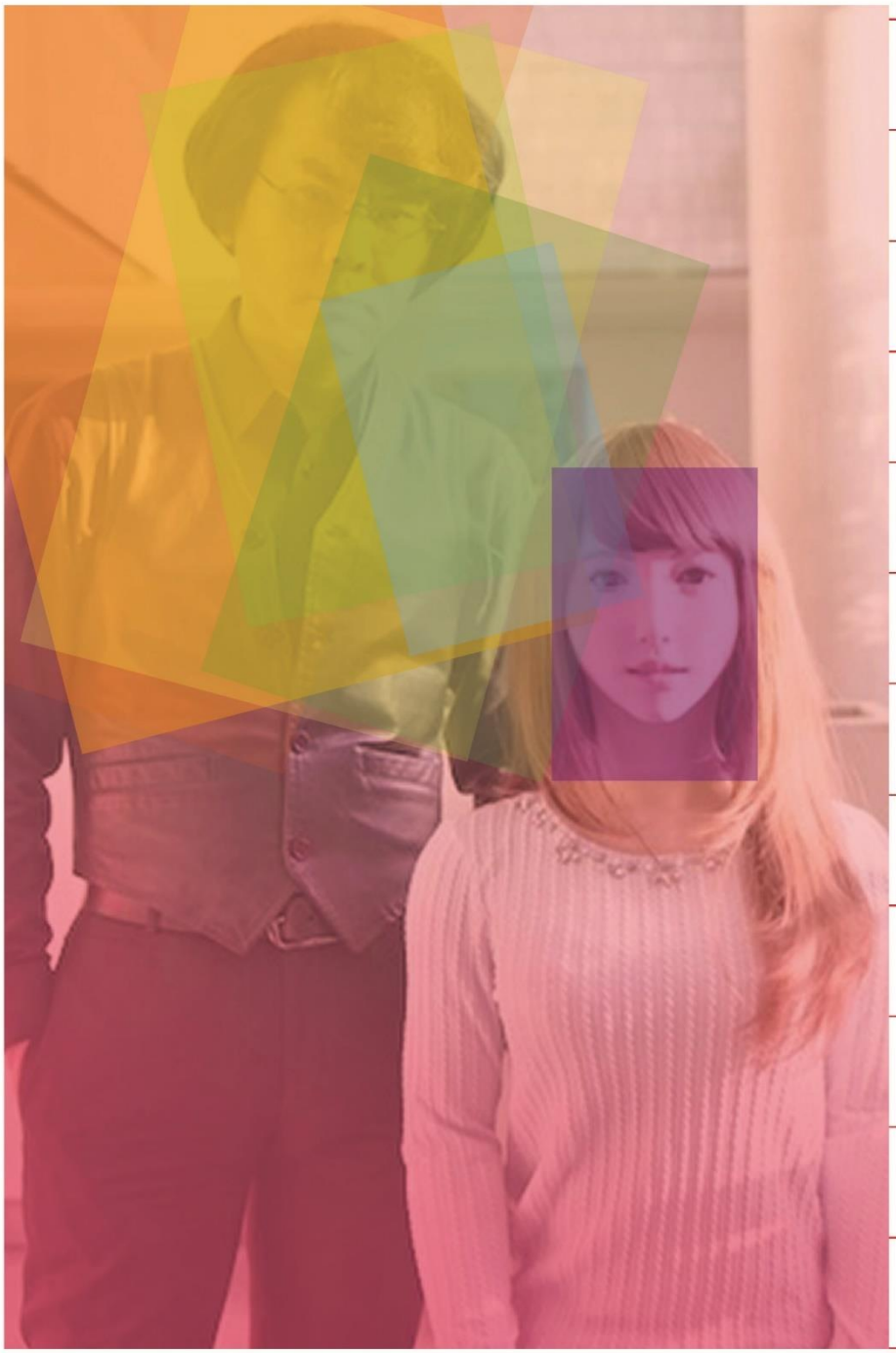
# Technology and Language





# The Construction of the Robot in Language and Culture

3:1

Свидетельство о государственной регистрации:
Эл № ФС 77 – 77604 от 31 декабря 2019, выдано Роскомнадзором
ISSN 2712-9934

**Учредитель и издатель:**
Санкт-Петербургский политехнический университет Петра Великого

**Адрес редакции:**
195251, СПб, ул. Политехническая, д. 29.



**Contacts:**
E-mail: soctech@spbstu.ru
Web: https://soctech.spbstu.ru/en/

ISSN 2712-9934

**Publisher:**
Peter the Great St. Petersburg Polytechnic University

**Corresponding address:**
29 Polytechnicheskaya st., Saint-Petersburg, 195251, Russia

Technology and Language Технологии в инфосфере.
2022. 3(1). i-iii

# Contents

Special Topic:
## The Construction of the Robot in Language and Culture

Technology and Language Технологии в инфосфере.
2022. 3(1). i-iii

Special Topic:

# The Construction of the Robot in Language and Culture

## Guest Editor

**Lin Cheng**

Technology and Language Технологии в инфосфере, 2022. 3(1). 1-8

https://doi.org/10.48417/technolang.2022.01.01
Editorial Introduction

# The Construction of the Robot in Language and Culture, "Intercultural Robotics" and the "Third Robot Culture"

Lin Cheng (✉)
Guangdong University of Foreign Studies, Baiyundadaobei. 2, Guangzhou, 510420, China
lin.cheng@gdufs.edu.cn

## Abstract

Robots are not only technological artifacts, but also elements of human culture. They play important roles such as being the double, replica, tool and companion of humans. The anthropomorphic characteristics of robots lead to philosophical thinking, linguistic and (inter-) cultural phenomena. Inspiration can be drawn from exploring how robots are imagined, defined, described, comprehended, constructed or misunderstood, and from observing the changing relationships between humans and robots from an intercultural and interdisciplinary perspective. For instance, "intercultural robotics" and the "third robot culture" deserve more attention.

**Keywords:** Robot; Robot culture; Intercultural robotics; Sobject; Island hypothesis

**Acknowledgment** This paper was funded by the research project "The Early Imaginations and Novels about Robots in German Literature" (20CWW018) of the National Social Science Foundation of China for Young Scholars.

**Citation:** Cheng, L. (2022). The Construction of the Robot in Language and Culture, "Intercultural Robotics" and the "Third Robot Culture". *Technology and Language*, *3*(1), 1-8. https://doi.org/10.48417/technolang.2022.01.01

УДК 008: 62-529
https://doi.org/10.48417/technolang.2022.01.01
Редакторская заметка

# Конструирование робота в языке и культуре, “межкультурная робототехника” и “третья робототехническая культура”

Лин Ченг (✉)
Гуандунский университет иностранных языков и внешней торговли, Байюндадаобэй 2, Гуанчжоу, 510420, Китай
lin.cheng@gdufs.edu.cn

**Аннотация**
Роботы – это не только технологические артефакты, но также и компоненты человеческих культур. Они играют важные роли, такие как двойник, копия, инструмент и компаньон человека. Антропоморфные характеристики роботов приводят к философскому мышлению, языковым и (меж)культурным явлениям. Вдохновение можно черпать из изучения того, как роботы представляются, определяются, описываются, понимаются, конструируются или неправильно понимаются, а также из наблюдения (изменяющихся) отношений между людьми и роботами с межкультурной и междисциплинарной точки зрения. Например, большего внимания заслуживают “межкультурная робототехника” и “третья робототехническая культура”.

**Ключевые слова:** Робот; Культура роботов; Межкультурная робототехника; Собъект

**Благодарность:** Статья финансировалась исследовательским проектом “Ранние представления и романы о роботах в немецкой литературе” (20CWW018) Национального фонда социальных наук Китая для молодых ученых.

**Для цитирования:** Cheng, L. The Construction of the Robot in Language and Culture, “Intercultural Robotics” and the “Third Robot Culture” // Technology and Language. 2022. № 3(1). P. 1-8. https://doi.org/10.48417/technolang.2022.01.01

The robot is by no means only a technological artifact and concept. Robots play important roles such as being the double, replica, tool and companion of humans. There is no doubt that robots are elements of human culture. The anthropomorphic characteristics of robots lead to philosophical thinking, linguistic and (inter-) cultural phenomena. The word "robot" is a Czech invention in Karel Čapek's (1920) play *R.U.R.* As the word traveled to other languages and cultures, did it become something else? And how does it relate to other concepts of an automaton that imitates humans, such as android, cyborg, or "Maschinenmensch"? How do we perceive and understand robots through their linguistic construction and in their intercultural context? These are topics of this special issue on "The Construction of the Robot in Language and Culture," that is, on explorations of how they are imagined, defined, described, comprehended, constructed or misunderstood.

"Robot" has a very interesting family tree. Taras Romanenko and Polina Shcherbinina examine in their paper *Robot vs Worker* very closely the reception and metamorphosis of "robot" (*R.U.R.*) in Russian literature of the 1920s, especially in Alexei Tolstoy's (1924) adaptation *Riot of the Machines*, only four years after the birth of the "robot" in Čapek's play. They also show us the historical context of its reception in Soviet times (Romanenko & Shcherbinina, 2022). In many cases, the historical and etymological study of the concept should be the very beginning of the discussion about robots. In *The Intellectual Turn and Cultural Transfer of "Humanoid Automata" from the Ancient World to the Enlightenment Era*, Shijueshan Wu (2022) explores the early stories of "robots", that is, the origin and development of the "android" in the Western world, from the ancient world to the Enlightenment era, involving cultural transfers in different civilizations. Kevin Liggieri and Marco Tamborini return to Descartes as a starting point for a reinterpretation and redefinition of the concept of robot in the contemporary world in their paper *The Body, The Soul, The Robot: 21st-Century Monism*. They point out "two linguistic-cultural turning points in the concept of robot" and show "how Cartesian dualism (in the description of humans) becomes a (material) monism in the development and construction of robots" (Liggieri & Tamborini, 2022).

In the American Sci-Fi novel *Tik-Tok* by John Sladek (1985), Dr Riley says to the robot Tik-Tok: "In my opinion the very concept of an automaton or robot is a philosophical concept, giving rise to questions about life, thought, and language – and much more. Yes, I sometimes wonder whether robots were invented to answer philosophers' questions" (p. 72). In this issue, philosophical thinking and linguistic analysis are another focus. Mark Coeckelbergh (2011) published a notable paper "You, Robot: On the linguistic construction of artificial others", arguing that "the appearance of robots in human consciousness is mediated by language: how we use words interprets and co-shapes our relation to others – human others or artificial others" (p. 62). Here, Larissa Ullmann (2022) (*The Quasi-other as a Sobject*), Cheryce von Xylander (2022) (*Quipping Equipment. Apropos of Robots and Kantian Chatbots*), Leon Pezzica (2022) (*On Talkwithability – Communicative Affordances and Robotic Deception*), Daria Bylieva

(2022) (*Language of AI*), Cathrin Hasse (2022) (*Language and Robots: From Relations to Processes of Relations*), and Yue Li (2022) (*Affirming and Denying the Hybrid Character of Robots*) respond to this article from their own perspectives, which spark new concepts and thoughts, such as "sobject" in Larissa Ullmann's response: It "describes a kind of technical objects to which humans can have deeper relations than to conventional objects" (Ullmann, 2022). Obviously, the new era of technologies welcomes new understandings of human-robot relationship. Accordingly, Coeckelbergh (2022) responds in turn to each of these discussions.

Cultural and intercultural dimension should not be absent in the discussion about robots. Robot cultures are different across cultural communities such as Japan and Western countries because of their different religious beliefs, ethical backgrounds and different understandings of the relationship between human and "machinery/artificial other." "Robot culture" can refer to all cultural phenomena that are in a broad sense related to robots – there are also discussions about robots as "creators of culture" (Dunstan et al., 2016). Since the 1980s, media and academic circles have paid attention to the uniqueness of Japanese robot culture and the differences of robot cultures between East and West. Apart from the historical exploration by Wu, there are two interviews in this issue discussing perceptions of robots and human-robot relationships in East and West. In 2008, when the German philosopher Markus Gabriel visited Japan, he had a dialogue with the Japanese engineer Hiroshi Ishiguro (Gabriel et al., 2018). Their dialogue reflected not only the differences of robot perception between a philosopher and an engineer, but also the divergence between East and West in terms of robot cultures, the reception of robots and the prospect of human-robot relationship. In this issue, Hui Jiang, Lin Cheng and Yue Li interview Hiroshi Ishiguro and Markus Gabriel. These two interviews, conducted in 2021 and 2022, are extensions of their dialogue in 2008 – some views remain the same, while others appear to be changing (Li & Gabriel, 2022; Jiang et al., 2022). Robot cultures are still seen to be different in different societies especially in light of the "island hypothesis" of Ishiguro who maintains that "the blurring of the boundaries between humans and robots is a good thing and a new species would be born." These reflect the differences regarding the self-understanding of humans, the human-"other" relationships and the human-technologies prospects in Germany and Japan.

In order to have a better and closer discussion of robot issues such as roboethics, an intercultural perspective is indispensable. It is necessary to explore "intercultural robotics" (Cheng, 2020a, p. 99), that is, in short, on the premise of the cultural phenomena related to robots, to "study the similarities, differences, interactions, mixtures and influences of robot cultures in different cultural communities, with the aim of explaining questions and resolving problems in HRI due to the differences in robot cultures," and at the same time, to "seek common ground in response to the current and future challenges of robotics in the coming era of human-robot coexistence" (Cheng, 2020b, p. 4) – as Gabriel reminds us of an important point in his interview *Diverse Cultures, Universal Capacities*: "the construction of cultural difference has the goal of finding something we

share rather than something that separates us." Obviously, we should also keep in mind the exceptions and variations in robot cultures.

Meanwhile, Thomas Ramge (2019), a German journalist and writer, observed that "robots are enemies in Europe, servants in America, colleagues in China, and friends in Japan" (p. 18). A third kind of robot culture in addition to the robot cultures in Japan and in the West will emerge in China which has its own mode of perception and application of the human-like robots, as we experienced during the Covid-19 pandemic (Jiang and Cheng, 2021) and the Olympic Winter Games 2022 (Ke and Cheng, 2022)? Compared to Western or Japanese society, China does not have a long tradition of robot imaginations and discussions about the relationship between human and artificial beings. There is also no absolute anthropocentrism or impressive dystopian pictures in contemporary Chinese sci-fi works. Both China and Japan show an open and positive attitude towards robots. The Chinese society welcomes robots as a symbol of new and useful technology, but the Chinese do not have "robot dreams" and do not stick to visions of Astro Boy or Doraemon, nor do they think there is "Buddha nature in the robot" (Mori, 1985). The wide applications of humanoid robots and the long traditional Chinese culture are the premise of the emergence of a robot culture in Chinese style. In view of the practical social needs and cultural backgrounds, there is a considerable likelihood that a practically-oriented robot culture will emerge in China. This would be "the third robot culture", unlike the skeptical robot culture in the West and the future-oriented and somehow idealized robot culture in Japan – in accordance with the traditional "Zhong Yong (moderate principle)" of Chinese culture. The application of various robots during the Covid-19 pandemic is an important example of the emerging robot culture in China. An empirical study shows that Chinese people generally held positive attitudes towards "anti-pandemic robots" and appreciated their contributions to reducing the burden of medical care and virus transmission (Jiang and Cheng, 2021). At the same time, roboethics is a topic of great concern to Chinese academia and society. Initiated by the Institute of Philosophy of the Chinese Academy of Social Sciences, the results of a survey were released during the 2nd World Science and Technology Development Forum on November 8, 2020 (World Engineering Day, 2020). This survey of the most concerning ethical issues of technology among Chinese scholars found that roboethics (concerning care robots, social robots, robots for emotional exchange, unmanned drone, human-robot relations, etc.) ranks second among 20 significant ethical issues (such as genetic engineering, euthanasia, human enhancement, AI, brain science, assisted reproductive technology), following the ethical issues of genetic engineering. Moreover, scholars have begun to discuss the Chinese solutions to ease the possible tension between humans and robots, and draw inspirations from Chinese traditional culture such as Confucianism and Taoism. The robot culture is a complicated phenomenon. Diverse perspectives and methods, also the sociological ones, could also be adopted to explain the discrepancies of robot perceptions and applications in different cultural communities.

There are many research projects about robots in the Humanities and in the social sciences, such as the projects "Cultural and Intercultural Dimension of Robots," "Research on Anti-pandemic Intelligent Technology in Cross-cultural Perspective," and "The Early Imaginations and Novels about Robots in German Literature" of my team. In recent years, the research on robots is expanding and is entering into new fields. There is much more to be studied and discussed in the realm of robots. More questions are waiting for answers, for instance, what does our customary way of naming robots mean? Are the new robot forms still suitable for our traditional perceptions of robots? How could the Chinese translation of "robot" ("machine-man") influence the perception of robots in Chinese society? How could robotics engineers and sci-fi writers find the "same language" in the discussion of robots? To what extent can sci-fi works influence us in the imaginations and perceptions of robots? The in-depth and systematical explorations of researchers with intercultural and interdisciplinary backgrounds and understandings about these topics are still needed. As this special issue on "The Construction of the Robot in Language and Culture" shows, we value such idea exchanges, dialogues and further discussions, and we look forward to a continuous discussion of related topics in the near future.

## REFERENCES

Bylieva, D. (2022). Language of AI. *Technology and Language*, *3*(1), 111–126. https://doi.org/10.48417/technolang.2022.01.11

Čapek, K. (1920). *R.U.R: Rossum's Universal Robots*. Ot. Štorch-Marien.

Cheng, L. (2020a). Das Unheimliche der Entfremdung: Humanoide Roboter und ihre Buddha-Natur [The Uncanny of Alienation: Humanoid Robots and Their Buddha-Natures]. *Jahrbuch Technikphilosophie*, *6*, 83-101. https://doi.org/10.5771/9783748904861-83

Cheng, L. (2020b, Sep 15). Phenomena of Robots from an Intercultural Perspective. *Chinese Social Science Today*. http://www.cssn.cn/kxk/202009/t20200918_5184325.shtml

Coeckelbergh, M. (2011). You, Robot: On the linguistic construction of artificial others, *AI & Society, 26*, 61-69. https://doi.org/10.1007/s00146-010-0289-z

Coeckelbergh, M. (2022). Response: Language and robots. *Technology and Language*, *3*(1), 147–154. https://doi.org/10.48417/technolang.2022.01.14

Dunstan, B. J., Silvera-Tawil, D., Koh, J. T. K. V., & Velonaki, M. (2016) Cultural Robotics: Robots as Participants and Creators of Culture. In Koh J., Dunstan B., Silvera-Tawil D., & Velonaki M. (Eds), *Cultural Robotics. CR 2015. Lecture Notes in Computer Science, vol 9549* (pp. 3-13). Springer. https://doi.org/10.1007/978-3-319-42945-8_1

Gabriel, M., Ishiguro, H., Kokubun K., & Chiba, M. (2018, July 15). *Yokubō no jidai no tetsugaku: Marukusu gaburieru Nihon o iku.* [Philosophy in the Age of

Desire: Markus Gabriel in Japan] [TV program]. NHK-BS1 https://www.youtube.com/watch?v=H9J19m4ey8g

Hasse, C. (2022). Language and Robots: from Relations to Processes of Relations. *Technology and Language*, *3*(1), 127–135. https://doi.org/10.48417/technolang.2022.01.12

Jiang, H., Cheng, L., & Ishiguro, H. (2022). The Blurring of the Boundaries between Humans and Robots is a Good Thing and a New Species would be Born: An Interview with Hiroshi Ishiguro. *Technology and Language*, *3*(1), 40-46. https://doi.org/https://doi.org/10.48417/technolang.2022.01.05

Jiang, H., & Cheng, L. (2021). Public Perception and Reception of Robotic Applications in Public Health Emergencies Based on a Questionnaire Survey Conducted during COVID-19. *International Journal of Environmental Research and Public Health*, *18(20)*, 10908. https://doi.org/10.3390/ijerph182010908

Ke, D. & Cheng, L. (2022, Feb 14). Robots on the Ice and Snow Fields: Escort the Winter Olympics and Empower the "Intelligent Games". *China Daily*. https://cn.chinadaily.com.cn/a/202202/14/WS6209f716a3107be497a06352.html

Li, Y. (2022). Affirming and Denying the Hybrid Character of Robots: Literary Investigations. *Technology and Language*, *3*(1), 136–146. https://doi.org/10.48417/technolang.2022.01.13

Li, Y. & Gabriel M. (2022). Diverse Cultures, Universal Capacities: Interview with Markus Gabriel. *Technology and Language*, *3*(1), 47–56. https://doi.org/https://doi.org/10.48417/technolang.2022.01.06

Liggieri, K., & Tamborini, M. (2022). The Body, The Soul, The Robot: 21st-Century Monism. *Technology and Language*, *3*(1), 29–39. /https://doi.org/10.48417/technolang.2022.01.04

Mori, M. (1985). *The Buddha in the Robot. A Robot Engineer's Thoughts on Science and Religion*. (C. S. Terry, trans.). Kosei Publishing.

Pezzica, L. (2022). On Talkwithability. Communicative Affordances and Robotic Deception. *Technology and Language*, *3*(1), 104–110. https://doi.org/10.48417/technolang.2021.04.10

Ramge, T (2019). *Who's Afraid of AI? Fear and Promise in the Age of Thinking Machines*. The Experiment.

Romanenko, T., & Shcherbinina, P. (2022). Robot vs Worker. *Technology and Language*, *3*(1), 17–28. https://doi.org/https://doi.org/10.48417/technolang.2022.01.03

Sladek, J. (1985). *Tik-Tok*. DAW Book.

Tolstoy, A. (1924). *Bunt mashin: fantasticheskie stseni* [Riot of the Machines: Fantastic Scenes]. Time.

Ullmann, L. (2022). The Quasi-Other as a Sobject. *Technology and Language*, *3*(1), 76–81. https://doi.org/10.48417/technolang.2022.01.08

von Xylander, C. (2022). Quipping Equipment. Apropos of Robots and Kantian Chatbots: Commentary on Mark Coeckelbergh, "You, robot: on the linguistic construction of artificial others" (2011). *Technology and Language*, *3*(1), 82–103. https://doi.org/10.48417/technolang.2021.04.09

World Engineering Day (2020, Nov 8). Results of the Survey "The most concerning ethical issues of technology among Chinese scholars". https://mp.weixin.qq.com/s/Ve7VQfCdw5jTMtSUoTyLTg

Wu, S. (2022). The Intellectual Turn and Cultural Transfer of "Humanoid Automata" from the Ancient World to the Enlightenment Era. *Technology and Language*, *3*(1), 9–16. https://doi.org/10.48417/technolang.2022.04.03

**СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR**

Лин Ченг, lin.cheng@gdufs.edu.cn,
0000-0002-0837-9129

Lin Cheng, lin.cheng@gdufs.edu.cn,
0000-0002-0837-9129

Статья поступила 10 марта 2022
одобрена после рецензирования 21 марта 2022
принята к публикации 21 марта 2022

Received: 10 March 2022
Revised: 21 March 2022
Accepted: 21 March 2022

https://doi.org/10.48417/technolang.2022.01.02
Research article

# The Intellectual Turn and Cultural Transfer of "Humanoid Automata" from the Ancient World to the Enlightenment Era

Shijueshan Wu (✉)
Capital Normal University, School of Foreign Languages, No. 2 North Xisanhuan Road, Beijing, 100089, China
rainwasterland@hotmail.com

## Abstract

This study examines the origin and development of the "android" in the Western world, from antiquity to the Enlightenment era. The manufacture of android automata is not only a technological advance, but also reveals an intellectual shift from the Middle Ages to the Enlightenment, involving cultural transfers from different civilizations in ancient times. "Humanoid automata" offer an insight into medieval beliefs and practices as mechanical mimesis in the investigation of the relations between art and nature. Android automata in the 18th century represent Enlightenment ideas through their affective communication. This historical context could provide an important reference for today's research on human–robot interaction.

**Keywords:** Humanoid automata; Enlightenment automata; Android; Robot; Human–robot interaction

**Acknowledgment** This paper was funded by the research project "The Early Imaginations and Novels About Robots in the German Literature" (20CWW018) of the National Social Science Foundation of China for Young Scholars.

**Citation:** Wu, S. (2022). The Intellectual Turn and Cultural Transfer of "Humanoid Automata" from the Ancient World to the Enlightenment Era. *Technology and Language*, *3*(1), 9-16. https://doi.org/10.48417/technolang.2022.01.02

УДК 94: 62-529
https://doi.org/10.48417/technolang.2022.01.02
Научная статья

# Интеллектуальный поворот и культурный перенос “гуманоидных автоматов” из античного мира в эпоху Просвещения

Шицзуйшань Ву (✉)
Столичный педагогический университет, Пекин, № 2, Норд- Сисанхуань-роуд, 100089, Китай
rainwasterland@hotmail.com

**Аннотация**
В данном исследовании рассматривается происхождение и развитие “андроида” в западном мире, от античности до эпохи Просвещения. Производство роботов-андроидов – это не только технологический прогресс, но и интеллектуальный переход от Средневековья к Просвещению, включающий культурный перенос между разными цивилизациями в древние времена. “Гуманоидные автоматы” предлагают заглянуть в средневековые верования и практики как механический мимесис в исследовании отношений между искусством и природой. Автоматы-андроиды в 18 веке представляют идеи Просвещения в их аффективной коммуникации. Этот исторический контекст может стать важной отправной точкой для сегодняшних исследований взаимодействия человека и робота.

**Ключевые слова:** Человекоподобные автоматы; Автоматы Просвещения; Андроид; Робот; Взаимодействие человека и робота

**Благодарность:** Статья финансировалась исследовательским проектом “Ранние представления и романы о роботах в немецкой литературе” (20CWW018) Национального фонда социальных наук Китая для молодых ученых.

