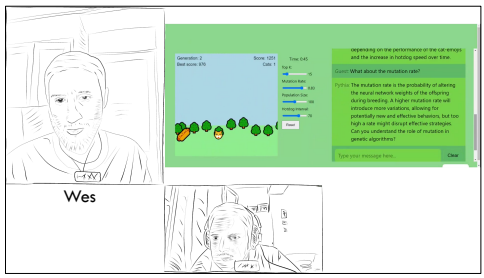
The procedure applied in the analysed sequence has been invented as a methodical tool to establish an evidentiality that is wholly divorced from the 'ground truth' of the system's workings (in actuality, the reset button merely returns all sliders to their default values, and restarts Artemis from scratch). The participants of our experiments used evidence from past encounters with the AI to make sense of new encounters. They applied what Garfinkel calls 'documentary method of interpretation' when specific instances are viewed as 'documents' of underlying patterns (whether they reflect the 'objective' reality or not) (Garfinkel 1967). Across our dataset, multiple sessions featured instances where participants 'instructed' the cats by moving the mouse or pressing the arrow keys on the keyboard. The events on the screen were then evidentially connected to the manipulations, resulting in the (factually incorrect) conclusion that the cats can be trained through direct user input.

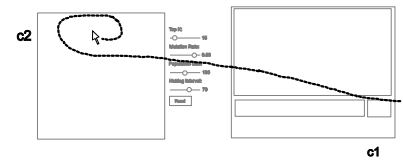
In sum, 'our' actual knowledge of the system's inner workings has tenuous bearing on a situated explanatory encounter and should not be seen as privileged. Instead, care must be taken to attend to the resources that were made available for making sense of the activity, and the possibilities afforded thereby.

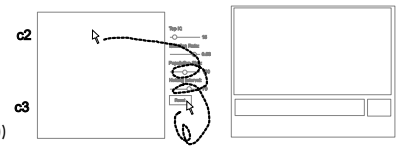
Sequentiality

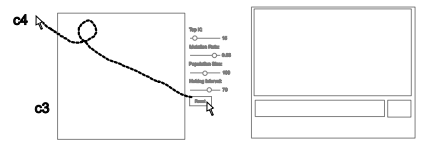
In our first fragment we have shown that the spatial configuration of the interface bears relevance for the distribution of the participants' attention. The temporal arrangement is no less important. Artemis generates observable, highlightable, and evidentializeable events contemporaneously with participant talk: cats are moving, scores are advancing, hotdogs are falling. In stark contrast, Pythia's window remains static unless interacted with. While the chatbox itself can be modified in real-time, a submitted query typically takes 10–30 seconds to generate a response. Pythia itself also never volunteers new responses. As such, Pythia appears to be a more asynchronous participant compared to all other items onscreen.

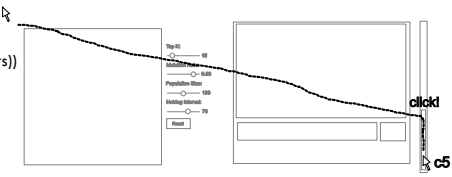
Fragment 3 starts with Wes (who controls the interface) submitting a query to Pythia. The response production creates a notable delay each time, thus giving birth to 'waiting turns' that fill the gap between first and second pair parts of the user-AI interaction (lines 1–5). As in ordinary conversation, prolonged silences can be filled with continuers (4), and there is a possibility to refocus attention to another part of the screen. This happens simultaneously in our case with Wes's cursor traveling to Artemis's window and then to the reset button. The movement c2–c3 projects his next verbal turn – the clarification of the reset function (7–8) established earlier in the interaction (and notably different from the previous case, fragment 2). This turn is augmented with a hand gesture which means that Wes has to disengage the mouse for a moment. As the new gap-filling action is already underway and Pythia's response appears silently, the participants' attendance to it is also delayed. The delay is noted by Dan who asks Wes to scroll down (11) (possibly

