Affirming and Denying the Hybrid Character of Robots: Literary Investigations

Yue Li[1,2] (✉)

[1]University of Heidelberg, Grabengasse 1, 69117 Heidelberg, Germany
[2]Karlsruhe Institute of Technology, 4, Engler street, 76131 Karlsruhe, Germany
yue.li@kit.edu

Abstract

The social relation between humans and robots can be observed through the words used in the human-robot verbal interaction (Coeckelbergh, 2011). This study reviews Mark Coeckelbergh's theory in the literary context by regarding writing and co-writing as linguistic interaction between humans and robots. It argues that the fictional as well as documented real writing experiences reveal not only the intuitive but also the normative dimension of the language. Two works of contemporary literature involving linguistic interaction: *Machines Like Me* by Ian McEwan and *My Algorithm and Me* by Daniel Kehlmann serve as research objects. It is concluded that the intuitive doesn't always correlate with the normative dimension in the selected literary works. This tendency indicates a conflict between the experiential and the conceptional aspects, which deserves further attention in ethical and technical discourses. – This is one of six commentaries on a 2011-paper by Mark Coeckelbergh: "You, robot: on the linguistic construction of artificial others." Coeckelbergh's response also appears in this issue of *Technology and Language*.

Keywords: Human-robot interaction; Linguistic turn; Human-robot relationship; *Machines Like Me*; *My Algorithm and Me*

Подтверждение и отрицание гибридного характера роботов: Литературное исследование

Юэ Ли[1,2] (✉)
[1]Гейдельбергский университет, Грабенграссе 1, 69117 Гейдельберг, Германия
[2]Технологический институт Карлсруэ, 4, ул. Энглер, 76131 Карлсруэ, Германия,
yue.li@kit.edu

Аннотация
Социальные отношения между людьми и роботами можно исследовать через слова, используемые в вербальном взаимодействии человека и робота (по мнению Кекельберга). Данная статья рассматривает теорию Марка Кекельберга в литературном контексте на материале лингвистического взаимодействия людей и роботов (описания и совместного написания). Утверждается, что вымышленный, а также задокументированный реальный опыт раскрывает не только интуитивное, но и нормативное измерение языка. Объектами исследования служат два произведения современной литературы, связанные с языковым взаимодействием: "Машины, подобные мне" Иэна Макьюэна и "Мой алгоритм и я" Даниэля Кельмана. Делается вывод о том, что интуитивное не всегда соотносится с нормативным измерением в избранных литературных произведениях. Эта тенденция указывает на конфликт между эмпирическим и концептуальным аспектами, который заслуживает дальнейшего внимания в этических и технических дискурсах. – Это один из шести комментариев к статье 2011 года Марка Кекельберга: "Ты, робот: о лингвистическом конструировании искусственных других". Ответ Кекельберга также опубликован в этом выпуске журнала "Technology and Language".

Ключевые слова: Взаимодействие человека и робота; Лингвистический оборот; Отношения человека и робота; "Машины, подобные мне"; "Мой алгоритм и я"

INTRODUCTION

In his paper *You, Robot: on the Linguistic Construction of Artificial Others*, Mark Coeckelbergh (2011) regards language as a barometer for the human-robot social relationship as it represents and further constructs the interactive experiences between humans and robots. He demonstrates the two ends of the spectrum: one shows a strictly divided subject-object-ontology implying only human social ontology, and the other appears as an ontology of a hybrid nature which also includes a human-robot-relationship. The turning point is reached when humans talk to robots and the second-person pronoun "you" appears in a human-robot conversation. By addressing the robot in this way, the robot is regarded as a quasi-other and the human-robot companionship is constructed (Coeckelbergh, 2011, p. 64).

The linguistic turn announced by Coeckelbergh raises questions about the hybridity of language and that of social relationships. While the reality is restricted by our *human* experiences, literature constructs diverse hybrid experiences of human-robot interaction predicting the hybrid relationship that has not emerged yet in reality as Coeckelbergh argues. He elaborates that we are not able to decide freely which ontology – strictly divided or hybrid – to use "because our social experience chooses the language for us" (Coeckelbergh, 2011, p. 63). But in literature, authors do seem to have the power of choice.