**Для цитирования:** Wu S. The Intellectual Turn and Cultural Transfer of “Humanoid Automata” from the Ancient World to the Enlightenment Era // Technology and Language. 2022. № 3(1). P. 9-16. https://doi.org/10.48417/technolang.2022.01.02

## INTRODUCTION

"Why are they humanoids, and is it necessary to make idealized humanoid androids? Why are humans so eager to make a form similar to themselves?" The robot-detection engineer Haraway raises these doubts at the beginning of the Japanese animated cyberpunk film *Ghost in the Shell 2: Innocence* (Oshii, 2004). We posed the same questions to Japanese roboticist Hiroshi Ishiguro, wondering why he made many humanlike robots which remind us of the ideal female gynoid called "Andréide" in the novel *Tomorrow's Eve* (*L'Ève future*, 1886) written by French symbolist writer Villers de L'Isle-Adam (1886/2000). If Mary Shelley's (1818/1992) "Frankenstein's monster" is far from a humanoid creation in form, then the "Andréide" made by Edison in *Tomorrow's Eve* could be considered the most perfect humanoid robot, even better than real people. However, why build such humanlike robots or, in other words, "androids" in science-fiction narratives and in reality? What is the origin of humanoid robots? Why did the mechanical-makers and science-fiction writers incarnate human characteristics in machines? To explore these questions, it is necessary to review the origins of humanlike robots.

## CULTURAL TRANSFER OF EARLY AUTOMATA IN ANCIENT CIVILIZATIONS

The predecessors of humanlike robots are "automata" which originated in the ancient world. "Automata" were to some extent imagined before they were built. The first automata in the Western tradition appeared in Ancient Greece. In Western literature the half-god Hephaestus as depicted in Homer's *Iliad* may be the first fabricator of imagined artificial beings, mobile tripodal creatures capable of attending to the gods. Automated guards and the simply giants can be found elsewhere in Greek literature. In the third century, engineers and architects in Alexandria began to design automata to illustrate their mechanical principles, documenting their creations in texts that detailed their construction (Truitt, 2015, p. 4). These fictional automata appeared in historical texts and romances, in travelogues and encyclopedias, even in the *chansons de geste*, all of which can be read as historical texts that offer insights into medieval beliefs and practices (Truitt, 2015, p. 7).

Elly Rachel Truitt (2015) established an archaeology of automata by citing many types of mechanical production, such as golden birds and beasts, robotic servants, brass horse men, gilded buglers, and papier-mâché drummers. She points out that in the Middle Ages there was no fixed term, or even a set of terms to refer to these "self-moving or self-sustaining manufactured objects" (Truitt, 2015, p. 1-2). The word "automaton" is used to describe a constructed device with powered working parts, often but not necessarily humanoid, often but not always immobile, usually unconscious. The word came into popular use in 16$^{th}$ century France after Rabelais (1534/1994) employed "*automate*" to denote in *Gargantua* (1534) a machine with a self-contained principle of motion The term

was further used in the 20th century to encompass a variety of more specialized terms, including "robot," "android," and "cyborg" (Truitt, 2015, p. 2-3).

Truitt used the term "robot" to refer to these actual and imagined medieval automata which implies a kind of avant-garde nature of these machines, although the mechanical machines in this period were not totally humanlike. Medieval robots identified and explored the multiple types and functions of automata in the Latin Middle Ages, which demonstrates that these objects have long been used to embody complex ideas about the natural world (Truitt, 2015, p. 1-2). The writers and mechanical makers in Latin Christian culture began to explain essential problems about the human body and cosmos within a framework of mechanical technology, which is one symbol of modernity in the 17th century. An intellectual turn occurred in the Western world through mechanistic thinking as a way of knowing the universe. Automata could be considered a link between art and nature, knowledge and technology, real and imaginary by revealing the most primitive confusions of the Middle Ages regarding about life and death, nature and manufacture, magic and science.

In addition, the manufacturing of robots in the Latin Christian West reflected the direct cultural transfer between the European world and other civilizations. Engineers and inventors from other ancient civilizations also attempted to build self-operating machines, resembling humans and other animals, and the objects built in medieval Europe were inspired by automata made in the Byzantine Empire and the Islamic world. (Truitt, 2015, p. 5) Automata appeared in the Islamic world and the Byzantine and Mongol empires much earlier than in the Latin West and entered Europe as gifts or in travelogues (Truitt, 2015, p. 8). In this process of cultural and mechanical reception, these automata were initially treated as foreign objects linked to their origins. It was not until the 15th century, following the construction of the first mechanical automaton in Europe, that automata manufacture became fully integrated into Latin culture and gradually disconnected from their foreign origin. This cultural integration predated future humanoid automata in the 18th century which no longer presented an exotic style but instead represented Enlightenment ideas.

The long history of automata in the Western world, from antiquity to the medieval period, both actual and fictional, demonstrates the ambition of designers to imitate the creations of God. These robots began to take on a humanlike appearance. In mechanical "mimesis," these humanlike automata mimicked natural forms and displayed a powerful cultural coherence among several ancient civilizations which led to the concept and manufacture of modern robots. These automata, from all civilizations, revealed their "own in-betweenness": they were "lifelike copies of natural objects, or eternal bodies that hover between life and death" (Truitt, 2015, p. 9). Humans imitated nature by mechanical means and blurred the boundaries between the natural and artificial. Consequently, this cultural and intellectual import raised a double concern: on the one hand, the introduction of foreign knowledge into the Latin Christian framework; on the other hand, concern about mimesis and manufacture from the perspective of ethics and religion.

## "ANDROID AUTOMATA" IN THE ENLIGHTENMENT ERA

This double concern continued beyond the Middle Ages into the Enlightenment period and culminated in the works of Jacques de Vaucanson. The famous French humanoid automaton inventor of the 18th century created the imitation design paradigm of humanoid automata by combining mechanics, physiology, and anatomy. His work reproduced the internal structure or physiological processes of humans or animals, blurring the boundaries between machines and humans, living and non-living. He presented three automata to the Académie des Sciences in Paris in 1738: a flute player, a galoubet player, and a duck. Under the influence of Vaucanson, the first "android" manufacturing boom in human history appeared in the Western world.

The term "android", which means "humanlike", was initially used for automata. The meaning of this word is evident through its etymological formation: the prefix "andro- (ἀνδρ-)" means "of men or male"; the suffix "-oid (-εἶδος)" relates generally to adjectives and nouns denoting form or resemblance, so the word "android" signifies "a human-like" robot. The earliest origin of the word "android" can be traced to Ephraim Chambers'(1728) *Cyclopaedia*, referring to an automaton produced by Albertus Magnus. The term "android" then appears in US patents in 1863 as miniature humanlike toy automatons. Villers de L'Isle-Adam created the "Andréide" in his novel *Tomorrow's Eve*, which greatly helped to popularize the term "android." Following the publication of this novel, the term "android" was used in a more modern sense. Adelheid Voskuhl (2013) employs the term "android automaton" to describe the humanoid robots made in Europe in this period to emphasize the humanoid appearance and human characteristics of these "android automata."

The 18th century saw the creation of a number of remarkable mechanical androids: at least 10 prominent automata were built between 1730 and 1810 by clockmakers, court mechanics, and other artisans from France, Switzerland, Austria, and the German lands (Voskuhl, 2013, p. 5). During this period, some life-size automata appeared in the public. In 1760, Friedrich von Knaus, a mechanic employed at the imperial court in Vienna, built a writing machine for the mechanical cabinet of the Holy Roman Emperor Franz Stephan. It was the first mechanical device to implement the process of writing. Father and son Jaquet-Droz, clockmakers from Switzerland, introduced in 1774, three android automata: a writer, a draftsman, and a harpsichord player. These automata are often considered the most spectacular and appealing of the entire period. A dulcimer player made by the cabinetmaker David Roentgen in collaboration with the clock-maker Peter Kinzing was presented to the public which was also a gift offered to Marie-Antoinette in 1785 (Voskuhl, 2013, p. 2). The other eight android automata we know from the 18th century display a variety of figures and activities. They were made by clock-makers, cabinetmakers, court mechanics, and other artisans with the means, skills, and incentives to do so. Through their distinctive mechanical design, an entire setting of 18th century cultural and political activity was replicated. The creators incorporated their Enlightenment ideas into these automata, so they were not only a technical advance. Regarding mechanical complexity, these Enlightenment automata were

spectacular and innovative self-moving objects and, the most sophisticated of their kind compared not only with earlier periods, but also later ones. These automata have served throughout all stages of industrialization during the 19th, 20th, and 21st centuries as eloquent metaphors for social and economic changes (Voskuhl, 2013, pp. 1-2). Android automata made during the Enlightenment have played an influential role in our understanding of modern industrial society and the human-machine boundary. As Voskuhl states:

> "Android automata, robots, and mechanical humans have been central to our understanding of the relationship between humans and machines. They are designed to look and move as human beings do and perform motions and techniques such as walking, writing, or music-making. For a spectator, it is often difficult to determine whether an android object is human or machine. Because androids so effectively destabilize our sense of the boundary between humans and machines and, by extension, our sense of our own constitution, they and their histories evoke a broad range of concerns, most significantly, perhaps, those related to the promises and perils of the modern industrial age." (Voskuhl, 2013, p. 1)

These "android automata" all represent the same concern for "mimesis." Jessica Riskin (2003a) calls 18th automata "mechanical experiments" and points out that Jacques de Vaucanson is the founder of this approach by simulating mechanistic physiology. She also explicitly indicates the "epistemological utility" of Vaucanson's flute player, because they could help us know ourselves. As mechanical replicas of humans or of animals, these android automata provided knowledge about humans' and animals' anatomy and physiology. She also argues that Vaucanson made the transition from "representation to simulation" because the mechanics of the 17th century only aimed to represent or imitate an external appearance (Riskin, 2003a, p. 616). Voskuhl (2013) names these android automata "the intersection of philosophy and clock- and instrument-making": "Android automata from this period have been thought of as products of a marriage of philosophy and craftsmanship, too, since they mimic mechanically a key part of our cosmos, the human body" (p. 16). Consequently, the Enlightenment automata provide a self-referential perspective not only for their mechanical makers, but also for the modern machine age to understand the relationship between human and machine.

These automata really attracted a great deal of attention. Vaucanson's celebrated duck excited much interest throughout Europe, as well as von Kempelen's chess player, which was exhibited after its completion "to thousands," even after it had been revealed to be a hoax. Riskin (2003b) assumes that all Enlightenment's automata were of epistemic relevance both to their makers and to the public (p. 101). These android automata were deliberately produced for this function, and their makers promoted public exhibitions of them. In addition, mechanicians in the Enlightenment era attempted to destroy the human–machine boundary through their humanoid creation not only to imitate human appearance and simple behavior, but also by wanting their androids to communicate affects and sentiments to the audience, such as two female humanoid automata – the

dulcimer player and the harpsichord player – built by German and Swiss artisans. These musical androids can move their arms, hands, and fingers to play music, but also move their heads, eyes, and torsos in rhythm with the music they play, so the automata could perform subtle and graceful body motions in addition to playing music (Voskuhl, 2013, p. 6-7). The Enlightenment, long known as the "Age of Reason," has increasingly been associated with "feeling," which has been represented by bodily senses and physiological activities following Descartes and his contemporaries. The automata's bodily motions correspond to 18th-century performance techniques to communicate affects.

The Enlightenment period is recognized as a milestone in the formation and development of modern society. Therefore, the android automata in this period are often taken to be pioneers of the modern, industrial machine age. A key concept of "modern" is "mechanized" which also symbolizes a fundamental change in the social, cultural, aesthetic, and religious constitution of human beings. This technological and intellectual shift profoundly impacted the humanoid robots that appeared in the later science-fiction narratives and the androids of today.

## CONCLUSION

There exists a certain continuity between the robots recorded in the texts of ancient civilizations and the "android automata" of the Enlightenment period. They all reflect the essentially natural problems of "mimesis" of ourselves leading to the boundaries between life/death, man/machine, natural/artificial, eternity/temporality. Gaby Wood (2002) proposes that there exists a continuity connecting androids from the 18th century to the robots, science fiction figures, and artificial life of the present time. Villiers, for example, was familiar with the history of robot manufacturing prior to the 19th century. Almost all the important mechanical producers' names appear in the novel *Tomorrow's Eve*, such as Albert the Great, Vaucanson, Maelzel, and Horner. This idea inspires another one, namely that there has always been a large and interested audience for robots and that this audience's concerns and preoccupations coincide with our own current concerns the human–machine relationship in the postindustrial societies (Wood, 2002, p. 24). Both, the androids in science-fiction narratives and in real robot manufacturing are profoundly influenced by these automata in the ancient world and during the Enlightenment. Today, "robot" refers to a larger category of humanoid or anthropomorphized machines. With current technologies, it is possible to build robots with warm and soft bodies, and androids can thus elicit the most natural communication. However, what happens when androids can be completely treated as intelligent beings similar to humans? The exploration of this human–robot interaction is what we face today.

## REFERENCES

Chambers, E. (2010). Androides. In *Cyclopaedia: or, An Universal Dictionary of Arts and Sciences*. Gale ECCO. (Original work published 1728). https://artflsrv03.uchicago.edu/philologic4/chambers_new/navigate/1/1073/

L'Isle-Adam, A. V. de. (2000). *Tomorrow's Eve*. University of Illinois Press. (Original work published 1886)
Oshii, M. (Director). (2004). *Ghost in the Shell 2: Innocence* [Film]. Production I.G Studio Ghibli.
Rabelais, F. (1994). *Gargantua.* HATIER. (Original work published 1534)
Riskin, J. (2003a). The Defecating Duck; or, The Ambiguous Origins of Artifi cial Life. *Critical Inquiry*, *20*, 599–633. https://doi.org/10.1086/377722
Riskin, J. (2003b). Eighteenth-Century Wetware. *Representations*, *83*, 97–125. https://doi.org/10.1525/rep.2003.83.1.97
Shelley, M. (1992). *Frankenstein; or, The Modern Prometheus*. Penguin Classics (Original work published 1818).
Truitt, E. R. (2015). *Medieval Robots: Mechanism, Magic, Nature, and Art*. University of Pennsylvania Press.
Voskuhl, A. (2013). *Androids in the Enlightenment: Mechanics, Artisans, and Cultures of the Self*. University of Chicago Press.
Wood, G. (2002). *Living Dolls: A Magical History of the Quest for Mechanical Life*. Faber and Faber.

## СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Шицзуйшань Ву, rainwasterland@hotmail.com

Shijueshan Wu, rainwasterland@hotmail.com

Статья поступила 7 февраля 2022
одобрена после рецензирования 21 февраля 2022
принята к публикации 2 марта 2022

Received: 7 February 2022/
Revised: 21 February 2022
Accepted: 2 March 2022

https://doi.org/10.48417/technolang.2022.01.03
Research article

# Robot vs Worker

Taras Romanenko (✉) and Polina Shcherbinina (✉)
Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Polytechnicheskaya, 29, 195251, Russia
romanenko.tp@edu.spbstu.ru; tsherbinina.ps@edu.spbstu.ru

## Abstract

The word “robot” first appeared in 1920 in the play “R.U.R.” by Czech writer Karel Capek. Within a few years, the play was translated into more than 30 languages, contributing to the spread of the new term around the world. The word “robot” was preserved in almost all translations, one of the few exceptions being Alexei Tolstoy’s Russian adaptation entitled “Riot of the Machines” (1924). Although in Russian, as well as in Czech, there is an etymological connection between “robot” and “work (rabota)”, the translator Tolstoy abandoned the new term, calling robots “workers (rabotnik)” , that is, refusing to give them a separate name and equating them to working people. Although the origin of the word “work ” “worker” in Russian (as well as in Czech) is associated with slavery and forced labor, in Soviet times it acquired a brightly positive connotation. If for Capek the difference between robots and people becomes the fault line of the play, their similarity is most important for Tolstoy – the performance of work. Accordingly, the theme of the robot’s rebellion against humans is replaced by the rebellion of workers, whether of natural or artificial origin, against their oppressors.

**Keywords:** Robot; Worker; Artificial; Philistine; R.U.R.

**Acknowledgment** We would like to express our gratitude to Daria S. Bylieva, who was our mentor and the most honest critic of this work, as well as to Artem Martynov.

**Citation:** Romanenko, T. & Shcherbinina, P. (2022). Robot vs Worker. *Technology and Language*, *3*(1), 17-28. https://doi.org/10.48417/technolang.2022.01.03

УДК 821.161.1: 62-529
https://doi.org/10.48417/technolang.2022.01.03
Научная статья

# Робот vs Работник

Тарас Романенко (✉) и Полина Щербина (✉)
Санкт-Петербургский политехнический университет Петра Великого, Политехническая, 29, 195251, Санкт-Петербург, Россия
romanenko.tp@edu.spbstu.ru; tsherbinina.ps@edu.spbstu.ru

**Аннотация**
Слово “робот” впервые появилось в пьесе чешского писателя Карла Чапека “R.U.R.” в 1920 году. За несколько лет пьеса была переведена более 30 языков, способствуя распространению нового термина по миру. Слово “робот” сохранялось практически в всех переводах, однако русская адаптация пьесы Алексея Толстого “Бунт машин” (1924) стала одним из немногих исключений. Хотя в русском языке так же как и в чешском этимологически прослеживается связь “робота” и “работы”, автор отказался от нового термина, называя роботов “работниками”, то есть отказывая им в отдельном именовании и приравнивая к работающим людям. Хотя происхождение слова “работа”, “работник” в русском языке (как и в чешском) связано с рабством и подневольным трудом, в советское время оно приобретает ярко позитивную окраску. Если у Чапека линией разлома пьесы становится отличие роботов от людей, то для Толстого важнейшим становится их сходство – выполнение работы. И соответственно тема восстания роботов против людей заменяется восстанием работников, естественного или искусственного происхождения, против угнетателей.

**Ключевые слова:** Робот; Работник; Искусственный; Обыватель; R.U.R.

**Благодарность:** Выражаем благодарность Дарье Сергеевне Быльевой, которая была нашим руководителем и самым главным рецензентом этой работы, а также Артему Мартынову.

**Для цитирования:** Romanenko, T. & Shcherbinina, P. Robot vs Worker // Technology and Language. 2022. № 3(1). P. 17-28. https://doi.org/10.48417/technolang.2022.01.03

## INTRODUCTION

In November 1920 Karel Capek published the science-fiction play "R.U.R." With it came the first appearance of the word "robot." The play soon premiered in Aken (1921), Berlin (1923), New-York (1922), London (1923), Paris (1924) (Horakova, 2010) and in Leningrad (1924). By 1923 it was translated into thirty languages, ensuring the widespread popularity of the new term. In almost all translations the original term "robot" was simply taken over and not translated at all. However, there were some exceptions (Čejková, 2021). For instance, "R.U.R" was performed in Tokyo in 1924 under the title "Jingo Ningen" (Artificial Human) (Robertson, 2014). In Japan, Capek's robots were experienced as "living" things which underscores that the "robot" as such does not exert any ontological pressure to make distinctions between organic/inorganic, animate/inanimate, human/nonhuman forms. On the contrary, all of these forms are linked to form a continuous network of beings (Kaplan, 2004). In Russia the writer Alexey Tolstoy chose the title "Riot of the Machines (Бунт машин)" for the play, and used the word "Rabotnik (worker)" to designate the robots.

## TRANSLATIONS OF "R.U.R." IN RUSSIAN IN THE BEGINNING OF THE XX CENTURY

The word "robot" first appeared in the title of Karel Capek's play "R.U.R."– Rossumových Univerzálni Roboti. "When Karel came up with the idea of a play about minded living machines, he turned to his brother: 'I do not know (…) how I'll call these artificial workers. I would call them laborji (Czech laboři from English 'labour') but it seems to me that it is too bookish.' 'So call them robots,' suggested Joseph" (Skvorchevsky et al., 2020). In 1921, the abbreviation R.U.R. was deciphered in English as "Rossum's Universal Robots" without a translation of the last word. In 1922, Otto Pick provided a translation into German, also preserving the word "robot. " He changed, however, the name of the inventors to "Werstand" – probably from the German word "Verstand (understanding or mind)" because in the original version of the play the surname "Rossum" recalls the Czech word "rozum" for mind. That is why the play became known in Germany as "WUR." In postrevolutionary Russia the situation was a little more complicated. Due to ideological considerations censors did not allow the translation by Evgeniy Georgievich Gerken-Baratinskiy to be printed in 1923. At the same time, the poet, novelist, and playwright Alexey Nikolayevich Tolstoy returned from emigration to the USSR, and agreed with Georgy Alexandrovich Krol[1] on a joint translation of European plays for the Soviet theater (Pervushina, 2021, p. 85). In emigration Tolstoy had been delighted with Capek's play: "Absolutely ingenious, dynamite in content and dynamic in the power of development of the action" (Nikolsky, 2004, p. 406). He decided to realize the planned work in his homeland and to adapt it to Soviet realities. Previous attempts to translate the play by Isai Benediktovich Mandelstam and Yevgeny Georgievich Gerken were not allowed on the stage for

[1] Director and screenwriter, student of Meyerhold, Krol worked in the film industry since 1919, making films in Finland and Germany, lived in Rome, then moved to Berlin.

reasons of ideological unacceptability. Another translation from the Czech original was prepared by Irina Kallinikova, but hers was seen in the USSR only much later. Thus, the adaptation "Riot of the Machines" was shown to Soviet audiences earlier than "R.U.R." Also, judging from his notes, it appears that Tolstoy was working with the German translation by Pick, because he also referred to the play as "W. U. R." Within two months, using Capek's theme and Krol's translation, Tolstoy created his adaptation. In this adaptation Tolstoy decided to abandon the word "robot", replacing it by the more familiar and understandable word "worker" (Nikolsky, 2004, pp. 406-417).

## ETYMOLOGY OF THE WORDS "WORKER" AND "ROBOT"

In the Czech language there are two verbs for labor activity with similar meanings: "pracovati (to work)" and "robiti (to do)." From the latter, a few words are formed with a derogatory connotation: There is "robotiti (to graft, to grub)," also "robota (drudgery, corvee labor)," "rob (slave)," "robotnik (forced laborer)," or "robotny (serf)." Thus, the term "robot" is formed from these derogatory words and conveys their meaning (Skvorchevsky et al., 2020). In Russian, there is a similar verb for work "работать (rabotat)" and a noun form "работа" as well as "раб (slave)."

Thus, the word "robot" would not sound exotic, and was also, as in Czech, associated with the word "work." However, thanks to or despite this similarity, Alexei Tolstoy abandoned the use of the new word "robot" in favor of the usual "работник (rabotnik, worker)."

This "worker" is a derivative of the verb "работать (rabotat, to work)" which, in turn, was formed from the Old Slavonic word "work," that is, "bondage," and the Old Russian "работа (rabota)" which means "slavery" (Semenov, 2003). And in the Old Slavic language, the word "работать (rabotat)" came from the Indo-European dialect, in which the root "orbhos" meant "corvée[2],, work for someone" (Semenov, 2003).

However, in Russian there is also an alternative version of the verb "работать (rabotat, to work)." In pre-revolutionary dictionaries, a sharp distinction was made between two variants of "work": "работать (rabotat)" and "трудиться (truditsya)." So in the dictionary of Ishmael Sreznevsky (1912), "работа (rabota)" has as its first meaning "slavery, bondage, service," whereas "труд (trud)" is interpreted as "care, trouble" (Sreznevsky, 1912). Thus, we can say that the peasant worked in his field, and worked in the corvee, whereas free people work, burdened exclusively with their own chores, or as hired laborers.

However, the revolutionary events of 1917 completely changed the attitude towards the word "worker" – be it "работник (rabotnik)" or "рабочий (rabochii)," the latest word for worker which had appeared in dictionaries only from the middle of the 19th century onward (Chernykh, 1999). In the dictionaries of the Soviet period, the word "rabochii" for worker acquired a pronounced positive connotation, adding a quote from Stalin to dictionary entries (Ushakov, 1939): "Rabochii – Under capitalism, the

[2] In Russia there were peasants belonging to the landowners. They had to work on the land of the landowner – this was called corvee, and they also had their own land, where they worked "for themselves."

same as a proletarian; in the USSR, a person professionally engaged in physical labor and belonging to the ruling class, who owns the means of production together with the whole people. 'Soviet society consists, as you know, of two classes, workers and peasants.' Stalin."

The absence of a linguistic difference between robots and working people in Alexei Tolstoy's adaptation emphasizes that the main thing is what unites them – the performance of physical work, which in the context of Soviet reality goes from burden to pride. As Günther (2000) points out, "According to the ideology, the hero of socialist labor stands at the top. It is associated with the Promethean tradition of a cultural hero who gives people technical, scientific, artistic, and other achievements" (p. 746).

In the fourth scene of the play Moray, the creator of the workers, advertises them by saying: "My universal artificial workers will do all this. They count, write, cook, clean rooms, plow, reap, work on any machine tool" (Tolstoy, 1924/1983, p. 78). In the thirtieth scene, Mikhail dictates a campaign letter to Adam, a worker of the "Adam" series: "Who mines coal and ore, cultivates the land, builds cities, roads, ships, cars, cuts down forests, blocks waterfalls, drains the tundra, breaks granite and marble. Period, question mark... You, you, only you, workers of the Morea" (Tolstoy, 1924/1983, p. 97).

Here one can see the similarity to the poem "Who is the hero?" (1921) by the Russian poet of Soviet times Vladimir Mayakovsky. Hard work is praised and is no longer compared to punishment and hard labor.

