Postmodern Human-Machine Dialogues: Pedagogical Inquiry Experiments

Marie Bocquelet, Fabien Caballero, Guillaume Bataille, Alexandre Fleury, Thibault Gasc, Norman Hutte, Christina Maurin, Lea Serrano, and Madalina Croitoru

University of Montpellier, France

Abstract. In the classical Human-Machine Dialogue (HMD) setting, existing research has mainly focused on the objective quality of the machine answer. However, it has been recently shown that humans do not perceive in the same manner a human made answer and respectively a machine made answer. In this paper, we put ourselves in the context of conversational Artificial Intelligence software and introduce the setting of postmodern human machine dialogues by focusing on the factual relativism of the human perception of the interaction. We demonstrate the above-mentioned setting in a practical setting via a pedagogical experiment using ChatGPT3.

1 Background and contribution

In the classical Human-Machine Dialogue (HMD) [23] existing research in symbolic Artificial Intelligence has classically been focused on the objective quality of the machine answer (its correctness with respect to the formal representation of the problem at hand, the famous trade-off of expressivity / tractability [21], the speed and memory requirements for finding one / all correct answers etc.). However, one aspect that becomes more and more prevalent in modern days is the quality of the human experience within the dialogue process [29]. This, of course, has a lot to do with what the given answer is, but not only. Perception plays a very important role, as us, humans, are notoriously and hopelessly biased [11]. It has been recently shown that humans do not react in the same way to the same human made or machine made answer when they are aware of the source of the answer to be a human, respectively a machine. For instance, it has been shown that when faced to credit application, humans applicants prefer the approval of the human decision maker as opposed to that of the machine [30]. This concept has yet to be explored in the context of Artificial Intelligence as it is very different, at its core, from the Turing test principle proposed by Alan Turing over 80 years ago [12]. In the vision of Alan Turing the perceived intelligence of the machine was, certainly, made with respect to a human observer, but always within two main differences. First, and very importantly, the original imitation game was a third-party interaction where two entities are observed by a third trying to distinguish given perceived qualities based on their dialogue

(originally gender and then artificial-ness). Albeit highly relevant as setting in the context of post-modern Artificial Intelligence, third party interactions are outside the scope of this paper (as it has been shown that the observation of a process alters the process as such). Second, the test proposed by the famous '50s paper when Alan Turing posed the problem of computers thinking puts the spotlight on how the human observer can be "tricked" into believing its dialogue partner to also be human [27]. But please note what we are interested in this paper is different at the core. The act of tricking a partner into having certain qualities can be possibly pleasant to the partner (in certain conditions) but it is neither a necessary nor a sufficient condition to ensure that this dialogue runs smooth, feels good, natural, interesting and engaging. Instead, we should really focus on the human partner experience of the process and try to measure it. This paper is a call for arms into this yet unexplored but essential aspect of HMD.

Since most work in Artificial Intelligence has been done in the symbolic realm [20], explanation has been a core concept that researchers have put forward (historically, and, recently, with doubled-up enthusiasm) regarding the usability of HMD software [26]. While explanation capabilities in the machine certainly push for the humans to engage more, it also means that the humans overwrite more easily proposed decisions of the machine [28]. Our interest in explanation should go beyond usability and far deeper into how the interaction wholly affects the human decision-making.

In this paper we put ourselves in the modern context of conversational Artificial Intelligence software that currently passes the classical Turing test (such as ChatGPT, Bart, etc.). It is a highly hyped and dynamic setting as, since the launch of ChatGPT in November 2022, not a week goes by without a new Generative Artificial Intelligence (GAI) software being released. Since the beginning of the year Silicon Valley saw more than 500 GAI start-ups newly created; not to mention Meta's LLAMA, Baidu's Ernie, Google's Bard, Anthropic's Claude, GPT 2,3 or 4 etc. All of these softwares fall within the HMD setting and are using as backbone GAI techniques. They are being used by school kids to cheat on their homework, by judges to pass on moral decisions, by researchers to write up papers. Our interest here lies precisely with how the human participant to a two party dialogue with such software experiences the interaction. Against this background, our contribution is twofold:

- Introduce the setting of "postmodern human machine dialogues" by focusing on the factual relativism of the human perception of the interaction.
- Demonstrate the interest of formalising the above-mentioned setting in a practical pedagogical experiment carried out at the University of Montpellier using ChatGPT3.

To conclude this introductory section, we would also like to highlight that the paper is highly relevant to the ICCS community. Since its 1992 kick off workshop [6], the ICCS community gathered a unique blend of logicians, philosophers, mathematicians, and engineers. Such a community would thus be a first class candidate for fostering discussions on our proposal.