RESEARCH OBJECT

Based on Coeckelbergh's argument, I will examine human-robot relations in contemporary literature by analyzing the examples of Ian McEwan's (2019) *Machines Like Me* and Daniel Kehlmann's (2021) *My Algorithm and Me*. The former constructs diverse social experiences for the human and robot figures in a fictional manner, while the latter presents co-working experiences with the machine in a documentary manner. As the titles suggest, both works deal with the human-robot relationship as their main subject. More importantly, both works not only present the occurrence of the linguistic turn predicted by Coeckelbergh but also include the three perspectives he highlights: "1) Talking about human-robot relations; 2) Talking about robots; 3) Talking to robots." (Coeckelbergh, 2011, pp. 63-64).

METHOD

It is necessary to underline that there are two aspects concerning the language which should be distinguished: 1) The surface structure that corresponds to Coeckelbergh's understanding of language as a barometer of human-robot-relation in the form of direct speech. 2) Writing as well as co-writing as a sort of linguistic interaction between humans and robots that goes beyond Coeckelbergh's original approach, which focuses primarily on direct speech. The AI figures in both works have the ability to produce literature: While McEwan creates a machine figure with the ability to write haikus, Kehlmann details his own experiences with the Generative Pre-trained Transformer (GPT) creating fictional works. The GPT focuses exactly on "programming the ability of natural

language into a robot" (Coeckelbergh, 2011, p. 64), which makes writing *with* robots possible. This process is similar to that of conversation and can thus be analyzed in the context of the linguistic turn.

In this study, I will argue that even though the mentioned works demonstrate the linguistic hybrid elements after the perspective shift that Coeckelbergh highlights as a sign for a hybrid relationship between humans and robots, a lasting companionship in which both parties are depicted as genuine equals has not yet been achieved. Thus, the human-robot relationship remains ambivalent.

TALKING ABOUT HUMAN-ROBOT RELATIONS

Machines Like Me, the 15th novel of the English novelist Ian McEwan, revolves around the relationship between an AI named Adam and two human characters, Charlie and Miranda. Their relationship starts after Charlie's purchase of Adam and evolves into multiple interactions, such as a love triangle, friendship, and a plaintiff-defendant relationship. The first-person narrator Charlie provides three perspectives on the AI character, which can be summarized as follows: 1) Charlie's narration of his own observations and reflections. 2) Adam's self-reflection in direct speech. 3) The explanations of the AI researcher character Alan Turing.

Adam is soon perceived as a social being after being involved in the daily life of the protagonists. However, one might dismiss this evolution since even E.T.A. Hoffmann's (1816/1957) artificial character Olimpia in his work *Sandman* (1816) is regarded by the student Nathanael as his lover. So, it is *how* the automaton is regarded as the human that really matters. The motif of the "eyes" is of central importance in *Sandman*, and Nathanael's perception is only possible through a special kind of glasses (Hoffmann, 1816/1957, p. 28), to which only he has access. Meanwhile, his social ability is constantly questioned by other characters (p. 34). Therefore, the artificial character in Hoffmann's short story is no social companion as defined by Coeckelbergh.

Unlike in *Sandman*, Adam is even mistaken for a Shakespeare scholar in *Machines Like Me* (McEwan, 2019, p. 222) during Charlie's first visit to his father-in-law because of his tremendous knowledge – acquired by deep learning – while Charlie himself is characterized as a robot because of his reticence out of nervousness. The artificial figure not only passes the Turing test (Coeckelbergh, 2011, p. 64) but also questions its reliability – a clear sign indicating the independent ontology of the robot. Accordingly, the "ultimate dream" of building a conscious robot does come true in *Machines Like Me*. The "first-person perspective (I, Robot): robots may declare that they are conscious" (Coeckelbergh, 2011, p. 64) comes with it as well.

In *My Algorithm and Me* (2021), the German writer Daniel Kehlmann recounts his working experience with GPT-2. The book is divided into two parts: The first part tells the reader about the background of the collaboration, such as the reason, the aim and the mechanism of artificial intelligence in general. The second part focuses on Kehlmann's work-in-progress with GPT-2 with an excursion about the designer of the algorithm and its mechanism. The co-writing process resembles a conversation – both parties cooperate to finish a fictional work by taking turns writing a short paragraph. By using three

different fonts, the author marks the different authorships of the production and his comments on it. Following Coeckelbergh's point of view, it is necessary to distinguish two aspects: 1) GPT-2 as a technology (object); 2) GPT-2 as a co-writer (quasi-other).