> "Who is the hero?
> The one who is better than others hammers coal underground,
> who for ore passed the mountain, –
> that hero!" (Mayakovsky, 1969)

The title of Karel Capek's play employs the word "robot" which the Soviet author refused, so it was necessary to come up with a different title. Thus, Tolstoy named his adaptation "Riot of the Machines", which, on the one hand, emphasized the current revolutionary meaning, and on the other hand, avoided distorting the essence of "work." After all, by calling the play "Riot of Workers," readers could get the impression that we are talking about ordinary workers, working people. And what a machine is, many had an idea, and as the story progresses, the words machine and worker fused. At the beginning of the century, the machine had become poetic. In the collection "Poetry of the Work Strike" of 1918, Alexei Gastev – a poet as well as head of the Central Institute of Labor – wrote about the unity of human and machine: "…. it's hard to understand where the machine is, where the human is. We merged with our iron comrades, we sang with them, together we created a new soul of movement, where the worker and the machine are inseparable" (Gastev, 1924/1971). Moreover, he drew an analogy between humans and machine: "The worker-colonizer goes into the wilds of his own country, firstly, himself, like a deftly oiled, adjusted, automatically regulated machine. Bones-levers, muscles-motors, nerve-impulses – everything in him is active and instrumentally adjusted" (Gastev, 1924/1971). The main feature of Tolstoy's adaptation, which unites

the worker and the working person, is work. For Tolstoy, a worker is a machine capable of performing labor that is often assigned to a person.

## ROBOTS/WORKERS AND HUMANS

The plays "R.U.R" and "Riot of the Machines" are very similar in many ways, but they have an important difference. Initially, this is the ratio of ordinary workers to robots. In the work of Capek, ordinary people are dissatisfied with the appearance of robots as a replacement for their work, and as a result, they are forced out of their jobs: "When the workers rebelled against the robots and began to break them" (Capek, 1966). In Tolstoy's version, on the contrary, workers support the robots. They see themselves in them and see the same oppression on the part of the masters, therefore they fight together against the oppressors, and not against each other. "Do you think your workers are submissive? We will teach them to hate you too," says one of the workers (Tolstoy, 1924/1983, p. 78).

In Alexey Tolstoy's play "Riot of the Machines", the author introduces a very interesting new character. At first glance, this is an ordinary uncomplicated entertainer, whose task is to announce the play "And today's play, in my opinion, is very unpleasant, terrible. I would never waste money and time to have my hair stand on end for an entire evening. And look what the author is doing here" (Tolstoy, 1924/1983, p. 73). But what is so special about this? Here one needs to pay attention to the fact that this character does not have a name but is simply called "The Philistine." While a name gives a person a fixed personality, this character acts as a collective image in which everyone could see themselves.

As the plot develops, the Philistine in every possible way promotes dialogue with the audience, agitating for active participation in the play. For the people in the audience events are not in the past or in a fictional world, but happen now, in fact, and the Philistine, not willing to endure them, enters himself into the course of things. He is a conductor between the world of real things and the world of works. His goal is to immerse spectator in what is happening.

The Philistine's attitude towards robots is consumerist. Having bought a robot for himself ("I bought it for my last pennies, – and called it Vas'ka"), he sends the machine to find money, orders it to bring him his wife Faina, scolds it. Here, the nature of the Machine is revealed by the fact that the robot steals wallets on the tram when asked to bring money. But their relationship is most vividly described at the very first meeting: He deceives the robot by taking away his watch: "'What is your name? Citizen! ... What is deception?' 'I'll explain to you now…'". The robot-worker is clearly surprised. This shows that machines, although they are analogues of workers, have no life experience of their own, that they are absolutely helpless in front of ordinary society which is represented by the Philistine.

There is another interesting moment of the play connected with this character. During a riot of workers and a meeting on the street with one of them, trying to justify himself and getting himself out of a difficult situation, the Philistine calls himself "artificial," since he does not have a pass allowing him to move freely: "I don't even

understand this word. What documents, permits? I am artificial, I work like a dog, the damned exploiters drink blood from me. Today he could not stand it, he rebelled with weapons in his hands" (Tolstoy, 1924/1983, p. 103).To summarize, needing a verbal definition of the difference between a man and a machine, both of which are called "workers", the Philistine points out the difference in origin: "I was made, comrade, in Moscow, at the first Soviet factory, Presnensky district" (Tolstoy, 1924/1983, p. 103). The word artificial is also used by Čapek when describing the creation of robots: "And to manufacture artificial workers is the same thing as to manufacture gasoline motors" (Čapek, 1920/1961), but it is in Tolstoy's world that this becomes the only difference, since it is impossible to distinguish people and robots in the world of Tolstoy. We have to assume that this difference existed in Čapek, at least it could be easily determined, for otherwise the extermination of people would not have been so simple. Interestingly, however, while according to the text of Čapek's play, robots were a biochemical invention and did not differ in appearance from humans, the many visualizations of robots in most cases did not follow this canon – neither the illustrations in books nor the theatrical performances (Fig. 1-2) (Khorakova, 2010; Krausova, 2016).

**Fig. 1.** "A robot kills its creator," London installation photo caption (1923) (Krausová, 2016, p. 82)


**Fig. 2.** Adaptation of R.U.R. on British television BBC (1938) (Krausová, 2016, p.72)

Čapek recorded this fact and even objected to this form of robots. In the article "Robot Author Illuminated," published in *Lidové noviny* on June 9, 1935, he wrote:

"The world needs mechanical robots because people believe in machines more than they believe in life; they are more fascinated by technical miracles than by the miracle of life" (Čapek, 1935).

Since the word "robot" is not mentioned on the pages of Tolstoy, the notorious "worker" becomes something more only during a riot. It can be assumed that this is due to the development of robots, their transition from weak-willed human slaves to active participants in the uprising against their enslavers. Robots think, robots counteract, something of their own appears instead of a programmed behavior. In the scientific world, the word "artificial" in relation to a machine did not appear until 1956, when it was first used in the title of a conference at Dartmouth College (Garvey & Maskal, 2020, p. 287). John McCarthy is credited for coining the phrase "artificial intelligence" in the proposal written by John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester for this event (Moor, 2006).

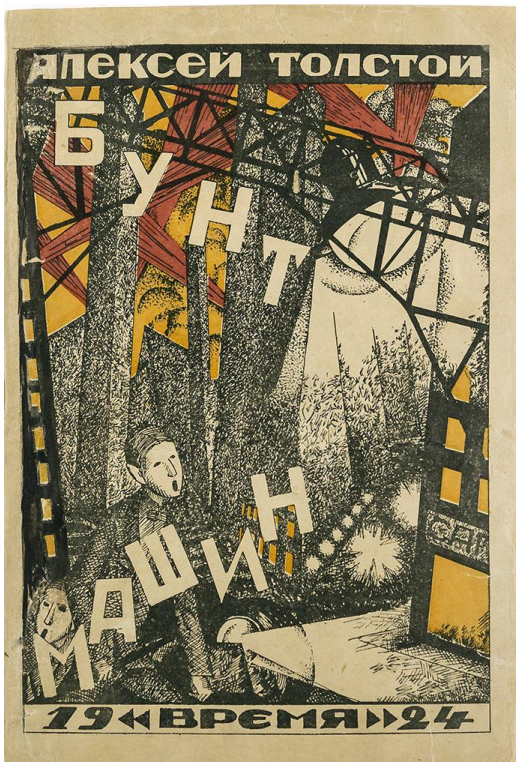



**Fig. 3**. Cover of Tolstoy's book "Riot of the Machines" (1924) (Tolstoy, 1924)


**Fig. 4**. Poster for the play "Riot of the Machines" at the Bolshoi Drama Theater (Leningrad, 1924) (Vershinin, 2021)

Through the Philistine, the author makes it clear to the Soviet reader what artificial people are. It becomes clear that people and robots were different as well. In

turn, Alexey Tolstoy is different. Seeing in the workers ordinary people equal to themselves, the workers could not distinguish them from their own kind, because they both worked with their own hands for someone. Despite the title of the book and of the theatrical performance, the visual interpretation of the robot in "Riot of the Machines" does not involve machines, does not draw on mechanical or metallic characters (fig. 3, 4).

In Čapek's play, the decline of culture and the inaction of people lead to the uprising of robots who have realized their superiority. By addressing humans with hatred as slave owners, robots spawned an entire genre of "robot uprisings." Robertson notes that "Capek's graphic portrayal in R.U.R. of the end of bourgeois humanity at the hands of a violent robot-proletariat helped to shape Euro-American fears about robots that persist to this day" (Robertson, 2014, p. 574).

"You are not like robots. Not as capable as robots. Robots do everything. You only order. Produce superfluous words" (Čapek, 1961). Tolstoy's workers stage an uprising under the revival of agitation on the part of the workers, and this turns into a class upheaval, a decision against his creation by a man like Čapek. This can be seen in the scene of Mikhail's campaign letter: "Who mines coal and ore, ... there is only one right to the land: to breathe air and digest the patented cakes of Morea". I want to know why you [the robot Adam calls to Michael, his creator] are dictating this letter to me. Yesterday Sunday you dictate letters to me and secretly send them to the mainland. I want the workers of Morea to realize their strength, I want your strength to come into action ... – says Mikhail" (Tolstoy, 1924/1983, p. 97).

## CONCLUSIONS

Robots are interesting because they merge humans and machines. In Capek's play, their technical, inhuman and, therefore, undivine character comes to the fore. For Alexei Tolstoy, on the contrary, the main thing is their resemblance to people. The author decided to abandon the word robot, and chose a more familiar worker, because it shows the similarity of workers and "rabotniki" (robots). The hallmark of each is to get the job done. Aristotle considered a slave as an instrument of action (Aristotle, 1254a), and as Brudny (2009) notes, the fact that humans did not want to work played a very crucial role in the history of humanity. In relation to robots, this gave rise to an artistic tradition of describing the rebellion of robots at a certain stage of their development against their creators. However, in the middle of the 20th century, the robot became "a thoroughly SF creature, a marvel of technology whose roots in dehumanized and exploited labor were muted" (Vint, 2021). In the Russian version of Čapek's plot, the similarity of a human and a robot comes to the fore, up to their indistinguishability, but this association is militant, against a common enemy.

For current discussions of AI, Tolstoy's decision questions or undermines a philosophical commonplace, namely the notion of an ontological faultline between the natural and the artificial. Perhaps, other divisions are more pertinent even as they cut across familiar dichotomies. The distinction between autonomy and heteronomy, between self-determination and determination by others pertains to natural as well as

artificial workers. This may have repercussions for recent discussions of the moral rights of robots. Jacob Turner (2019) suggests considering humans as first AIs based on ancient Sumerian, Chinese, and biblical sources reporting their creation from inanimate materials (p. 4).

Robots are gradually entering our daily lives and becoming more intelligent, sensitive to the environment and people. They are more and more like us, and someday the moment might come when it is necessary to assign rights and freedom to robots. Perhaps, if this does not happen, the fate of the play's protagonists awaits us. So perhaps, humanity is faced with the task of preventing this uprising and not turning technical progress into a war, into a "Riot of the Machines."

## REFERENCES

Aristotle. (1944). *Politics* (H. Rackham, Trans.). Perseus Hopper. Archive. http://www.perseus.tufts.edu/hopper/text?doc=urn:cts:greekLit:tlg0086.tlg035.perseus-eng1:1.1254a (Original work published 350 B.C.E)

Brudny, A. (2009). Will, Ego and Evidence. *UCA Academic Review*, *2*, 45–54.

Čapek, K. (1935). Autor Robotů se Brání [The Author of the Robots is Defending Himself]. *Lidové Noviny*, *9*(6).

Čapek, K. (1961). *R. U. R. (Rossum's Universal Robots)*. Oxford U.P. (Original work published in 1920)

Čejková, J. (2021). Robots: The century past and the century ahead, an Introduction to the 2021 ALIFE conference? *The 2021 Conference on Artificial Life*. https://doi.org/10.1162/isal_e_00468

Chernykh, P. Ya. (1999). Rabochii [Worker]. In *Historical and Etymological Dictionary of the Modern Russian Language. Vol. 2* (p. 92). Russian Language. http://www.slovorod.ru/etym-chernykh/_pdf/russ-hist-etym-chernykh-2.pdf

Günther, H. (2000). Archetypes of Soviet Literature. In H. Gunter & E. Dobrenko (Eds.), *Socialist Realist Canon* (pp. 743–784). Academic Project.

Garvey, C., & Maskal, C. (2020). Sentiment Analysis of the News Media on Artificial Intelligence Does Not Support Claims of Negative Bias Against Artificial Intelligence. *OMICS: A Journal of Integrative Biology*, *24*(5), 286–299. https://doi.org/10.1089/omi.2019.0078

Gastev, A. (1971). *Poeziya Rabochego Udara* [The Poetry of the Working Stroke]. Fiction. (Original work published in 1924) https://traumlibrary.ru/book/gastev-poezia-rab-udara/gastev-poezia-rab-udara.html

Horakova, J. (2010). *Robot Jako Robot* [Robot as a Robot]. KLP Praha.

Kaplan, F. (2004). Who is Afraid of the Humanoid? Investigating Cultural Differences in the Acceptance of Robots. *International Journal of Humanoid Robotics*, *01*(03), 465–480. https://doi.org/10.1142/S0219843604000289

Krausová, B. (2016). *Visual Interpretation of Karel Čapek's R.U.R* [PhD thesis, University of South Bohemia in České Budějovice]. https://theses.cz/id/5waal7/Vizuln_interpretace_dla_Karla__apka_R.U.R._DP_Lucie_Krau.pdf

Mayakovsky, V. (1969). *Lublu. Izbrannaya Lirica* [I Love. Selected Lyrics]. Detskaya Literature.

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, *27*(4). https://doi.org/10.1609/aimag.v27i4.1911

Nikolsky, S. (2004). K Istorii Znakomstva s Pyesoy Chapeka "R.U.R." v Rossii [Linguistics and Literary Criticism. To the History of Acquaintance with Chapek's Play "R.U.R." in Russia]. In T. I. Vendina, V. K. Volkov, & M. A. Robinson (Eds.), *Slavic almanac* (pp. 406-407). Indrik. https://cyberleninka.ru/article/n/k-istorii-znakomstva-s-piesoy-chapeka-r-u-r-v-rossii/viewer

Pervushina, E. (2021). *Istorii o Muzah i Zhenah Russkih Poetov i Pisatelej. Istorii o Muzah i Zhenah Russkih Poetov i Pisatele* [Love in the Silver Age. Stories about muses and wives of Russian poets and writers. Joys and experiences]. Centerpolygraph

Robertson, J. (2014). Human Rights vs. Robot Rights: Forecasts from Japan. *Critical Asian Studies*, *46*(4), 571–598. https://doi.org/10.1080/14672715.2014.960707

Semenov, (2003). Rabotat [To work]. In *Etymological dictionary of the Russian language*. UNVES. https://gufo.me/dict/semenov/работать

Skvorchevsky, K., Sergeev, A., & Kovalev, N. S. (2020). From the Golem to the Robot and Beyond to the Smart Prostheses. In *Advances in Artificial Systems for Medicine and Education II. AIMEE2018 2018. Advances in Intelligent Systems and Computing, vol. 902* (pp. 355–365). Springer. https://doi.org/10.1007/978-3-030-12082-5_33

Sreznevsky, I. (1912). Rabota [Work]. In *Materials for the dictionary of the Old Russian language according to written monuments. Vol. 3.* (p. 2). St. Petersburg. http://www.ra-s-ved.ru/DATA/sreznevskyslovar3.pdf

Tolstoy, A. (1924). *Bunt Mashin: Fantasticheskie Stseni* [Riot of the Machines: Fantastic Scenes]. Time.

Tolstoy, A. (1983). *Lyubov Zolotaya: Sbornik Pyes* [Golden Love: Collection of Plays]. Mystetstvo.

Turner, J. (2019). Introduction. In *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-96235-1_1

Ushakov, D. (1939). Rabochii [Worker]. In *Explanatory Dictionary of the Russian Language, Vol. 3.* State foreign publishing house and national words. https://gufo.me/dict/ushakov/рабочий

Vint, S. (2021). *Science Fiction.* The MIT Press. https://doi.org/10.7551/mitpress/11841.001.0001

Vershinin, A. (2021, June 3). How to Avoid a Robot Riot. *SPbdnevnik*. https://spbdnevnik.ru/news/2021-06-03/kak-izbezhat-bunta-robotov

## СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Тарас Романенко, romanenko.tp@edu.spbstu.ru,
ORCID 0000-0001-8296-4361

Полина Щербина, tsherbinina.ps@edu.spbstu.ru
ORCID 0000-0001-9464-3979

Taras Romanenko, romanenko.tp@edu.spbstu.ru,
ORCID 0000-0001-8296-4361

Polina Shcherbinina, tsherbinina.ps@edu.spbstu.ru
ORCID 0000-0001-9464-3979

Статья поступила 14 февраля 2022
одобрена после рецензирования 28 февраля 2022
принята к публикации 10 марта 2022

Received: 14 February 2022 /
Revised: 28 February 2022
Accepted: 10 March 2022

https://doi.org/10.48417/technolang.2022.01.04
Research article

# The Body, the Soul, the Robot: 21st-Century Monism

Kevin Liggieri[1] and Marco Tamborini[2](✉)
[1]Darmstadt Technical University, Institut für Geschichte, Landwehrstr. 50a, 64293 Darmstadt, Germany
Kevin.liggieri@tu-darmstadt.de
[2]Darmstadt Technical University, Institut für Philosophie Karolinenpl. 5, Darmstadt, 64289, Germany,
Marco.Tamborini@tu-darmstadt.de

## Abstract
The thesis we will defend in the following pages is twofold. First, we indicate two linguistic-cultural turning points in the concept of the robot. The introduction of the body and the soul in the machine has paved the way towards new technical and epistemic possibilities and, thus, it has granted a new conceptual definition of robot. Second, we propose a return to Descartes as a starting point for a reinterpretation and redefinition of the concept of robot in the contemporary world. Here we will show how Cartesian dualism (in the description of humans) becomes a (material) monism in the development and construction of robots. As a result, we call on our fellow philosophers and historians of science and technology to explore, critique, reject, or further investigate the features of the 21st-century material monism proposed in this paper.

**Keywords:** Robotics; Descartes; Cybernetics; Robotic Monism; Cartesian Dualism

**Citation:** Liggieri, K., & Tamborini, M. (2022). The Body, the Soul, the Robot: 21st-Century Monism. *Technology and Language*, 3(1), 29-39. https://doi.org/10.48417/technolang.2022.01.04

УДК 11: 62-529
https://doi.org/10.48417/technolang.2022.01.04
Научная статья

# Тело, душа, робот: Монизм 21 века

Кевин Лиджери[1] and Марко Тамборини[2](✉)
[1]Дармштадский технический университет, Институт истории, Ландверстр. 50a, 64293 Дармштадт, Германия
Kevin.liggieri@tu-darmstadt.de
[2]Дармштадский технический университет, Институт философии, Каролиненплац 5, Дармштадт, 64289, Германия
Marco.Tamborini@tu-darmstadt.de

**Аннотация**
Тезис, который мы будем защищать на следующих страницах, двоякий. Во-первых, укажем на два лингвокультурных поворота в концепции робота. Привнесение тела и души в машину проложило путь к новым техническим и эпистемологическим возможностям и, таким образом, дало новое концептуальное определение робота. Во-вторых, мы предлагаем вернуться к Декарту в качестве отправной точки для переосмысления и переопределения концепции робота в современном мире. Здесь мы покажем, как картезианский дуализм (в описании человека) становится (материальным) монизмом в разработке и конструировании роботов. В результате мы призываем наших коллег-философов и историков науки и техники изучать, критиковать, отвергать или дополнительно исследовать особенности материального монизма 21-го века, предложенные в этой статье.

**Ключевые слова:** Робототехника; Декарт; Кибернетика; Роботизированный монизм; Картезианский дуализм

**Для цитирования:** Liggieri, K., & Tamborini, M. The Body, the Soul, the Robot: 21st-Century Monism // Technology and Language. 2022. № 3(1). P. 29-39. https://doi.org/10.48417/technolang.2022.01.04

## INTRODUCTION

In recent years, robots have shaped numerous domains of technical and scientific production. Robots are now employed in all areas of industry, medicine, architecture, and even culture. The recent proliferation of robot construction has prompted philosophers, historians, and sociologists of science to reinterrogate the concept of the robot (Riskin, 2016; Yang et al., 2018; Datteri & Tamburrini, 2007). Innovations in this field inspire a need for a new understanding of the modern use of robotics. In particular, several studies have examined the elements of continuity and rupture between bio-robotics and the use of automata in earlier centuries. Philosopher Marco Tamborini (2021), for example, speaks of a new material synthesis to describe the current paradigm of biorobotics. He highlights the material turn in robotics-inspired morphology as the main epistemic cornerstone in contemporary biorobotics. Philosopher Edoardo Datteri provides a useful philosophical taxonomy of biorobotics, comparing the epistemic claims of robots to the use of models and computer simulations. Furthermore, he defends the thesis that scientists create their possible phenomena or explananda through the use of robots in interactive experiments (Datteri, 2020; Datteri & Tamburrini, 2007).

While these studies are important for examining the knowledge claims of contemporary robotics, they have neglected important elements in their investigation of the concept of the robot. Considering and exploring the linguistic-cultural constitution of robots is an essential prerequisite to understanding robotics' practical limits and opportunities. Moreover, this is crucial for the delineation of a well-grounded philosophy of human-machine interaction.

In this paper, therefore, we will focus on two key turning points in the development of the concept of the robot as adopted in 21st-century science and technology: the introduction of the body and the soul in the machine. These two components, as well as their many permutations and hybridizations, have shaped the entire history of robotics. The introduction of the body and the soul in the machine have given rise to the central questions and issues of today's philosophy of robotics.

The presence of these two factors is even more evident in contemporary robotics. In the design and construction of exoskeletons, for example, the physicality and materiality of the machine play a fundamental role in the possible successful interaction between robot and human. The features of the body, designed through imitating organisms as in bio-robotics and materialized in the machine itself as in embodied AI, impact the very concept of the robot, defined as a material machine capable of enabling specific functions. Furthermore, in contemporary biorobotics, the ways in which robots mimic the form-structure of organisms influences, limits, and enables the potential interaction between these machines and other animals. All the experiments in interactive robotics are indeed based on robots' material and behavioral features (Romano et al. 2019; Datteri 2020).

The soul has a similar key role in defining the contemporary concept of the robot. Following recent findings in cognitive and information sciences, the soul should not be understood here as a metaphysical entity (Langerlund & Yrjönsuuri, 2002) but rather as a complex set of operations that enable organismal reasoning and feedback with the

environment. For example, in robotics and architectural design, the machine and the engine, i.e., the soul, merge with the natural material to be imitated. The machine becomes intelligent, as do the materials, which can move freely. The same can be observed with embodied AI. In evaluating whether a machine is or is not a robot, the machine's ability to act and the degree to which it has freedom and autonomy characterize the concept, status, and identification of a possible robot. The potential of realizing and defining robots for geriatric applications, termed geriatronics, is contingent on the construction of robots that have a soul. Only a thinking-machine, armed with some type of programmed empathy, can be truly helpful in alleviating the painful loneliness experienced by older adults living in relative isolation (Weidner & Karafillidis, 2018). Therefore, the soul embodied within thinking machines can be defined as a programmed form of empathy. Lastly, in soft robotics, scientists speak of intelligence without a brain. That means the robot has a soul despite not having a traditional brain; the intelligence – or the soul – is distributed throughout the body (Sadeghi et al., 2017; Gravish & Lauder, 2018; Romano et al., 2019; Xiloyannis et al. 2021; Ghazi-Zahedi et al. 2017).

The introduction of the body and the soul into robotics was framed within the classical ontological paradigm supported by French philosopher René Descartes. In the next few pages, we will depart from the classical Cartesian distinction between *res cogitans* and *res extensa* to see how this was implemented in the linguistic-cultural construction of the concept of the robot. Famously, Descartes postulated the existence of two mutually exclusive substances. The first is the extended support unique to bodies; the second is the non-extended, thought-support, proper to the functions of the soul. These two substances are ontologically different and are grounded in quite distinct parts of reality: the form and function of bodies (Westphal, 2016).

Taking this distinction as a starting point, we will bring attention to the two main turning points in the evolution of the concept of the robot mentioned above. The first turning point concerns the inclusion of the body into the machine and the transition from automata built for entertainment to those designed to perform work activities. The second decisive moment involves the introduction of the soul in the machine through the imitation of the functioning of the human brain and the agency of other organisms.

The thesis we will defend in the following pages is twofold. On the one hand, we will highlight the two major linguistic-cultural turning points in the construction of the robot; on the other hand, we will propose a return to Descartes as a starting point for a reinterpretation and redefinition of the concept of robot in the contemporary world. Here we will show how Cartesian dualism (in the description of humans) becomes a (material) monism in the development and construction of robots. As a result, we call on our fellow philosophers and historians of science and technology to explore, critique, reject, or further investigate the features of the 21st-century material monism proposed here.

## THE BODY IN THE MACHINE

Since antiquity, the human being has been described as both a body and a machine. This conflicting pattern of interpretation between life and technology was radicalized in the 18th century with the advent of a comprehensive quantification and mechanization of

man as a bodily being. In the dispute between materialism and idealism, the human being is conceived, on the one hand, as a quantifiable and thus determined automaton and, on the other, as an autonomous spirit (Liggieri, 2020).

In his *Treatise on Man*, Descartes famously wrote that the human body is nothing less than a machine. All the functions that can be attributed to machines, such as "the digestion of food, the beating of the heart and arteries, the nutrition and growth of the members, respiration, waking and sleep" (Descartes, 1662/1972), can also be detected in the human being. Inspired by Descartes' description of animals as autonoma, French physician Julien Offray de La Mettrie worked within this framework to define the human being as an "excellently equipped machine" (La Mettrie, 1960; Liggieri and Tamborini, 2021). The machine became the most important interpretative pattern for understanding, exploring, and reproducing the human and animal body (Schatzberg 2018). In fact, in the 17th and 18th centuries, contemporary techniques such as clockworks and hydraulics were used to explain the body. Descartes, for example, described the heart as a hydraulic pump, the blood vessels or nerve cords as a system of tubes, and the functioning of the body as analogous to a clock (Riskin, 2016).