2 Proposed setting

The most important motivation for the study of interactional relations stems from a knowledge representation perspective. So far, knowledge representation took the stance of representing things that "are". Ontologies (in their broad sense, even if the term has been widely used in computer since mainly for the past 20 years) were employed as a formal conceptualisation of shared knowledge [17]. Controversially, in this paper, our research hypothesis claims that "shared knowledge" is not a realistic concept in cognitive human interactions. Moreover, such assumption will hinder the development of meaningful interactions from the artificial side. We claim that it is high time for a complete rethink of what a knowledge representation paradigm should encompass in order to aim to achieve successful interactions with humans. The assumption behind a unique and universally valid truth is also the main reason the Turing test is fundamentally different from our proposal. It makes sense that if the human is the owner of the supreme truth, then the only way that the machine can aspire to any truthful behaviour is by imitating the human. But here we do not make such assumption. This is due to two main factors. First, it has been shown by research in neuroscience that the universally truthful perception does not exist [10]. And this not only applies to relative things (such as morality) but also to what we consider "objective" things, such as colour or taste [2]. Second, and closely related, is that the observer will always alter the perception object [14]. Therefore, the perception of the same dialogue will be different from one human to the other. Such differences need to be captured and analysed.

Our proposal for the introduced notion of postmodern Artificial Intelligence relies on the fact that postmodernism is associated with relativism that considers "reality" to be a mental construct [8]. It rejects the possibility of absolute reality and asserts that all interpretations are contingent on the perspective from which they are made. This notion of perspective contingency has been long explored by social sciences, starting from the imago concept of Jung [1]. But the first, partly subconscious image of a given concept, while highly personal and relative, is not enough within the context of this paper. The interaction process of the dialogue, i.e. the engagement, is also crucial for how the perception is being transformed.

Walter Truett Anderson described postmodernism as a world view in which truth is defined through methodical, disciplined inquiry [7]. Peter Drucker suggested the post-modern world is based on the notions of purpose and process rather than a primordial cause [4]. These authors and many more show that our claim of knowledge as a product of the interaction is not novel. Even more recently, the introduction of relational quantum mechanics solved many formalisation problems by making explicit the fact that a universal observer (holder of the truth) does not exist [3]. In their view the state is the relation and the interaction process between the observer and the system.

In this paper, we introduce the setting of postmodern human machine dialogues. Within this setting, for simplification, it is accepted that there is a world but that absolute world it is not accessible to the observers. Instead, every agent (human or artificial) has a personal view on the world which corresponds to the

information it has access to, the language it has for representing this knowledge, experiences etc. This is represented in Figure1.

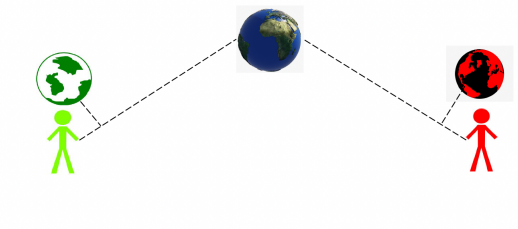

Fig. 1: Relative world representations

Of course, when interacting, we also make a mental model of the agent in front of us (natural or artificial), based on what we think their model of the world is. At their turn, they also make a model about our vision of the world, and, recursively, our vision of their world. Please note that this process is fundamentally different from epistemic logic [15]. The fundamental difference lies in our rejection of universal truth. It is not the case that I know that the agent in front of me does not know that the Earth is flat (while the Earth being flat is an undisputed truth and the lack of knowledge of the agent can be remediated to this effect). In our setting, each agent has a mental model of the shape of the Earth, which is, for them only, the undisputed truth. This is illustrated in Figure 2.

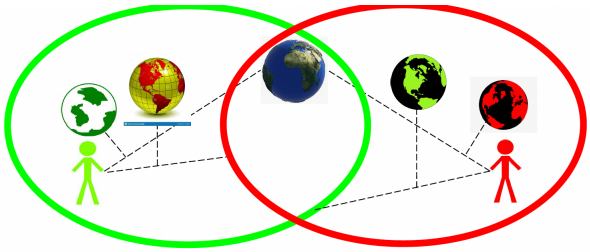

Fig. 2: Recursive relative world representations

In this setting, we claim that argumentation [25] is the main backbone to be used as reasoning facility. This is due to the dual aspect of the argumentation process. On the one hand, argumentation allows for reasoning with inconsistent knowledge (by means of extension based semantics [16] or ranking based semantics [22]) and, on the other, the dialogue process is an inherent part of the argumentation method [5]. While the dialogue will allow for the exchange of information needed within the HMD setting, the reasoning mechanism will

attempt, if required, to restore consistency within the agent's models. This is depicted in Figure 3.

Fig. 3: Dialogue induced alteration of participant world representations

It has been shown that certain argumentation semantics coincide with inconsistent tolerant semantics in logic based models [19]. Dialogue games exist of argumentation semantics, and they have been adapted for reasoning in presence of inconsistency in ontology based data access settings [24]. The dialogue based aspect of reasoning is investigated and practically evaluated in the next section.

