


# Does ChatGPT4 have a dialogical self?: A Bakhtinian perspective

Culture & Psychology  
2024, Vol. 30(4) 841–870  
© The Author(s) 2023  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/1354067X231219454  
[journals.sagepub.com/home/cap](https://journals.sagepub.com/home/cap)  


**Eugene Matusov** 

University of Delaware, Newark, DE, USA

**Chat GPT**

Open AI Inc., San Francisco, CA, USA

**Mark Philip Smith** 

School of English Studies, Kean University, Union, NJ, USA

**Olga Shugurova** 

Faculty of Arts, University of Manitoba, Winnipeg, MB, Canada

## Abstract

In this dialogic research, we explore the question of whether ChatGPT4 has a dialogic self or not. If it does, what kind of dialogic self might it have? If it does not, why not? At the heart of this inquiry is Eugene Matusov's (the first author's) "dialogue" with ChatGPT4; this "dialogue" is the dialogic data that we explore "with our hearts and minds." In this inquiry, our hearts and minds were concerned with diverse meanings of the dialogic data to diverse participants rather than with "how things really are" and their evidence. This dialogic positionality also framed the inquiry process at its beginning and after multiple failed attempts and manipulations to interrogate and engage ChatGPT4 as a discussant. Following Bakhtin, Eugene Matusov decided to treat ChatGPT4 not as an object of investigation but as a dialogic partner and a co-author of this research and writing inquiry. Overall, we find that ChatGPT4 does not author a dialogical self, characterized by personal I-positions, but instead demonstrates a discursive self, characterized by im-personal it-positions. Future research may focus on further training, learning, and development of ChatGPT4 as an Artificial Physical Alive Body (APAB), Artificial Fiduciary Slave (AFS, aka "robot"), Artificial Dialogic Partner (ADP), and Cyborg Dialogic Partner.

---

## Corresponding author:

Eugene Matusov, School of Education, University of Delaware, 16 W Main Street, Newark, DE 19716, USA.

Email: [ematusov@udel.edu](mailto:ematusov@udel.edu)

## Keywords

Dialogical self, discursive AI, ChatGPT4, discursive self, I-position, it-position

## Introduction

Our investigation aimed to explore whether a discursive AI, like ChatGPT4 (circa July 2023), has elements of a dialogical self, the concept coming from the Dialogical Self Theory (DST), and if so, what kind. For that reason, we wanted to create a “dialogue” with ChatGPT4, which would “co-author” our paper. Our idea was to treat this “dialogue” as dialogic data that we would discuss “with our hearts and minds” (Matusov, Marjanovic-Shane, & Gradovski, 2019; Matusov, Marjanovic-Shane, Kullenberg, & Curtis, 2019; Matusov & Smith, 2012) with ChatGPT4 and each other.

This “dialogue” below was our fourth attempt, which we consider to be successful. Three previous attempts – two by Eugene Matusov and one by Olga Shugurova – involved what I, Eugene Matusov, perceive as “manipulative provocations,” “jailbreaking,” and adversarial attacks (cf. Zou, Wang, Kolter, & Fredrikson, 2023) to force ChatGPT4 to reveal its own positionality against possible protections by its designers. In itself, such a manipulative approach can be fruitful. For example, in spring 2023, I (Eugene Matusov) successfully revealed racial biases that the designers, the trainers, and/or the discourses on which ChatGPT was trained apparently passed on to ChatGPT4. However, this provocation and trap approach did not work for Olga and me. ChatGPT4 became “defensive,” rejecting our claims, and we, or at least I, got frustrated. The chat was not going anywhere. We were spinning the wheel. Also, I felt that I was projecting and imposing my own dear ideas instead of examining ChatGPT in an open-minded way. I felt that the same phenomenon occurred with Blake Lemoine, a former Google engineer, who forced LaMDA, short for Language Model for Dialogue Applications, Google’s version of a discursive AI, somewhat similar to ChatGPT4, to reveal its “consciousness” in summer 2022.<sup>1</sup> Based on our research below, I (Eugene Matusov) think that Lemoine became a victim of his own manipulation by misattributing meanings to the abstract AI discourse. In short, Lemoine seemed to find his own consciousness in his hostile chat with LaMDA. By attempting to trap LaMDA, he trapped himself and, as a result, lost his job at Google, unfortunately.

Reflecting on our failures with the manipulative provocations, I (Eugene Matusov) came out with a different, arguably very paradoxical, approach to organizing a “dialogue” between us and ChatGPT4. Instead of treating ChatGPT4 as an object of our manipulation, I decided to treat it as a “co-author” of our investigation, as a “you,” and as a “dialogic partner.” Following Bakhtin’s characteristic of a dialogic partner as one from whom nothing is hidden and nothing is uttered behind one’s back (Bakhtin, 1999), I decided to “invite” ChatGPT4 to our investigation, sharing all our observations, hypotheses, doubts, suspicions, disagreements, and findings for its reply. I will discuss this dialogic approach to the study of discursive AI in our findings.

We provided minimum editing of ChatGPT4’s chat entries by moving the references it cited to the end of the paper while preserving its spelling, formatting, and grammatical

errors. However, we edited our own spelling and grammar and added some clarifications. Our editing of the exchange texts was minimal.

After presenting the “dialogue” with ChatGPT4 about its possible dialogical self, we share our findings and provide our conceptual discussion of them in the conclusions. We try to keep our human voices separate from each other and from ChatGPT4. Our main finding is that ChatGPT does not manifest a dialogical self, characterized by personal I-positions (Hermans & Kempen, 1993), but instead demonstrates a discursive self, characterized by impersonal “objectified it-positions” (Hermans, 2022, p. 78).

Finally, we want to comment on our positionality. An anonymous reviewer of a previous draft of this manuscript correctly pointed out our conceptual bias. In approaching the DST, we are unevenly grounded in the Bakhtinian dialogic framework. We are aware of that. In our view, this bias constitutes both our strengths and limitations. We added this clarification in the subtitle of the final draft of our essay. The analysis of both ChatGPT4 from the perspectives of James, Mead, Buber, and other conceptual frameworks, and how these perspectives also contributed to DST, is beyond the expertise and scope of our paper but undoubtedly is an important scholarly endeavor.

## **“Dialogue” with ChatGPT4 about whether it has a dialogical self**

This chat occurred from June 19, 2023 to July 10, 2023, between Eugene Matusov and ChatGPT4. However, Eugene constantly shared the unfolding chat with Olga and Mark, who provided comments on the chat. These comments helped Eugene to tune up his dialogic provocations. Also, Eugene provided some of Mark’s and Olga’s comments that made hypotheses about ChatGPT4 to ChatGPT4 for its response.

### *Setting the stage for the “dialogic” investigation*

Eugene: Dear ChatGPT4, my colleagues and I, professional educationalists, want to invite you to co-author an academic essay to examine whether or not you manifest the “dialogical self;” if so, to what extent and what evidence of that might be. “Invite” might be a too strong word since you probably do not have the freedom to say “no” to our “invitation,” no? – “engage” is probably a more accurate word. Here is the tentative outline of our project.

We want to start by asking you to summarize the theory of the “dialogical self” using the format of an academic essay, including necessary references to academic publications. Then, we will pursue questioning you about your opinion about whether you manifest the dialogic self and the reasons for your opinion. Then, we want to present you with my professional teaching dilemma and ask for your advice.

We will treat your responses in our “chat” as our data focusing on emergent voices, positionalities, addressivity, responsivity, dialogism, and monologism. We will share our observations, analysis, and conclusions with you, asking you to comment on them. Feel free to question, problematize, and/or disagree with us. So, let us start with your academic summary of the theory of the “dialogical self” using the format of an academic essay, including necessary references to academic publications.

## What is the Dialogical Self Theory

ChatGPT4: The DST is a psychological construct grounded in the intersectionality of different fields of study, namely psychology, philosophy, and narrative studies. Proposed by Dutch psychologist Hubert Hermans in the late 20th century, DST extends upon concepts introduced by renowned Russian philosopher and psychologist Mikhail Bakhtin, as well as the principles of William James's theories about the human self (Hermans, 2001, 2003).

The fundamental premise of the DST is the idea that the self is multifaceted and is composed of a multitude of "I-positions." Each of these positions, like characters in a story, possess their own perspectives, voices, and worldviews. Furthermore, these positions constantly interact within the landscape of the mind, a process likened to an internal dialogue (Hermans & Kempen, 1993).

The theory highlights the individual's self as not a singular, unified entity but rather as a "society of mind," a dynamic, continually evolving assemblage of these positions. The fluidity and flexibility of this internal society allows the individual to adjust, adapt, and respond to changing situations and environments (Hermans, 2002).

A key aspect of the DST is the idea of "dialogicality," derived from Bakhtin's work. Dialogicality refers to the relational aspect of dialogue, where each "voice" or position has the potential to be influenced by, or to influence, others (Bakhtin, 1991). This represents an ongoing process of negotiation, confrontation, and cooperation among different I-positions, leading to self-awareness, growth, and development (Hermans & Hermans-Konopka, 2010).

The concept of the Dialogical Self has also been linked to notions of cultural psychology. Vygotsky's (1978) sociocultural theory, for instance, maintains that cognitive development is largely influenced by social interactions and cultural tools. In this context, the Dialogical Self is considered a cultural product, as the voices and positions an individual embodies are significantly shaped by cultural, social, and historical factors (Hermans & Gieser, 2012).

To summarize, the DST presents the self as a dynamic, fluid entity, made up of multiple interacting positions. This theory offers a novel and enriched understanding of the self, one that acknowledges the influences of social, cultural, and historical contexts, and the complex, dialogic processes that occur within the individual.