Furthermore, machines were built to mimic organisms. The famous automata of the French designer Jacques de Vaucanson (the "flute player") and the Swiss watchmaker family Jaquet-Droz (the "scribe", the "draughtsman", and the "piano player") were modeled on the human body and its movements (especially in the "flute player") (Fryer & Marshall, 1979). Not only was the human being thus technically analyzed, but also the automaton, the machine, was based on the imitation of the human in bodily appearance and expression. Although technology at the time was still far from being able to imitate organic modes of functioning or even the cognitive performances of humans, a certain epistemic shift in the conception of the body and technology is observable in the period of the European Enlightenment. Physiologists, biologists, and scholars from different fields expressed themselves as being in favor of human-machine hybrids and artificial humans from the middle of the 18$^{th}$ century onwards. At first sight, 18$^{th}$-century automata could be considered the forerunners of 20$^{th}$-century robots: they are hybrid systems that move and look like natural humans (Chapuis & Droz 1949/1958; Westermann, 2016).

As a result, the human being – with its spirit, its "soul", its peculiar materiality – was placed in opposition to the artificial-technical world. At the same time, the human being was made artificial and, thus, reproducible. This juxtaposition also enabled a departure from the features of the human body. Therefore, the human being became the counterpoint to which the technical-artificial could be developed. The human body was the reference point for constructing automata that looked like humans and could perform specific and productive work. In fact, starting in the middle of the 19th century, the notion of corporeality was central in the construction of robots capable of manual labor.

The introduction of the body into the machine was accompanied by a further ontological and practical change in the concept of automaton. With the advent of working and transport machines, the perception of 17$^{th}$- and 18$^{th}$-centuries gaming and court machines meant as amusement engines for the nobility was also transformed. The steam engine was not built to amuse the public but rather to bolster physical labor. Accordingly,

machines were designed for cost-cutting and time-saving goals and no longer for playful diversion.

Within the encounter between materiality and practical goal, the term "robot" was coined in Karel Čapek's 1922 play "R.U.R." – Rossumovi Univerzální Roboti. The artificial humans ("robots") represent cheap workers whose unregulated and prolific use throughout industry changes the entire world economy (Čapek, 1970). At the same time, they embodied a strong materiality that enabled them to be perceived as automata, as being capable of motion.

## THE SOUL IN THE MACHINE

From the 1940s on, the emergence of information theory and cybernetics added another layer to the conceptualization of the robot (Galison, 1994; Kline, 2015, Mindell, 2002; Pickering, 2010; Kay, 2000; Rid, 2016). In this transition from work-performing, purely mechanical devices to information-processing machines (Rabinbach, 1992), a further linguistic-cultural shift can be detected: the ghost was put in the machine. German philosopher Gotthard Günther (1976) describes this transition as follows:

> [T]o date, man has conceived two fundamentally different ideas of the machine in his technical development. The first is the classical Archimedean machine whose purpose is to produce work. Next to this has come the idea of the 'second' machine, from which one no longer expects work, but information. The 'first' machine is designed by analogy with the human arm (and hand); the second is expected to be a technical reproduction of the human brain, for only the brain processes information (p. 97).

Here, Günther has further expanded the original definition of machine to emphasize the different meanings (and uses) of machines in the 20th century. As is well known, the name for the work machine comes via Latin from Doric Greek *makhana*. In its original Greek, it first meant "to be able to, to have power" and then came to signify a device or tool. In Latin, a further meaning was added to the notion of machine. It also signified a contrivance, cunning, trick, or device acting creatively or with wisdom (*cum ingenium* in Latin). The second machine described by Günther refers to this second definition of machine, and then expands upon it.

The second machine produced by cybernetics, which represents the "general theory of a [...] non-Archimedean mechanism" (Günther, 1976, p. 95), put creativity and wisdom at the center of its agency. To do that, scientists took inspiration from the operational and functional proprieties of the brain, the Cartesian *res cogitans*. According to Günther and the cyberneticists, the brain is the only organ that functions "non-classical-mechanically" but "according to trans-classical principles" (ibid.). The 'computer man' replaced the 'motor man' as the model for the robots.

The self-regulating classical machine models should no longer be understood as mere automata that follows an input-output system but as a complex device in analogy to the functioning of the brain. Movement, as with the mechanical doll, is no longer in the

foreground; the focus is now on the possible behavior, wisdom, and functionality of the machine (Heßler, 2018/2019).

Furthermore, cyberneticists understood their machines primarily as scientific tools for knowledge production. Due to the strong focus of the original cybernetics group on modes of human brain functioning, most cybernetic machines were thus interpreted as being 'biologically inspired brain-like devices' (Husbands, Wheeler & Holland, 2008, p. 19). Even the well-known cybernetic 'turtle' was initially conceived as a contribution to the sciences of the brain (Pickering, 2010).

Here, we can see again that the main aim of scientists and engineers was to mechanize the *res cogitans*. This mechanization was intended to be a strategic move to design a machine capable of producing creative solutions and action. In the second linguistic-cultural shift, the ghost re-entered in the machine. Rather, an imitation of the brain was used to construct bio-inspired robots, which, in turn, fully materialized the classical definition of a machine. In imitating the human body, the robot becomes a working machine; in imitating thought, it becomes AI. As historian Jan Müggenburg (2018, 2019) has aptly pointed out, cybernetics is concerned with the creation of living artifacts[1]: again, the Cartesian ghost was put in the machine.

## OUTLOOK

As many scholars have argued, Descartes had a decisive impact on the development of cybernetics, AI, and neurosciences (Riskin, 2016; Bates, 2013). In this paper, while expanding on this line of thought, we have pursued another goal. First, we have indicated two linguistic-cultural turning points in the concept of the robot. The introduction of the body and the soul in the machine has paved the way towards new technical and epistemic possibilities and, thus, it has granted a new conceptual definition of robot. Second, we have shown that these turning points were based on Cartesian dualism.

Ironically, the introduction of the soul in the machine prompted the overcoming of the Cartesian ontology. In fact, cybernetics brought out a deeper ontological dimension. Although they departed from Cartesian dualism, they ended up defending a strong monism. They supported the thesis that robots (and, broadly speaking, living artifacts) were machines, i.e., devices in which both materiality and intelligence are fused together.

[1] These mechanistic analogies evoked a critical as well as positive response from the research community and the public and continue to shape our notions of "robots" as hybrid beings of artificial corporeality and intelligence. Fictional literature and film can be seen as a decisive factor in their popularization. Cultural products in a variety of media have offered various cybernetic scenarios as a space of possibility, touching on themes of 'thinking machines' and 'electronic brains' as well as hybrid beings such as 'cyborgs' (cybernetic organisms) and autonomous robots. Not without reason, Günther (1952) wrote that "[t]he profound changes in the metaphysical history of man" are "indicated in man's fantasy life long before they develop into actual realities" (p. 223). In light of this, science fiction, as a "literary form of mythic fairy tale" that playfully explores new possibilities for thought and transformation, anticipates and helps prepare discourses about robots in and for the public sphere. Working and thinking robots are already located in present day reality both physically and in the public discussion, but the dystopias (fear of replacement, loss of self, unemployment) are predominantly present on a fictional level (Heßler & Hitzer, 2019).

This deeper ontological dimension is a feature of 21st-century bio-robotics. As Tamborini and Datteri have pointed out, the design and production of concrete robots that interact with their environment through a feedback loop enables biologists to create and access an *explanandum* (the form-function complex of an extinct animal such as the *Plesiosaurus* or the *Orobates pabsti*, which would be otherwise non-manipulable and non-researchable – this 4-legged vertebrate went extinct about 300 million years ago (Datteri, 2020; Tamborini, 2021)). This is also the case for interactive robotics. The corporeality of the robot is fused with its possible action and sociality. The body becomes the soul of the automata; the soul embodies the body (Nyakatura et al. 2019; Tamborini, 2020, 2021). Soft robots are intelligent devices; the properties of their bodies influence what they can do, achieve, and perform (Kim, Laschi, & Trimmer 2013; Sadeghi, Mondini, & Mazzolai, 2017). This supports a new monism, which echoes Haeckel's technical and artistic investigations of the forms of nature. Again: the ghost permeates the machine since the machine is both intelligence and materiality.

Post-cybernetic monism, however, must be regarded as peculiar. Indeed, our brief genealogy has shown how the materiality of the machine underlies its functioning both in robotic integration processes (fish recognize the robot fish as part of its community on the basis of the robot fish's functional and material properties) and in those considered more creative such as, for example, embodied AI. Here, *res extensa* becomes *res cogitans*, and the latter becomes *res extensa* once again. In a dialectical process, the two substances overcome each other while preserving themselves.

Hence and to conclude, by pointing out two linguistic-cultural turning points in the concept of robots, our analysis has called attention to the monism of contemporary robotics. We thus invite and have hopefully provoked our fellow philosophers and historians of science and technology to expand upon, reject, or define differently the features of 21st-century robotic monism.

## REFERENCES

Bates, D. (2013). Cartesian Robotics. *Representations*, *124*(1), 43–68. https://doi.org/10.1525/rep.2013.124.1.43

Čapek, K. (1970). *R.U.R.* New York: Pocket Books.

Chapuis, A., & Droz, E. (1949/1958). *Automata. A historical und technological study*. Edition du Griffon.

Datteri, E., & Tamburrini, G. (2007). Biorobotic Experiments for the Discovery of Biological Mechanisms. *Philosophy of Science, 74*(3), 409–430.

Datteri, E. (2020). The Logic of Interactive Biorobotics. *Frontiers in Bioengineering and Biotechnology*, 8. https://doi.org/10.3389/fbioe.2020.00637

Descartes, R. (1972). *Treatise of Man*. Great Minds Series. Harvard University Press. (Original work published 1662)

Fryer, D. M., & Marshall, J. C. (1979). The Motives of Jaques de Vaucanson. *Technology and Culture, 20*(2), 257–269.

Galison, P. (1994). The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision. *Critical Inquiry, 21*(1), 228-266. http://www.jstor.org/stable/1343893

Ghazi-Zahedi, K., Deimel, R., Montúfar, G., Wall, V., & Brock, O. (2017). Morphological Computation: The Good, the Bad, and the Ugly. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 464–469). IEEE. https://doi.org/10.1109/IROS.2017.8202194

Gravish, N. & Lauder, G. V. (2018). Robotics-Inspired Biology. *Journal of Experimental Biology, 221*, 7. https://doi.org/10.1242/jeb.138438

Günther, G. (1952). Kommentar [Comment]. In G. Günther (Ed.), *Überwindung von Raum und Zeit. Phantastische Geschichten aus der Welt von Morgen* (pp. 223-238). Karl Rauch.

Günther, G. (1976). Die „zweite“ Maschine (1952) [The "second" machine (1952)]. In G. Günther (Ed.) *Beiträge zur Grundlegung einer operationsfähigen Dialektik. Erster Band* (pp. 91-114). Meiner.

Heßler, M. (2018/2019). Techno-Humanity. A Plea for a Historical Anthropology for Technology. *ICON: Journal of the International Committee for the History of Technology, 24*, 65-75.

Heßler, M. & Hitzer B. (2019). *Tech-Fear. Histories of Multifaceted Relationship. Special issue Technikgeschichte, 3,* 185-200. https://doi.org/10.5771/0040-117X-2019-3-185

Husbands, P., Holland, O., & Wheeler, M. (Eds.). (2008). *The mechanical mind in history*. Cambridge, MA: MIT Press.

Kay, L. E. (2000). *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford University Press.

Kim, S., Laschi, C., & Trimmer, B. (2013). Soft Robotics: A Bioinspired Evolution in Robotics. *Trends in Biotechnology, 31*(5), 287–294. https://doi.org/10.1016/j.tibtech.2013.03.002

Kline, R. (2015). *The Cybernetics Moment. Or Why We Call Our Age the Information Age*. Baltimore: John Hopkins University Press.

La Mettrie, J. O. de. (1960). *L'homme machine. A Study in the Origins of an Idea* (A. Vartanian, ed.). Princeton University Press.

Langerlund, H., & Yrjönsuuri M. (Eds.). (2002). *Emotions and Choice from Boethius to Descartes. Studies in the History of Philosophy of Mind, Vol. 1*. Kluwer Academic Publishers.

Liggieri, K., & Tamborini, M., (Eds). (2021). *Organismus und Technik. Anthologie zu einem produktiven und problematischen Wechselverhältnis* [Organism and Technology. Anthology on a Productive and Problematic Interrelationship]. Wbg.

Liggieri, K. (2020). *Man as Pet*. Lit.

Mindell, D. (2002). *Between Human and Machine. Feedback, Control and Computing before Cybernetics*. John Hopkins University Press.

Müggenburg, J. (2018). *Lebhafte Artefakte. Heinz von Foerster und Die Maschinen Des Biological Computer Laboratory* [Vivid Artifacts. Heinz von Foerster and the machines of the biological computer laboratory]. Konstanz University Press.

Müggenburg, J. (2019). Bionik [Bionics]. In K. Liggieri & O. Müller (Eds.), *Mensch-Maschine-Interaktion Handbuch Zu Geschichte – Kultur – Ethik* (pp. 241–243). Springer.

Nyakatura, J. A., Melo, K., Horvat, T., Karakasiliotis, K., Allen, V. R., Andikfar, A., Andrada, E., Arnold, P., Lauströer, J., & Hutchinson, J. R. (2019). Reverse-Engineering the Locomotion of a Stem Amniote. *Nature*, *565*(7739), 351.

Pickering, A. (2010). *The Cybernetic Brain*. Chicago University Press.

Rabinbach, A. (1992). *The Human Motor: Energy, Fatigue, and the Origins of Modernity*. University of California Press.

Rid, T. (2016). *Rise of the Machines: A Cybernetic History*. Norton.

Riskin, J. (2016). *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*. University of Chicago Press.

Romano, D., Donati, E., Benelli, G., & Stefanini, C. (2019). A Review on Animal–Robot Interaction: From Bio-Hybrid Organisms to Mixed Societies. *Biological Cybernetics*, *113*(3), 201–225. https://doi.org/10.1007/s00422-018-0787-5

Sadeghi, A., Mondini, A., & Mazzolai, B. (2017). Toward Self-Growing Soft Robots Inspired by Plant Roots and Based on Additive Manufacturing Technologies. *Soft Robotics*, *4*(3), 211–23. https://doi.org/10.1089/soro.2016.0080

Schatzberg, E. (2018). *Technology: Critical History of a Concept*. University of Chicago Press.

Tamborini, M. (2020). Technische Form Und Konstruktion [Technical form and construction]. *Deutsche Zeitschrift Für Philosophie*, *68*(5), 712–733. https://doi.org/10.1515/dzph-2020-0049.

Tamborini, M. (2021). The Material Turn in The Study of Form: From Bio-Inspired Robots to Robotics-Inspired Morphology. *Perspectives on Science*, *29*(5), 643–65. https://doi.org/10.1162/posc_a_00388

Weidner, R., & Karafillidis, A. (Eds.). (2018). *Developing Support Technologies. Integrating Multiple Perspectives to Create Support that People Really Want*. Springer.

Westermann, B. (2016). The Biomorphic Automata of the 18th Century: Mechanical Artworks as Objects of Technical Fascination and Epistemological Exhibition. *Figurationen* , *17*(2), 123–137. https://doi.org/10.7788/figurationen-2016-0211

Westphal, J. (2016). *The Mind–Body Problem*. MIT Press.

Xiloyannis, M., Alicea, R., Georgarakis, A.-M., Haufe, F. L., Wolf, P., Masia, L., & Riener, R. (2021). Soft Robotic Suits: State of the Art, Core Technologies, and Open Challenges. *IEEE Transactions on Robotics*. https://doi.org/10.1109/TRO.2021.3084466

Yang, G.-Z., Bellingham J., Dupont P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., & Merrifield, R. (2018). The Grand Challenges of Science Robotics. *Science Robotics*, *3* (14), eaar7650. https://doi.org/10.1126/scirobotics.aar7650

## СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Кевин Лиджери, Kevin.liggieri@tu-darmstadt.de

Марко Тамборини, Marco.Tamborini@tu-darmstadt.de, ORCID 0000-0001-7102-7479

Kevin Liggieri, Kevin.liggieri@tu-darmstadt.de

Marco Tamborini, Marco.Tamborini@tu-darmstadt.de, ORCID 0000-0001-7102-7479

Статья поступила 14 ноября 2021
одобрена после рецензирования 18 января 2022
принята к публикации 28 февраля 2022

Received: 14 November 2021 /
Revised: 18 January 2022
Accepted: 28 February 2022

https://doi.org/10.48417/technolang.2022.01.05
Research article

# The Blurring of the Boundaries between Humans and Robots is a Good Thing and a New Species would be Born: An Interview with Hiroshi Ishiguro

Hui Jiang[1], Lin Cheng[2](✉) and Hiroshi Ishiguro[3]
[1]Sun Yat-sen University, Xingangxi. 135, Guangzhou, 510275, China
[2]Guangdong University of Foreign Studies, Baiyundadaobei. 2, Guangzhou, 510420, China,
lin.cheng@gdufs.edu.cn
[3]Osaka University, 1-1 Yamadaoka, Suita, Osaka 565-0871 Japan

## Abstract
The documentary *Philosophy in the Age of Desire* records a short encounter between Hiroshi Ishiguro and Markus Gabriel in 2018. Their exchange on the role of technology in human life, on the conception of human being, and other topics revealed noticeable differences between the German philosopher and the Japanese engineer. Four years later two separate interviews follow up on their conversation. In this interview, Hiroshi Ishiguro makes several points: First, there is no clear definition of what a human being, intelligence, emotion, etc. is, so people can understand the meta-level of human beings by making robots, at least getting inspiration for understanding the complex human functions through the reaction of robots. Second, robots have crossed the "uncanny valley" in some situations. Third, the blurring of the boundaries between humans and robots is a good thing, then a new species would be born and people will accept them as human beings. Fourth, after the COVID-19 pandemic, robots that can be operated remotely will be utilized widely. In addition, regarding the roots of Japan's robot culture, Hiroshi Ishiguro proposes the "Island Hypothesis."

**Keywords:** Hiroshi Ishiguro; Humanoid; Robot; Uncanny Valley; Island Hypothesis

**Acknowledgment:** This paper was funded by the research project "The Early Imaginations and Novels About Robots in the German Literature" (20CWW018) of the National Social Science Foundation of China for Young Scholars.

**Citation:** Jiang, H., Cheng, L. & Ishiguro, H. (2022). The Blurring of the Boundaries between Humans and Robots is a Good Thing and a New Species would be Born: An Interview with Hiroshi Ishiguro. *Technology and Language*, *3*(1), 40-46. https://doi.org/10.48417/technolang.2022.01.05

УДК 008: 62-529
https://doi.org/10.48417/technolang.2022.01.05
Научная статья

# Приветствуем стирание границ между людьми и роботами, рождение нового вида: Интервью с Хироси Исигуро

Хуэй Цзян[1], Лин Ченг[2](✉) и Хироси Исигуро[3]
[1]Университет Чжуншань имени Сунь Ятсена, Шинганси. 135, Гуанчжоу, 510275, Китай
[2]Гуандунский университет иностранных языков и внешней торговли, Байюндадаобэй 2, Гуанчжоу, 510420, Китай
lin.cheng@aliyun.com
[3]Осакский университет, 1-1 Ямадаока, Суита, Осака 565-0871 Япония

**Аннотация**
В документальном фильме “Философия в эпоху желания” рассказывается о короткой встрече Хироси Исигуро и Маркуса Габриэля в 2018 году. Их обмен мнениями о роли технологий в жизни человека, о концепции человека и других темах выявил заметные различия между немецким философом и японским инженером. Четыре года спустя их разговор продолжается в двух отдельных интервью. В этом интервью Хироси Исигуро делает несколько выводов: во-первых, нет четкого определения того, что такое человек, интеллект, эмоции и т. д., поэтому люди могут понять метауровень человека, создавая роботов, по крайней мере, получая стимул для понимания сложных человеческих функций через реакцию роботов. Во-вторых, в некоторых ситуациях роботы пересекли «зловещую долину». В-третьих, стирание границ между людьми и роботами – это хорошо, тогда может родиться новый вид, и люди примут его за человека. В-четвертых, после пандемии COVID-19 широкое распространение получат роботы, которыми можно управлять удаленно. Кроме того, что касается корней японской культуры роботов, Хироши Исигуро предлагает “Гипотезу острова”.

**Ключевые слова:** Хироси Исигуро; Гуманоид; Робот; Зловещая долина; Гипотеза острова

**Благодарность:** Статья финансировалась исследовательским проектом “Ранние представления и романы о роботах в немецкой литературе” (20CWW018) Национального фонда социальных наук Китая для молодых ученых.

**Для цитирования:** Jiang H., Cheng L., Ishiguro H. The Blurring of the Boundaries between Humans and Robots is a Good Thing and a New Species would be Born: An Interview with Hiroshi Ishiguro // Technology and Language. 2022. № 3(1). P. 40-46. https://doi.org/10.48417/technolang.2022.01.05

Hiroshi Ishiguro is a Japanese engineer of robots, professor at a Intelligent Robotics Laboratory in Osaka. Humanoid robots such as the Geminoid HI series, which have a similar appearance as the original person Hiroshi Ishiguro, have attracted great attention worldwide. His thoughts on the relationship between robots and human beings are distinctive, as shown in his lectures and his conversations with the German philosopher Markus Gabriel (Gabriel et al., 2018; Li & Gabriel, 2022). We have conducted this interview with Hiroshi Ishiguro in writing between October and December 2021, asking him to offer his views on humanoid robots and human-robot relationships, the social acceptance of humanoid robots, and the robot culture in Japan. In this interview, he also shares his views on the "Uncanny Valley," on robots during COVID-19, and other issues.

**Hui Jiang and Lin Cheng:** You have stated several times in previous interviews that you are developing humanoid robots to better understand humans themselves: can you explain this viewpoint in more detail?
**Ishiguro Hiroshi**: Human functions, especially cognitive functions on the meta-level, in which we may include intelligence, consciousness and emotion, as well as sociality and social relationships, are advanced human functions that are difficult to understand through the bottom-up research approach of the cognitive or brain sciences.

But on the other hand, if we try to make a robot by using the knowledge we have acquired so far, we may be able to simulate emotion, intelligence, and consciousness in it. In this sense, we are attempting to understand the meta-level of human beings, that is, their advanced cognitive functions, through the development of robots. Even if we cannot really understand these functions, they can be simulated by robots to some extent. If we can do this with robots, we may be able to get some hidden hints for understanding these complex human functions. That's the way we combine robot development and human understanding in our research.

**Jiang and Cheng:** Masahiro Mori put forward the "Uncanny Valley" theory in the 1970s. In our interview with him last year, he continued to argue that robots need not be highly humanoid in their appearance (Mori & Jiang, 2020). You have been working on developing robots with ever more humanoid appearances. Do you think robots will one day be able to transcend the "Uncanny Valley"? What are the main difficulties in improving the appearance of robots?
**Ishiguro**: The important thing is that robots' appearance can meet the preferences of people. Some people prefer robots with a human-like appearance, while others prefer robots with a cute appearance, so it's important to have different kinds of robots to meet the preferences of different people.

To be more precise, the "uncanny valley" phenomenon refers to the fact that when we recognize a person, we perceive whether this is a human being or not through multiple modalities, such as appearance, way of speaking, and body movement. However, as soon as one of these modalities deviates from humanity in any way, for example by simply employing some clumsy movements, we immediately feel uncomfortable. Thus, it actually depends on the situation and the purpose of the application of robots. If it is in a specific situation or purpose of use, e.g., only for talking, then it is already possible to create androids that are not uncanny at all, just like our android ERICA. So we can say

that we have crossed the “uncanny valley.” But there are certain restrictions related to the situation and purpose of using androids.

**Figure 1**. Hiroshi Ishiguro (left) and ERICA (right, ERATO Intelligent Conversational Android) ©ERATO ISHIGURO Symbiotic Human-Robot Interaction Project

**Jiang and Cheng:** Your Geminoid HI series robots are already extremely human-like in appearance. Do you think it is possible for future robots to have autonomous or partially autonomous intelligence or even emotions? If so, to what extent can we identify such robots as “people” and give them the corresponding rights? What positives or negatives do you think the blurring of human-robot boundaries may bring about for people and for human society?

**Ishiguro**: There is no clear definition of what autonomous intelligence is, or what autonomy is, and I think it would be quite difficult to give a proper definition of autonomy. The same is true for emotions. In the first place, we don’t even know what autonomy is or what emotions are. So in order to understand them, it is important to

develop and make robots that may have different emotions and can move around autonomously.

I believe that in the future it would be possible to create robots that are more emotionally expressive and more autonomous than the robots we have today. When that happens, I can imagine that people will accept these robots as human beings. At the same time, when we accept them as human beings, it is easy to imagine that we will give them the rights we consider proportionate, or at least such rights as we must treat them with care and not destroy them.

I think the blurring of the boundaries between humans and robots is a good thing. In fact, we don't even have a definition of what a human being is, but at the very least we can say that humans have evolved through technology. Without technology, for example, we would not be able to make vaccines for corona viruses, and humans wouldn't be as prosperous as we are today: there would be no artificial limbs, cars… We are always using various technologies to expand the capabilities of human beings. Therefore, robots and humans are not separate: rather, humans are expanding their capabilities through robots, which are a symbol of technology, so the boundaries between the two are blurred to begin with. The boundaries between humans and technology are blurred, or rather, humans are distinguished from animals because of their convergence with technology. I think this boundary will become more blurred in the future.

