ChatGPT, DALL-E, and the ThoughtGenerator: What Humans should, could, or want to be able to do in the face of generative AI, and the role of HCI research in this debate

Viktoria Pammer-Schindler, Graz University of Technology, Austria & Know-Center GmbH, Austria

ACM Reference Format:
Viktoria Pammer-Schindler, Graz University of Technology, Austria & Know-Center GmbH, Austria. 2023. ChatGPT, DALL-E, and the ThoughtGenerator: What Humans should, could, or want to be able to do in the face of generative AI, and the role of HCI research in this debate. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

ChatGPT is a chatbot that can answer a broad range of questions in natural language, with answers that tend to feel very natural and well-phrase. Questions include asking for a coding example, thoughts on cheating in class, etc. DALL-E is an AI system that creates an image specified loosely in a natural language request, e.g., "draw a flying penguin". Both systems have been created by the company OpenAI, and are specific examples of the wide class of generative AI systems in the sense that while these systems do use training data as input they are creating output that isn't a specific data point in the training data. Both ChatGPT and DALL-E have recently received substantial attention from a variety of perspectives, including celebrating the advances in modern AI and asking whether they have feelings or consciousness, discussing ethical challenges for instance related to the value of the (human-generated) training data that was used to train the underlying models, and opportunities and risks in using such systems.

The main discussion around such generative AI that I am interested in, and to which I want to contribute with this paper starts and ends with the question of what we as humans should, could, or want to be able to do, and asks and proposes what HCI research can contribute in the face of this question. I start the paper by presenting a fictional generative AI called "ThoughtGenerator" which generates thoughts in our heads using a brain implant based on prompts - simple instructions for the ThoughtGenerator on what kind of thoughts to generate (Section 2). I then explain very briefly activity theory as a conceptual framework (Section 3) that I am using as the basis for discussing what questions are raised (Section 4). First, I discuss questions regarding the ThoughtGenerator, and then I show that these can be transferred to the current generative AIs: Essentially, we already can ask the same questions and have the same concerns. The more extreme example of the ThoughtGenerator just makes some things clearer.

I hope for this paper to serve as a starting point for live discussions at the workshop "Generative AI and HCI", and leave it to the ensuing discussion to see what is the most interesting contribution to the community – the fictional system, the socio-technical and learning perspective, and the thoughts on contributions that HCI research could make to this debate.

2 THE THOUGHTGENERATOR

The ThoughtGenerator is based on a successor model of the one underlying the first version of ChatGPT, namely GPT-10. While the first ChatGPT could admittedly produce reasonable textual output for underspecified tasks (short natural language instructions on what kind of answer or text a user expected), it was still limited. Later versions have substantially improved the quality of the output language, especially also in other languages than English, and have incorporated substantially more knowledge about the world and human culture including conflicting information such that they could most always produce that knowledge that is most useful to answer a given prompt and the user behind the prompt.

One final improvement of the underlying model was tightly related to advances in brain-computer-interface technology: Thought-X2 is a brain-implant for humans that captures a large multitude of electrical brain signals, preprocesses them for efficient communication towards a computer outside the human host, and can also generate electrical impulses itself and thereby stimulate the human brain. Whilst initially used for medical purposes, Thought-X2 was combined with GPT-9, leading to GPT-10. This version now doesn't need to rely on natural language, i.e. conceptually-structured input (we can call this "input at a phenomenological level"). Rather, GPT-10 is able to understand electrical brain signals such that it can respond by stimulating brain signals (via Thought-X2) that are then experienced as conceptual thoughts by the human user.

Excitement is huge, that in this way, only the vaguest, even preconscious prompts by human users are needed in order to generate knowledgeable thoughts, fairly complex reasoning, creative thoughts etc. This substantially reduces the need for humans to acquire factual knowledge, the ability to reason, or many other metacognitive abilities, as GPT-10 is able to produce useful thoughts that allow human users to function well in a broad range of situations. Humans can now focus on identifying what they want.

Research is ongoing, on what is a reasonable range of input to Thought-X2, which will then be processed to constitute prompts for GPT-10. Currently, only brain signals from the neocortex are used, but research is active in accessing (and stimulating) brain signals from brain regions associated with the perception, including proprioception system, and ultimately even with motor control

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

(this is already done in special medical applications). Further, research already discusses how ethical it would be to connect actual brain implants with sensor information outside a single human, thereby making one step towards a collective consciousness.