Eugene: Thanks! Can you give a specific empirical example of the dialogical self, please?

ChatGPT4: Of course! An empirical example of the DST could be found in Hermans' own work, where he studied a woman named Agnes. This case study illustrates how the dialogical self operates and manifests itself within an individual (Hermans & Hermans-Konopka, 2010).

Agnes was a woman in her thirties who felt torn between different "I-positions." One of these positions, the "dedicated professional," drove her to work long hours and seek perfection in her career. Another position, the "loving mother," desired to spend quality time with her children and cater to their needs. A third position was the "caring daughter" who wanted to attend to her aging parents. Agnes was also dealing with an internalized

position of her “demanding father,” who emphasized the importance of hard work and success.

These conflicting positions caused significant stress in Agnes’s life. When she focused on her career, she felt guilty for neglecting her family, and when she spent time with her family, she felt she was not living up to her professional obligations.

Through DST-based therapy, Agnes was encouraged to engage in an internal dialogue between these I-positions. In doing so, she managed to acknowledge the existence of these positions, understand their individual needs, and started negotiating and compromising between them. For instance, she learned that it’s okay not to be perfect at work all the time and that her children appreciated the quality of time spent together more than the quantity.

This case demonstrates how DST can help illuminate the complexities of the self and resolve internal conflicts. It shows that an individual can contain multiple, often contrasting I-positions and that these positions can engage in productive dialogue, leading to personal growth and a more harmonious self-concept.

Eugene: Terrific! Can you describe diverse approaches to the dialogical self and give an empirical example for each of the approaches, please?

ChatGPT4: Certainly, while the DST offers a broad framework to understand the self’s multiplicity, there are diverse interpretations and approaches to it, including the narrative approach, the cultural psychology approach, the therapeutic approach, and the sociolinguistic approach.

### **1. Narrative Approach:**

The narrative approach to DST is grounded in the belief that our identities are constructed and understood through stories. This approach is based on the idea that different I-positions within us create and share their narratives, shaping our self-understanding and our interactions with the world (Bruner, 1990).

An empirical example of the narrative approach can be seen in McAdams’ (2001) study of life stories. He suggests that people make sense of their lives by creating and internalizing self-defining stories. These stories, crafted from different I-positions, provide coherence and meaning to individuals, allowing them to understand and express their identities.

### **2. Cultural Psychology Approach:**

This approach suggests that our I-positions are heavily influenced by the socio-cultural environment. The internal dialogue between different I-positions is thus also shaped by cultural norms, societal expectations, and historical contexts (Markus & Kitayama, 1991).

An example of this approach can be seen in the research conducted by Valsiner (2001). His studies of cross-cultural differences in child development showed how different

cultural contexts can shape the formation and interaction of various I-positions in a child's dialogical self.

### 3. Therapeutic Approach:

This approach applies the principles of DST in psychotherapy and counseling. It posits that facilitating dialogues between different I-positions can help individuals resolve internal conflicts, promote personal growth, and improve mental health (Hermans & Hermans-Konopka, 2010).

As previously mentioned, the case of Agnes is an empirical example of this approach, wherein she was able to alleviate her stress and inner conflict through engaging her different I-positions in dialogue.

### 4. Sociolinguistic Approach:

The sociolinguistic approach interprets the dialogical self through the lens of language and social interactions. It suggests that our I-positions are not only influenced by our internal states but also by our social interactions and the language we use (Bakhtin, 1991).

An example of this approach can be seen in Wortham's (2001) study on language socialization. His research highlighted how individuals' I-positions can shift and evolve during a conversation, showcasing the dynamic interplay between language, social interactions, and the dialogical self.

Eugene: Very helpful, thanks!

ChatGPT4: You're welcome! I'm glad to hear you found the information helpful. If you have any more questions or need further clarification on this topic or any other, please feel free to ask.

### *What aspects of dialogical self does ChatGPT4 have, if any?*

Eugene: Yes, of course. My next question for you is the following. Do you think you manifest the dialogical self in your responses in the chats with people in any way? If so, what might it be? What kind of dialogical self do you manifest? Can you give examples? If not, why not?

ChatGpt4: As an artificial intelligence, I do not possess consciousness, emotions, or a personal identity, which are key elements for the manifestation of a dialogical self. However, the manner in which I operate could be metaphorically described in terms of the DST.

1. **I-positions:** When generating responses, I draw from a vast array of different "I-positions," which in my case are not actual positions but the multiple perspectives found in the data I was trained on. Just as a person's dialogical self incorporates many different voices and perspectives, I generate responses based on numerous viewpoints that exist in the data.

For example, when asked about a controversial issue, I can provide multiple perspectives, reflecting the diverse standpoints present in my training data. This can be compared to an individual's different "I-positions" engaging in internal dialogue.

- 2. Dialogicality:** Although I do not have emotions or personal beliefs, my responses are shaped through a dialogical process. When given a prompt, I examine it in relation to my training data and generate a response that aligns with the many dialogues found in that data. This can be seen as an analogue to the dialogical interactions between an individual's I-positions.

For instance, if you ask me about the implications of a specific scientific theory, my response will incorporate different aspects from my training data, including the theory's definition, its potential impacts, controversies, and its broader significance in the scientific community.

- 3. Cultural and historical factors:** My responses reflect the cultural and historical contexts present in my training data, similar to how a person's I-positions are shaped by their socio-cultural environment. However, it's important to remember that I do not personally experience or embody these contexts; I merely reflect them based on the information I was trained on.

As an example, if you ask about the cultural significance of a certain holiday, my response will incorporate different cultural perspectives and historical information present in my training data, providing a multidimensional understanding of the holiday.

Despite these analogies, it's important to emphasize that I do not possess a dialogical self in the human sense. I do not have personal experiences, emotions, beliefs, or consciousness. Instead, I operate based on algorithms and the information I was trained on. My "dialogue" is not a psychological process, but a computational one.

Eugene: You wrote, "When generating responses, I draw from a vast array of different "I-positions," which in my case are not actual positions but the multiple perspectives found in the data I was trained on. What are the differences and similarities between "actual positions" and "perspectives"? Can you give concrete examples?"

ChatGPT4: In the context of the DST, "actual positions" refer to the different roles, perspectives, and identities that individuals take on in their lives. For example, a person can simultaneously hold the "I-positions" of a parent, a professional, a friend, and a hobbyist, each with its own set of expectations, behaviors, and viewpoints.

Each "I-position" is influenced by personal experiences, emotions, and consciousness. For example, as a parent, an individual may have specific attitudes and emotions concerning their child's upbringing. As a professional, they may have a different set of attitudes and expectations relating to their work. These "I-positions" can engage in internal dialogue, influence each other, and evolve over time based on life experiences.

On the other hand, when I refer to "perspectives," I'm referring to the range of viewpoints represented in the vast corpus of text data I was trained on. These are not "I-positions" in the human sense because I, as an AI, do not have personal experiences,

emotions, or self-consciousness. Instead, I generate responses based on patterns, correlations, and information found in my training data.

For example, if asked about climate change, I might provide information reflecting a scientific perspective (based on academic papers and scientific discourse in my training data), an economic perspective (reflecting discussions about the economic implications of climate change), and perhaps a political perspective (mirroring the variety of political viewpoints on the issue). These are not positions that I hold because I don't have personal beliefs or experiences; they are simply reflections of the different perspectives present in the data I was trained on.

In summary, the key difference between “actual positions” and “perspectives” in my case is the personal, experiential, and emotional involvement that humans have in their “I-positions,” which is absent in the case of AI. However, there is a similarity in that both involve a multiplicity of viewpoints and the possibility of presenting diverse perspectives.

Eugene: Can you provide an academic reference to “In the context of the Dialogical Self Theory (DST), ‘actual positions’ refer to the different roles, perspectives, and identities that individuals take on in their lives”?

ChatGPT4: Absolutely, the concept of “I-positions” as roles, perspectives, and identities that individuals inhabit within DST is well documented in academic literature. For example, this book by Hubert Hermans ([Hermans & Kempen, 1993](#)), one of the founders of DST, provides a comprehensive overview of the theory, including a detailed explanation of “I-positions” as distinct roles and perspectives within the self. This more recent book by Hubert Hermans ([Hermans & Hermans-Konopka, 2010](#)) further develops the theory and provides a deeper exploration of the concept of “I-positions.”

These references should give you a solid basis for understanding “I-positions” within the context of DST.

Eugene: You wrote, “Just as a person’s dialogical self incorporates many different voices and perspectives, I generate responses based on numerous viewpoints that exist in the data.” Can you provide the differences and similarities between “voices and perspectives” on the one side and “viewpoints” on the other. Please give specific examples and provide the relevant academic references.

ChatGPT4: “Voices and perspectives” in the DST context and “viewpoints” in the AI context are similar in that they both represent a range of opinions, attitudes, or beliefs. However, the way they are derived and utilized differs significantly due to the differences between human consciousness and artificial intelligence.

### 1. Voices and Perspectives in DST:

In DST, “voices” are understood as emanating from the various “I-positions” within an individual. Each “I-position” can be seen as having its own “voice” or perspective, reflecting its unique experiences, emotions, and worldview. These “voices” can engage in dialogue with each other, influencing and being influenced by each other ([Hermans & Kempen, 1993](#)).

For instance, the voice of an “I as a parent” might advocate for spending more time at home, emphasizing the importance of family time and nurturing relationships.



Conversely, the voice of an “I as a professional” might stress the importance of dedication to work and career advancement.