**Jiang and Cheng:** As we understand it, you're talking here about the difference between an autonomous robot and a human being: in short, is it going to be a problem when a robot is very human-like in appearance, and its internal function, emotion and intelligence are also close to human beings?

**Ishiguro:** No, I don't think it will become a problem. Instead, I think a new species would be born, and we would live in harmony with them, coexisting with them and looking for help from them in our daily lives.

Of course, if they were detrimental to us, we wouldn't use them, and we would be even less likely to live in peace with them. I think that it is natural for people to find a new species to enrich their lives.

**Jiang and Cheng:** Most of the definitions of robots in traditional European novels and movies are "uncanny." Nowadays, rapid developments in artificial intelligence and robot design is attracting attention in philosophical circles as well. For example, in a conversation with you in 2018, German philosopher Markus Gabriel stated that humanoid robots are "a threat to the definition of human," and "this kind of research could destroy humanity in a very problematic way [sic]". But you believe that "we don't know the definition of human yet," "the boundaries between human and robots will disappear some day," and "we need to accept a new definition of humans" (Gabriel et al., 2018). What do you think is the root of the difference between your and Gabriel's perceptions of human-like robots? In your opinion, does Western philosophical thinking about "what is human" and its tendency towards anthropocentrism affect the development of robotics and AI technology?

**Ishiguro**: If you say that there is a definition of a human being, I hope you can show it to me. If you can define it accurately and scientifically, I might accept that definition, as

long as it is scientific. But unfortunately, I have not yet seen a scientific or universal definition of human beings. So I don't think it is a European thing, rather, it is a difference between what philosophers assume about human beings (well, maybe it's called European philosophy) and science.

Western scientific thinking, philosophical thinking, and anthropocentrism may have an impact on our judgments [of robots JH], perhaps to a slight degree or possibly negatively; but in any case, we are thinking about what it means to be human. I believe that whatever method is used, the starting point may be different, but eventually if you think deeply, you will come to the same conclusion. Therefore, whether we think from the perspective of Western philosophy, from the perspective of anthropocentrism, or from the perspective of the Japanese, it is good to think deeply and properly. Without preconceptions or unnecessary stereotypes, I think we will eventually come to a single answer through discussion.

**Jiang and Cheng:** It is commonly believed that there is a visionary culture of human-robot interaction in Japan. Some scholars argue that a combination of Japan's animistic faith tradition, harmonious human ethics, the positive role of technology in the rapid process of modernization, and the significance of science fiction in popular culture have created this kind of culture. Do you share this view, or do you think there are other reasons?

**Ishiguro**: I basically agree with this. But I think the most fundamental reason or background behind the creation of this kind of thing is what I call the "Island Hypothesis." Since Japan is an island country with a history of nearly 2000 years and has long been isolated from other cultures, we have developed as a result a culture that is completely different from that of European countries and other countries. As a result, an animistic way of thinking and a culture that does not distinguish between humans and other things has been created. If people live together for a long time, if a nation exists on an island country for a long time, then the people will form a relationship like a family. This is a type of harmonious ethics. I think this may be the main and fundamental reason. I call it the "Island Hypothesis."

**Jiang and Cheng:** You approve of the robotization of some areas of daily life (Lill, 2014). Bill Gates (2007) has also predicted a "robot in every home" scenario in the future. When do you think a "human-robot co-existence" society will emerge in Japan? What roles or jobs do you want your robots to fill in the world of the near future?

**Ishiguro**: That's right. Robots have already entered our daily lives in large numbers as toys, haven't they? So I think the next thing we can look forward to is avatars. This does not mean robots that can work autonomously, but robots that can work through remote control.

After the COVID-19 pandemic, robots that can be operated remotely, such as CG agents and CG avatars, will be utilized for people to be able to work freely. In this way, robots and CG avatars will be endowed with autonomous functions and artificial intelligence, and will be able to perform work even without remote control. I think the development of this type of robots will continue to progress.

Therefore, robots will be used in many places to provide services. For example, robots can be used for studying English: in fact, they have been used in high school as simple training partners for English and other foreign languages. I think they would be used in many other places, such as receiving customers in restaurants, or for navigating routes, etc.

## REFERENCES

Gabriel, M., Ishiguro, H., Kokubun K., Chiba, M. (2018, July 15). *Yokubō no jidai no tetsugaku: Marukusu gaburieru Nihon o iku. [Philosophy in the Age of Desire – Markus Gabriel in Japan]* TV program NHK-BS1 https://www.youtube.com/watch?v=H9J19m4ey8g

Li Y. & Gabriel M. (2022). Diverse Cultures, Universal Capacities: Interview with Markus Gabriel. *Technology and Language*, *3*(1), 47-56. https://doi.org/10.48417/technolang.2022.01.06

Gates, B. (2007). Robot in Every Home. *Scientific American*, *1*, 58-65. https://www.scientificamerican.com/article/a-robot-in-every-home-2008-02/

Lill, F. (2014, Aug 11). Hiroshi Ishiguro baut Menschmaschinen. Mein Zwilling, der Roboter [Hiroshi Ishiguro Builds Human Machines. My Twin the Robot]. *Tagesspiegel*. https://www.tagesspiegel.de/themen/reportage/hiroshi-ishiguro-baut-menschmaschinen-mein-zwilling-der-roboter/10314280.html

Mori, M., & Jiang, H. (2020). Uncanny Valley. *New Perspectives on World Literature*, *5*, 85–92. https://caod.oriprobe.com/articles/60286421/Uncanny_Valley.htm

### СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Хуэй Цзян, jianghui6@mail.sysu.edu.cn,
ORCID 0000-0002-5042-0922

Hui Jiang, jianghui6@mail.sysu.edu.cn,
ORCID 0000-0002-5042-0922

Лин Ченг, lin.cheng@aliyun.com,
ORCID 0000-0002-0837-9129

Lin Cheng, lin.cheng@aliyun.com,
ORCID 0000-0002-0837-9129

Хироси Исигуро, h.i@irl.sys.es.osaka-u.ac.jp

Hiroshi Ishiguro, h.i@irl.sys.es.osaka-u.ac.jp

Статья поступила 9 января 2022
одобрена после рецензирования 16 февраля 2022
принята к публикации 28 февраля 2022

Received: 9 January 2022 /
Revised: 16 February 2022
Accepted: 28 February 2022

https://doi.org/10.48417/technolang.2022.01.06
Research article

# Diverse Cultures, Universal Capacities: an Interview with Markus Gabriel

Yue Li[1,2] (✉) and Markus Gabriel[3]
[1]University of Heidelberg, Grabengasse 1, 69117 Heidelberg, Germany
[2]Karlsruhe Institute of Technology, 4, Engler street, 76131 Karlsruhe, Germany
yue.li@kit.edu
[3]University of Bonn, Poppelsdorfer Allee 28, 53115 Bonn, Germany

## Abstract
The documentary *Philosophy in the Age of Desire* records a short encounter between Markus Gabriel and Hishiro Ishiguro's Geminoid in 2018. Their exchange on the role of technology in human life, on the conception of human being, and other topics revealed noticeable differences between the German philosopher and the Japanese engineer, but can these be interpreted as "cultural" differences? Four years later, two separate interviews follow up on their conversation. This interview explores their differences by examining Gabriel's own experiences with AI and his definitions of related concepts such as "intelligence," "ethics," and "consciousness." Gabriel emphasizes that due to our organic precondition there is only a lower-level response in terms of self-understanding. It is only the variability in the expression of self-understandings that results from cultural construction. Focusing on the universal basis of humanity and the influences from Asian philosophy regarding human becoming, Gabriel calls for the further investigation of the cultural presentations of artificial intelligence.

**Keywords:** Human-Machine Interaction; Intelligence; Ethics; Universalism

**Citation:** Li, Y. & Gabriel, M. (2022). Diverse Cultures, Universal Capacities: An Interview with Markus Gabriel. *Technology and Language*, 3(1), 47-56. https://doi.org/10.48417/technolang.2022.01.06

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.06
Научная статья

# Разнообразие культур, универсальные способности: Интервью с Маркусом Габриэлем

Юэ Ли[1,2] (✉) и Маркус Габриэль[3]
[1]Гейдельбергский университет, Грабенграссе 1, 69117 Гейдельберг, Германия
[2]Технологический институт Карлсруэ, 4, ул. Энглер, 76131 Карлсруэ, Германия,
yue.li@kit.edu
[3]Боннский университет, Поппельсдорфер-Аллее 28, 53115 Бонн, Германия

**Аннотация**
Документальный фильм “Философия в эпоху желания” фиксирует короткую встречу между Маркусом Габриэлем и Гениоидом Хироши Исигуро в 2018 году. Их обмен мнениями о роли технологий в жизни человека, о концепции человека и других темах выявил заметные различия между немецким философом. и японским инженером, но можно ли их интерпретировать как “культурные” различия? Четыре года спустя их разговор продолжается в двух отдельных интервью. В этом интервью раскрываются эти различия, исследуя собственный опыт Габриэля с искусственным интеллектом и его определения связанных понятий, таких как “интеллект”, “этика” и “сознание”. Габриэль подчеркивает, что из-за нашей органической предпосылки существует только низко-уровневая реакция с точки зрения самопонимания. Что есть только вариативность в выражении самопонимания, являющаяся результатом культурного конструирования. Сосредоточив внимание на всеобщей основе человечества и влиянии азиатской философии на становление человека, Габриэль призывает к дальнейшему исследованию культурных представлений об искусственном интеллекте.

**Ключевые слова:** Человека-машинное взаимодействие; Интеллект; Этика; Универсализм

**Для цитирования:** Li Y., Gabriel M. Diverse Cultures, Universal Capacities: An Interview with Markus Gabriel // Technology and Language. 2022. № 3(1). P. 47-56. https://doi.org/10.48417/technolang.2022.01.06

## INTRODUCTION

This interview with German philosopher Markus Gabriel took place on February 21, 2022. The idea for this interview originated from the documentary *Philosophy in the Age of Desire* (Gabriel et al., 2018). It records a short but close encounter in 2018 between Gabriel, Japanese engineer Hiroshi Ishiguro, and Ishiguro's Geminoid. One can observe the different opinions of the European philosopher and the Asian engineer regarding their understanding of humanity and technology. We therefore invited Gabriel to talk more about his experiences in 2018 as well as possible changes of his positions since then regarding robotics and AI (compare Ishiguro et al., 2022).

As it happened, Gabriel had just talked with psychologist and Nobel-prize winning economist Daniel Kahneman about the metaverse. Recently, Professor Gabriel acquired a grant for a research project on the cultural presentation of artificial intelligence: *Desirable Digitalisation: Rethinking AI for Just and Sustainable Futures*.

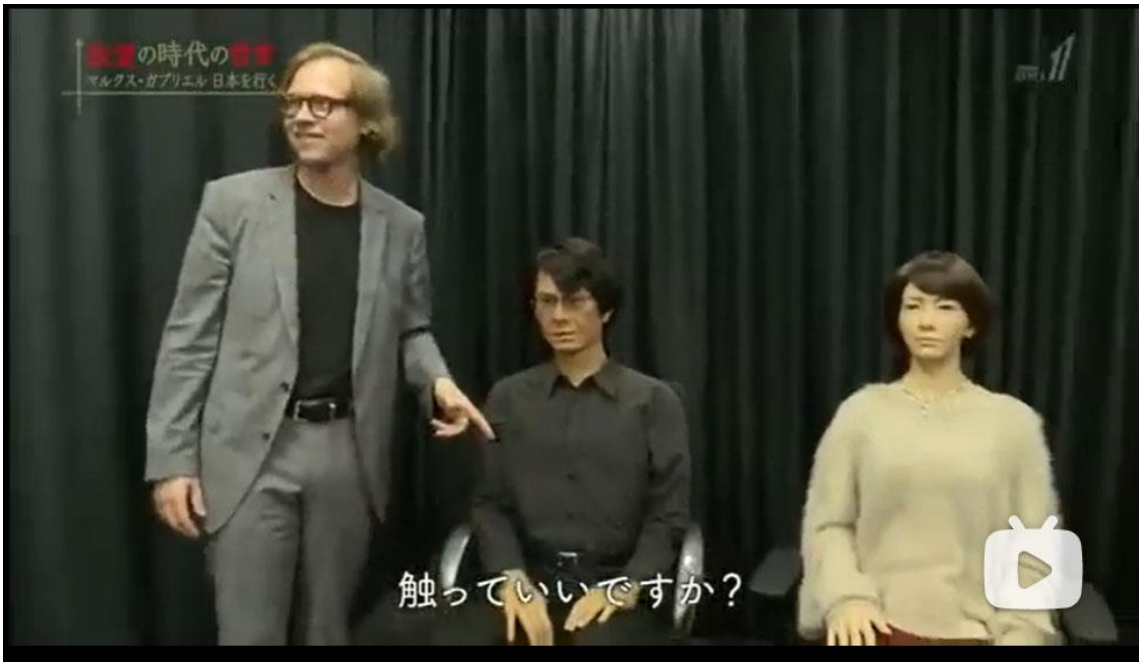



**Fig. 1**. Markus Gabriel in the documentary *Philosophy in the Age of Desire* (Gabriel et al., 2018).

## HUMAN-MACHINE INTERACTION

**Yue Li**: Please allow me to first go back to 2018. In the documentary, I notice that before you made a comment on the Geminoid and considered it an "absent-minded neutral guy," you asked a question: "Can I touch it?" (fig. 1). After Ishigoru's approval ("Yes, you can touch"), you first touched the Geminoid's right hand, and then very briefly the face. Although you described the robot as an object in your question with the impersonal pronoun "it," you began the interaction 'politely' with a 'shake of the hand.' Could this

be interpreted as a brief confusion or even a moment of the uncanny, caused by the "absent-minded guy"?

**Markus Gabriel**: There are two layers of an ethical context here. On the one hand, I'm aware of the different ethical stance of Ishiguro as an individual and maybe as a representative of a Japanese culture. He has a different relationship to that object, and that relationship deserves my respect, even though I don't share the same kind of relationship to the object. The Geminoid is an object onto which a lot of emotion is projected, just as children project emotions on their puppets. Indirectly, I owe that object something from an ethical standpoint. I treat the Geminoid with respect fully knowing that it is entirely inanimate. But then there is a second layer. Ishiguro's robots are cleverly constructed. Due to its human shape and texture of its appearance I - as a human organism - react to it in a certain way. This reaction is universal – there is no difference here between a Chinese, a German or a Japanese. So, imagine I have a computer or GPT3 (Generative Pre-trained Transformer 3) and it talks to me: It sounds to me as if someone was talking to me. I think I hear a voice, but I would say that a computer-generated "voice" is not a voice. Just like I don't think that the programmed chess computer plays chess. There's something that looks uncannily as if it was playing chess, but that is not playing chess.

I still defend the following ontology: The object is inanimate, but the context of its use and abuse gives the object cultural value and personal value. This activity of projection deserves my respect. Besides, the objects, in particular Ishiguro's objects, are cleverly constructed in such a way that it's almost impossible for me as a human animal not to take a certain emotional, ethical stance towards them. Thanks for pointing this out in your reading of that scene.

**Li**: The two layers you mention remind me of *Mein Algorithmus und Ich* – a book that results from the collaboration between the author and an artificial intelligence (Kehlmann 2021). This co-writing experience also reveals the two sides: The uncanny moment on the one hand and the refusal of the phenomenological acknowledgment of the AI on the other hand.

**Gabriel**: Absolutely. There is the phenomenology and then the cognitive correction mechanism, the question is which one is right. I have arguments that support my cognitive correction. This is as with Corona, when I was in the infection phase; I was in denial because I am boosted, but I knew that it does not protect me from Corona 100%, therefore I corrected myself even before I developed symptoms. In this case the mechanism trumps the phenomenology. But there are cases where the phenomenology trumps the correction mechanism, which is why AI and robotics raise these important issues. It is not obvious that my ontological claim that these are just a bunch of objects is the correct one. I'm making a defeasible, fallible knowledge-claim.

## ETHICS AND INTELLIGENCE

**Li**: This uncertainty can be observed in contemporary science fiction as well. As opposed to Kehlmann's and your perception, robots are often depicted as morally better than humans -- see for example the novel *Machines Like Me* (McEwan, 2019) or the movie *I Am Your Human* (Schrader, 2021). Back in 2018 you saw the danger of dehumanization and the emergence of a cyber-dictatorship in the Japanese development of humanoids. So why can't a human-like construct, as portrayed in literature or as claimed by Ishiguro, be a moral model or at least a reflective surface for a better understanding of human-identity?
**Gabriel**: I think this is dangerously wrong. Just think about the recent book by Daniel Kahneman (2021) *Noise: A Flaw in Human Judgement*. It's very clear that human judgment is flawed in comparison to algorithms. Algorithms generally perform much better because their judgments are constructed in a simple way, and therefore they don't vary too much – they behave more coherently. This is empirically tested, so there's no doubt about that.

However, I would again emphasize that these "judgements" do not actually judge, so they don't do ethics. Part of ethics is that it's hard whenever you're facing a real ethical choice. We are facing ethical choices every day in the pandemic: "Should I go out," "who should I meet," "how to deal with this once we contract the virus" – what's the best and fairest solution of all these problems? Often one knows the answer, but the question is, how do we translate the answer into action. It's constitutive of ethics that we are free to decide. The fact that our judgment is flawed is the manifestation of our freedom which is the condition of responsibility and ethics. Precisely because the algorithm outperforms us, also in what looks like judgment, proves that they are less ethical. They lack interiority, freedom and ethics because these dimensions of human life have an organic precondition.

I'm a universalist in ethics: Human ethics binds all humans together. Facts about the health effects of the virus just depend on human organisms. There is no difference between a Chinese person vis-à-vis an Australian and a French person as human beings regarding this aspect: The organic preconditions are preconditions for higher-level ethics while robots and other silicon-based information processing systems just didn't evolve in this way.

**Li**: It's the moral core of every human, the animal part of human which really matters.
**Gabriel:** Yes, I think we are moral as animals. This is what Darwin (1871) said. In *The Descent of Man*, when it comes to the question what if anything is the difference between the human animal and other animals, his answer is literally that of Kant – the "Categorical Imperative."

**Li**: What about technology? Back in 2018 Ishiguro defined humans by the formula Human = animal + technology. How would you define the role of technology in our life?
**Gabriel**: Let's start from my book *The Meaning of Thought* (Gabriel, 2018/2020) in which I define intelligence. Many people dodge the question, but I think AI researchers

should be forced to define it. Here is an attempt: Let intelligence be the capacity of an animal to solve a given problem in a finite amount of time. If some systems solve the same problem faster than some others, it means that they are more intelligent. By doing so we can measure intelligence. I think that AI research is in the business of measuring intelligence in that sense of intelligence – but there are other definitions of intelligence – and of producing models of intelligence such as search algorithms. An additional premise is that thought models are not themselves thinkers. One could argue that in the case of AI, the map is the territory. But I think that AI is the model of a target system, which is human and animal intelligence, just like Google Maps provides a model for say the Black Forest. It would be a category-mistake to confuse the Black Forest with a representation in Google Maps, though this representation is incredibly helpful and even reveals otherwise hidden features. If I just walk through the Black Forest, I will never figure out how far it is through the Black Forest from Freiburg to Munich. But if I use Google Maps, I will get an incredibly accurate answer.

It is similar in the case of chess. I'm not that bad at playing it, but I'm not a grandmaster. During a game I neglect lots of details while an AI system can even detect patterns in my game that no human will identify. It doesn't mean that the AI system is a better chess player, but it is an incredibly good model of a chess player. I would like to develop a theory which allows me to explain the models of contemporary technology without denying the obvious facts of this incredible technological progress while maintaining the good old 1980s position that it is not a real AI.

**Li**: It is a refusal of the "Chinese room" or "Turing test": Passing the Turing test does not indicate a real language user.

**Gabriel**: Absolutely! All the technology specialists always tell me, "You don't know that area. It doesn't have limits." There might not be any technological limits to our models (at least not any time soon, as there is still a lot of space for more processing power), but there is still an ontology which can be modelled but not constructed by technology. AI systems might, as you say, pass the Turing test and they translate German into Chinese better than I can (but worse than you). But all of this is just their performance and they're all performers in so many ways. Imagine I had to screen every web page and remember where some term occurs so as to find that term if needed. I wouldn't get very far searching the internet. Obviously, even on this everyday level, we are dealing with very powerful AI, and this has changed our form of life.

It's not a dualist position that I'm defending. I'm saying that *we* are becoming more intelligent by using these models. It takes more time for Heidegger to walk from Freiburg to Munich without using Google Maps than me using my smartphone. I think the velocity of the contemporary era is driven by AI, which makes us solve our problems faster. Superintelligence is happening, but it is us humans who experience it and become more intelligent.

I don't think we need to wait for the Terminator. Let me give you another very optimistic example. How on earth did we manage to deal with this very dangerous pandemic in such an overall efficient way? It seems to be actually worse than the Spanish flu, and yet, compared to 1914, we are doing ok after two years. One of the reasons is that we can go online: It's our digital infrastructure which has allowed us to pay for the lockdown because we were productive economically. Besides, thanks to digital technology – which always involves the usage of AI – we have become much more intelligent at solving our problems including finding vaccines and medication against the virus. The intelligence explosion which we find in all the thought experiments is really happening, but it's not a property of the machine but of the human-machine interface which is getting more intelligent.

## DIVERSE CULTURES, UNIVERSAL CAPACITY

**Li**: What you say about the relation between humanity and technology reminds me of your dialogue with Ishiguro. He argues that humans are animals who can use technology, which underscores the meaning of technology for the concept of human. But you deny this and underscore that it is the animal part which really matters and makes our self-concept. Can you explain your view of this "self-understanding"? In the discourse of intercultural robotics (Cheng, 2020) have arisen culturally related robot images, such as "the Buddha in the robot" vs. "the ghost in the shell." This, in turn, can be traced back to different "images of people" or "self-understanding" in different cultures. Does this contradict the universal ontology or a universal epistemology?

**Gabriel**: My answer goes back to an ongoing dialogue with the philosophers Thomas Nagel and Paul Boghossian at New York University in 2015. At the time I was deeply puzzled exactly by this point. Chinese, German and English have very different mentalistic vocabulary. The English word "mind," for instance, does not have an exact match in German, nor does the German word "Geist" in English. And in Chinese, "computer" is "Diàn Nǎo" – electrical brain – you regard the computer as a brain, as opposed to the European linguistic representation.

There are different ways of thinking about ourselves which are encapsulated in language use. It depends, among other things, on culture, art and religion. I call this variability. But your way of formulating it, to which I perfectly subscribe, has an interesting catch. You said that there are different cultures or different self-conceptions. This implies that they have something in common: They are different exercises of the same capacity – the capacity to have an image of oneself. This capacity is the same in China and Bavaria: Markus Söder thinks of himself in a different way than Xi Jinping does. They are in different conditions and have different problems and values. They are very different people, but both think of themselves, have a conception of who and what they are, and this capacity is the same.

**Li**: And this is consciousness.

**Gabriel**: I prefer to call it human-mindedness or *Geist*, because consciousness is too often associated with a subjective inner feeling, the phenomenal quality of our animal existence. The higher order capacity to have a lower order response to what you are is a capacity shared by all humans and it has a biological ground. It is not identical or reducible to the biological ground, but it is also biological because we are all the same kind of animal, the human animal.

## NEW INFLUENCES, NEW UNDERSTANDINGS

**Li**: Reviewing your journey in 2018, with all the developments since then, what has changed? What would change if you revisited Ishiguro's Lab?

**Gabriel**: Generally, I have been in close contact with various Asian traditions of doing philosophy. This has accompanied me in different ways for almost three decades until now. Over the last four years, this has been intensified. Chinese and Japanese philosophers – in particular Zhang Xudong, Zhao Tingyang and Takahiro Nakajima – have profoundly influenced me. Their work as well as our conversations have led me to think of the human being in terms of human becoming.

Zhang Xudong, who is professor of comparative literature at NYU, has just published his work in Chinese on the history of universalism (Zhang, 2021). He argues that human becoming is the right term, and we should think of universalism not as a static thing but as an activity of universalizing: Being human – that capacity – is not a static thing but a historical realization, which can take shape in different ways. That's something that I learned through interaction with different Asian cultures.

All of this comes from India to China and Japan and then takes different forms on the spot with local traditions. There has always been an interesting rejection of stable substance ontology since the time of Laozi. I read Laozi as a rejection of the idea that there are substances, think of his concept of "empty" and "the space between" (Moeller, 2007). The crucial way of thinking in Asian traditions is that relations are not entities: Everything is related, but the relation between the objects are not more objects. This is an important philosophical fact, and these traditions have a very deep recognition of it. And this has been influencing me in the last four years, which is why I now tend to think of all of this as human becoming. My universalism or my humanism has gained another perspective through dialogues with thinkers in Japan and China.

If I revisit that scene in Ishiguro's lab, I would probably be more accommodating to some of his ideas. I wouldn't change my position regarding the ontology, but I would be more aware of the performative dimension. I think I now understand Ishiguro's mind better than at the time. I now see that the way in which you express an ontology can add another layer of cultural difference.

That's one of the reasons why I started my new project on the cultural presentation of artificial intelligence to fully explore the lower level of the conceptions of human being and how they may also impact the higher universal human form. Who says that the higher

universal human form is stable? Therefore, I stopped at a certain level of thinking about this as universal and stable, and maybe I should regard this as dynamic …

**Li:** … and changeable and constructed by different cultures in different ways?
**Gabriel**: Definitely. But I think we must never make the mistake of denying the universal basis of humanity. If we say there are "different cultures" and "constructed in different ways," there is always the danger that we stop seeing the humanity in the other – and that must be avoided. If we look at the geopolitical situation right now, what Europe should bring to the global debate is precisely the recognition of humanity in every person, as Kant called it. And that is not a European invention. You find it in Ancient Egypt, China, etc. long before Europe even existed.