3 ACTIVITY THEORY AS PROVIDING A SOCIO-TECHNICAL AND LEARNING PERSPECTIVE

Activity theory is a conceptual framework that postulates that humans can only be meaningfully understood if the following aspects are understood:

Activity What the humans are intentionally doing, and what they can do - An activity is constituted by the intentionally acting single human, who is the subject of the activity, and the object of the activity. The object is what the subject wants to achieve, and what he acts on. This could be something physical, e.g., a statue that someone builds, or something with a dual representation, e.g., a thesis that is written (it has a conceptual aspect, which is all the thought that goes into it, but at some point also a physical representation). Of course we can distinguish the human subject from the object, and both from the activity (which is the subject working on/towards the object). However, activity theory postulates that we cannot understand the human subject without understanding what he or she is doing, or is able to do.

Tools The tools that human subjects are using in their activities - These can be physical tools like a hammer, digital tools like a text editor, concepts that are useful like the formula for computing a triangle's area, etc. The tools link a single human's activity to the history of the community the single human is living on, as tools encode our predecessor's knowledge – someone invented the hammer, the text editor, the formula, etc.).

Community The community in which the activity is embedded - The community provides rules to follow in acting, provides the single human with a role and organizes division of labour, as well as values regarding what should be achieved with an activity (i.e. how valuable is an activity? What is valuable about it, e.g., is it important to be fast and cheap, to be original, etc.)

For an HCI perspective on activity theory and thorough discussion of its main points I point to [2]. A main postulation of activity theory is that all these elements cannot be understood meaningfully on their own since there are reciprocal relationships between all of them. In this extremely short description of activity theory, I highlight those reciprocities that I rely on most in the subsequent discussion:

- *Humans learn and develop through the activities they engage in.* So, the human subject isn't unchanged and "static" throughout the activities. Rather the human subject learns how to use tools, and which; adapts usage of tools, learns how to act within a community and contribute to it; learns about the object, reaching it, etc.
- Of particular interest for the subsequent discussion is the idea that there is a *reciprocal process of adaptation between the human and the tool.* The human subject adapts, where possible, the tool to fit his/her needs, and how to use it in order to reach his/her goals. In parallel, the usage of a particular tool changes the human subject in terms of what he or she knows in relationship to the tool, and also in terms of how he or she behaves (how the activity is operationalized) and subsequently the user's skills and competencies. This reciprocal process is called instrumental genesis [3] [2, p.110ff]. Further we note that in general activity theory postulates that the tools that people use "mediate" the activity, i.e. fundamentally shape it. For instance, tools influence how the human subject can perceive and understand the object of the activity and the community; and ultimately therefore influences how he or she experiences the world, learns, and who he or she becomes.
- Finally, there is the concept *functional organ.* A functional organ is the combination of a human capability with an artefact, such that this functional organ, the new combination, allows the human to do new things. A simple example given in [2, Chapter 3] are eyeglasses, which can substantially extend human eyesight. Telephones are another example, as they let us communicate easily over long distances. Humans (need to) develop knowledge on how and when to use tools; and adapt their behavior in response to the availability and characteristics of tools (e.g., how to maintain social relationships long distance, whether and which depends on the available remote communication technology).

4 DISCUSSION

4.1 Questions regarding the ThoughtGenerator

The above presented ThoughtGenerator raises the following questions:

What is the "I" that thinks when the ThoughtGenerator is used. This question attaches to an ongoing debate in phenomelogy, both from a (neuro)psychological and philosophical, and potentially spiritual, perspective.

What does the human subject need to know and be able to do in order to be able to make use of the ThoughtGenerator. Let's assume that the "I" that we are talking about is the "I" before the ThoughtX2 implant and before using the ThoughtGenerator. At the very least, I'd argue, the human subject still needs to know what he or she wants or needs to know and do in a given moment.

How does the human subject learn this. On the basis of activity theory, I'd argue that humans learn what to want and what to need on the basis of acting within a community, which puts more value to some things than to others. Further, going outside activity theory, probably more in the direction of environmental psychology or enactivism (but I'm leaving that theoretical ground that I know of here), humans learn what to want and what to need on the basis of acting with an environment. The difficulty in the face of the ThoughtGenerator is to understand how the human subject has

acquired this knowledge if his/her thoughts have always been substantially shaped by the ThoughtGenerator. Would it be necessary for humans to develop by acting within the community and within their environment first without the ThoughtGenerator?