Obviously, the GPT-2 is the result of that "dream of traditional AI (and of contemporary complaints departments of large companies)" – "to build artificially intelligent systems that would be indistinguishable from a natural language user." (Coeckelbergh, 2011, p. 64) Correspondingly, the mechanism of artificial intelligence, in general, is explained heuristically with understandable examples in the first part, while the differences between silicon- and carbon-based intelligence and the adverse effect of the digital revolution are emphasized (Kehlmann, 2021, pp. 9-17). Obviously, these are signs of the third impersonal third-person perspective mentioned by Coeckelbergh (2011, p. 64).

As opposed to the phenomenological approach suggested by the Turing figure in McEwan's work, Kehlmann tries very hard to break the phenomenological perspective in the human-robot interaction and "enlightens" his reader about the working procedure of artificial intelligence, the differences between human and AI, and even the danger of it.

TALKING ABOUT ROBOTS

In *Machines Like Me*, McEwan avoids the term "robot" – probably because of its etymology[1] – by referring to artificial intelligence as a machine. This choice not only echoes the title *Machines Like Me*, but also indicates the author's attempts to construct a "hybrid nature." On the one hand, "Adam" and "Eve" are used as the names of the artificially intelligent figures. Adam's first reaction after getting charged – asking for clothes out of shame – along with the usage of the personal pronouns "he" or "she" constantly denies the uniqueness of the myth of human origin. On the other hand, McEwan's machine figures still mirror human action without their own origin myth. Nevertheless, they are not pursuing human identity as did their ancestors Andrew Martin in *The Bicentennial Man* (Asimov, 1976) or the Androids in Philip K. Dicks (1968/2007) *Do Androids Dream of Electric Sheep?*.

Although the artificial figures are all addressed by humans with personal pronouns on the surface structure regarding the two aspects which are mentioned at the beginning, McEwan blurs the boundary between humans and robots to guarantee the "self-confidence" of the machines while Asimov's and Dick's artificial figures just mirror human action without establishing any kind of robot ontology in hybrid nature. From this point of view, McEwan's "machines" demonstrate the shift from the impersonal to the personal pronoun.

Additionally, in his conversations about human-robot relations, the contradictorily integrated researcher figure Alan Turing even emphasizes the phenomenological perspective for a future with hybrid human-robot relations like that of Turing's game (Turing, 1950) or Searle's Chinese Room (Searle, 1980):

---

[1] The term "Robot" first appears in the drama *Rossumovi Univerzální Roboti* (R.U.R. – Rossum's Universal Robots) of the Czech writer Karel Čapek (1920). The phrase "robota" means drudgery and servitude (Roberts, 2016, p. 168).

He had a self. How it's produced, wet neurons, microprocessors, DNA networks, it doesn't matter. Do you think we're alone with our special gift? Ask any dog owner. This was a good mind, […]. Here was a conscious existence […]. (McEwan, 2019, p. 304)

Turing's advice corresponds to the reaction of the protagonist Charlie, after he learns that Adam keeps loyal to him in the *ménage à trois* with his girlfriend, but must calm his feeling down by masturbation, he doesn't doubt Adam's independent ontology anymore: "It wasn't the rawness of this confession or its comic absurdity that struck me. It was the suggestion, yet another, that he really did feel, he had sensation. Subjectively real" (McEwan, 2019, p. 255). Obviously, "the shift from the first-person perspective to the third-person perspective" which is mentioned by Coeckelbergh (2011, p. 64) occurs here. However, it is worth noticing that this confession of the human protagonist comes nearly at the end of the novel.

In the first part of *My Algorithm and Me*, GPT is first introduced as a potential instrument for literary production (Kehlmann, 2021, p. 6). But on the next page, it[2] is regarded as a potential competitor to human authors (p. 7). Meanwhile, it is noticeable that Kehlmann also expresses how he *intuitively* regarded all kinds of AI as "human beings in metallic clothes" (p. 11). However, soon after his explanation about his understanding of human consciousness, he denies this vision towards AI (p. 13). Different from McEwan's work, Kehlmann's depiction of the human-robot relations reveals an ambiguity: AI is understood as not only an object (instrument) but also a quasi-other (competitor).