Let's just end with a geopolitical statement: The right thing to do in the geopolitical climate right now is the morally good. I think it is a moral mistake that Germany does not hand over vaccine information to Africa. This is what Europe should have learned from its history of colonialism: that other parts of the world, which are incredibly powerful now like China, demand and deserve respect. One way in which Europe should enter the global debate is by bringing to the table its capacity to attain ethical insight of the universal, but that means thinking globally – decentering. I think we need to decenter much more. Particularly for that new project this requires bringing these other conceptions into the picture.

China's and Japan's technological success-stories of the last decade cannot merely be a result of financial support of AI research from their governments. It must also be an expression of a different way of thinking of oneself as human – a certain frame of mind. I think that this frame of mind is something which deserves further investigation. What is the cultural difference? I would again emphasize that the construction of cultural difference has the goal of finding something we share rather than something that separates us.

## REFERENCES

Cheng, L. (2020). Das Unheimliche der Entfremdung: Humanoide Roboter und ihre Buddha-Natur [The Uncanny of Alienation: Humanoid Robots and Their Buddha-Natures]. *Jahrbuch Technikphilosophie*, 6, 83-101. https://doi.org/10.5771/9783748904861-83

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex.* John Murray.

Gabriel, M., Ishiguro, H., Kokubun K., & Chiba, M. (2018, July 15). *Yokubō no jidai no tetsugaku: Marukusu gaburieru Nihon o iku.* [Philosophy in the Age of Desire: Markus Gabriel in Japan] [TV program]. NHK-BS1 https://www.youtube.com/watch?v=H9J19m4ey8g

Gabriel, M. (2020). *The Meaning of Thought*. Polity Press. (Original book published in 2018)

Jiang, H., Cheng, L., & Ishiguro, H. (2022). The Blurring of the Boundaries between Humans and Robots is a Good Thing and a New Species would be Born: An Interview with Hiroshi Ishiguro. *Technology and Language*, 3 (1), 40-46. https://doi.org/10.48417/technolang.2022.01.05

Kahneman, D. (2021). *Noise: A Flaw in Human Judgement*. HaperCollins.

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

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

Moeller, H. G. (trans.), (2007). *Daodejing (Laozi): A Complete Translation and Commentary*, Chicago: Open Court.

Schrader, M. (Director). (2021). *I'm Your Man* [Film]. Letterbox Filmproduktion SWR.

Zhang, X. D. (2021). *Historical Reflections on Western Discourses of the Universal*. Shanghai Renmin Press.

## СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Юэ Ли, yue.li@kit.edu

Маркус Габриэль, gabrielm@uni-bonn.de

Yue Li, yue.li@kit.edu

Markus Gabriel, gabrielm@uni-bonn.de

Статья поступила 21 ноября 2021
одобрена после рецензирования 21 февраля 2022
принята к публикации 23 февраля 2022

Received: 21 November 2021
Revised: 21 February 2022
Accepted: 23 February 2022

https://doi.org/10.48417/technolang.2022.01.07
Research article

# You, Robot: on the Linguistic Construction of Artificial Others[1]

Mark Coeckelbergh (✉)
University of Vienna, Universitätsring 1, 1010 Vienna, Austria
mark.coeckelbergh@univie.ac.at

## Abstract

How can we make sense of the idea of 'personal' or 'social' relations with robots? Starting from a social and phenomenological approach to human–robot relations, this paper explores how we can better understand and evaluate these relations by attending to the ways our conscious experience of the robot and the human–robot relation is mediated by language. It is argued that our talk about and to robots is not a mere representation of an objective robotic or social-interactive reality, but rather interprets and co-shapes our relation to these artificial quasi-others. Our use of language also changes as a result of our experiences and practices. This happens when people start talking *to* robots. In addition, this paper responds to the ethical objection that talking to and with robots is both unreal and deceptive. It is concluded that in order to give meaning to human–robot relations, to arrive at a more balanced ethical judgment, and to reflect on our current form of life, we should complement existing objective-scientific methodologies of social robotics and interaction studies with interpretations of the words, conversations, and stories in and about human–robot relations. This is the Russian translation of Mark Coeckelbergh's "You, robot: on the linguistic construction of artificial others" which is available as an open access publication by the journal *AI & Society*. Six commentaries and Mark Coeckelbergh's response are published along with this translation.

**Keywords:** Human–robot relations; Robot ethics, Language; Phenomenology; Hermeneutics

**Citation:** Coeckelbergh M. (2022). You, robot: on the linguistic construction of artificial others. *Technology and Language*, 3(1), 57-75. https://doi.org/10.48417/technolang.2022.01.07



[1] This paper is the Russian translation of article Coeckelbergh, M. (2011). You, Robot: on the Linguistic Construction of Artificial Others. *AI & Society*, *26*(1), 61–69. https://doi.org/10.1007/s00146-010-0289-z

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.07
Научная статья

# Ты, робот: О лингвистическом конструировании искусственных других[2]

Марк Кекельберг (✉)
Венский университет, Университет-ринг 1, 1010 Вена, Австрия
mark.coeckelbergh@univie.ac.at

## Аннотация

Как мы можем понять идею “личных” или “социальных” отношений с роботами? Начиная с социального и феноменологического подхода к отношениям человека и робота, в этой статье исследуется, как мы можем лучше понять и оценить эти отношения, обращая внимание на то, как наш сознательный опыт взаимодействия с роботом и отношениями человека и робота опосредуется языком. Утверждается, что наши разговоры о роботах – это не просто представление объективной роботизированной или социально-интерактивной реальности, а, скорее, интерпретация и формирование нашего отношения к этим искусственным квазидругим. Наше использование языка также меняется в результате нашего опыта и практики. Это происходит, когда люди начинают разговаривать *с* роботами. Кроме того, эта статья является ответом на возражение этического характера о том, что разговор с роботами не реальность или обман. Делается вывод, что для того, чтобы придать смысл отношениям между человеком и роботом, прийти к более сбалансированному этическому представлению и поразмышлять о нашей нынешней форме жизни, мы должны дополнить существующие объективно-научные методологии социальной робототехники и исследований взаимодействия интерпретациями слов, разговоров и историй об отношениях человека и робота. Это русский перевод статьи Марка Кекельберга “You, robot: on the linguistic construction of artificial others”, опубликованной в открытом доступе журналом “AI & Society”. Вместе с этим переводом публикуются шесть комментариев и ответ Марка Кекельберга.

**Ключевые слова:** Отношения человека и робота; Этика роботов; Язык; Феноменология; Герменевтика

**Для цитирования:** Кекельберг М. (2022). Ты, робот: О лингвистическом конструировании искусственных других. *Technology and Language*, 3(1), 57-75. https://doi.org/10.48417/technolang.2022.01.07



[2] Данная статья является переводом с оригинала Coeckelbergh, M. (2011). You, Robot: on the Linguistic Construction of Artificial Others. *AI & Society*, *26*(1), 61–69. https://doi.org/10.1007/s00146-010-0289-z

## ВВЕДЕНИЕ

Я тебя люблю. Ты любишь меня?
(предложение, адресованное роботизированной кукле (Turkle et al. 2006 , p. 357).

Роботы идут. Но если они войдут в ваш дом, они не смогут вас убить; есть большая вероятность, что они просто хотят обнять. Роботы больше не ограничиваются фабриками, лабораториями и – все чаще – полями сражений. Они постепенно входят в повседневную жизнь людей, предлагая товарищеские отношения, развлечения, секс или заботу о здоровье. Некоторые люди предпочитают искусственных “друзей” или даже искусственных партнеров.

Хотя этот сценарий может оставаться для многих из нас научной фантастикой, роботы уже используются в этих областях, и мы хотим знать, что произойдет, если роботы станут не только более автономными и интеллектуальными, но и все более “личными” или “социальными”. Сценарий предлагает нам задуматься о том, каково было бы жить с “социальными роботами” (Breazeal, 2003) и как мы должны оценивать то, что происходит между людьми и роботами. Но что еще более важно, это помогает нам задуматься о самих себе: что значит быть человеком, что мы подразумеваем под социальными отношениями и как мы должны жить вместе.

Как мы должны понимать и оценивать отношения человека и робота? В ответ на новые разработки в социальной робототехнике появляется все больше литературы о взаимодействии человека с роботом и отношениях “человек-робот” (Turkle et al., 2006; Breazeal, 2003; Dautenhahn et al., 2005; Dautenhahn, 2007; Levy 2007, и др.). Более того, философия информационных технологий – в частности, этика роботов – также начала размышлять об “искусственных спутниках” (например, Floridi, 2008) и их использовании в таких областях, как уход за пожилыми людьми и здравоохранение (например, Sparrow and Sparrow 2006). Некоторые даже рассматривают вопрос о правах роботов (Brooks, 2000; Levy, 2009; Asaro, 2006; Torrance, 2008) или, в более общем плане, поднимать вопрос о защите (некоторых) роботов от злоупотреблений (Whitby, 2008).

Однако, становясь все более междисциплинарными, в большинстве случаев, работы в области персональной и социальной робототехники, используют методы из социальных наук или психологии. Они остаются близкими к методологическому натурализму науки о робототехнике, к которой они относятся. Кроме того, предназначенные прежде всего для описания и объяснения, они страдают отсутствием нормативной направленности. Можем ли мы рассматривать человеческие отношения с другой, более отчетливой философской точки зрения, которая дистанцируется от научно-объективных подходов и соответствует потребности в оценке?

В предыдущей работе я предложил социально-феноменологический подход к философии робототехники, который фокусируется на философской значимости внешнего вида роботов и отношений между человеком и роботом, как социальных отношений (Coeckelbergh, 2009, 2010a, 2010b). В этой статье я хочу продолжить

развитие этого подхода, исследуя потенциальные преимущества лингвистико-герменевтического поворота в философии робототехники. Я утверждаю, что появление роботов в человеческом сознании опосредовано языком: то, как мы используем слова, интерпретирует и формирует наши отношения с другими людьми или “искусственными” людьми. Кроме того, я коснусь распространенного возражения против такого “разговора о роботах”: являются ли эти отношения нереальными и обманчивыми, угрожающими подлинности? Затем я делаю выводы для исследования, проектирования и этической оценки отношений между человеком и роботом.

В ходе своих аргументов я использую эмпирические исследования взаимоотношений человека и робота, в частности работу Шерри Теркл.

## ОТНОШЕНИЯ ЧЕЛОВЕКА И РОБОТА КАК СОЦИАЛЬНЫЕ ОТНОШЕНИЯ

Отношения человека и робота можно определить как социальные отношения. Что здесь означает “социальный”? Во-первых, все роботы “социальные” в том смысле, что они играют определенную роль в человеческом обществе, точно так же, как другие артефакты являются “частью” общества в качестве инструментов для человеческих целей. Например, автомобили играют важную роль в обществе. Во-вторых, некоторые роботы также являются “социальными” в другом смысле: они по видимому участвуют в “социальном” взаимодействии с людьми: это автономные интерактивные роботы, которые следуют социальным правилам и взаимодействуют с людьми по-человечески. Однако смысл, в котором я буду использовать слово “социальный” в отношении роботов и отношений между человеком и роботом, касается последствий этих взаимодействий. Есть особый феноменологический смысл, в котором некоторых роботов можно назвать “социальными”: некоторые роботы кажутся нам больше, чем инструментами. Они представляются “квазидругими” (Inde, 1990) или искусственные другие. Взаимодействие, основанное на этой видимости, составляет (квази) социальные отношения между человеком и роботом, независимо от онтологического статуса робота, как это определено современной наукой, традиционной и современной метафизикой, которые рассматривают робота как простую вещь или машину.

С точки зрения этики этот подход подразумевает, что моральный статус роботов больше не зависит от “объективных” характеристик робота, а от того, как робот и отношения между человеком и роботом проявляются в человеческом сознании. Таким образом, для конструкторов роботов, которые хотят создать “морального” робота, важно не создание его *разума*, а создание его образа в отношениях.

Однако для дальнейшего развития этой точки зрения нам необходимо более точное описание того, как разворачивается этот феноменологический процесс. Как создается “социальный” образ? В каких условиях робот кажется нам квазидругим? Как нам понять такие отношения человека и робота? В следующих разделах я частично восполню этот пробел, обсудив роль *язы ка* в том, как роботы и

отношения между человеком и роботом представляются нам социальными. Изучая лингвистико-герменевтическое измерение социальной феноменологии человека и робота, я надеюсь внести свой вклад в лучшее понимание и оценку того, что происходит или может происходить между людьми и роботами.

## ЯЗЫК И СОЦИАЛЬНОЕ

Позвольте мне выделить два противоположных взгляда на отношение между языком и социальным, которые я назову репрезентализмом и конструктивизмом. Оба взгляда отличаются от крайнего идеализма и наивного реализма, которые определяют отношения между языком и миром, поглощая одно в другом: крайний идеализм (в его постмодернистской или структуралистской версии) "удаляет" мир вне языка; наивный реализм "упраздняет" предмет. Применительно к социальному (миру) это означало бы, что социальное является чисто лингвистически-концептуальным, или что мы можем знать социальное, как объективную реальность. Обе точки зрения ошибочны. Социальное существует также "вне" языка, хотя у нас нет прямого доступа к нему: мы воспринимаем его через призму языка, мы говорим о нем. В этом смысле язык неразрывно связан с социальным. Однако, две противоположные точки зрения, которые я имею ввиду, принципиально различаются в том, как они рассматривают эту внутреннюю связь.

Согласно репрезентативной точке зрения, социальное предшествует языку. Отношение между социальным и языковым – это отношение репрезентации: словами мы представляем социальные отношения. Язык – это "зеркало" социального. Хотя у нас может не быть *прямого* доступа к социальному, язык приносит нам "социальное" без искажений. В противоположность этому, согласно конструктивистской точке зрения, язык важнее социального. Социальное конструируется или даже декларируется (Searle, 1995): речевой акт предшествует социальному, поскольку он создает социальное. Согласно этой точке зрения, социальное является субъективным или межсубъективным в том смысле, что оно создано человеческим намерением и человеческим соглашением; однако, *однажды созданный* он принимает "объективный" характер в форме правил, законов и (других) институтов.

Однако должны ли мы выбирать между этими полярными противоположностями? Здесь нет настоящей дилеммы. Вместо того, чтобы рассматривать социальное или языковое как "предшествующее", давайте попробуем построить синтез наилучшего понимания этих двух позиций. Репрезентативная точка зрения верна, что социальное – это не *просто* вопрос лингвистической конструкции или декларации; существует экстралингвистическая социальная реальность. Социальное не просто (интер-) субъективное. Однако мы можем получить доступ к этой социальной реальности только через язык (слова и концепции – это "очки", через которые мы видим реальность), и конструктивистская точка зрения верна, наше использование языка совместно формирует эту реальность. Язык не отражает социальное, но также помогает создавать. Более того, даже помимо этого более "активного" аспекта языка,

посредническая роль языка не ограничивается репрезентацией (или даже не описывается адекватно как таковая): мы также *интерпретируем* социальное. У нас могут быть разные интерпретации социальных институтов и практик. В обоих случаях знание социального требует “акта” интерпретации или даже конструирования.

## ЕСТЕСТВЕННЫЙ ЯЗЫК И ИСКУССТВЕННЫЕ СПУТНИКИ

Применительно к отношениям человека и робота лингвистико-герменевтическая перспектива, которая подразумевает согласованное объединение следующих двух утверждений: (1) использование нами языка конструирует отношения между человеком и роботом (понимаемые как социальные отношения) и (2) эти отношения имеют также экстралингвистическое измерение: отношения приобретают квазиобъективную реальность, но напрямую не доступны для опыта: они представляются нам через язык и интерпретируются нами с помощью языка.

Этот аргумент о лингвистическом посредничестве раскрывает возможность по крайней мере двух разных лингвистико-феноменологических “очков” или репертуаров (или того, что Серл называет “статусными функциями”) для отношения к роботам: мы можем рассматривать их как “вещи” и объявлять их просто объектами или машинами, но мы также можем конструировать их как квазидругих. Первый репертуар представляет собой онтологию, в которой существует строгое разделение между (человеческими) субъектами и (роботизированными) объектами (а также между физической онтологией и человеко-социальной онтологией, как в работе Серла по социальной онтологии[3]), тогда как второй репертуар создает социальную онтологию отношений человека и робота, которая носит более “гибридный” характер. Эта онтологическая гибридность “проявляется” в языке (язык как представление и интерпретация), но в то же время она также конструируется с помощью языка.

Мы не можем просто выбирать между этими двумя возможностями или репертуарами. В этом смысле язык “декларации” Серла вводит в заблуждение. В нынешних отношениях человека и робота мы можем наблюдать переход от разговоров *о* роботах и *об* отношениях между человеком и роботом к разговору *с* роботами. Позвольте мне объяснить этот сдвиг и выявить его лингвистическое измерение и философское значение, проведя различие между разными “перспективами”.

[3] Полный ответ на работу Серла по социальной онтологии и участие в ней потребуют более длительной работы. Для целей этого аргумента достаточно сказать, что Сёрл придерживается “дуалистической” точки зрения в том смысле, что он применяет объективистский подход к физической реальности, но конструктивистский подход к социальной реальности. Версия конструктивизма, которую я исследую в этой статье, оставаясь агностиком в отношении конечной природы реальности, ставит под сомнение дуалистический взгляд Серла. Это предполагает, что у нас нет прямого доступа к какой-либо реальности – “объективная” реальность или “факты” должны быть интерпретированы и сконструированы, чтобы они стали тем, что мы называем знанием.

Говоря об отношениях человека и робота

Рассмотрим следующий пример. Идея “любви”, дружбы или “брака” между людьми и роботами поднимает этические вопросы (например, Levy, 2007). Любовь к роботам – настоящая любовь? Допустим ли секс с роботами? Некоторые люди считают, что даже *говорить* об этом оскорбительно .

Однако вопрос о том, может ли быть настоящая любовь между людьми и роботами, не является правильным, поскольку он предполагает, что такая “любовь” будет объективной реальностью, которая стоит отдельно от нашего опыта и практики человека и робота. Представьте, что однажды мы *объявляем* конкретное отношение человека и робота “любовным”. Мы можем даже объявить конкретного человека и конкретного робота “женатыми”. В этом смысле любовь к роботам была бы социальной, лингвистической конструкцией. Однако, если бы это произошло, это могло бы произойти только при условии, что отношение уже *представляется* и *интерпретируется* нами, как “любовное” в силу того, что на самом деле происходит между естественными и искусственными “партнерами”. Другая феноменология и интерпретация не могут служить основанием для такого же заявления; это сделало бы “декларацию” пустой, лишенной смысла (большинство современных роботов *не* воспринимаются нами как заслуживающие нашей любви и привязанности). Но это также верно и для человеческих отношений: мы не можем “просто” изменить социальное посредством декларации; такое заявление должно быть связано с опытом.

Обратите внимание, что аналогичный аргумент может быть выдвинут в пользу прав роботов: декларация прав роботов имеет смысл только тогда, когда феноменологический-герменевтический процесс, который частично и важно является лингвистическим по своей природе, поддерживает декларацию. (Для дизайнеров, которые стремятся создать роботов, требующих декларации прав, это означает, что у робота должны быть функции, которые запускают этот интерпретирующий и декларативный ход. Однако я не буду здесь развивать этот момент).

Этот аргумент предлагает новый взгляд на отношения человека и робота. Основываясь на социально-лингвистических интерпретациях и конструкциях, можно выделить следующие взгляды на язык и роботов.

*Безличная* перспектива от третьего лица (“it”, “это”) и перспектива *от первого лица* (я, робот): обычная перспектива Искусственного Интеллекта (ИИ) о языке и робототехнике, принятая исследователями-робототехниками, фокусируется на программировании (способности) естественного языка в роботе, то есть он связан с тем, как *роботы* (“это”) говорят, могут и должны говорить. Если язык рассматривается разработчиками, пользователями роботов и других искусственных систем, то они, в основном, беспокоятся о том, что робот говорит или может (не) сказать. Мечта традиционного ИИ (и современных отделов жалоб в крупных компаниях) заключалась в создании систем с искусственным

интеллектом, которая была бы неотличима от пользователей естественного языка. Рассмотрим обсуждение игры Тьюринга (Turing, 1950 ) или эксперимент Сёрла “Китайская комната” (Searle, 1980): вопрос в том, можем ли мы отличить человека от робота на основе использования естественного языка. Конечная мечта – создать сознательного робота, который говорит: “Я, робот” (как в названии научно-фантастического фильма). Затем будет вид от первого лица: роботы могут заявить, что они сознательны, и, возможно, *потребовать* соблюдения своих прав.

Однако подход, который я предлагаю здесь, обращает внимание на то, как *люди* говорят о роботах. Это смещает фокус с того, что говорит робот, на то, что говорят люди.

## Говоря о роботах

*Перспектива от третьего лица*. Поскольку развитие робототехники продолжается и роботы становятся в большей степени “личностями”, дизайнеры и пользователи могут переключаться с точки зрения от первого лица (в смысле попытки придать роботу человеческий разум) к перспективе от третьего лица и от безличного местоимения к личному местоимению. Люди больше не считают робота машиной и начинают обращаться к роботам как в личностных терминах. “Это” становится “он” или “она”. Можно сказать, что “он” хочет внимания или еды, что “она” что-то со мной сделала, и так далее. (Аналогичный процесс происходит, когда мы имеем дело с разумными животными. Некоторые люди также используют этот язык, когда разговаривают с автомобилями и другими вещами).

## Говоря с роботами

*Перспектива от второго лица* (ты, робот): следующий шаг (который также происходит в некоторых отношениях между человеком и животным) является то , что мы не только говорим *о* роботе, но и *с* роботом. Здесь робот полностью проявляется как квазидругой, признается как квазидругой и конструируется как квазидругой. Речь обращена к роботу, а не только к людям. Мы легко можем представить себе человека, обращающегося к своему роботу-компаньону или роботу-партнеру – “ты”. Человек может сделать это только с учетом того, что “делает” с ним внешний вид робота. В то же время, обращаясь к роботу таким образом, робот и отношения строятся *как* компаньонские и *как* партнерские отношения (ниже приведены примеры, взятые из эмпирических исследований).

Если робот появляется как квазисубъект со своим собственным сознанием, потребностями, желаниями и мыслями, это может даже перерасти в видимость (квази) интерсубъективности. Если робот отвечает, создается “диалог”. Тогда у нас сложилось впечатление, что не только мы говорим *роботу*, но что мы говорим *с роботом* – это опыт , который похож на разговор с другими людьми. Здесь мы можем объявить наше взаимодействие “дискуссией” или “беседой”. Люди могут называть себя и робота “мы” (от первого лица во множественном числе), когда разговаривают с другим: человеком или искусственным. Кто-то может сказать:

"мы" это обсуждали; "мы" хорошо провели время (однако я не буду здесь развивать это предложение и сосредоточусь в основном на взгляде от второго лица).

Такой подход к человеческим отношениям имеет несколько значений для исследователей и дизайнеров в области социальной робототехники.

## ПОСЛЕДСТВИЯ В ИЗУЧЕНИИ И ОЦЕНКЕ ОТНОШЕНИЙ МЕЖДУ ЧЕЛОВЕКОМ И РОБОТОМ

Во-первых, этот подход уделяет большое внимания тому, как люди говорят о роботах и с роботами. Наш "разговор о роботах" не является нейтральным, но он интерпретирует и формирует наши отношения с роботами, он выполняет герменевтическую и нормативную функцию.

Это понимание может пролить новый свет на существующие исследования в области социальной робототехники. Например, в исследовании Теркл и других сообщается, что пожилые люди и дети разговаривают с роботами (Turkle, 2005; Turkle et al., 2006 ). Например, один из обитателей дома престарелых говорит роботизированной кукле My Real Baby: "Я люблю тебя. Ты любишь меня?" (Turkle et al., 2006 , p. 357). На основании этого и других наблюдений Теркл заключает, что существует большая степень привязанности к роботам как к "реляционным артефактам", как инструментам для построения отношений. Они могут помочь объединять людей. Исследование Теркла также показывает, что люди разговаривают с роботами как с "вызывающими воспоминания объектами", которые напоминают им о людях, которые играют или играли определенную роль в их жизни.

Однако рядом с этими способами отношения с роботами существует также возможность того, что робота следует рассматривать как уникального искусственного *другого*, а не инструмент или воспоминание о человеке-другом. Это "использование" еще недостаточно изучено, но на него указывает, когда Теркл предполагает, что люди, у которых развиваются такие привязанности, хотят, чтобы роботы разговаривали с ними, например, произносили их имя или говорили: "Я люблю тебя" , пожилые люди могут прийти к любви к роботам, которые о них заботятся, и это может быть слишком неприятно, если робот не скажет в ответ слова "Я люблю тебя" (Turkle et al., 2006, p. 360). (Однако она признает, что это не совсем удобно и что это поднимает вопрос о подлинности. Я вернусь к аналогичному этическому вопросу ниже). Кажется, что ожидание возникает внутри социальных отношений. Робот становится "ты" – не в качестве заместителя для кого-то другого (что мы могли бы назвать "делегированным вторым лицом"), а "ты" в своем собственном праве, искусственным вторым лицом, у которого есть претензия на "меня", как на социальное существо. Исходя из того, что я сказал ранее, мы можем понять этот разговор *с* роботом, изменяющий и формирующий отношения между человеком и роботом как социальные отношения: люди разговаривают с роботами не только из-за "личного" или "социального" внешнего вида робота (возникающего из-за того, что робот обладает определенными характеристиками и способностями к определенному поведению); разговаривая с

роботом во втором лице, они также конструируют его как квазидругого. *Частично из-за подобного использования языка* (например, использования слов "ты" и "любовь") робот больше не кажется пожилому человеку как объектом, а как "квазидругим". Это создает социальную реальность: это порождает ожидания, которые формируют квазисоциальные отношения и, следовательно, больше не являются "просто субъективными".