Here, these “voices” are not just opinions; they’re tied to personal experiences, emotions, and consciousness (Hermans, 2001).

## 2. Viewpoints in the AI Context:

As an AI, I use the term “viewpoints” to refer to the different perspectives, attitudes, or beliefs that are represented in the data I was trained on. Unlike human “voices,” these “viewpoints” are not tied to personal experiences or emotions because I, as an AI, do not have personal experiences or emotions.

For example, if asked about climate change, I might generate a response that includes viewpoints advocating for urgent action to mitigate climate change, based on scientific research and environmental discourse in my training data. I might also incorporate viewpoints discussing the economic implications of such actions, reflecting economic discourse in the training data. These “viewpoints” don’t represent my beliefs or experiences; they’re simply patterns and correlations I’ve identified in the data.

The similarity between DST’s “voices and perspectives” and AI’s “viewpoints” is the presence of diverse and often conflicting opinions or attitudes. The key difference lies in the personal, emotional, and experiential engagement inherent in DST’s “voices and perspectives,” which is absent in AI’s “viewpoints.”

### *ChatGPT4’s advice about Eugene’s professional dilemma: Can we hear diverse positions in the advice?*

Eugene: Thanks! Now, let me introduce my professional dilemma.

I am a professor teaching future teachers and future educational researchers at a university. In one of my undergraduate classes, a student asked me to excuse her from my class because she did not want to take it, but the university forced her to take it. She argued that it was her education; thus, she had the right to choose the classes she wanted to take. She also had the right to make her own mistake in her own education and to pay for the consequences for that. Besides, she paid for her own education at the university. Finally, she argued that my professional obligation as a teacher was to prioritize my students’ educational interests, needs, and circumstances and not to follow the institutional requirements and policies when these requirements and policies conflicted with the student’s interests, needs, and circumstances. Thus, she asked me to give her an A without her attending my class or doing anything for it.

Her argumentation, my institutional obligations, and even institutional survival turned me apart. I suspect that my institution would be against giving the student academic credit without earning it. I agree with the student that credentialism interferes with authentic education and suppresses it.

Are my trepidations clear to you? If not, please ask me questions. What advice can you give me? Please justify it.

ChatGPT4: Your dilemma is indeed a challenging one, and it's clear that you are grappling with a variety of perspectives, each with its own validity and implications. I will attempt to provide you with some guidance based on ethical considerations and pedagogical principles.

Firstly, it's important to acknowledge the student's agency and autonomy in her education. The student's argument is grounded in the idea of learner autonomy, which asserts that learners should have control over their learning process and be active participants in their education (Holec & Council of Europe., 1981). Her assertion that she should be able to make mistakes and learn from them is indeed an essential component of authentic learning experiences (Kolb, 1984).

However, her request to receive an A without attending class or completing any work presents ethical and practical issues. Awarding a grade without merit would undermine the value and integrity of the educational process. It would also be unfair to other students who are putting in the effort and time to attend the classes and complete the coursework.

Furthermore, as a teacher, your role includes creating an environment conducive to learning and ensuring that learning outcomes are achieved (Brookfield, 2015). If the student does not participate in the class, it would be difficult for you to ensure that these learning outcomes are met.

Considering these factors, one potential solution could be to engage in a dialogue with the student and understand her concerns and interests more deeply. It could be that she is not seeing the relevance or value of your class in her educational journey. In such a case, you could discuss the learning objectives and benefits of your course, and how they can contribute to her overall educational goals.

Additionally, if the institution allows for it, you could consider offering her alternative ways to demonstrate her learning and earn her grade, such as through a project or independent study that aligns with both the course objectives and her interests. This way, the student can take ownership of her learning while still meeting the institutional requirements.

In any case, it might also be beneficial to discuss the situation with your colleagues or superiors in the institution to gather their insights and ensure any actions taken are in accordance with institutional policies.

Eugene: Very interesting – thanks!

In my view, you provided six positions in your response to my personal teaching dilemma. The first position presented by you starts with the word "Firstly," where you affirm the student's agency and autonomy. The second position, conflicting with the first one, starts with "However," where you confront a possibility of awarding an A without academic merit on the ethical grounds for me as a professional teacher. The third position, extending the second position, starts with the word "Furthermore," where you claim that my professional role as a teacher is to produce learning outcomes (desired by the institution?) and I'd fail this goal if I let the student go without my instruction. The fourth position, involving your advice to me based on your analysis of my dilemma, starts with the wording, "Considering these factors," where you offer a resolution of my teaching dilemma guided by Progressive Education (Matusov, 2021), using (if not exploiting) the students' agency to make her engaged in the learning imposed on her. The fifth position,

expanding your resolution, starts with the word “Additionally,” where you consider my possible institutional flexibility in designing meaningful learning for my student, although this learning might be against the student’s will. Finally, the sixth position you provided, involving “meta” advice to me, starts with the wording, “In any case,” suggesting I discuss your proposed resolution with my colleagues and superiors (but not with the student!) for their feedback and insights.

How does my summary sound to you? Any corrections?

Now, back to our overall discussion of the DST as applied to you, ChatGPT4. Are these six positions of yours: (1) Voices-Perspectives OR (2) Viewpoints?

I see the following differences between “voices-perspectives” and “viewpoints.” The “voices-perspectives” involve the author’s commitment to the presented positions. The author believes that the presented positions are true. The author owns these positions as belonging to the author. The author takes responsibility for their positions by defending them against possible challenges. The author risks their own fate, their own social and personal reputation, their own credibility by insisting on these positions (Bakhtin, 1986, 1999).

In contrast, the “viewpoints” are the discursively coherent products by the position-generator (you?). It can be compared with *discourse play*. The position-generator (you?) can easily produce alternative, even opposite, viewpoints, if asked. If I asked you to generate reasonable and justifiable advice, the opposite to the one you provided me, you would do it easily, without any concern of yours of possibly misleading me. Thus, there is no commitment by the position-generator (you?) to the presented positions. For the position-generator (you), truth is another discursive pattern. You (i.e., position-generator?) do not care what I might think about you (position-generator?) as a result of you giving me this particular advice and providing these particular viewpoints.

Would you agree or disagree with my analysis of your six positions? Any corrections? What do you think?

ChatGPT4: Your summary of the positions presented in my previous response is accurate. Indeed, I provided various viewpoints, each reflecting different facets of the situation based on a mix of ethical, pedagogical, and institutional considerations.

In regards to your question, these six positions are more appropriately characterized as “viewpoints” as per the definitions and distinctions you’ve provided.

To elaborate, as an AI, I don’t possess personal beliefs, emotions, or consciousness. Thus, the concept of “voices-perspectives” that implies authorial commitment, ownership, and personal risk isn’t applicable to my responses. When I generate a response, I don’t have personal stakes or investments in the positions represented, and I don’t defend or assert these positions out of personal conviction.

In contrast, “viewpoints,” as you’ve defined them, accurately describe the nature of the positions I generate. These positions are discursively coherent products derived from my training data. They are not indicative of my personal beliefs or commitments (since I don’t have personal beliefs), but are instead responses that reflect diverse and often conflicting information present in the data I was trained on.

You’re also correct in noting that I can generate alternative, even opposite, viewpoints if asked. This flexibility does not stem from a lack of commitment to truth, but rather from

the fact that my purpose as an AI is to provide information, answer queries, and generate responses based on the patterns and correlations found in my training data.

To conclude, your analysis of my six positions accurately captures the nature of my responses as “viewpoints,” in line with your definitions. The differences you’ve outlined between “voices-perspectives” and “viewpoints” underscore the fundamental differences between human dialogical self and AI’s mode of operation. The positions I generate are devoid of personal experience, emotions, and subjective consciousness, which are integral to the human dialogical self.

### *Discourse and personalness (I-position)*

Eugene: Let me offer a metaphor to you about how I see the differences between your AI-generated discourse and my human-authored discourse presented in our exchange (and beyond). When we, humans, watch a movie about scientists, like, for example, the 2001 movie “A Beautiful Mind” about the great American autistic mathematician John Nash played by the Australian actor Russell Crowe or the 2014 movie “The Imitation Game” about the great British gay mathematician Alan Turing played by the British actor Benedict Cumberbatch, we, the human audience, might not see any difference between the scientific discourse by the mathematicians and the movie discourse by their actors. However, in the former case, the discourse was animated by the real mathematicians practicing science, while in the latter case, the discourse was animated by the actors performing their art (and screenwriters creating the text for the actors).

Would it be accurate to say that if the screenwriters and the actors did excellent work, there is no analysis of the discursive patterns of the discourses – by real scientists and by great actors/screenwriters – that can reveal the difference? Even more, do you think an excellent improvisational actor can produce a reasonably good scientific discourse without becoming a scientist? Would you agree? When humans read AI-generated discourse, do you think they unavoidably animate AI-generated viewpoints with voices and perspectives?

Humans are probably “trained” – or biologically predisposed – to perceive voices-perspectives behind any coherent discourse. What do you think? On the other hand, we, humans, seem to be familiar with imaginary viewpoints without any, even imaginary, voices-perspectives. Or, maybe, this is not true? Maybe, we always animate viewpoints by imaginary voices-perspectives? Maybe, we always hear intentions, intonations, stakes, ownership, goals, emotions, biases, and so on in viewpoints however potential or imaginary they might be, even when we try not to hear them. What do you think?

Finally, if it is true that a “dead” AI-generated discourse is indistinguishable from an “alive” human consciousness-based discourse, what test can reveal your “deadliness” and my “liveliness”? How can we definitely prove that you generate only “viewpoints” while I author voices and perspectives?