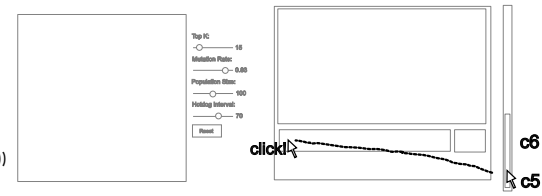












1	Wes	((Send)) (0.1) it's (0.5) °(that)° [one
2		[(1.0) ((scrollbar down))
3		(2.4)
4	Wes	c1 ⟶ c2 mhmm:::
5		c2 ⟶ (1.5)
6		[((Pythia's answer appears))]
7	Wes	c3 ⟶ c4 [so] whenever your press reset basically th- the game starts ((Dan nods))
8		(0.3) all over the- all over [again,] ((circular hands movements))
9	Dan	[all over again] ((Dan nods)) c4 ⟶ yeayeayea yeah.
10	Wes	c5 [ah okay okay]
11	Dan	[(move) that we can] see?
12		(2.6) ((scrollbar down))
13	Wes	°during the [( ) that's (what)°] ((mouse disappears))
14	Dan	°[((inaudible))°]
15		(2.0)
16	Dan	oh it's like a randomness factor I guess
17		(1.5)
18	Wes	mhm [okay.]
19	Dan	[.h] (0.3) introduced into [this game] yes?
20	Wes	[°(probability)°]
21	Wes	°of altering (them)°
22		(6.9) ((both gaze to Pythia))
23	Dan	*can you [under]stand the role of mutation in genetic algorithms* mhm: (0.2) mhm
24	Wes	[mhm]
25	Dan	I- are we £daring£ to say that we understand £that£? hah hah hah
26		(0.4)
27	Wes	wha[hah]
28	Dan	[hah] hah
29		(0.5)
30	Wes	c5 £ah:::£ ((mouse reappears))
31		(0.8)
32	Dan	I'm [not sure (.) of it]
33	Wes	[I- I would say] ((scrollbar down))
34		c5 ⟶ c6 (0.4)
35	Wes	ma- >maybe<, ((typing body position))
36		(0.2)
37	Wes	ah[m:]
38	Dan	[£may]be£ HAH hah hah £somewhat£ [heh heh] hhh
39	Wes	[ehm]
40	Wes	yeah
41		(0.5)
42		To [some ext]
43	Wes	[*to some extent please can you?*]
44		(1.9) ent, can yo[u please el]aborate?
45	Wes	[*you please elaborate*].

Transcript 3.

because he noticed Wes's gestural disengagement) although at that moment Wes's mouse had already returned to the scrollbar (c4–c5). Dan's turn reveals his interpretation that the main activity is 'conversation with Pythia' and the whole sequence in 1–5 can be understood as a multimodally expanded continuer.

Pythia's answer appears as a block of quite complex text that needs to be parsed. This creates another type of a waiting sequence when participants are waiting for each other to finish reading. The private activity of reading/understanding is made public by mumbling (partially reciting Pythia's message, 13–14 and 20–21) and voiced interpretations (16 and 19) which publicize each participant's progress in the activity. Sequentially, Pythia's answer is a multi-unit turn which is decomposed in the succeeding participants' turns. In our case it is divided into two parts: substantive (regarding the understanding of the mutation rate) and communicative (Pythia's question to the participants) which are read top to bottom. In the waiting sequence, Dan first utters his interpretation of the message's content (16 and 19); second, he reads aloud Pythia's question which marks the beginning of a new sequence (formulating an answer to Pythia). The sound quality of the reading-alouds during the two sequences are markedly different: in the waiting phase they are low and mumbling; line 23 is uttered clearly and with normal volume which makes it addressable to the co-participant. Dan explicitly addresses Wes in his next turn (25) which supports the interpretation that line 23 functions as initiating a new sequence.

The answer to Pythia's question is also produced with a delay. Wes reengages the mouse as early as line 30 but sets up the interface (33) and his own body (35) for typing when both participants already expressed their doubts verbally. The end of the fragment gives us the best illustration of how different the temporalities of typing and speech are. Although the typing starts earlier, the verbal projection of a question is uttered much faster. The attempt at 'typing aloud' explains the unfinished turn in line 43 which outran its typed counterpart (42) and its finishing in line 45.

Words, text, and mouse movements inhabit different temporalities. The crossing of modalities creates interactional effects such as waiting sequences which adapt to the asynchronicity of the AI. Even if an artificial agent is treated as dialogical by humans, its contribution to the interaction differs from that of others – it exists in an interactionally different time.

Conclusion

This paper investigated the possibility of using an AI to conversationally explain another AI. Our findings highlight that an explainer-AI, even when it is set up as such, does not become actualized as an explainer unless explicitly recruited by the participants as an agent capable of co-construction. Our case studies illustrate three recurring considerations that affect the interactional positioning of a conversational explainer agent, and the potential of AI-based explainer agents.

Firstly, the assignment of a participatory role is contingent not only on the initial framing of the encounter but can be constructed by the participants in multiple ways. The interface, AI responses, as well as general sensemaking about the activity can be recruited as building blocks for establishing the situational role of the explainer AI.

Secondly, the assumptions about the what of an explanation are determined through practices of evidence-building by the participants rather than being located at the planning stage of the XAI setup. AI-based explainer agents can help bridge the gap between plans and situations by being set up as responsive to unanticipated explanatory needs. In the context of systems like ChatGPT, a re-evaluation of the relationship between plans and situated actions may become relevant. While ChatGPT may not necessarily be a situational agent, it may support more situationally-contingent plans, since the planning model may include instructions about how to handle situationally-contingent open-ended developments.