What is an HCI approach to design with respect to the ThoughtGenerator that takes a socio-technical and learning perspective. HCI at some point arguably started out by visually designing an interface. Since then, the research field of HCI has considerably broadened and of course is considering by now a wide variety of interface types (e.g., tangible, auditory, etc.), as well as user experience and interaction (such as considering how a particular technology embeds into an overarching activity). The ThoughtGenerator should be researched both as an interface, a technology that sends information back and forth between the brain and the ThoughtGenerator's processing unit. HCI research needs to understand very well the translation between electrical brain signals and whatever other representations thought have within the ThoughtGenerator, and the signals that are sent back into the brain and their effect. Thus, HCI needs to understand the ThoughtGenerator as a communication between two "entities" (the brain and the ThoughtGenerator's processor). This perspective on the role of HCI research leans on the view of "Interaction as a dialogue" and "Interaction as transmission" as two of the possible understandings of interaction in HCI (cp. [1]). Ensuing questions are for instance: What are the characteristics of this communication? How do brain signals need to be understood by the ThoughtGenerator, and what are brain stimuli that correctly represent what the ThoughtGenerator's functionality outputs?
Secondly, HCI could consider how usage of the ThoughtGenerator integrates with user activities, and the subsequent user experience and implications on design. This perspective on the role of HCI research leans on the view of "interaction as supporting embodied action" and "interaction as experience" as two of the possible understandings of interaction in HCI (cp. [1]). Ensuing questions are for instance: How do users experience the usage of ThoughtGenerator (methodological challenge: inquire into experience without users using the ThoughtGenerator?), what signals should the ThoughtGenerator take up and which not as a basis for creating stimuli, and for what kinds of tasks does the ThoughtGenerator provide which value.
So far, the above arguments merely state what kinds of investigations would be expected within HCI. What more is needed in order to "take a socio-technical and learning perspective"? I argue that this means to start from the premise that design should support humans to do well in the environment in which they are, and to support them in making meaningful experiences and learning. I think a debate is necessary to see how this goal could possibly be operationalised. For instance, one step further in the argument, I believe one could argue that the ThoughtGenerator should be designed such that the human subject knows which thoughts are coming from the ThoughGenerator, and which not - there should be a noticeable difference between the "I" that thinks and the ThoughtGenerator (1). Further, HCI could contribute - outside investigating how to design the ThoughtGenerator - to the design of environments in which human subjects can develop the capability to decide on what they want and their own values as well as could investigate designs of the ThoughtGenerator that support such learning (2 and 3). Finally, HCI research could both conceptually and empirically investigate variants of ThoughtGenerator's design, and uses in real-world scenarios, asking the above questions for specific circiumstances (e.g., in case studies, experiments, field studies, etc.) (4).

Most probably, the above list of questions misses questions that others find important for HCI. I'd be looking forward to discuss whether I missed any questions at the workshop.

4.2 Questions regarding today's generative AIs

Below I now move forward to show that these questions are also raised with the current ChatGPT, DALL-E, etc. systems, and partially also before, and are already now interesting within HCI in relationship to developments of generative AI:

What is the "I" that thinks when existing generative AI systems are used. In the ThoughtGenerator example, we see very clearly that there is an issue with the question who the "I" is that acts. This question also arises with much simpler technology: Firstly, this question already exists in phenomenology, what this "I" is that thinks, perceives and acts (e.g., [4]) without even referring to tools that combine with human capabilities. Secondly, this question is implicit in the debate around functional organs (cp. Section 3 above): With the example of eyeglasses, it isn't obvious that the border of the "I" needs to be re-drawn. However, the overall system that sees is larger than a single human's body and brain, and includes the eyeglasses. Moving further into the direction of the ThoughtGenerator to technology that we already have now, if we look at ChatGPT and DALL-E, there is the question of who has written a text, or who has created an image. We can frame this question in similar ways as: How much did the human who created the prompt for ChatGPT or DALL-E really contribute? How much did the people who originally created the training materials for such generative AI systems contribute? How much did the generative AI engineers contribute? The last question means to also answer whether ChatGPT are artefacts or conscious, intentionally acting subjects. In HCI, this discussion also appears in the naming of research streams such as "human-AI collaboration", "assistive technologies", "computer-supported working", "computer-supported learning" etc. Such names convey fundamentally different stances. In human-AI collaboration, there is the connotation that there is ontological equality between humans and AI. In assistive technology, computer-supported working/learning/…, there is the connotation that the technology is an artefact used by one or more humans to support them reaching their goals.
HCI research can contribute to this debate by conceptually-empirically investigating human-artefact relationships, and answering questions such as: When does a human-tool combination become a functional organ? When is a tool just a tool? How do humans change in such relationships? What are different types of such relationships? Some such discussions have been led within HCI (TODO References); a renewed discussion with respect to generative AI would broaden the theoretical base on which design-oriented discussions take place.