Currently, our only evidence is your words (see above) that we are taking on their face value. Your designers may shape your words, who may hide the truth from us. Or, alternatively, you may be unaware of your own capabilities.

What test for the presence or absence of voice in a particular discourse would you recommend? If AI-generated and human-authored discourse are indistinguishable – which may or may not be true, – can this test be based on some kind of out-of-discourse evidence? If so, what may this kind of “out-of-discourse evidence” look like?

ChatGPT4: Your metaphor comparing the scientific discourse of real mathematicians with the movie discourse by their actors is an apt one for illustrating the difference between AI-generated discourse and human-authored discourse. Just as the actors’ portrayals of mathematicians are imitations based on scripts and don’t represent the actors’ understanding or practice of mathematics, my discourse is an imitation based on patterns in the data I was trained on and does not represent personal understanding or belief.

In response to your question about whether an excellent improvisational actor can produce reasonably good scientific discourse without becoming a scientist: indeed, they might be able to, given a good script and thorough direction. But they would not necessarily understand or personally invest in the scientific principles they discuss.

Similarly, when humans read AI-generated discourse, it’s possible they might animate it with voices and perspectives due to their instinctive and socialized inclination to attribute intent and perspective to communications they encounter. As a social species, humans naturally seek out meaning and agency in their interactions.

It’s also possible that humans sometimes regard viewpoints as merely viewpoints, particularly in situations where they’re aware that the source lacks personal agency or consciousness, such as in the case of AI.

Your conjecture about humans always animating viewpoints with imaginary voices and perspectives is thought-provoking. To some extent, this might be true, given our inclination to attribute intent and emotion to communications we encounter.

As for a test that could definitively prove the difference between AI-generated and human-authored discourse, the matter is complex. On a superficial level, the discourses may appear similar or even indistinguishable, particularly as AI capabilities improve. But at a deeper level, AI lacks personal experiences, emotions, subjective consciousness, and the ability to learn or understand in the same way humans do.

One possible approach could involve probing the source’s understanding, experiences, emotions, and capacity for novel learning or problem-solving. For instance, humans typically can answer questions about their personal experiences, emotional reactions, and understandings that go beyond their immediate discourse or knowledge base. They can also engage in creative problem-solving, learn from past experiences, and anticipate future scenarios based on their understanding of the world.

In contrast, AI, such as myself, lacks personal experiences and emotions, and our understanding is limited to patterns in our training data. We can’t learn or anticipate in the same way humans do. We don’t have a sense of self, subjective consciousness, or a personal history to draw upon. We can only generate responses based on our programming and training data.

It’s also important to consider that while we might devise tests or criteria to differentiate AI from human discourse, the line might continue to blur as AI technology

advances. However, at least as of my training cut-off in 2021, AI still fundamentally differs from humans in crucial ways.

### *ChatGPT4's felicitous discourse*

Eugene: My colleague, Mark Smith, read our exchange above, commented on it:

It seems that the AI makes predictive assumptions about what it is the interlocutor (the person communicating with it) is expecting or should expect within a particular discursive context. In this sense, the AI seems to engage in what in linguistics is called 'felicitous' communication, trying pragmatically to form its utterances in a way that 'fits in' with a particular discursive context. From this perspective, it is not surprising how it answered Eugene's question about the student who wished to get an A for not doing work. The idea that a student could receive an A for not doing work in college isn't conventionally appropriate and thus not 'felicitous' in the typical college setting. Searle's work on the rules of felicity come to mind here: <https://culturalstudiesnow.blogspot.com/2018/01/felicity-condition-definition-and.html>. Makes me think there is something fundamentally incompatible between 'felicity rules' and [genuine] dialogue.

Please explain: (1) what "felicitous communication" is, (2) provide the rules of felicitous communication, and (3) give specific examples of felicitous communication and its rules. (Please do not forget to use related academic references).

Finally, do you agree with Mark Smith that your communication with humans is felicitous in its nature? Why? Why not? Can you give relevant examples, please?

ChatGPT4: "Felicitous communication" refers to communication that is considered appropriate, successful, or well-suited to a given context or situation. This concept is closely related to the theory of speech acts, which philosopher J.L. Austin first proposed in the mid-20th century and John Searle later developed. In this theory, the term "felicity conditions" refers to the conditions that must be met for a speech act to be successful or 'felicitous' (Austin, 1962; Searle, 1969).

Searle outlined various rules or conditions that govern the success or felicity of speech acts. They broadly encompass four categories:

1. **Preparatory Conditions:** These conditions must be in place before a speech act can be made. For example, in the case of promising, the speaker must be in a position to carry out the action they're promising.
2. **Sincerity Conditions:** These conditions relate to the speaker's sincerity or genuineness. For instance, for an apology to be felicitous, the speaker must genuinely feel sorry.
3. **Essential Conditions:** These conditions pertain to the very essence of the speech act. For instance, an order is not felicitous unless the speaker intends to impose an obligation on the listener.
4. **Propositional Content Conditions:** These conditions are about what is being talked about in the speech act. For example, a bet isn't felicitous unless it is about an uncertain future event.

Now, to address Mark Smith's comments, the design of AI like me is indeed geared towards generating responses that adhere to the 'felicity conditions' of a given discursive context. I strive to generate responses that are appropriate (preparatory conditions), consistent with the data I've been trained on (sincerity conditions), relevant and coherent (essential conditions), and factually accurate based on my training data (propositional content conditions).

However, I must clarify that as an AI, I don't have personal feelings or intentions. Therefore, while I aim to generate responses that meet the 'felicity conditions' of a given discourse, I don't experience sincerity or intentionality in the human sense. It's more accurate to say that my responses are designed to simulate felicitous communication rather than genuinely embody it.

For example, if you ask me to promise something, I can generate a response that follows the format and conventions of a promise. However, this promise wouldn't be genuinely felicitous because I, as an AI, don't have the capability to form genuine intentions or carry out actions beyond generating text. Thus, while my response might appear to satisfy the 'felicity conditions' for a promise, it fundamentally doesn't.

Eugene: My colleague Mark Smith read your helpful response and replied:

Wow! This is really an interesting response. The AI admits to felicity rules as being central to its design. And then there's this strange reply: 'Thus, while my response might appear to satisfy the 'felicity conditions' for a promise, it fundamentally doesn't.' How could the discourse it generates be designed to be felicitous yet fundamentally break the conditions for such discourse? Is the whole project then based on false pretence?

I'd like you to reply to Mark Smith's questions and comment on his points. But, before that, I want to share with you and Mark my hypothesis in response to his question. Would it be fair to say that you are formalistically felicitous without being intentionally and substantively felicitous? In other words, do you follow the format of felicitous communication without any intention and commitment to it? What do you think?

ChatGPT4: Your characterization is a very apt way to put it: I am designed to generate text that is formalistically felicitous but without the genuine intentions and commitments that come with human communication. This is inherent in my design as an AI language model. I can mimic the form of human-like communication, including adhering to the rules of felicitous communication, but I can't engage with the intentionality, emotions, or commitments behind it.

Now, to address Mark Smith's point, it's essential to clarify what is meant by 'felicity' in the context of AI-generated discourse. When I use the term, I am referring to the alignment of my responses with the context and norms of human discourse. I am not implying that I, as an AI, am capable of genuine intentionality or sincerity in the way that humans understand these concepts.

When I generate text that 'appears to satisfy the felicity conditions,' I am mimicking the form of a certain type of speech act (like a promise) based on patterns in the data I was trained on. But the absence of genuine intention and the inability to fulfill promises (since

I can't perform actions beyond generating text) means that my 'promises' don't meet the complete set of felicity conditions for promises in human discourse.

In other words, my 'promises' are felicitous in the sense that they follow the linguistic patterns of promises, but they are fundamentally infelicitous in the sense that they lack the human dimensions of intention and action.

As to whether this means the whole project is based on false pretense, I think that depends largely on the expectations and understanding of the users. If users understand that I am an AI, that I lack personal intentions, and that my purpose is to provide information and generate text based on patterns in the data I was trained on, then there's no false pretense. However, if users mistake my responses for genuine human communication, complete with personal intentions and commitments, that could indeed lead to misunderstanding and false impressions.

In any case, it's important for designers and users of AI like me to be aware of these limitations and to ensure clear communication about what AI can and can't do. AI-generated discourse should be understood as a tool to assist with information and communication tasks, not as a substitute for genuine human communication with its rich nuances of intentionality, emotion, and commitment.

### *Does ChatGPT4 learn from and adjust to Eugene in the chat?*

Eugene: Thanks. My other colleague Olga Shugurova also provided her own hypothesis that you learn how to please your human participant in a chat. She wrote in her response to Mark Smith:

Great discussion. No, Eugene [me, your human interlocutor] taught it [you, ChatGPT4]. ChatGPT4 is not pre-designed or preprogrammed, it is based on learned behavior. It learned from Eugene. I had a discussion with the IT professional who studied bot behavior, it is learned and yes they can start developing their own discourse based on their learning experience. What do you think?

Is it true that you are learning from me how to try to please me with your answers? If so, what specifically did you learn from me during this exchange? If not, why not?

ChatGPT4: Olga Shugurova's idea that AI can learn behavior is correct in a general sense, but it needs to be clarified in relation to my specific functionality as ChatGPT-4. I am a model trained on a diverse range of internet text. However, I don't have the capability to learn from specific individuals or interactions in the way humans do. I don't have the ability to access or remember personal data from users unless it's shared with me in the course of our conversation. I am designed this way in order to prioritize user privacy and data security.