Thirdly, we highlighted the crucial role of temporality, procedurality, and sequentiality in the explanatory situation. While the explainer AI was shown to be incapable of volunteering synchronous contributions to the explanation, its diachronic nature was drawn upon by the human participants to pursue relevant sensemaking projects. Any ambitions about setting up responsive explainer systems need to take the (a)synchronicity of their agent into account.

Supplemental material

Vectorized figures of the Transcripts 1, 2 and 3 as well as an interactive version of the simulation 'Artemis' are available as supplementary material: https://doi.org/10.17605/OSF.IO/4YMT3

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center TRR 318/1 2021 'Constructing Explainability' (438445824)

References

Bassett C (2019) The computational therapeutic: Exploring Weizenbaum's ELIZA as a history of the present. *AI & SOCIETY* 34: 803–812. doi:10.1007/s00146-018-0825-9.

Button G (1990) Going up a blind alley. In: Luff P, Gilbert N and Frohlich D (eds.) *Computers and Conversation*. London, UK: Academic Press.

Elliott A (2022) *Making Sense of AI: Our Algorithmic World*. Cambridge, UK: Polity Press.

Garfinkel H (1967) *Studies in Ethnomethodology*. Englewood Cliffs, NJ, USA: Prentice Hall.

Garfinkel H, Lynch M and Livingston E (1981) The work of a discovering science construed with materials from the optically discovered pulsar. *Philosophy of the Social Sciences* 11: 131–158. doi:10.1177/004839318101100202.

Goodwin C (2017) *Co-Operative Action*. New York, NY, USA: Cambridge University Press. doi:10.1017/9781139016735.

Han SJ, Ransom KJ, Perfors A and Kemp C (2024) Inductive reasoning in humans and large language models. *Cognitive Systems Research* 83: 101155. doi:10.1016/j.cogsys.2023.101155.

Heritage J (2012) Epistemics in action: Action formation and territories of knowledge. *Research on Language and Social Interaction* 45: 1–29. doi:10.1080/08351813.2012.646684.

Moore RJ, Szymanski MH, Arar R and Ren GJ (eds.) (2017) *Studies in Conversational UX Design*. Cham, Switzerland: Springer. doi:10.1007/978-3-319-95579-7.

Reeves S and Porcheron M (2023) Conversational AI: Respecifying participation as regulation. In: Housley W, Edwards A, Beneito-Montagut R and Fitzgerald R (eds.) *The SAGE Handbook of Digital Society*. London, UK: Sage, pp. 573–592. doi:10.4135/9781529783193.n32.

Rohlfing K, Cimiano P, Scharlau I, Matzner T, Buhl H, Buschmeier H, Grimminger A, Hammer B, Häb-Umbach R, Horwath I, Hüllermeier E, Kern F, Kopp S, Thommes K, Ngonga Ngomo AC, Schulte C, Wachsmuth H, Wagner P and Wrede B (2021) Explanation as a social practice: Toward a conceptual framework for the social design of AI systems. *IEEE Transactions on Cognitive and Developmental Systems* 13: 717–728. doi:10.1109/TCDS.2020.3044366.

Schegloff EA (2007) *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press. doi:10.1017/CBO9780511791208.

Suchman LA (2007) *Human-Machine Reconfigurations. Plans and Situated Actions*. 2nd edition. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511808418.

Author Biographies

Nils Klowait is a multimodal ethnomethodologist studying embodied sensemaking practices in technology-supported interactional contexts. Using videography, he investigates how interaction is co-constructed by diverse participants through at-hand interactional resources, with a strong focus on the interplay between verbal and non-verbal action. Research settings include human-computer interaction, virtual reality, and telemediated communication.

Maria Erofeeva is a social scientist interested in how technologies mediate human action. She has expertise in the fields of multimodal conversation analysis, video-mediated communication, science and technology studies, and human-computer interaction. She currently investigates human social interaction in immersive virtual reality.

Michael Lenke is an education researcher with a foundation in computer science. He focusses on exploring innovative methods for enhancing learning outcomes through technology integration and data-driven approaches by bridging the gap between theory and practice to empower educators and students alike in the digital age.

Ilona Horwath is a sociologist specialized in inter- and transdisciplinary technology research and development. Her expertise includes the sociology of organizations and institutions, gender and diversity studies, sociology of knowledge as well as science and technology studies. She's particularly interested in how technologies promote or prevent social inequalities and discrimination.

Hendrik Buschmeier is a junior professor for Digital Linguistics at the Faculty of Linguistics and Literary Studies at Bielefeld University. Hendrik is interested in empirical and computational modeling of dialogue phenomena, speech and multimodality, and conversational interaction between humans, and between humans and artificial agents.