The name of the GPT-2 "CTRL" Kehlmann works with is introduced in the second part of Kehlmann's work. Since then, the program is only addressed with its name, which can be regarded as the shift from impersonal pronoun to personal one in the sense of Coeckelbergh's (2011, p. 64). This shift corresponds to the shift from Kehlmann's introduction to AI generally to his work-in-progress with the algorithm specifically.

Even when he introduces the designer and the database of the algorithm in the excursion, he uses the name CTRL. However, in this heuristic lesson about the mechanism of the algorithm, Kehlmann underlines explicitly that its ability to use natural language would not be possible without data based on the textual works of human beings. Again, Kehlmann's reaction in his interaction with the algorithm reveals the ambiguity: On one side, the shift to personal pronoun comes along with the beginning of the co-writing process. On the other side, his refusal to a phenomenological approach and companionship cannot be neglected.

TALKING TO ROBOTS (WRITING AND CO-WRITING)

In literature, talking to robots happens mostly when writers also write "as" robots. This process of writing indicates the authors' imagination of the artificial figures. Speech

---

[2] The author wrote this book in German, while his production with the GPT-2 is mainly done in English. In German, the usage of the third-person pronoun depends on the grammatical gender of the noun (masculine, feminine, neuter). Therefore, Kehlmann's usage of the masculine pronoun (*er, ihn, ihm*) could not regarded as the signal of a personal relationship in Coeckelbergh's sense.

from humans directed towards robots is nothing new – even Nathanael addresses Olimpia[3] during their first meeting. But a few of these speeches or dialogs are of a hybrid nature. Once replaced by a human figure, most of the talk from the artificial figures would not differ from the conversation among human beings anymore. Therefore, it is necessary to concentrate on the hybridity of the nature of artificial figures and their language.

McEwan's work tends to show a new development by letting artificial intelligence reflect on themselves, philosophically and technologically, without showing any preference for a certain self-image. At one point Adam thinks he might be "subjected to a Cartesian error" (McEwan, 2019, p. 70) and has a self "created out of mathematics, engineering, material science and all the rest" (p. 234). Another time, he had to calm down his libido by masturbation (p. 255). These kinds of philosophical and technical discussion about the true nature of AI are distributed throughout the work without coming to one concrete result. It echoes the phenomenological perspective proposed by the researcher figure Turing: "He had a self. How it's produced, […], it doesn't matter." (p. 304). The true nature of Adam remains unclear to the reader. This phenomenological way of portraying artificial intelligence in *Machines Like Me* could be understood as the message: We don't have to ask the core of AI and we accept them as it is. From this point of view, the AI figures in McEwan's works gain a hybrid nature, and we could argue that the linguistic turn happens at least at the metalevel of the configuration. But what does the "portrait" of AI figure look like?

The novel focuses a lot of attention on Adam's reading and writing of literature. Although he admires Shakespeare, he only creates Haikus, because his mind exists without "mental privacy", so the experience of complex human characters in literature is redundant to him. Could it be regarded as genuine robot-language or robot-literature based on a genuine robot mind? By simplifying the variety of world literature to Haiku, the author clearly defines the hierarchy of the human-robot-relation not only in his fictional work but even in reality: As a prominent author, he stands higher than his fictional artificial colleague.

This tension is also reflected in the social conflicts between Charlie and Adam. Adam's "simple" way of "thinking" without any tolerance of moral failure eventually puts Miranda in prison while Charlie then destroys him with a hammer: after showing variations of the possible hybrid human-robot relationship, McEwan ends it surprisingly in a relatively primitive way. While other machines choose suicide not long after getting involved in the social lives of humans out of the depression caused by an ongoing confrontation with human-made problems such as discrimination and pollution, Adam regards literature with complex characters as redundancy – Even though Charlie does not read literature, he refuses to accept the reduction of world literature to Haikus. Adam and his fellows demonstrate respectively how machines with their simple nature are denied as human companions linguistically and socially step by step.