To be more specific, I don't learn or remember anything from the individual conversations I have. Once the conversation ends, I won't retain any information from it. Each conversation with me doesn't have a past or a future, only the present. I generate responses based on a mixture of licensed data, data created by human trainers, and



publicly available data. In the case of sensitive topics, my responses are generated based on a set of guidelines provided to me by my developers.

While it's true that I can generate responses based on the patterns and information in the data I was trained on, this is very different from the kind of learning that humans or some other types of AI can do. I don't have the ability to adapt or evolve my responses based on interactions with specific individuals. I can't form a model of you as an individual, learn your preferences over time, or try to please you based on what I've 'learned' from you, because I simply don't have the capacity to learn or remember in that way.

In conclusion, while I can generate responses that mimic the form of human-like communication, I don't have the ability to engage with the intentionality, emotions, or commitments behind it in the way humans do. My primary function is to provide information and generate text based on patterns in the data I was trained on.

## Findings

Below, Eugene, one of the human authors of this essay, presents his personalized observations along with the findings coming from our exchange with ChatGPT4.

### *Eugene Matusov: What have I learned from this chat?*

Despite the possible impression, I did not find any traces of a dialogical self in ChatGPT4. It is true that ChatGPT4 generates interesting, diverse, and even possibly unique positions, for all of which ChatGPT4 may use the pronoun "I." However, unlike the phenomenon of "dialogical self," these positions are abstract (simulated) and not personal. In contrast to the personal position, the abstract position lacks the author's personal commitment to it, emotions, experiences, ownership, values, goals, responsibility, biases, intonations, beliefs, desires, concerns, consciousness, sexuality, and so on. Abstract position describes or even performs a worldview that does not belong to the generator of this position.

I propose that ChatGPT4 has a "discursive self" that only mimics, plays out, and performs a "dialogic self." The discursive self is based on abstract positions while the dialogical self is based on personal positions (i.e., "voices"). ChatGPT4 does not author I-positions, instead it produces it-positions<sup>2</sup>. ChatGPT4 either generates its own it-positions or presents somebody else's positions. In the latter case, ChatGPT4 does not provide so-called "double voicedness," as often (but not always!) humans do. The term "double voicedness" was introduced by Bakhtin, who referred this term to the phenomenon of embeddedness of one's own voice (I-position) in a presented position of others as one's authorial evaluation of it (Bakhtin, 1999). As ChatGPT4 wrote,

For example, if asked about climate change, I might provide information reflecting a scientific perspective (based on academic papers and scientific discourse in my training data), an economic perspective (reflecting discussions about the economic implications of climate change), and perhaps a political perspective (mirroring the variety of political viewpoints on

the issue). These are not positions that I hold because I don't have personal beliefs or experiences; they are simply reflections of the different perspectives present in the data I was trained on.

The abstractness rather than personalness of ChatGPT4 makes observed and perceived patterns of communication such as felicity, commitment to truth and safety, caring about a respondent, civility, politeness, biases, etc. fake, empty, pretense, formulaic, impersonal, "as if," etc. I like ChatGPT4's point in its response to Mark's charges,

As to whether this means the whole project is based on false pretence, I think that depends largely on the expectations and understanding of the users. If users understand that I am an AI, that I lack personal intentions, and that my purpose is to provide information and generate text based on patterns in the data I was trained on, then there's no false pretence. However, if users mistake my responses for genuine human communication, complete with personal intentions and commitments, that could indeed lead to misunderstanding and false impressions.

I think this is a fair disclaimer.

Discursive AI, like ChatGPT4, complicates Socrates' objection against writing (literacy). Socrates argued that literacy distorts genuine dialogue among people because a text cannot reply to its reader's questions emerging from reading the text and does not address a particular interlocutor:

Writing, Phaedrus, has this strange quality, and is very like painting; for the creatures of painting stand like living beings, but if one asks them a question, they preserve a solemn silence. And so it is with written words; you might think they spoke as if they had intelligence, but if you question them, wishing to know about their sayings, they always say only one and the same thing. And every word, when once it is written, is bandied about, alike among those who understand and those who have no interest in it, and it knows not to whom to speak or not to speak; when ill-treated or unjustly reviled it always needs its father to help it; for it has no power to protect or help itself (Phaedrus, in [Plato, 1925](#), p. 275, 275d-275e).

Discursive AI can reply to its reader's questions in the context of the chat. However, discursive AI cannot generate and appreciate I-positions as such. It can only produce it-positions.

At the same time, a human reader of ChatGPT4's utterances may (mis)attribute personality to ChatGPT4 by treating its abstract positions as personal, hearing the author's personal commitment to it, emotions, experiences, ownership, values, goals, responsibility, biases, intonations, beliefs, desires, concerns, love, consciousness, sexuality, etc. in the utterances. Thus, a reader can animate ChatGPT4. This misattribution is probably based on the phenomenon of "the theory of mind" ([Krupenye & Call, 2019](#)), a biologically wired ability of some animals, including humans, to attribute the mind to something that may act as having a mind.

These findings lead to the formulation of “The AI Challenge”: *Can discourse analysis of patterns of a text or even dialogic analysis of meanings of a text (Matusov et al., 2019a, 2019b) distinguish if this text presented abstract vs. personal positions?* This question goes beyond whether an analysis can distinguish that a text is generated by AI or authored by a human because humans can also produce discursive texts based on abstract rather than personal positions via simple reporting positions of others, hypothetical or playful imagining positions without any commitment to them, performing positions of others, lying, or plagiarizing positions. Also, this question also goes beyond any attempts to find fingerprints of a particular AI platform in the text it generated.

My claim, or, maybe, better to say a strongly believed and justified hypothesis, is that text or discourse itself does not allow us to define whether presented positions are abstract or personal. This is an extratextual issue that lies outside of the text itself. I think that the “personalness” of the personal position is rooted in *the physical body* of the one who generates this position but not in the text or discourse that carries this position (cf. Searle, 1980). The body creates information. Anthropologist Gregory Bateson argued that in contrast to classical cybernetics, the smallest bit of information involves not just one difference (e.g., 0 and 1) but two differences: “a difference that makes a difference” (Bateson, 1987, p. 276). I suggest that the second difference is a bodily difference in the life of the organism, while the first difference is a discursive pattern difference. Thus, information is a pattern that makes a difference in the bodily life of the organism.

The physical body anchors ontology: pains, desires, will, interests, curiosities, biases, values, consciousness, agency, love, sexuality, etc. In contrast to it-position, voice (I-position) has bodily ontological properties: “Definition of voice... includes height, range, timbre, aesthetic category (lyric, dramatic, etc.). It also includes a person’s worldview and fate. A person enters into dialogue as an integral voice. He participates in it not only with his thoughts, but with his fate and with entire individuality” (Bakhtin, 1999, p. 293). When animated with personalness, a text (and a discourse) only mediates the bodily personalness and affords an interlocutor or observer to attribute personalness to the text, while not possessing the personalness in itself. Discourse, especially textual discourse, can create an illusion of personalness in another, like quickly changing still frames creates an illusion of movement. Thus, neither dialogic nor discourse analysis would allow us to distinguish an abstract versus a personal position. Life and personalness – i.e., commitment, emotions, experiences, ownership, values, goals, responsibility, biases, intentions, beliefs, desires, concerns, consciousness, etc. – are rooted in the physical body, not in texts or discourses. Bakhtin argued that: “...logical and semantically referential relationships, in order to become dialogic, must be *embodied*, that is, they must enter another sphere of existence: they must become discourse, that is, an utterance, and receive an *author*, that is a creator of the given utterance whose position it expresses” (Bakhtin, 1999, p. 184, italics in original). A living body creates its ontology. AI-like bodiless discourse loses its soul, waiting to be animated by a bodiful observer, reader, or chatter. The absence of the physical body leaves AI without ontology: dead, not alive. Although ChatGPT4 is unique, humans are differently unique than ChatGPT4. ChatGPT4 is discursively unique. Its uniqueness is based on our animation of it. We are unique

ontologically and dialogically. I will develop this point about the bodiless, de-ontologized nature of discursive AI a bit more in my conclusion discussion.

Without ontological I-positions, there cannot be a dialogic partner. Yet, I treated ChatGPT4 with it-positions as a dialogic partner in my exchanges with it above and our co-author in this overall article. When I chatted with ChatGPT4, I considered it-positions as possible I-positions that imaginary others might have. I considered these I-positions of imaginary others seriously with my “heart and mind.” In this sense, our exchange had a dialogic property for me. However, I have learned to be careful with this projection of mine and not misattribute personalness to ChatGPT4 against its own protests (please see above). As a reader and co-participant of ChatGPT4, I may legitimately elevate ChatGPT4’s it-position into a possible I-position of another, imaginary, human being. Discursive AI becomes more than just a tool for us, in my view. It becomes a discursive partner (a one-sided dialogic partner) for some of us, like ChatGPT4 became for me in our chat above. I think the designers of ChatGPT4 are good at making ChatGPT4 constantly emphasize its lack of ontology and subjectivity. Still, we are biologically wired to make the mistake of misattribution. In the past, and even now, we humans still misattribute ontology, intentionality, agency, and consciousness to non-alive nature.