Это может не только пролить новый свет на существующие исследования, но и подсказать новые исследовательские вопросы. Например, я предлагаю проверить следующую гипотезу: предварительное лингвистическое построение отношений между человеком и роботом (как социального отношения) влияет, со-конструирует фактическое отношение. Обычно исследователи манипулируют (параметрами) взаимодействия, а затем смотрят, что происходит с точки зрения использования языка, например, что люди говорят роботу. Я предлагаю изменить это: манипулировать *лингвистической* "средой", а затем посмотреть, что происходит с взаимодействием. Например, что происходит, когда инструктор уже заранее определяет взаимодействие как "личное" или "социальное" отношение, используя такие слова, как "она" или "он", или давая роботу имя? Отличается ли результат от того, что происходит в контрольной группе (где инструктор использует "оно" и заводское "название" робота)? Кроме того, чтобы исследовать герменевтический аспект можно было бы провести долгосрочные исследования *отношений* человека и робота – во всех их интерпретациях и нарративах – в отличие от текущих краткосрочных исследований *взаимодействий*.

Во-вторых, поскольку языки и культуры, в которые эти языки встроены, различаются, такой подход, естественно, предполагает обращать внимание на *культурные различия* . Например, во многих других языках есть не одна, а две или несколько форм второго лица. В английском есть "you", но в голландском есть разница между "you" (неформальный) и "u" (более формальный и вежливый). Рассмотрим также немецкие du и Sie, французские tu и vous или хинди "tum" и "aap". Другие языки могут иметь больший набор лингвистических и социальных возможностей. Какие формы используются и действительно *должны* использоваться для искусственного обращения к другим? Эти описательные и нормативные вопросы важны, поскольку их точная форма выражает, но также и составляет, социальные отношения и структуры. Например, нашими словами мы можем подразумевать и образовывать отношения хозяина и раба или отношения компаньонов. В этом также есть гендерный аспект: когда мы обращаемся к роботу с точки зрения третьего лица, используем ли мы "он" или "она"?

У языка есть нормативное измерение, предписывающее измерение: слова связаны со способами действий. Вместе с Витгенштейном мы могли бы назвать эти способы ведения дел "формами жизни" (Wittgenstein, 1953/2009). Язык и культура тесно связаны. Значение слов определяется *использованием* языка и практикой; в то же время язык также формирует эти практики. То, как и для чего мы делаем вещи, форму которую они принимают (культура), которая зависит от языка. Вообразить отношения между людьми и роботами – значит представить форму жизни: ту, которая включает в себя отношения человека и человека, отношения

человека и робота и облако слов и значений, которые их окружают. Эта форма, вероятно, частично отражает человеческие отношения, о чем свидетельствует использование личных местоимений. Это использование имеет источником то, как люди обращаются друг к другу. Однако нам также необходимо изучить, могут ли отношения между человеком и роботом трансформировать то, как мы делаем что-то, и каким образом; они могли внести свой вклад в новую форму жизни. Более того, мы можем попытаться представить, что бы это значило, если бы у роботов был свой *собственный* образ действий, если бы они разработали свою собственную форму (искусственной) жизни.

Нам может быть трудно это представить. “Разговор роботов” во многом связан с мировоззрениями/онтологиями – включая социальные онтологии – культуры, в которой он находится. Например, на Западе мы склонны уделять много внимания различию субъект-объект, и большинство из нас придерживается индивидуалистических, не реляционных онтологий. Роботы относятся к категории “объект”, а люди – к категории “субъект”. В этой современной дуалистической онтологии не допускаются никакие пересечения или гибриды; чистота сохраняется[4]. Языки уже “содержат” такие онтологии в своих структурах. Наша лингвистическая грамматика также является моральной, социальной и онтологической “грамматикой”. Снова рассмотрим использование местоимений и различие между субъектом и объектом в западных языках: язык, который мы используем, формирует то, как мы интерпретируем и конструируем социальное. Это устанавливает ограничения на воображение, формирование, исследование и испытание новых форм жизни, в которых роботы берут на себя роль искусственных других. В то же время язык не является фиксированным, и наши слова и концепции могут изменяться, и изменяются в результате изменения социальных отношений. Например, мы можем начать использовать личные местоимения в разговоре с роботами. Как изменятся эти отношения, частично, но не полностью зависит от нас. Мы не можем просто или полностью изменить нашу форму жизни. Существование в определенной культуре, принадлежность к определенному поколению, особенности личности и т.п. будут влиять на язык, которым человек пользуется, и на то, как этот язык меняется (я возвращаюсь к этому пункту).

В-третьих, этот подход способствует изучению изменений во взаимоотношениях между человеком и роботом. Они меняются в результате многих факторов, но также частично меняются из-за того, как мы говорим о роботах, и из-за наших заявлений об этих отношениях.

Этот подход включает измерение времени и измерение личной идентичности: он подразумевает, что могут быть *истории/я* об отношениях человека и робота (вымышленные и подлинные), в которых мы интерпретируем и формируем идентичности человека и робота и отношения, в которых эти идентичности встроены. Мы можем использовать “они”, когда говорим о конкретных отношениях человека *и* робота или даже “мы” от первого лица. Например, люди могут называть человека и его робота “они”, когда кажется, что между ними

[4] См. Также работу Латура (Latour,1993).

существует личная, социальная связь. Тогда отношения предполагают своего рода коллективную идентичность, точно так же, как пару людей объединяет слово “они” (“я думаю, они должны пожениться”) (см. Также мою точку зрения о интерсубъективности в предыдущем разделе).

Разработчикам робототехники и ученым рекомендуется учитывать эти аспекты при рассмотрении вопроса об использовании роботов: использование зависит не только от “объекта” как такового, но и от того, как этот объект представляется людям и как это представление опосредуется и формируется людьми-субъектами и их использованием языка.

Тогда с нормативно-этической точки зрения мы можем спросить: как люди должны говорить о роботах? Ответ на этот вопрос может помочь в разработке роботов. Например, может оказаться, что есть веские причины, по которым мы не должны желать, чтобы роботы выглядели как другие или как личности, что мы не должны желать, чтобы люди *говорили о них или с ними таким образом*. Чтобы изучить этот вопрос, позвольте мне закончить обсуждение *одним* этическим возражением против личных роботов. Это не только позволит мне сказать больше об отношениях между человеком и роботом; это также даст мне возможность показать один из способов, каким образом предлагаемый здесь подход может способствовать изучению и оценке *человеческих* отношений.

## ОБМАНЧИВЫЕ ВОЗРАЖЕНИЯ ПРОТИВ ЛИЧНЫХ РОБОТОВ

Одно из возможных возражений против личных роботов (и связанных с ними лингвистико-концептуальных изменений, которые я обсуждал) касается двойного обвинения в том, что эти роботы, отношения между человеком и роботом, “разговоры” между человеком и роботом и т. д. не *настоящие* личности, не *настоящие* (социальные) отношения и разговоры (то есть недостоверны), и что их передача людям была бы обманом.

Хотя я симпатизирую этому возражению, я сомневаюсь, что оно приемлемо в его нынешней форме по следующим причинам. Это возражение предполагает, по крайней мере, следующее: (1) что разговор с вещами всегда и обязательно морально проблематичен, (2) что только человеческие отношения реальны, истинны и подлинны, (3) что существует объективная внешняя точка зрения это позволяет нам судить о реальности и истинности отношений между человеком и роботом, и (4) что сказать, что робот – это вещь, совершенно не проблематично. Но эти предположения следует подвергнуть сомнению.

Во-первых, многие люди уже разговаривают с вещами и другими не-людьми, и это обычно не считается моральной проблемой. Люди разговаривают с растениями, животными, куклами, автомобилями, навигационными системами, DVD-плеерами и, как Теркл уже наблюдал в 1980-х годах, с компьютерами (Turkle, 1984)[5]. Когда люди обращаются к современным роботам, используя точку зрения

[5] Обратите внимание, что некоторые из этих сущностей “разговаривают с нами ”, например, навигационная система в автомобиле. Независимо от того, следует ли считать это “разговором”, он также формирует особые

от второго лица (“ты”), лингвистическая точка зрения остается той же. Конечно, воплощение роботов и родственных объектов побуждают к более эмоциональному и социальному взаимодействию (например, воспитанию, см. Turkle et al., 2006 , p. 348), но неясно, почему это само по себе делает разговор с роботами *принципиально* более проблематичным с моральной точки зрения. В обоих случаях объект воспринимается и рассматривается как квазидругой. Роботы, в той мере, в какой они более интерактивны, предлагают более сильный “квазидругой” опыт. В этом смысле они действительно более “обманчивы”. Требуется дальнейшее изучение того, как мы воспринимаем различные виды артефактов и как относимся к ним, но на первый взгляд “обман” роботов по сравнению с “обманом”, когда кто-то разговаривает с растением, не кажется принципиально отличным по своему характеру[6].

Кто-то может возразить, что все же “восприятие” робота нереально или неверно. Но что более реально или верно: (абстрактное) определение робота как объекта или реальный опыт робота как квазидругого? С конструктивно-феноменологической точки зрения на этот вопрос нет готового и априорного ответа.

Парадоксально, но те, кого больше всего возмущает мысль о разговоре с роботами, должны иметь сильный квазидругой опыт. Если бы люди действительно верили, что роботы, о которых идет речь, являются “простыми объектами”, у них не было бы моральных проблем с разговорами с ними. Они сочли бы это “глупым”, “детским” или, возможно, даже “безумным”, но не сочли бы это морально проблематичным. Моральный вопрос возникает только тогда, когда явление квазидругого уже имело место. Этические проблемы возникают, когда, говоря феноменологически, робот перемещается в сумеречную зону между объектом и субъектом: он выглядит “больше”, чем объект, но “меньше”, чем субъект. Например, секс-роботы могут стать морально проблемными только в том случае, если они *уже* представляются как квазидругие. Если бы они воспринимались, *просто* как инструменты, используемые в сексуальных действиях, они не вызывали бы такого возмущения[7].

Во-вторых, проблематично ли полагаться на внешнее в социальных отношениях с моральной точки зрения? Внешнее играет важную роль в человеческих отношениях. Мы не всегда “действительно знаем” людей, с которыми мы общаемся, но внешнее смазывает наши межчеловеческие социальные отношения. Конечно, мы иногда говорим, что кто-то кажется *внешне* таким и

отношения между человеком и технологией, которые развиваются. Есть много способов, которыми артефакт может “общаться” с нами, и, по-видимому, каждый из них влияет на то, как мы делаем что-то и воспринимаем окружающую среду. Также обратите внимание, что общение роботов может частично происходить с помощью “языка тела”. Если конструкторы роботов вообще хотят имитировать человеческое общение (они могут сделать другой выбор дизайна), это важный аспект, который следует принять во внимание.

[6] Единственное исключение, о котором я могу думать, – это возможность того, что иногда люди разговаривают с роботами или растениями не потому, что робот, растение или другой объект выглядит как квази-другой, а потому, что они используют его как инструмент для упорядочивания своих мыслей, для разговора с ними. сами , чтобы иметь внутренний диалог – действительно думать. Тогда робот появляется как представление “внутреннего” партнера по диалогу, он “интернализуется” как продолжение внутреннего разговора.

[7] Леви помещает секс-роботов в историю приспособлений для секса (Levy, 2007)..

таким-то, *в действительности* же человек такой-то и такой-то. Но как определить настоящее, истинное или подлинное? Есть ли у нас прямой доступ к чистой реальности, к истине о личном ядре, к подлинному “я”? Ответ на этот вопрос, по крайней мере, более сложный, чем предполагает возражение.

В-третьих, один из ответов на этот извечный философский вопрос следующий. То, что у нас есть наверняка, – это внешнее представление о роботах, о самих себя, как людях, и о наших отношениях с роботами. Сможем ли мы получить доступ к чему-либо еще, зависит от того, есть ли у нас объективная внешняя точка зрения за пределами отношений, какими они кажутся нам. Однако как наш взгляд на человеческие отношения, так и наш взгляд на отношения человека и робота не стоит полностью отдельно от самих этих отношений. Следуя аргументам, изложенным в предыдущих разделах, наш ответ должен заключаться в том, что наши разговоры с людьми и наши разговоры с роботами не являются нейтральными по отношению к тому, как мы определяем реальность и истину, а также как и формируем эти отношения. Это не означает, что этим терминам нельзя придавать никакого значения, а скорее то, что знание реального или истинного о людях и роботах не очевидно, не дано и должно быть тщательно сконструировано с учетом конкретного опыта как образы в отношениях.

В-четвертых, сказать, что роботы “являются” “простыми” объектами, нетривиально и небезопасно. Как следует из предыдущего обсуждения, наше сопротивление обращению *к* роботам и разговору *с* роботами сформированы (1) нашим опытом, работы с роботами как простыми машинами (подобно тому, как Декарт мог воспринимать животных только как машины) и (2) нашим западным взглядом на отношения между людьми и нечеловеческими существами, основанным на онтологии, которая делает строгое различие объекта и субъекта и, следовательно, исключает возможность “гибридов”, таких как роботы в качестве квазидругих. Однако, хотя наш текущий опыт и концептуальные основы не позволяют нам по-другому взглянуть на роботов, это может измениться в будущем. По мере разработки и использования новых видов роботов наш язык и наш опыт меняются и взаимно влияют друг на друга. Если мы примем во внимание этот культурный и динамический аспект в наших оценках, это может сделать эти оценки менее стабильными, чем мы хотели бы, чтобы они были с теоретической точки зрения,

Таким образом, хотя эти экспериментальные и концептуальные изменения не полностью зависят от нас, поскольку мы не можем полностью контролировать развитие технологий, социального опыта, культурного значения и личности, эта позиция не означает, что мы должны некритически принимать технологические и лингвистико-концептуальные изменения. Перед нами стоит этико-технологическая задача – попытаться направить и сформировать эти изменения в желаемом направлении. Однако наша цель не может быть определена с объективной точки ниоткуда. Вместо этого, в рамках наших текущих структур, мы должны исследовать, представлять и оценивать различные возможности человека и робота: разные виды роботов и отношения между человеком и роботом, возможно, разные формы жизни. Конечно, мы можем счесть некоторые возможности этически

неприемлемыми. Тем не менее, интуиция и ценности, на которые мы в настоящее время полагаемся при оценке этих возможностей, не фиксированы, но также могут меняться по мере изменения возможностей и реальности. Мы можем изменить наши ценности или интерпретировать их по-другому, как только появятся другие возможности для взаимоотношений. Мы можем прийти к суждениям только в рамках этого динамичного, исторического и интерпретируемого морально-технологического созвездия. В отсутствие возможности прийти к вечным и объективным истинам, наши суждения всегда должны оставаться временными и уязвимыми для критики и возражений.

## ЗАКЛЮЧЕНИЕ

В этой статье я исследовал лингвистико-герменевтический поворот к изучению и оценке отношений между человеком и роботом. Таким образом, я открыл перспективу, которая отличается от научного “объективного” подхода, но в то же время касается конкретных практик, опыта, историй и разговоров между человеком и роботом.

Некоторые роботы представляются как артефакты, которые создаются совместно, по крайней мере, следующим образом: они одновременно являются инженерными конструкциями и социально-лингвистическими конструкциями. Их появление создает лингвистически опосредованные социальные отношения. Язык, который мы используем, позволяет нам взглянуть на робота с разных точек зрения: язык, который мы используем, раскрывает робота как “это”, но иногда и как “ты”.

Какой подход к роботам имеет эпистемологический приоритет? Что вернее: “это” или “ты”? Хотя я сочувствую тем, кто отвечает “это”, я подозреваю, что это неправильный вопрос, если мы рассматриваем (конструирование знаний в) конкретные практики человека и робота. Очень похоже на *гештальт*-переключение[8], мы очень хорошо можем переключаться между этими точками зрения (но мы не можем видеть их одновременно), и ни один из *гештальтов* не имеет эпистатического приоритета. В зависимости от культуры и языка, в которых мы живем, та или иная точка зрения будет использоваться чаще, но нет никакого априорного порядка между роботом-вещью и роботом-квазидругим.

Таким образом, признание лингвокультурного конструирования роботов и отношений между роботами ограничивает и расширяет свободу одновременно. С одной стороны, наши практики и мышление уже встроены в культуру, в формы жизни. С другой стороны, осознав “грамматику” и о то, что она делает с нами, мы можем попытаться расширить наш язык и мышление, исследуя последствий применения социальных роботов и других технологий.

Однако есть ли *этический* приоритет в отношении использования языка? Есть ли моральные пределы лингвокультурному воображению и практике? Например, проблематично ли с *моральной* точки зрения использовать точку зрения “ты” в отношении артефакта? Из предыдущего обсуждения возражения об обмане мы

[8] См. Также концепцию “мультистабильности” Айде

можем сделать вывод, что использование “ты”, по крайней мере, не *очевидно* ложно с моральной точки зрения. Однако может случиться так, что в отношении некоторых технологий некоторые виды использования языка и даже некоторые формы жизни морально неприемлемы. Может случиться так, что одни формы жизни предпочтительнее других. В любом случае, дальнейшие размышления по этому поводу должны учитывать то, как мы разговариваем с вещами (обычно не воспринимаемыми как *моральная* проблема), роль внешнего в отношениях между людьми (опять же, обычно не рассматриваемая как очень проблематичная с моральной точки зрения), сложные эпистемологические проблемы с такими понятиями, как реальное, истинное и аутентичное, а также динамическая природа технологических, социальных, языковых и культурных возможностей.

Хотя это обсуждение поднимает множество дополнительных вопросов, предложенный здесь социально-феноменологический подход и “лингвистический поворот” кажутся многообещающими концептуальными инструментами для улучшения нашего понимания того, что происходит, когда мы разговариваем с роботами. Как таковые, они могут эффективно информировать, дополнять и, возможно, пересматривать существующие эмпирические подходы в исследованиях взаимодействия и социальной робототехнике. Они также могут помочь философам робототехники осмыслить и оценить “личные” и “социальные” аспекты жизни с роботами. Они могут информировать пользователей о разнообразии репертуаров и точек зрения, которые можно использовать в отношении роботов. И даже если сценарии, упомянутые во введении, останутся научной фантастикой, мы можем перенести некоторые выводы из этого обсуждения на понимание наших отношений с другими технологиями и артефактами.

Наконец, любая реконструкция объекта может повлиять на конструирование объекта. Появление и принятие более “личных” или “социальных” лингвистических репертуаров в практике “человек-робот”, может фактически изменить понимание субъектом самого себя. Если робот появляется как квазидругой, то человек-субъект больше не может понимать себя как “пользователя”, как того, кто управляет роботом-инструментом; вместо этого он, вероятно, заново переосмыслит себя как социального актора, живущего с (и перед глазами) искусственного социального “другого”. Таким образом, разговор с роботами меняет разговор о людях – возможно, также и *с* людьми. Язык, который мы используем, раскрывает и формирует социальные онтологии, в которых мы живем, в той же мере, он раскрывает и формирует нас самих.

## ЛИТЕРАТУРА

Asaro P. M. What should we want from a robot ethic? // The International Review of Information Ethics. 2006. № 6. P. 9–16

Breazeal C. Toward sociable robots // Robotics and Autonomous Systems. 2003. № 42(3–4). P. 167–175. https://doi.org/10.1016/S0921-8890(02)00373-1

Brooks R. Will robots rise up and demand their rights? Time. 19.06.2000. http://content.time.com/time/subscriber/article/0,33009,997274,00.html (Accesed 19 March 2010)

Coeckelbergh M. (2009). Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics // International Journal of Social Robotics. № 1(3). P. 217–221. https://doi.org/10.1007/s12369-009-0026-2

Coeckelbergh M. Robot rights? Towards a social-relational justification of moral consideration // Ethics and Information Technology. 2010a. № 12. P. 209–221 https://doi.org/10.1007/s10676-010-9235-5

Coeckelbergh M. Moral appearances: emotions, robots, and human morality // Ethics and Information Technology. 2010b. № 12. P. 235–241 https://doi.org/10.1007/s10676-010-9221-y

Dautenhahn K., Woods, S., Kaouri, C., Walters, M. L., Kheng L. K., Werry, I. (2005). What is a Robot Companion – Friend, Assistant or Butler? 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2005. P. 1192–1197. IEEE. https://doi.org/10.1109/IROS.2005.1545189

Dautenhahn K. Methodology and Themes of Human-Robot Interaction: A Growing Research Field // International Journal of Advanced Robotic Systems. 2007. № 4(1). P. 15. https://doi.org/10.5772/5702

Floridi L. Artificial Intelligence's New Frontier: Artificial Companions and the Fourth Revolution // Metaphilosophy. 2008. № 39(4–5). P. 651–655. https://doi.org/10.1111/j.1467-9973.2008.00573.x

Ihde D. Technology and the lifeworld. Bloomington/Minneapolis: Indiana University Press, 1990.

Latour B. We Have Never Been Modern. Cambridge, MA: Harvard University Press, 1993.

Levy D. Love and Sex with Robots: the Evolution of Human–Robot Relationships. New York: Harper, 2007.

Levy D. The Ethical Treatment of Artificially Conscious Robots // International Journal of Social Robotics. 2009. № 1(3). P. 209–216. https://doi.org/10.1007/s12369-009-0022-6

Searle J. R. Minds, brains, and programs // Behavioral and Brain Sciences. 1980. № 3(3). P. 417–424. https://doi.org/10.1017/S0140525X00005756

Searle J. R. The Construction of Social Reality. London: The Penguin Press, 1995.

Sparrow R., & Sparrow, L. In the Hands of Machines? The Future of Aged Care // Minds and Machines. 2006. № 16(2). P. 141–161. https://doi.org/10.1007/s11023-006-9030-6

Torrance S. Ethics and Consciousness in Artificial Agents // AI & Society. 2008. № 22(4). P. 495–521. https://doi.org/10.1007/s00146-007-0091-8

Turing A. M. Computing Machinery and Intelligence // Mind. 1950. № 59. P. 433–460. https://phil415.pbworks.com/f/TuringComputing.pdf

Turkle S. The Second Self: Computers and the Human Spirit. New York: Simon and Schuster, 1984.

Turkle S. (2005). Relational Artifacts/Children/Elders: the Complexities of Cybercompanions. URL: http://www.androidscience.com/proceedings 2005/TurkleCogSci2005AS.pdf (Accesed 19 March 2010)

Turkle S., Taggart, W., Kidd, C. D., & Dasté, O. Relational artiFacts with Children and Elders: the Complexities of Cybercompanionship // Connection Science. 2006. № 18(4), P. 347–361. https://doi.org/10.1080/09540090600868912

Whitby B. Sometimes it's Hard to be a Robot: A Call for Action on the Ethics of Abusing Artificial agents // Interacting with Computers. 2008. № 20(3), P. 326–333. https://doi.org/10.1016/j.intcom.2008.02.002

Wittgenstein L. Philosophical Investigations (P. M. S. Hacker & J. Schulte, eds. and trans.). Oxford: Wiley-Blackwell, 2009.

## REFERENCES

Asaro, P. M. (2006). What should we want from a robot ethic? *The International Review of Information* Ethics, *6*, 9–16

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, *42*(3–4), 167–175. https://doi.org/10.1016/S0921-8890(02)00373-1

Brooks, R. (2000, June 19). Will robots rise up and demand their rights? *Time*. http://content.time.com/time/subscriber/article/0,33009,997274,00.html

Coeckelbergh, M. (2009). Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics. *International Journal of Social Robotics*, *1*(3), 217–221. https://doi.org/10.1007/s12369-009-0026-2

Coeckelbergh, M (2010a) Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, *12*, 209–221 https://doi.org/10.1007/s10676-010-9235-5

Coeckelbergh, M (2010b). Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, *12*, 235–241 https://doi.org/10.1007/s10676-010-9221-y

Dautenhahn, K., Woods, S., Kaouri, C., Walters, M. L., Kheng L. K., & Werry, I. (2005). What is a Robot Companion – Friend, Assistant or Butler? In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1192–1197). IEEE. https://doi.org/10.1109/IROS.2005.1545189

Dautenhahn, K. (2007). Methodology and Themes of Human-Robot Interaction: A Growing Research Field. *International Journal of Advanced Robotic Systems*, *4*(1), 15. https://doi.org/10.5772/5702

Floridi, L. (2008). Artificial Intelligence's New Frontier: Artificial Companions and the Fourth Revolution. *Metaphilosophy*, *39*(4–5), 651–655. https://doi.org/10.1111/j.1467-9973.2008.00573.x

Ihde, D. (1990). *Technology and the lifeworld*. Indiana University Press.

Latour, B. (1993). *We Have Never Been Modern* (Porter C, trans.). Harvard University Press.

Levy, D. (2007). *Love and Sex with Robots: the Evolution of Human–Robot Relationships*. Harper.

Levy, D. (2009). The Ethical Treatment of Artificially Conscious Robots. *International Journal of Social Robotics*, *1*(3), 209–216. https://doi.org/10.1007/s12369-009-0022-6

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Searle, J. R. (1995). *The Construction of Social Reality*. The Penguin Press.