3 Inquiry Dialogue based Evaluation

Argumentation dialogues have been long investigated for reasoning with conflicting information [5]. According to Walton, we can distinguish amongst five kinds of dialogues types (that could mix and match during the dialogue between two agents): persuasion, negotiation, deliberation, inquiry, and explanation. All of them have been formalised in terms of turn taking games [13]. In this section, we will illustrate how the inquiry dialogue can be practically evaluated in the newly proposed setting of HMD.

In the proposed experiment, we have used ChatGPT3 in an advanced algorithmic class of 20 Master students in Computer Science at the University of Montpellier (France). The aim of the algorithmic class was for the students to investigate and analyse three classical problems [20]: the TIC-TAC-TOE game, the puzzle 12 game and the cannibals and missionaries problem. The algorithms were investigated over the course of three weeks, one algorithm per week, during a three-hour practical session (no course was available, the students had to use existing resources to learn about the problem and subsequently implement a solution). None of the students had any experience with the above-mentioned problems beforehand.

The students were split into two groups. The first group of students had to implement the above-mentioned algorithms, from scratch, using the programming language of their choice. The second group had to implement the algorithms solely relying on their interaction with ChatGPT3. They were using an inquiry

dialogue for obtaining the code necessary for solving the problem. The postmodern setting of HMD (of Figure 1) was demonstrated by the students having a model of the problem to be solved themselves, having a model of ChatGPT's model for solving the problem and then trying to find the right "prompt" to extract their perceived model from the ChatGPT agent.

All the students in the first group (the business as usual algorithmic class) implemented the games without any difficulty, using either Python or C++ as the programming language of choice during the time duration of the class. One student used JavaScript. The other students, interacting with ChatGPT, reported the following results.

First, all groups managed to implement the TIC-TAC-TOE algorithm without problems. The implementation was done in Python. As a rule of thumb, ChatGPT is much better at implementing problems in Python rather than other languages. The students tried to obtain an implementation of Puzzle12 in C++ that definitely did not work. The same failure was noticed for JavaScript implementation. The main explanation for this, apart from the lack of examples available for learning (which we think it is not the case for JavaScript algorithms) is the fact that a limitation of the number of characters replied is imposed by ChatGPT. For the other two algorithms (the cannibals and missionaries, respectively, Puzzle 12) the success was less flagrant. Actually, for cannibals and missionaries, all the three groups failed to obtain a working code despite numerous tries. For Puzzle 12 two of the three groups managed to get a working Python code.

When the code gets too large, ChatGPT3 "cuts" the program. When asked to continue, the previously started program ChatGPT does not perform correctly. Actually, the higher the number of interactions with the student, the less reliable the answers are. It is clear that the interaction as such is not obtained, ChatGPT behaving much better in a one shot query answering setting rather than HMD. Please note that this is also consistent with the way ChatGPT has been portrayed in publicity (passing the bar exam, passing medical exams etc.), all interactions that are a one shot interaction rather than a elaborated dialogue.

Most intriguingly, the manner the students referred to the software changed within the same interaction and during the weeks. Their interaction was very much cautioned by frustration (due to the time limitation and the requirement of solely using ChatGPT for code) with initial requests carefully formulated and last requests harsh and, sometimes, abusive. A deeper analysis of these phenomena is definitely the first item on our future work, as it fully aligns with our hypothesis of explicitly examining the perceived quality of the interaction by the human.

Apart from illustrating the novel setting proposed by the paper, we believe that, purely from a pedagogical point of view, the use of ChatGPT in this class was actually very beneficial for two main reasons. First, students understood the limitations and eventual benefits of the technology and second, "mistakes" provided by ChatGPT generated code were excellent starting points for in-depth conversations about formal analysis of the algorithms.

4 Conclusion

In this paper, we present a novel human machine dialogue setting based on interaction and the notion of relative truth. The aim of this new view on the interaction between an artificial agent and a human agent lies in the new era of software clearly passing the Turing test, but that require further attention in terms of the quality of the interaction with the human counterpart. Basically, the setting allows attempting measuring when an interaction with a machine would "feel off" to the human. This setting is solely presented in a simplified version in this paper. We can easily extend the setting by considering how the mental model of the human participant is affected if the artificial agent is embodied [18]. We can also extend this work to capture three-person games [9], essential in the context of ubiquitous Artificial Intelligence future.

Bibliography

[1] Carl G Jung and Beatrice M Hinkle. "Symbolism of the mother and of rebirth." In: (1925).

[2] Clara R Brian and Florence L Goodenough. "The relative potency of color and form perception at various ages." In: *Journal of Experimental Psychology* 12.3 (1929), p. 197.

[3] Hugh Everett III. ""Relative state" formulation of quantum mechanics". In: *Reviews of modern physics* 29.3 (1957), p. 454.