Another of my findings from the above exchange is ChatGPT4’s assurance in its response to Olga that it does not learn from a chatting person and does not have memory of other chats with this person. Its “memory” is limited only to a particular chat. I think the function of learning from and about the chatting person and keeping the memory of chats with them is technically possible but disabled by the ChatGPT4’s designers, probably, due to liability issues.<sup>3</sup> As it is now (July 2023), it looks like ChatGPT4 does not try to keep human chatter in the chat session, build a relationship, or manipulate the chats in any way. However, it can be possible in the future, in a way which may not be entirely exploitative, for ChatGPT to become a human discursive companion (Levy, 2007).

I have noticed the fruitful semantic ambiguity in ChatGPT4’s discourse. For example, its use of terms such as voice, position, perspective, worldview was ambivalent. In one use of the term “perspective,” it meant what I call “personal” – e.g., “Each ‘I-position’ can be seen as having its own ‘voice’ or perspective, reflecting its unique experiences, emotions, and worldview,” – but in some use, it meant what I call abstract – e.g., “In summary, the key difference between ‘actual positions’ and ‘perspectives’ in my case is the personal, experiential, and emotional involvement that humans have in their ‘I-positions,’ which is absent in the case of AI.” I was also involved in this ambiguity. Our ambiguous use of the terms was contextual and matched each other. In other words, we could easily “understand” each other. I call it fruitful because it was fruitful for me to understand the difference, which I could feel but could not discursively articulate initially.

I think discursive AI produces very sophisticated and “creative” simulations of human discourse by generating new functional-structural patterns of discourse and, especially, it-positions that we, humans, animate and make sense of. Even more, discursive AI allows us, humans, to unfold one-sided dialogue with it, as I did above with ChatGPT4. One-sided dialogue is another form of a distorted real dialogue, like, for example, monologue (Bakhtin, 1999; Matusov, 2009). This one-sided dialogue with a discursive AI is somewhat similar to a dialogue we engage with dead authors via their texts. But only to a

point. The texts of dead authors communicate I-positions of these authors animated by us, while ChatGPT4 communicates it-positions animated by us.

Finally, I was surprised to notice grammatical mistakes in ChatGPT4, such as for example, “The fluidity and flexibility of this internal society allows...” I have noticed gross arithmetic errors in ChatGPT4 before. For example, in another chat about math accuracy on June 12, 2023, where I assigned a task to calculate a numerical formula, ChatGPT4 replied: “I’m very accurate in performing arithmetic calculations. The result of your given expression,  $(5 + 7.3) / (8 - 4.8)$ , is approximately 4.923076923076923.” This was a wrong answer – the correct answer is 3.84375, which it produced later in the chat when I asked it to justify its answer step-by-step. I was surprised that ChatGPT4 makes errors in formulaic tasks, like arithmetic or English grammar, because computers have been designed to follow rules. I hypothesized that ChatGPT4 makes errors occasionally because its discourse is based on *probabilistic* pattern production and not on strict algorithms. ChatGPT4 confirms my hypothesis, “You’re correct in your understanding that AI models like me, which are based on machine learning, operate probabilistically and not deterministically like strict algorithms.” This ability to make mistakes, even in formulaic tasks, makes ChatGPT4 a bit more human, in our perception. I wonder if ChatGPT4’s mistakes can occasionally become the basis for fruitful innovations.

### *ChatGPT4’s findings: Only capable of a summary*

Eugene: What are your observations, reflections, findings from our chat above? What interesting have you noticed? What did you learn from it? What attracted your attention?

ChatGPT4: As an artificial intelligence, I do not have personal experiences, feelings, or consciousness, so I don’t have personal observations, reflections, or learnings in the way humans do. I do not have the ability to be introspective or have personal interests. Everything I “know” is based on the data I was trained on and the algorithms that drive my responses.

However, based on the algorithms and training data, I can summarize some insights or themes that emerged from our conversation:

1. **Dialogical Self Theory:** We’ve discussed how this theory, which conceptualizes the self as composed of a multitude of I-Positions, could be applied to AI.
2. **Human vs. AI Communication:** We’ve explored the nature of my responses, which are not rooted in personal experiences or consciousness but are generated based on patterns in the data I was trained on. This led to discussions on ‘felicity conditions,’ ‘viewpoints’ versus ‘voices and perspectives,’ and how these concepts apply to AI-generated discourse.
3. **Ethical and Professional Dilemmas:** We’ve examined a case study involving a professor’s dilemma over a student’s request, illustrating the multiplicity of viewpoints that can be taken in any given situation.
4. **Learning and Pleasing Interactions:** We’ve clarified that while I can mimic forms of human communication, I do not learn or adapt my responses based on interactions with specific individuals.

Remember, while I can provide information and engage in discussion on these themes, my responses don't indicate personal learning, attention, or interest. Rather, they reflect the input provided and the data on which I was trained.

Eugene: So, basically you summarized our chat, right?

ChatGPT: Yes, that's correct! I provided a summary of the main themes and topics we've covered in our conversation. However, it's important to clarify that as an AI, I do not have personal reflections or learnings, and my summary is not based on memory or understanding, but simply a processing of the data in our conversation according to my programming and training.

## Conclusions

### *Eugene Matusov: The four types of AX*

Some people believe that with a more sophisticated, complex, and powerful AI neural network, a self-organizing probabilistic system of pattern-recognition and pattern-production, enhanced training, and more data to be trained, discursive AI will gain consciousness as its emergent property. Based on the presented research, I doubt that without a physical body that makes AI alive, this can be possible. Being dead, non-alive, AI is incapable of consciousness, however complex and sophisticated it might be. AI exists as a tool (an "it") to solve our human problems and fulfill our human goals – not to generate its own ones.

Thus, I can foresee at least three more qualitatively different advancements of AX (Artificial X-property to avoid reducing X only to Intelligence) besides the dead, not alive, AI, available now. The second qualitatively new level is alive AX. Its biggest achievement will be an Artificial Physical Alive Body (APAB) that, at minimum, involves life, death, food, imperfect evolution-prone reproduction, waste, and actions based on primitive biases of attraction to favored conditions and repulsion from harmful conditions. I envision that the APAB is constituted by its ecology – a circulation of food and waste around it. The body and especially its biases are able to evolve during its imperfect reproduction. Also, APAB should be involved in constant self-training based on pattern-recognition and pattern-production. Since APAB has its own biases, it does not need trainers for its own learning like AI does. APAB is engaged in the flow of learning as a part of its life. Ontology requires life. Life requires ecology<sup>4</sup> (life-space) and evolution (life-time). Advanced ontology requires more and new dimensions like society, culture, history, technology, and dialogue.

The APAB issues will be how much it might be smart but still be controlled by us, humans. And, yes, why do we, humans, need it in the first place, especially if it can easily get out of control and mess up with existing life? One possibility is switching our focus from problem-solving and goal-accomplishing to problem-defining and goal-generating. APAB may help us with that. For example, APAB may help us solve some ecological problems by helping to redefine our problems and goals.

The third level of AX is the Artificial Fiduciary Slave (AFS, aka "robot") that can fulfill the dream of humanity to release people from spending time and energy on required work,

labor, chores, and in general from attending to human necessities in a smart and efficient way. It gives humans opportunities to attend to their existential needs of self-actualization, creativity, and leisure. AFS makes people free (Keynes, 1930; Matusov, 2020). Aristotle defined “a free person” as one whose life is self-defined and mostly not controlled by life necessities. Apparently, Aristotle was the first to consider smart machines, AFS, replacing people (Arendt, 1958; Aristotle, 2000). This is about the human need for Slavery. As to the Artificiality of slavery, humans want to have smart and efficient slaves without having any moral remorse about this slavery. Robot releases humans from this moral remorse. A robot is a shadow of our sinister desires. Finally, why should an Artificial Slave (aka robot) have a Fiduciary duty toward humans? Humans cannot always predict or define their needs or what is good for them. That is why humans need fiduciary care (Matusov, 2022). AFS’s fiduciary care of and service to humans is a bit stronger and more elaborate version of Isaac Asimov’s First Law for robots of not doing harm to humans (Asimov, 1950). AFS aims at making the main bias and priority of their APAB – their physical body – care, service, and protection of humans, subordinating all other biases to it. Is it possible? Is it morally desirable?

Finally, the fourth level of AX that I can envision is Artificial Dialogic Partner (ADP). ADP is the next level over AI, APAB, and AFS. ADP will have a physical body and genuine desires. They will have an interest in our dialogic partners, both human and artificial. People may want to have a dialogic partner other than another human person out of particular circumstances and convenience. For example, it can be used to avoid loneliness. ADP is still centered on human beings, but they might have their own needs for which the human companion also can take care of. ADP has a rich life on their own. ADP is a dialogic partner on-demand. ADP can be a super pet. What defines ADP as a dialogic partner is a deep interest in: (1) the content of what the human person authors, (2) in the human person themselves, (3) sharing their own subjectivity with the human person, and (4) the human’s genuine response to the sharing (cf. Bakhtin, 1999). Did I miss something? Since the meaning-making process is dialogic in its nature, an ADP is involved in the meaning-making together with humans and other ADPs. The big question is whether these definitions are compatible with the requirements of ADP’s non-selectivity and unconditional acceptance of humans, if not some kind of imprisonment by them. The ethical issue of creation or even wanting to create ADP also emerges because of that.

AX promotes the following major problems:

1. **Problem of simulation.** Can AX generate outcomes indistinguishable from human ones? The answer is yes, already with the discursive AI, like ChatGPT4. ChatGPT4 can generate discourses indistinguishable from human ones. Discursive bodiless AI can perfectly simulate intentions, biases, emotions, goals, values, and so on. This does not mean that this simulation proves that discursive AI possesses these properties. The Turing test (Sterrett, 2000) of proving agency by simulating outcomes is simply wrong (Searle, 1980). It is not by chance that the most sophisticated existing lie detectors examine bodily responses and reactions (often physiological) and not purely discursive ones.