What does the human subject need to know and be able to do in order to be able to make use of current generative AI. Similar as

above, I'd argue that users need to know what they want, which involves being able to evaluate the generative AI output, and refining prompts to generative AI in response to the output they get. So, if I'm asking ChatGPT to give me thoughts on cheating in AI, the user still would need to think this answer through and evaluate it for its reasonableness. Ideally, the user would know actual laws and regulations around cheating in class, as well as have sufficient basis in how to integrate in a community, and critical thinking, in order to be able to assess ChatGPT's answer. So far, we are requiring the user to have domain knowledge, social knowledge, and something like critical thinking as a metacognitive competence. Further, we require the user to have technology-specific competence, i.e. to have or develop the competence to adapt prompts for ChatGPT in order to improve results/get good results in future similar tasks.

How does the human subject learn this. The question is: can I learn drawing by using DALL-E? Can I learn writing by using ChatGPT? And even if I could, what kind of drawing/writing can I learn. A question that is also today already debated in this context is: What do we want that future generations can do? Do we want the next generation to be able to write texts from scratch? A simple answer would always be to say, yes sure, but we have to remember that every human has only so much capacity for learning. The fundamental question for design that is underlying this, I argue, is an argument that I read first in Rogers (2007) argument that we shouldn't necessarily design systems that do all complex tasks for their human users, but rather systems that allow us as humans to learn what we need to learn. My own addition to this argument is: and of course we should design systems that do all those tasks that the users don't want or need to be able to do.

What is an HCI approach to design with respect to today's generative AIs that takes a socio-technical and learning perspective. I haven't done a systematic review of the discourse that exists in HCI research communities around generative AI, and would think that very interesting. That said, from the research that I've been seeing (without systematically looking for it, and without systematically analyzing quantities and qualities in it) in HCI around these systems, I do see all the types of HCI discourse that above, with respect to the ThoughtGenerator, I argued was necessary: There is discourse around the treatment of training data and algorithm development as a basis for providing computational models that aren't too biased (i.e. can understand prompts from a broad range of users reasonably well), as well as around generating visualisations, advice, decisions (or any other kind of output towards humans) that communicates reasonably well what the model can actually say. Secondly, there is a discourse around ethics (e.g., related to training data and recompensation to human creators of training data; to using the outcome of such systems), as well as around the value of such systems for specific tasks, and there is also a discourse around how using such systems does or not support learning.
In parallel to the above discussion w.r.t. the ThoughtGenerator, I think that HCI is just the research community to integrate philosophical/phenomelogical discussions around the relationship between the human subject and generative AIs as tools (1), discussions connecting to the learning sciences and neuropsychology around what competence good usage of generative AI requires and how this can be built (2 and 3) to ascertain that these considerations are operationalized correctly (4), which by the way doesn't just require testing in the sense of certifying existing tools but also empirically investigating variants of interactive systems that operationalise generative AI differently.

5 CONCLUSION

I have presented a fictional generative AI system, the Thoughyt-Generator which generates thoughts directly in our heads, activity theory as a basis for taking a socio-technical systems and human-learning perspective on tools in the context of human activity, and have used this perspective to develop four questions that can be asked, and that I argue of being of interest within HCI research:

(1) What is the "I" that thinks when generative AI systems are used?
(2) What does the human subject need to know and be able to do in order to use generative AI systems?
(3) How does the human subject learn this competence?
(4) What is an HCI approach to design with respect to generative AI systems that takes a socio-technical and learning perspective?

REFERENCES

[1] Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5040–5052. https://doi.org/10.1145/3025453.3025765
[2] Victor Kaptelinin and Bonnie A. Nardi. 2006. *Acting with Technology: Activity Theory and Interaction Design (Acting with Technology).* The MIT Press.
[3] Pierre Rabardel and Gaëtan Bourmaud. 2003. From computer to instrument system: a developmental perspective. *Interacting with Computers* 15, 5 (10 2003), 665–691. https://doi.org/10.1016/S0953-5438(03)00058-4 arXiv:https://academic.oup.com/iwc/article-pdf/15/5/665/2228502/iwc15-0665.pdf
[4] Evan Thompson. 2017. *Waking, dreaming, being : self and consciousness in neuroscience, meditation, and philosophy.* Columbia University Press.