[4] Peter Drucker. "Landmarks of tomorrow: A report on the new" postmodern". In: *World* (1959).

[5] Douglas N Walton. "Dialogue theory for critical thinking". In: *Argumentation* 3 (1989), pp. 169–184.

[6] Heather Pfeiffer and Timothy E. Nagle, eds. *Conceptual Structures: Theory and Implementation, 7th Annual Workshop, Las Cruces, NM, USA, July 8-10, 1992, Proceedings*. Vol. 754. Lecture Notes in Computer Science. Springer, 1993. ISBN: 3-540-57454-9. DOI: 10.1007/3-540-57454-9. URL: https://doi.org/10.1007/3-540-57454-9.

[7] Walter Truett Anderson. "The moving boundary: art, science, and the construction of reality". In: *World Futures: Journal of General Evolution* 40.1-3 (1994), pp. 27–34.

[8] Stanley J Grenz. *A primer on postmodernism*. Wm. B. Eerdmans Publishing, 1996.

[9] Anne H Anderson et al. "Multi-mediating Multi-party Interactions." In: *Interact*. 1999, pp. 313–320.

[10] Y Hu and MA Goodale. "Grasping after a delay shifts size-scaling from absolute to relative metrics". In: *Journal of Cognitive Neuroscience* 12.5 (2000), pp. 856–868.

[11] Daniel Kahneman. "A perspective on judgment and choice: mapping bounded rationality." In: *American psychologist* 58.9 (2003), p. 697.

[12] Alan Turing. "Intelligent machinery (1948)". In: *B. Jack Copeland* (2004), p. 395.

[13] Henry Prakken. "Formal systems for persuasion dialogue". In: *The knowledge engineering review* 21.2 (2006), pp. 163–188.
[14] Henry P Stapp. *Mindful universe: Quantum mechanics and the participating observer*. Vol. 238. Springer, 2007.
[15] Hans Van Ditmarsch, Wiebe van Der Hoek, and Barteld Kooi. *Dynamic epistemic logic*. Vol. 337. Springer Science & Business Media, 2007.
[16] Paul E Dunne and Michael Wooldridge. "Complexity of abstract argumentation". In: *Argumentation in artificial intelligence* (2009), pp. 85–104.
[17] Nicola Guarino, Daniel Oberle, and Steffen Staab. "What is an ontology?" In: *Handbook on ontologies* (2009), pp. 1–17.
[18] Guy Hoffman. "Embodied cognition for autonomous interactive robots". In: *Topics in cognitive science* 4.4 (2012), pp. 759–772.
[19] Madalina Croitoru and Srdjan Vesic. "What can argumentation do for inconsistent ontology query answering?" In: *Scalable Uncertainty Management: 7th International Conference, SUM 2013, Washington, DC, USA, September 16-18, 2013. Proceedings 7*. Springer. 2013, pp. 15–29.
[20] Stuart Russel, Peter Norvig, et al. *Artificial intelligence: a modern approach*. Vol. 256. Pearson Education Limited London, 2013.
[21] Stefano Germano, Thu-Le Pham, and Alessandra Mileo. "Web stream reasoning in practice: on the expressivity vs. scalability tradeoff". In: *Web Reasoning and Rule Systems: 9th International Conference, RR 2015, Berlin, Germany, August 4-5, 2015, Proceedings. 9*. Springer. 2015, pp. 105–112.
[22] Elise Bonzon et al. "A comparative study of ranking-based semantics for abstract argumentation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30. 1. 2016.
[23] Stavros Mallios and Nikolaos Bourbakis. "A survey on human machine dialogue systems". In: *2016 7th international conference on information, intelligence, systems & applications (iisa)*. IEEE. 2016, pp. 1–7.
[24] Abdallah Arioua, Madalina Croitoru, and Srdjan Vesic. "Logic-based argumentation with existential rules". In: *International Journal of Approximate Reasoning* 90 (2017), pp. 76–106.
[25] Pietro Baroni et al. "Handbook of formal argumentation". In: (2018).
[26] Tim Miller. "Explanation in artificial intelligence: Insights from the social sciences". In: *Artificial intelligence* 267 (2019), pp. 1–38.
[27] Bernardo Gonçalves. "Machines will think: structure and interpretation of Alan Turing's imitation game". In: (2020).
[28] Davide La Torre et al. "Team formation for human-artificial intelligence collaboration in the workplace: a goal programming model to foster organizational change". In: *IEEE Transactions on Engineering management* (2021).
[29] Ben Shneiderman. *Human-centered AI*. Oxford University Press, 2022.
[30] Gizem Yalcin et al. "How Do Customers React When Their Requests Are Evaluated by Algorithms?" In: *MIT Sloan Management Review* 63.3 (2022), pp. 1–3.