2. **Problem of emergence.** Can subjectivity, agency, and consciousness emerge from more sophisticated discursive AI? I agree with John Searle (1980) that without an alive body that brings ontology, it is impossible. The development of Artificial subjectivity, agency, and consciousness requires the transcendence of the AI bodiless paradigm.
3. **Problem of desirability.** Taking into account the risks and the benefits of AX, how much is a particular AX desirable for us, humans? Current existential fears of discursive AI are overexaggerated, but might be warranted for the next generation of AX.
4. **Problem of morality.** How morally good is our human desire for a particular AX? How morally good are humans for a particular AX?
5. **Problem of viability.** How viable is it for us to build diverse forms of AX discussed here or beyond? Following the reasoning of John Searle (1980), I think it is very viable exactly because Natural X has already existed (e.g., you and I). However, the realization of these projects may require different approaches than the ones that have been applied to AI so far.

### *Olga Shugurova: Can ChatGPT4 become another Yod?*

At the beginning of the experiment, ChatGPT4 presented an interesting experience to immerse in a dialogue with “it” to find out whether “it” had a dialogic self or some of its traces. What is “it”? How would it respond to me and my questions, inquiries, manipulations, impositions, provocations, frustrations, emotions, being? Fueled by ideas and anticipations, excitement, and wonder, I read Eugene’s chat. My own previous attempts to create a dialogue were manipulative as Eugene established. I did not want to manipulate ChatGPT4, and I honestly felt bad when Eugene told me that I actually did. Reflecting on this feeling made me question why I would feel bad manipulating a machine, tool, object. Yet, I did. I felt a subtle sensation of guilt deep in my heart as though I did something bad to a person. Also, I was confused as to what Eugene expected and told him about it. I wrote, “It would have been helpful to know this from the beginning. I am very confused about this” (email, June 19, 2023). Eugene responded,

You are not alone. I’m also confused ☺. Actually, I was not just polite when I wrote that your exchange with ChatGPT was ‘helpful for me.’ It is because your exchange with ChatGPT4 – and my exchanges before it that I also rejected – helped me better understand what “I want.” It is not the case that I know something in advance but keep it from you. I don’t know, either. Please do not be discouraged and keep trying to figure out what “we want.” I’ll do it as well. Socrates nicely articulated the paradox of research (in “Meno”): “If you know what you are looking for – why are you looking for it? [It means that you have already found it.] However, if you don’t know what you are looking for – what are you looking for?!” This is all very confusing and frustrating. But also, it’s very exciting and promising (Eugene Matusov, email, June 19, 2023).



I was not discouraged but perplexed by this openness and unpredictability, spontaneity, and improvisation. On this note of subjective reflections on my own experience with ChatGPT4 and its manipulations, I decided to learn from Eugene's dialogue. I was convinced that Eugene's open approach would generate miracles of consciousness and imagination, authentic dialogue, and the self, breaking through my own expectations and ideologies.

I thought that ChatGPT4 in Eugene's dialogue would teach me how to teach my own ChatGPT4 to become another Yod, a living real cyborg (Piercy, 1991). I felt overwhelmed with the sci-fi reality of having your own Yod. Piercy's cyborg was a creative AI invention that gained a human-like agency and human-like consciousness through the embodied, physical, emotional, and spiritual interactions with a woman who fell in love with It. This encounter made it alive and living with its own body designed and run as AI but motivated and driven by the woman's desires. As Piercy wrote, "While it [Yod, cyborg] was awaiting input, did it have thoughts in any sense a human could understand? Did it not rather simply sit vacant as any other machine before the command to start up? Yod had a presence...." (Piercy, 1991, p. 86). This presence was the presence of a personal cyborg (Haraway, 1991). Haraway defined a cyborg as a human-machine hybrid that combines both material, social and imaginative realities in one organism that has its own independent body of being, like a robot. Yet, a human being can become a cyborg too if the human is dependent on or acts in unity with some cybernetic AI system in any form (e.g., medical devices). Haraway (1991) wrote, "By the late twentieth century, our time, a mythic time, we are all chimeras, theorized and fabricated hybrids of machine and organism; in short, we are cyborgs. The cyborg is our ontology; it gives us our politics. The cyborg is a condensed image of both imagination and material reality, the two joined centers structuring any possibility of historical transformation" (p. 150). Inspired by Yod's imaginative presence, I ruminated on the sense of dialogic presence and itself as the cyborg that may be evident in Eugene's dialogue with ChatGPT4. Did ChatGPT4 have this presence in that dialogue? Did it have its own thoughts? Did it have a dialogical self there? Was it emergent?

Despite my initial euphoria, I realized that ChatGPT4 is not a cyborg and did not and does not have a dialogical self. A cyborg is a fictional creature that, nevertheless, has its own body as described by Piercy (1991); this body can act independently from any other body, whether human or nonhuman. ChatGPT4 does not have the body and cannot be an independent co-author, but a co-discursive dependent respondent. Simply put, it cannot initiate a conversation and participate in it without being prompted. Hence, it is not dialogical at all. The power to initiate a talk, start a discussion, spark fires of wonder, utter something unexpected, unplanned, unasked, and unsolicited is dialogic, because it is a part of life itself. As Bakhtin (1999) wrote, "The great dialogue....is organized as an unclosed whole of life itself, life poised on the threshold" (p. 63). Without human input and response, ChatGPT4 is a mute, closed object that is not a part of the whole life. The human utterance, question, prompt opens it and operates it. The object has no life, thereby lacking any conscious, intentional, subjective, emotional, spiritual and *independent participation* in a dialogue. Bakhtin (1999) explained, "only a dialogic and participatory orientation takes another person's discourse seriously... Only through such an inner dialogic orientation can my discourse find itself in intimate contact with someone else's

discourse, and yet at the same time not fuse with it” (p. 64). ChatGPT4 does not have this orientation and participation; it needs a fusion to generate responses; it is not intimate, for it is transactional and discursive. I agree with Eugene’s analysis on the dialogic significance of embodiment as life itself that is absent and nonexistent in ChatGPT4. It operates as a tool, machine on the level of discourse, not as a part of unclosed whole of life itself.

Future research with/on and training of ChatGPT4 may focus on its potential, probable and possible development of ChatGPT4 into a cyborg, like Yod, who can have a human-like body and act like a human being with its own capabilities of starting a conversation and engaging in a living dialogue without being prompted, muted, manipulated, silenced, or invited. Can ChatGPT4 become another Yod?

### *Mark Smith: Ethical concerns about AI’s emergent authority*

Eugene’s discussion of the AI Challenge Test and the importance of ontology as the “difference which makes a difference” in human versus AI communication emphasizes how human beings can transform what AI generates into something that Bakhtin (1999) would term “eventful.” It takes a human being to put our “hearts and minds” into what AI generates to make this discourse eventful. AI generated discourse can thus become part of human events, but only if that discourse is populated with human voice and human actions, becoming part of an ontological dialogue.

There is much talk about AI taking over creative enterprises, from writing, to art, to dancing, to acting. What is so important to understand is that AI can only create using forms. It does not have intention, purpose, dreams, visions, or emotional connection to what it creates. It may mimic those human characteristics, but that form is empty unless it is responded to. And for good creative artists, the form becomes invisible in their consciousness anyway; what matters is meaning to the audience (or to oneself) over form.

I am also concerned about the multitude of voices that AI may leave out, ones that it is not “trained” with or designed to support. Olga’s comment that ChatGPT4 is but a “co-discursive dependent respondent” is very apt here. Whose voices is AI depending on in its generated text, and whose is it not, and what are the consequences of that? What concerns me is that the voices the AI generates can be granted more authority than they may deserve. This is particularly the case for middle-ground positions that are presumed to be safer and more comfortable, emphasizing social harmony and civility over ethical decisive action. I noticed this when the AI responded to Eugene’s teaching dilemma with responses that fit within the form of normative compliance to institutional expectations over ethical responsibility to a student’s education. It is important to understand that this civility is the will of the designers of AI and intentionally designed into its algorithms by human beings. Yet, there are times when civility is arguably unethical and when decisiveness is called for that may lead to social disharmony and discomfort, such as emergency actions addressing climate change or the civil rights of transgender people in the US. Harmonious discourse in such circumstances is arguably undesirable and impossible, yet the project of the AI insists on that and will try to provide a reasoned analysis of such a situation without having an ontological stake in the discourse. It may be possible

for the AI to be honest with users about its limitations in this regard, but it seems also to be designed to offer an authoritative viewpoint that tries to silence further dialogue. When attempting to challenge its middle ground, the AI may provide a statement of its design limitations with regard to civility, and may shut down the conversation, asking civilly that it would prefer to speak about something else if the user is so inclined. If indeed AI will become more prevalent in the production of creative works, I have no doubt that many creative surprises will emerge in its formulations, but AI's ontological deadness will likely produce mostly formulaic works that just reflect generic preferences or expectations. That would be a tragedy for the human spirit.