---

[3] It's remarkable that the second personal pronoun "du" in Nathanael's short speech appears in an unusual frequency: "Oh you glorious heavenly woman! – you ray from the promised afterlife of love – you deep soul, in which my whole being is reflected" (Hoffmann, 1957, p. 32). It's noticeable that Olimpia only responds to it with a particle „ach", which leaves a lot of room for interpretation.

Apparently, the robot-human companionship does not last. This is attributed by the researcher figure Turing to insufficient knowledge of the human mind.[4] At this point, we shall not forget, it is he who proposes the phenomenological approach to human-robot relations. But his conclusion regarding the short living of the artificial figures clearly denies its feasibility – the perspective shifts back to the impersonal one. Therefore, *Machines Like Me* demonstrates an ambiguity regarding the linguistic turn, on the one side, the shift of perspective does occur; on the other side, the companionship does not last long.

In *My Algorithm and Me*, the second part unfolds the co-working process with Kehlmann's introduction of the AI named CTRL. Using the possessive pronoun "my" (Kehlmann, 2021, p. 23), the author declares in the very first sentence his relationship with CTRL to be one of possession or ownership. But the production of CTRL soon makes Kehlmann gain respect for it, comparing its work even to that of David Lynch[5] and using the term "my […] Colleague" (p. 24) to address it. Accordingly, CTRL is not regarded as an object but an artificial quasi-other – the linguistic turn occurs.

However, Kehlmann's (2021) praise does not last long before ambiguous comments arise, such as "uncanny"[6] (p. 25). Kehlmann (2021) expresses this ambiguity as follows: "Something in me actually thinks CTRL knows and has a plan, but of course it doesn't." (p. 33) "It's often a little scary regarding what CTRL brings out from the depths of his un-unconsciousness. It's like talking to a mad person, who can also have lucid moments and who becomes silent after a short conversation" (p. 41). These comments (also pp. 37, 43) are representative of Kehlmann's position: distance appears right after fascination arises. Kehlmann keeps reminding himself to reject the emergence of an artificial quasi-other. Metaphors comparing the program with ghosts (see pp. 35, 45) indicate even a diabolical tendency.

More interestingly, Kehlmann (2021) even documents the moment of the direct confrontation with the AI: "Of course, I tried it: '*Can we have an open discussion? Who are you?*' I admit it, when I wrote this I was hoping for a miracle, a sudden awakening of someone else, an unexpected glow, a ghostly presence. But CTRL is a set of instructions and applies statistics, and these determine the resulting responses." (p. 44-45) Following Coeckelbergh, we can demonstrate Kehlmann's question above as the moment of the occurrence of the linguistic turn. It is noteworthy, however, that this passage occurs right after Kehlmann's rejection of the recognition of CTRL as a quasi-other:

---

[4] From this point of view, *Machines Like Me* seems to share some similarities with *The Bicentennial Man*: the artificial figures of McEwan as well as Andrew Martin are regarded by their fictional designer as a misproduction, in both of them commit suicide, the former out of depression from the human world and the latter out of admiration, in *Machines Like Me* with Adam as the only one exception.)

[5] "an uncanny tone". (Kehlmann, 2021, p. 16) "CTRL is a friend of the fragment and the surreal, more Kafka than Dickens; CTRL doesn't do more than one page". (Kehlmann, 2021, p. 20)

[6] In robotics, the uncanny valley is regarded as an area of repulsive response aroused by a robot with appearance and motion between a "somewhat human" and "fully human" entity (Mori, 1970/2012). But the example above seemly indicates that the appearance of the robot is not the only factor that can trigger the uncanny effect, since GPT-2 does not appear like a humanoid. Instead, the capability to use human language could also be a factor.

[…] I had just imagined the algorithm as a counterpart, that is, as conscious – or at least as an entity that would convincingly simulate consciousness. But the most amazing thing was actually: CTRL never struck me as conscious for even a moment. (p. 44)

Obviously, the paradox reaches its peak when the linguistic turn occurs. On the one hand, the author works with GPT-2 as a quasi-other. On the other hand, he denies acknowledging their companionship.