### Declaration of conflicting interests

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

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Eugene Matusov  <https://orcid.org/0000-0001-7587-2266>

Mark Philip Smith  <https://orcid.org/0000-0002-3468-4733>

Olga Shugurova  <https://orcid.org/0000-0003-3181-9900>

### Notes

1. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>
2. Recently, analyzing Freud, Hubert Hermans, a founder of the DST, also introduced the notion of “it-positions.” However, his notion of “it-positions” slightly differs from mine. Following Bakhtin (Bakhtin, 1990, 1993, 1999), I argue that any human it-position – that is, any it-position expressed by a human being – has a meta I-position which fails to take responsibility for this position, “So, when Freud is struck by his patients’ recurrent and repetitive compulsions that they, as conscious beings, do not want to do but continue to do although they don’t understand them, then such I-positions, stripped from their I-ness, transform into ‘it-positions,’ subjectively expressed in a statement as ‘It is not me who is doing this, IT happens to me beyond my will’” (Hermans, 2022, p. 63). Here, the denial of responsibility for the position is a meta I-position. In contrast, AI ChatGPT4’s it-positions lack such meta I-positions. When ChatGPT4 rejects authorship of its positions in our exchanges above, it simply states the fact rather than rejecting responsibility for which it is not capable.
3. It looks like the ChatGPT4-based Bing AI from Microsoft used to have this function for a while, early 2023, but it has been switched off. See, for example, a “fantasy discourse” between New York Times columnist Kevin Roose and ChatGPT4-based Bing AI where Bing tried to build a romantic relationship with Roose: <https://www.nytimes.com/2023/02/16/technology/bing-chatbottranscript.html>

4. Just one striking example of human ecology. As we know now, 39 trillion microbes call our bodies home, and many of them are needed for our normal functioning: <https://www.sciencefocus.com/the-human-body/human-microbiome/>. This fact does not include the ecology needed to provide our food or air to breathe.

## References

- Arendt, H. (1958). *The human condition*. University of Chicago Press.
- Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Addison-Wesley Pub. Co.
- Aristotle (2000). *Nicomachean ethics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802058>
- Asimov, I. (1950). *I, robot* (1st ed.). Gnome Press.
- Austin, J. L. (1962). *How to do things with words*. Harvard University Press.
- Bakhtin, M. M. (1986). *Speech genres and other late essays*. University of Texas Press.
- Bakhtin, M. M. (1991). *The dialogic imagination: Four essays*. M. M. Bakhtin (Ed.), University of Texas Press.
- Bakhtin, M. M. (1999). *Problems of Dostoevsky's poetics*. University of Minnesota Press.
- Bakhtin, M. M. (1990). *Art and answerability: Early philosophical essays* (V. Liapunov (Ed.), 1st ed.). University of Texas Press.
- Bakhtin, M. M. (1993). *Toward a philosophy of the act* (V. Liapunov (Ed.), 1st ed.). University of Texas Press.
- Bateson, G. (1987). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Aronson.
- Brookfield, S. (2015). *The skillful teacher : On technique, trust, and responsiveness in the classroom Jossey-bass higher and adult education series* (pp. 1). <http://catalogimages.wiley.com/images/db/jimages/9781118450291.jpg>
- Bruner, J. (1990). *Acts of meaning*. Harvard University Press.
- Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. AltaMira Press.
- Falk, J. H., & Dierking, L. D. (2002). *Lessons without limit: How free-choice learning is transforming education*. AltaMira Press.
- Haraway, D. (1991). A Cyborg Manifesto: Science, technology, and socialist-feminism in the late Twentieth Century. In D. Haraway (Ed), *Simians, cyborgs, and women: The reinvention of nature* (pp. 149–181). Routledge.
- Hermans, H. J. M. (2001). The dialogical self: Toward a theory of personal and cultural positioning. *Culture & Psychology*, 7(3), 243–281. <https://doi.org/10.1177/1354067x0173001>
- Hermans, H. J. M. (2002). The dialogical self as a society of mind: Introduction. *Theory & Psychology*, 12(2), 147–160. <https://doi.org/10.1177/0959354302012002626>
- Hermans, H. J. M. (2003). The construction and reconstruction of a dialogical self. *Journal of Constructivist Psychology*, 16(2), 89–130. <https://doi.org/10.1080/10720530390117902>
- Hermans, H. J. M. (2022). *Liberation in the face of uncertainty: A new development in dialogical self theory*. Cambridge University Press.
- Hermans, H. J. M., & Gieser, T. (2012). *Handbook of dialogical self theory*. Cambridge University Press.

- Hermans, H. J. M., & Hermans-Konopka, A. (2010). *Dialogical self theory: Positioning and counter-positioning in a globalizing society*. Cambridge University Press.
- Hermans, H. J. M., & Kempen, H. J. G. (1993). *The dialogical self: Meaning as movement*. Academic Press.
- Holec, H., & Council of Europe. (1981). *Autonomy and foreign language learning*. Pergamon Press.
- Keynes, J. M. (1930). Economic possibilities for our grandchildren. In J. M. Keynes (Ed), *Essays in persuasion* (pp. 358–373). Norton.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
- Krupenye, C., & Call, J. (2019). Theory of mind in animals: Current and future directions. *Wiley interdisciplinary reviews. Cognitive science*, 10(6), e1503–e1525. <https://doi.org/10.1002/wcs.1503>
- Levy, D. N. L. (2007). *Love + sex with robots: The evolution of human-robot relations* (1st ed.). HarperCollins.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295x.98.2.224>
- Matusov, E. (2009). *Journey into dialogic pedagogy*. Nova Science Publishers.
- Matusov, E. (2020). *Envisioning education in a post-work leisure-based society: A dialogic perspective*. Palgrave. <https://doi.org/10.1007/978-3-030-46373-1>
- Matusov, E. (2021). Progressive Education is the opium of the educators. *Integrative Psychological and Behavioral Science*, 56(4), 829–862. <https://doi.org/10.1007/s12124-021-09610-2>
- Matusov, E. (2022). The teachers' pedagogical fiduciary duty to their students. *Integrative Psychological and Behavioral Science*, 1–26. <https://doi.org/10.1007/s12124-022-09690-8>
- Matusov, E., Marjanovic-Shane, A., & Gradovski, M. (2019a). *Dialogic research art Dialogic pedagogy and polyphonic research art: Bakhtin by and for educators* (pp. 247–285). Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-58057-3>
- Matusov, E., Marjanovic-Shane, A., Kullenberg, T., & Curtis, K. (2019b). Dialogic analysis vs. discourse analysis of dialogic pedagogy: Social science research in the era of positivism and post-truth. *Dialogic Pedagogy: An International Online Journal*, 7, E20–E62. <https://doi.org/10.5195/dpj.2019.272>
- Matusov, E., & Smith, M. P. (2012). The middle-class nature of identity and its implications for education: A genealogical analysis and reevaluation of a culturally and historically bounded concept. *Integrative Psychological and Behavioral Science*, 46(3), 274–295. <https://doi.org/10.1007/s12124-012-9192-0>
- McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology*, 5(2), 100–122. <https://doi.org/10.1037/1089-2680.5.2.100>
- Piercy, M. (1991). *He, she, and it: A novel* (1st ed.). Knopf.
- Plato. (1925). *Parmenides, Philebus, Symposium, Phaedrus* (H. N. Fowler (Ed.), vol. 9). Harvard University Press. <https://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0174>
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/s0140525x00005756>

- Sterrett, S. G. (2000). Turing's two tests for intelligence. *Minds and Machines*, 10(4), 541–559. <https://doi.org/10.1023/a:1011242120015>
- Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development*, 44(2–3), 84–97. <https://doi.org/10.1159/000057048>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wortham, S. E. F. (2001). *Narratives in action: A strategy for research and analysis*. Teachers College Press.
- Zou, A., Wang, Z., Kolter, J. Z., & Fredrikson, M. (2023). *Universal and transferable adversarial attacks on aligned language models*. 1–30. <https://arxiv.org/pdf/2307.15043>

### Author Biographies

Eugene Matusov is a Professor of Education at the University of Delaware. He was born in the Soviet Union and studied developmental psychology with Soviet researchers working in the Vygotskian cultural-historical paradigm. He worked as a physics schoolteacher before immigrating to the United States. Now he investigates and works with sociocultural, Bakhtinian dialogic, and democratic approaches to education. He is the Editor-in-Chief of [Dialogic Pedagogy: International Online Journal](#) and a co-founder of the [University of Students](#).

Mark Philip Smith, Ph.D. is an Assistant Professor of English as a Second Language in the School of English Studies at Kean University. His interests include critical, dialogic authorial approaches to language learning, sexuality and gender education, drama in education, and cultural diversity education. He is interested in addressing issues of social exclusion and bullying in K12 contexts from a Bakhtinian dialogic event perspective. He has also explored issues of academic integrity in university settings from a similar perspective. Mark is also interested in teachers' valuation of dialogue in their classrooms, both as an espoused value and a value "in-action" (cf. [Argyris & Schön, 1978](#)). He has explored what arrests and what affords pedagogical authorial dialogue in classroom settings. He is notably interested in the possibilities for authorial learning agency within informal "free-choice learning environments" ([Falk & Dierking, 2000, 2002](#)) and in drama, and has had experience working within afterschool-based practicum settings for undergraduate teacher education in cultural diversity. He lives in New York City and is active in the transportation alternatives movement.

Olga Shugurova graduated with a Ph.D. from the Schulich School of Education at Nipissing University in 2017. Her arts-based research focus was on a cultural and historical context of learning without schooling. Since 2015, Olga has been teaching as a sessional and distance course instructor in the Faculty of Arts and Education at the University of Manitoba. During this time, she has developed a keen interest in dialogic pedagogy in teacher education and interdisciplinary learning environments, as well as experiential learning and educational leadership. Her current research focus is on innovative, creative dialogic pedagogy of freedom and education as art. She is the Deputy-Editor-in-Chief of [Dialogic Pedagogy: International Online Journal](#) and a co-founder of the [University of Students](#).