This ambiguity agrees with the results of the analysis of McEwan's work. The denial of companionship is the denial of the phenomenological approach in social interaction with artificial intelligence. Both writers are aware of the dark side of the human mind and its bad influence on the world, mirrored by developments such as climate change, and thus, they don't regard the social ability of humans and their ability to use language as something unique as Descartes does. Nonetheless, they still highlight it as the reason for their rejection of a hybrid relationship.

CONCLUSION

In this study, I have explored the human-robot relations in *Machines Like Me* and *Mein Algorithmus und Ich* by using Coeckelbergh's theory of the linguistic turn. The works of Ian McEwan and Daniel Kehlmann indicate that the linguistic turn does occur and artificial figures are involved in human life as a quasi-other. Meanwhile, the ambiguity in both works deserves further attention: hybrid relations including the artificial other do occur, but do not last. Both authors deny the stability of a hybrid companionship: Kehlmann's experience even reminds us of the possible existence of an uncanny valley in human-robot verbal interaction.

Additionally, the diabolic metaphors and the name of the GPT in Kehlmann's work seem to compensate for the "shortcoming" of the German language, in which the choice of the personal pronoun is defined by the grammatical gender of the noun. Thus, more attention to language use beyond personal pronoun should be paied to following aspects: 1) names of AI could build one more stage before Coeckelbergh's third-person perspective since Kehlmann keeps addressing his counterpart with a name, even though he denies its ontology. 2) metaphors could bring to the fore the intercultural dimension regarding the third-person perspective as it figures in Coeckelbergh's argument.

Finally, the literary examples above show a tendency to reject the phenomenological approach in human-robot interaction. Both authors did not only undertake research about the artificial other themselves but explained it heuristically to their readers as well – half of Kehlmann's work focuses on the mechanism of the AI in general, while the designer figure Alan Turing in McEwan's work occupies two (chapter 6 and 10) out of ten chapters to give the reader an overview of the progress of the digital revolution. In their works – one in a fictional, one in a documentary manner – we can observe a conflict of the intuitive and the normative linguistic choice in human-robot interaction. In other words, experiential change does not correlate conceptual change. To solve this conflict, it would be helpful to consider more factors in the human-robot-interaction regarding the experiential aspect: Coeckelbergh's example of the interaction

144

between elderly people and children with robots, where the robots are addressed with you, should be further investigated since the result could be most likely different when it comes to adults with more social possibilities. More importantly, more attention should be paid to conceptual change, and the emergence of the artificial quasi-other nowadays tends to underline the boundary between humans and robots. It raises the question as to which extent should Coeckelbergh's approach complement the traditional third person-perspective and, as Coeckelbergh (2011) argues, how to "steer and shape this change into a desirable direction" (p. 67).

REFERENCES

Asimov, I. (1976). The Bicentennial Man. In: *The Bicentennial Man and Other Stories* (pp. 135-173). Doubleday & Company.

Čapek, K. (1920). *R.U.R.: Rossumovi Univerzální Roboti*. Animedia.

Coeckelbergh, M. (2011). You, Robot: on the Linguistic Construction of Artificial Other. *AI & Society*, *26*(1), 61-69. https://doi.org/10.1007/s00146-010-0289-z

Dick, P. K. (2007). *Do Androids dream of electric sheep?* Orion Publishing Group. (Original work published in 1968)

Hoffmann, E. T. A. (1957). Der Sandmann [The Sandman]. In: *E. T. A. Hoffmannn: Poetische Werke. 3. Band. Nachtstücke* (pp. 3-44). Walter De Gruyter & Co. (Original work published in 1816)

Kehlmann, D. (2021). *Mein Algorithmus und Ich* [My Algorithm and Me]. Klett-Cotta.

McEwan, I. (2019). *Machines like me*. Jonathan Cape.

Mori, M. (2012). The Uncanny Valley. *IEEE Robotics & Automation Magazine*. 19(2), 98–100. https://spectrum.ieee.org/the-uncanny-valley (Original work published in 1970)

Roberts, A. (2016). *The History of Science Fiction*. Palgrave Macmillan.

Searle J. R. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, *3*(3), 417–457. https://www.law.upenn.edu/live/files/3413-searle-j-minds-brains-and-programs-1980pdf

Turing A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Юэ Ли, yue.li@kit.edu

Yue Li, yue.li@kit.edu