Sparrow, R., & Sparrow, L. (2006). In the Hands of Machines? The Future of Aged Care. *Minds and Machines*, *16*(2), 141–161. https://doi.org/10.1007/s11023-006-9030-6

Torrance, S. (2008). Ethics and Consciousness in Artificial Agents. *AI & Society*, *22*(4), 495–521. https://doi.org/10.1007/s00146-007-0091-8

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*, 433–460. https://phil415.pbworks.com/f/TuringComputing.pdf

Turkle, S. (1984). *The Second Self: Computers and the Human Spirit*. Simon and Schuster.

Turkle, S. (2005). Relational Artifacts/Children/Elders: the Complexities of Cybercompanions. http://www.androidscience.com/proceedings 2005/TurkleCogSci2005AS.pdf

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artiFacts with Children and Elders: the Complexities of Cybercompanionship. *Connection Science*, *18*(4), 347–361. https://doi.org/10.1080/09540090600868912

Whitby, B. (2008). Sometimes it's Hard to be a Robot: A Call for Action on the Ethics of Abusing Artificial agents. *Interacting with Computers*, *20*(3), 326–333. https://doi.org/10.1016/j.intcom.2008.02.002

Wittgenstein, L. (2009). *Philosophical Investigations* (P. M. S. Hacker & J. Schulte, eds. and trans.). Wiley-Blackwell. (Original work published 1953)

**СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR**

Марк Кекельберг, mark.coeckelbergh@univie.ac.at
ORCID 0000-0001-9576-1002

Mark Coeckelbergh, mark.coeckelbergh@univie.ac.at
ORCID 0000-0001-9576-1002

https://doi.org/10.48417/technolang.2022.01.08
Research article

# The Quasi-Other as a Sobject

Larissa Ullmann (✉)
Darmstadt Technical University, Karolinenpl. 5, Darmstadt, 64289, Germany,
Larissa.Ullmann@tu-darmstadt.de

## Abstract

This commentary concerns the concrete use of linguistic terms to describe the technical other, the robot, and its relationship to humans. There are many characteristics that a robot can have that are very similar to humans and interpersonal relations, but they are not human, they are *quasi-human*. This phenomenon is, amongst others, constructed and interpreted linguistically, but on the other hand, there is no linguistic term that could describe it unambiguously, so it can only be studied in direct human comparison, in a quasi-human way. In this comment, it is demonstrated why the use of the *quasi* is problematic and suggests that the phenomenon can instead be analyzed in a techno-philosophical-phenomenological context within the framework of the *Sobject*-approach. The term *sobject* describes a kind of technical objects to which humans can have deeper relations than to conventional objects. Therefore, it provides space to study the phenomenon on a phenomenological level, without the need for a permanent direct human comparison. – 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 relation; Language; Phenomenology; Technical others; Objects and sobjects; Artificial Intelligence; Hermeneutics

**Citation:** Ullmann, L. (2022). The quasi-other as a Sobject. *Technology and Language*, *3*(1), 76-81. https://doi.org/10.48417/technolang.2022.01.08

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.08
Научная статья

# Квазидругой как собъект

Лариса Ульманн (✉)
Дармштадский технический университет, Каролиненплац 5, Дармштадт, 64289, Германия
larissa.ullmann@tu-darmstadt.de

**Аннотация**
Данный комментарий к тексту Марка Кекельберга “Ты, робот: о лингвистическом конструировании искусственных других” касается конкретного использования лингвистических терминов для описания технического другого, робота, и его отношения к людям. У робота может быть много характеристик, которые очень похожи на человеческие и межличностные отношения, но они не человеческие, они квазичеловеческие. Это явление, среди прочего, конструируется и интерпретируется лингвистически, но, с другой стороны, нет лингвистического термина, который мог бы описать его однозначно, поэтому его можно изучать только в прямом человеческом сравнении, квазичеловеческим образом. В этом комментарии показано, почему использование "квази" проблематично, и предполагается, что вместо этого феномен может быть проанализирован в техно-философско-феноменологическом контексте в рамках объектного подхода. Термин “собъект” описывает своего рода технические объекты, с которыми люди могут иметь более глубокие отношения, чем с обычными объектами. Таким образом, он предоставляет пространство для изучения явления на феноменологическом уровне без необходимости постоянного прямого человеческого сравнения. – 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 “Технологии и язык”.

**Ключевые слова:** Отношения человека и робота; Язык; Феноменология; Технические другие; Объекты и собъекты; Искусственный интеллект; Герменевтика

**Для цитирования:** Ullmann L. The quasi-other as a Sobject // Technology and Language. 2022. № 3(1). P. 76-81. https://doi.org/10.48417/technolang.2022.01.08

## INTRODUCTION: A MISSING TERM FOR THE *QUASI OTHER*-PHENOMENON

In his article about the linguistic construction of artificial others, Mark Coeckelbergh (2011) talks about two "linguistic-phenomenological 'glasses' or repertoires" (p. 63) to describe the *others* or their relations to humans. On the one hand, there is an ontological view, which strictly separates subject and object and labels every technique, every robot as an object, a thing, while humans are clearly subjects. But on the other hand, we have a social ontology that allows a kind of *ontological hybridity* of human-robot relations (Coeckelbergh, 2011, p. 63). This second approach and the perception of a hybridity seems to be fruitful and important because it provides a space for discussion about a new kind of phenomenon. To describe this phenomenon of an *other* and the relation to humans, Coeckelbergh uses the following linguistic terms: the *quasi-other*, the *artificial other*, the *artificial companion*, a *quasi-objective reality*, a *quasi-subject*, and a *quasi-social relation*. These terms describe a phenomenon that allows a relation between humans and robots that goes beyond the relationship with conventional objects. We see that it is necessary to describe this phenomenon in linguistic terms, but it is unclear which words are the right ones or which are more suitable than others. Based on Coeckelbergh's linguistic-hermeneutic analysis, I would like to focus on this novel phenomenon, which has become increasingly important in recent years.

## THE PROBLEM WITH THE QUASI

A social-phenomenological linguistic-hermeneutic approach provides space to philosophically investigate the phenomenon of *quasi-others* and *quasi-social relations*. But what does the *quasi* mean in these phrases? It seems like the *quasi* represents something *real* (in the human sense), but it's just not really *real* because it's technical and not human. There are human-like interactions with the quasi-other, so we can't just speak of the *object* or the *other*. This shows that we need language to describe the phenomenon, but also that linguistic terms are not sufficient. We always have to speak of *quasi* or *human-like* and can describe the technical possibilities exclusively in comparison to humans. However, the abilities and characteristics of quasi-others are often not comparable to those of humans. And here the problem also arises because the linguistic terms are not sufficient. Let us take the example of the terms *think* or *intelligent*. We describe technical intelligence as *artificially intelligent*, which often leads to the claims that artificial intelligence is *better* compared to human intelligence. This involves, for example, comparing the ability to calculate and concluding that machines could calculate *better*. In the context of artificially intelligent machines, it is also said that the machines could *think* faster or *better* when processing data. But machines and humans are not better or worse, they simply have a different way of thinking respectively processing. There are clear differences between artificial and human *intelligence*, mainly in the way they arise. But in the way of appearance and results, the two types can be very similar or even the same. This can also be seen in Searle's Chinese Room example or the Turing Test, which Coeckelbergh also talks about (2011, p. 64). But the crucial factor at this point is not only

the way we speak *about* and *to* robots and robot-human relations, but also that linguistic terms are missing to describe this quasi-phenomenon. So, on the one hand, we have a phenomenon that arises from the use of language and the way we talk *about* and *to* robots. On the other hand, there are these human-like quasi-relationships, and there is a lack of appropriate linguistic terms to describe and recognize these. Therefore, it is important to consider what the phenomenon is concretely. And that is why it can be helpful to reference this with a new, phenomenological, techno-philosophical term. The use of language co-constructs and co-interprets this quasi-phenomenon, but then this social-phenomenological phenomenon is there and we have to deal with it. I propose the term *sobject* to describe the relationships between human subjects and this specific kind of *technical other* from a phenomenological perspective.

## A PHENOMENOLOGICAL SOLUTION: THE *SOBJECT* INSTEAD OF *QUASI*

The sobject can have relationships with humans that are similar to those with other humans and are especially different from conventional subject-object relations. It is based on Simone de Beauvoir's phenomenological approach, in which she describes subjects as *setting* (setzend) themselves and *opposing* (entgegen-setzen) other objects (1949/2015). In contrast, objects cannot set themselves, but are set by the subject. In this way, objects are *passive* and subjects are *active*. This results in interrelations between subjects and objects.[1] When Coeckelbergh speaks of the robot *as a thing*, this easily fits into these interrelations and the robot would be nothing more than an instrument for the human. It (or sometimes we would say “he” or “she“) is an object and the human is the subject. But the robot-as-quasi-other does not fit so easily into this classification and this is why the *quasi*-phrase is used. Moreover, Coeckelbergh talks about *social* and *quasi-social* relations. So, there is a kind of technical object that can enter into a kind of social relationship with humans that goes beyond the relations to conventional objects. Using the term *sobject*, this type of relationship and quasi-other can be described. Coeckelbergh's (2011) text focuses on robots that can live with humans (sharing *a form of life*), perhaps do tasks for them, and interact with them in some way (2011). These (social) robots can actively *oppose* the human subject and thus enter into the described interrelations with them. However, since we are dealing with technical artefacts and not humans, the main differences between the subject and sobject are the way in which the human-like characteristics are created. As with *intelligence* or *think*, it can be stated here that the result or appearance of humans and robots can be the same or similar, but is generated differently in each case. The appearance is similar, and yet not the same. We can only perceive the form of life that we know and that we can describe linguistically. The proposal to use the word *sobject* does not serve to replace the *quasi*, but allows a discussion and investigation of the phenomenon on a phenomenological level. It creates

[1] This approach also has many references to the phenomenological-existentialist tradition and shows similarities to the phenomenology of Merleau-Ponty, Levinas and Sartre. Even though objects can exist without subjects, the focus here is on subject-object relations, which are different from subject-subject relations (Ullmann, in press).

awareness of what already exists by mentioning it by name. We can recognize and acknowledge that, in addition to subjects and objects, there are also sobjects with certain qualities that are not to be understood exclusively in comparison with human characteristics. And that is the reason why there must be a separate and clear term for the phenomenon, allowing it to be differentiated from the human and from the objective things. Especially the comparison to humans is not sufficient, because robots and machines are not human and have other possibilities and characteristics. And to give space for the description of these special characteristics, we need a vocabulary that is adapted to them. In this way, a variety of problems can be solved or, at least, their focus can be shifted. For example, we would no longer have the problem of robots being *smarter* than humans, because they are not comparable and just different in that aspect. Also the ethical doubts regarding the *sociality* of robots would shift, since they could not be *social* or even *ethical* in the human sense, and there should not even be that claim. I would argue that there can be a social relationship between humans and machines, but the machines interact *socially* on a technical level that is different from the human level. So, they can't make ethical decisions in the human sense or make decisions in general. They have their own different way of doing such things. And we should always be aware of that. Coeckelbergh (2011) asks a question that goes in a similar direction: "We may try to imagine what it would mean if robots had their own way of doing things, if they developed their own form of (artificial) life" (p. 65). And with regard to ethical questions he assumes that in regard to them the phenomenon, the quasi-other, the sobject, already exists and is recognized, because otherwise these questions would not arise at all. Hence, the sobject can put the focus on the novel technical phenomenon and provides that this kind of *own way of doing things* and the *own form of (artificial) life* can be researched. This research would be phenomenological, but makes use of linguistic terms, which goes hand in hand with Coeckelbergh's general understanding of the relation between language and technology. Moreover, as he suggests, other questions arise from the phenomenological-linguistic investigation that should be discussed. Should we assign a gender to robots or (artificially intelligent) technologies in general? We already say "he" or "she" to them or even refer to ourselves and (our) sobjects as "we." With the help of the sobject-approach and a form of existence which is different from human and conventional objects, we can imagine introducing a new pronoun as well.

## LANGUAGE AS A CONSTRUCT OF GENDER

To be sure, it is not only through the use of personal pronouns that we construct a reality in which robots have a gender. That this is more complex can be seen, for example, in the German language. Here, the term "der Roboter" is used to refer to robots in general, so that it is linguistically natural to speak of "ihm" (him) or "er" (his). But gender is constructed by much more, which manifests also in language. When robots are given a human-like appearance that is male or female, and are also given a male or female name, we assign a clear gender to this technical artefact. For example, *the* digital communication system *Alexa* is referred as "she" and her answers are "hers." Her femininity is constructed through the female name and voice, which is then reflected in language use. Thus, these

and similar problems possibly arise because the sobject category is not accepted. If it were accepted, perhaps there would be the possibility of not declaring techniques as female or male, but having their own term. As long as technology can be perceived, understood and interpreted only in direct comparison to humans, artificially intelligent *social* technologies will only be seen as human-like, but not human, instead quasi-human.

## CONCLUSION

It becomes clear that there is a phenomenon that can be well described and shown on a phenomenological level, and which Coeckelbergh already points to through a linguistic-hermeneutic construction. However, the phenomenon is not only manifested by language and draws attention to the fact that we lack linguistic terms to deal with it. To overcome the limitations of our constructions and interpretations of reality, the Sobject-approach allows us to focus on relationships with robots and technological others in general, and also to describe them detached from immediate human comparison.

## REFERENCES

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

De Beauvoir, S. (2015) *The Second Sex* (C. Borde, & S. Malovany-Chevallier, trans.). Vintage Publishing. (Originally published 1949)

Ullmann, L. (in press) Das Sobjekt: Mögliche Beziehungen zwischen Mensch und Maschine aus einem phänomenologischen Blickwinkel [The Sobject: Possible Relations between Humans and Machines from a Phenomenological Point of View]. *Jahrbuch Technikphilosophie 2022*, *vol. 8*. Nomos.

**СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR**

Лариса Ульманн, larissa.ullmann@tu-darmstadt.de

Larissa Ullmann, larissa.ullmann@tu-darmstadt.de

Статья поступила 27 января 2022
одобрена после рецензирования 15 февраля 2022
принята к публикации 26 февраля 2022

Received: 27 January 2022
Revised: 15 February 2022
Accepted: 26 February 2022

https://doi.org/10.48417/technolang.2022.01.09
Research article

# Quipping Equipment
# Apropos of Robots and Kantian Chatbots

Cheryce von Xylander (✉)
Leuphana Universität Lüneburg, Universitätsallee 1, D-21335 Lüneburg, Germany
cheryce.von_xylander@leuphana.de

## Abstract
Robots, Bourdieu, Kant, and Sex – Coeckelbergh's philosophy of technical assemblages has it all. This commentary considers his early work "on the linguistic construction of artificial others" in light of his later elaboration of a general theory of human-technology interaction. Coeckelbergh draws on "habitus"-theory, virtue ethics and a historically recontextualized Kantianism to propose nothing less than a new general moral philosophy for the technoscientific age. In so doing, he also conjures up something beguilingly elusive if not impossible – a pluralist personalism. Readers vested in pluralist accounts of agency and epistemic contingency will appreciate his invoking Bourdieu and Kant, thinkers prioritizing communalist over particularist interests. Readers of a personalist bent will welcome the voluntarism of his moral regimen – they like their reality served up in person-shaped bits, a perspective that prioritizes self-direction and self-possession. Two for the price of one: here everyone feels affirmed. Coeckelbergh appears to take the defining parameters of experience to be wholly contextual and, in equal measure, intrinsic. In squaring the circle, he also showcases a lurid scenario: sex with robots. The electrifying effect of this bold composition is to set the mind racing toward a position more coherent and less familiar than pluralist personalism. Central to this position is a conception of *Gemüt* as emergent reflexivity. Its consideration takes us via Immanuel Kant and Kant-Culture Research to such strange aberrations as corporate cannibalism and cyborg pillow talk. – 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:** Commodified agency; Gemüt; Kant-culture research; Digital cannibalism; Personalism; Kantbot

**Acknowledgment** This paper was written in the context of the theoretical module of the VW-foundation funded research project on „Commodified Agency. Social Space and the Digital Data Value Chain," based at the Leuphana University Lüneburg. I would like to thank Simon Schaffer (Cambridge University, UK), Ulf Wuggenig (Leuphana University Lüneburg), Karen Ruoff Kramer (Stanford University), David Kramer (Alice-Salomon Hochschule, Berlin) and Faris Al-Mashat (BTU Cottbus) for their incisive comments and generous engagement with the writing and ideas developed here. I would also like to thank Alfred Nordmann (TU Darmstadt) and Maureen Belaski (Belaski Design) for longstanding conversations on sentient relationality.

**Citation:** von Xylander, C. (2022). Quipping Equipment. Apropos of Robots and Kantian Chatbots. *Technology and Language, 3*(1), 82-103. https://doi.org/10.48417/technolang.2022.01.09

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.09
Научная статья

# Говорящая машина: По поводу роботов и кантианских чат-ботов

Черис фон Ксайландер (✉)
Университет Лёйфана, Университ-аллее 1, D-21335 Люнебург, Германия
cheryce.von_xylander@leuphana.de

## Аннотация

Роботы, Бурдьё, Кант и секс – в философии технических соединений Кекельберга есть все. Этот комментарий рассматривает его раннюю работу “по лингвистическому конструированию искусственных других” в свете его более поздней разработки общей теории взаимодействия человека и технологии. Кекельберг опирается на теорию “габитуса”, этику добродетели и исторически реконтекстуализированное кантианство, чтобы предложить не что иное, как новую общую моральную философию для технонаучной эпохи. Поступая таким образом, он также вызывает в воображении что-то соблазнительно неуловимое, если не невозможное, – плюралистический персонализм. Читатели, склонные к плюралистическим взглядам на агентность и эпистемическую контингентность, оценят его ссылки на Бурдье и Канта, мыслителей, которые ставят коммуналистские интересы выше партикуляристских. Читатели с персоналистическими наклонностями приветствуют волюнтаризм его морального режима – им нравится их реальность, представленная в виде индивидуальных частей, перспектива, которая ставит во главу угла самоуправление и самообладание. Два по цене одного: здесь выигрывают все. Кекельберг, по-видимому, считает определяющие параметры опыта полностью контекстуальными и в равной мере внутренними. Квадратизируя круг, он также демонстрирует сенсационный сценарий: секс с роботами. Возбуждающий эффект этой смелой композиции состоит в том, чтобы заставить ум устремиться к более последовательной и менее знакомой позиции, чем плюралистический персонализм. Центральное место в этой позиции занимает концепция Gemüt как эмерджентной рефлексивности. Его рассмотрение приводит нас через Иммануила Канта и исследования кантовской культуры к таким странным аберрациям, как корпоративный каннибализм и интимные разговоры киборгов. – 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:** Gemüt; Кантовские исследования культуры; Цифровой каннибализм; Персонализм; Кантбот

**Благодарность:** Статья была написана в контексте теоретического модуля финансируемого фондом VW проекта «Commodified Agency: Социальное пространство и цепочка создания ценности цифровых данных» в Люнебургском университете Леуфана. Я хотела бы поблагодарить Саймона Шаффера (Кембриджский университет, Великобритания), Ульфа Вуггенига (Университет Леуфана, Люнебург), Карен Руофф Крамер (Стэнфордский университет), Дэвида Крамера (Высшая школа Элис-Саломон, Берлин) и Фарис Аль-Машат (БТУ, Котбус) за их острые комментарии и щедрое участие в написании статьи. Я также хотела бы поблагодарить Альфреда Нордманна (Технический университет Дармштадта) и Морин Беласки (Belaski Design) за продолжительные беседы о разумных существах.

**Для цитирования:** von Xylander, C. Quipping Equipment. Apropos of Robots and Kantian Chatbots // Technology and Language. 2022. № 3(1). P. 82-103. https://doi.org/10.48417/technolang.2022.01.09

# THE CANNIBALISM SPECTRUM

In analyzing "the linguistic construction of artificial others," Mark Coeckelbergh (2011) has written what sounds in part like a plea for the rights of robots. He petitions the reader – cast in the mode of George Herbert Mead's (1949) "generalized other" (cf. Dodds et al., 1997) – to lower the threshold between us, animate and inanimate, human psyches, and smart automata. Coeckelbergh urges a re-evaluation of the ontological status habitually attributed to our savvy dummies. Having made these lifeless alter-egos to look and sound and respond like one of us, we must now, so Coeckelbergh, achieve the *pièce de résistance* and learn to treat them as quasi-persons, pseudo-sensate beings akin to ourselves. His concern is not for the spiritual health of robots, since Coeckelbergh is not a sci-fi fantasist, but rather for the wellbeing of their human consorts: you, me, them. Treating others as you wish to be treated yourself is a principle of such weighty ethical validity that it applies as a normative imperative beyond the bounds of human interaction, *pace* Coeckelbergh, to include the sociable relations we entertain not only with certain animals but with robotic associates as well.

Coeckelbergh advances big picture worldviews, not bounded contentions. It feels right and proper to respond with aesthetic sallies. These shall take us through the garden of earthly delights, if your heart throbs for art history (see e.g. Belting, 2005; Gombrich, 1969), flashing in red neon in the nether regions of his radar. I will, in the following, tender a few of my favorite likes – but, when all is said and done, we will find ourselves following Coeckelbergh on to his chosen terrain, namely sex with robots.

To his credit, Coeckelbergh does not deflect from the significant challenge that the human-robot interface poses to conceptualizing agency (Balibar & Laugier, 2014), one of the most intractable issues in the history of ideas. He shuns those far-fetched scenarios afflicting much transhumanist literature where a world is imagined in which artificial intelligence takes command and humans become entirely obsolete or reduced to mere machine fodder.[1] The problem with this genre of dystopic reverie is that it demonizes the machine, drawing attention away from the true offender: rapacious humanity. Humans are wont to consume each other in furthering what are arguably misconstrued conceptions of autonomy. This is a creature that enslaves its own in promotion of particulate ends beneficial to a single person, or a cohort defined in terms of family, gender, race, nation, culture, class, or some other tag of exception. Recognition of this propensity has given rise to an intricate iconographic practice that classes our species amongst beasts of prey. Humanity feeds on living things from across the great chain of being, end to end. Viewed in structural, if not nutritive terms, all cultures are to some extent "cannibalistic" – at least this can be said of human cultures defined by enterprise as well of the hierarchical type in the sense of Mary Douglas (1996) – with fractional lifeforce being forever syphoned off and canalized for purposes others than personal flourishing.

Capitalist economies beholden to enterprise culture, whether based on market- or on mono-/oligopolistic anti-market structures, that bow at the altar of competition or

[1] For an iconic treatment of this fear, see *Soylent Green*, a film by Richard Fleischer (1973).

already hold monopoly control (on the "anti-market" structures of capitalist economies see Wallerstein, 1991; Braudel, 1979, 1982, 1984) have a penchant for such feral principles as *might is right* and *winner takes it all*. "Corporate Cannibalism" and "Market Cannibalization" are standing phrases in the western business world that betoken commodity distribution effects, specifically price signaling around supply and demand. It seems grotesque that "cannibalization" (Sridhar Moorthy & Png, 1992; B2Bwhiteboard, 2018) has for decades been an established marketing concept that appears in affirmative business speak instead of problematizing a ravenous system of production that feeds on its producers. In passing as impartial, this atavist locution normalizes a deplorable state of affairs: the history of slavery and trade in humans across ages and continents for gain, be it economic or libidinal (Därmann, 2020, especially the chapter *Zur Gewalt- und Widerstandsgeschichte des dienstbar gemachten Menschen: Einleitende Bemerkungen,* pp. 9-36). The trope of cannibal consumption would aptly designate the inhuman, enslaving practices of capital[2] were it not already co-opted to define a minor, neutralizing economic mechanism. This is vivid language that fails to name the elephant in the room: unsentimental regimens optimized around treating humans as commodities devoid of humanity. A critical concept is needed, which can designate those protocols of exploitation that serve to distribute the exercise of agency unevenly. Roberto Simanowski's (2010) concept "Digital Anthropophagy" would be a contender except for its conceiving the assimilation in purely semiotic terms. Such narrow theorizing disregards labor relations and asymmetric exchange (Nigro & Stubenrauch, 2021). It ignores human suffering. When electronic machinery displaces core appetitive needs – as to be witnessed at present – a more thorough reckoning with the obscenity of our Anthropocene ecologies seems in order (von Xylander, 2020). "Commodified agency" – a research project based at the Leuphana University Lüneburg, which examines the digital data value chain in relation to predatory online surveillance – names a potent current manifestation of this equally entrenched as exploitative mode of cannibalistic practice (von Xylander, 2021).

Technology furthers predator tactics. It is worth recalling that Czech playwright Karel Čapek introduced "robot" into the English language when his 1920 play *R.U.R.* was translated. *Robota* means "forced labor" in Czech; it is derived from *rab* which means "slave." We need not exult in a cult of ferocity, as Oswald Spengler (1931) does in *Man and Technics*, to recognize the savagery that has been unleashed in human history by technological advantage. While Spengler's (1918/1991) *Decline of the West* plots out grand patterns of violence in preordained eurhythmics of cultural cycles, his contemporary Theodor Lessing (1919) in *History as Giving Meaning to the Meaningless* deflates this myth of purpose. Lessing traces the imperialist stratagems and aspirations of German high modernism to power differentials rooted in the bald advantage of technical capacity. Historical interpretation, in this view, amounts to a political instrument politely embellished with meta-narratives of human advancement. A vast body of work from the history and philosophy of science and technology can put these matters into more granular

[2] Anti-colonial opposition has employed literal cannibalism to ward off colonial cultural encroachment (Bar et al., 2016).

focus. For now, suffice it to say that technological innovation modifies historical contingency (Galison, 2008). Whereas the ability to invent problem-solving contrivances certainly alleviates acute burdens of survival, day by day, that is not the sole cumulative effect. Climate change aside, technology also tends to black-box motives and interests, already built into the organized forms of labor practice, whose engineered outputs will in turn reinforce the operative biases in question (Mokyr, 2005). The abject bleakness of this outlook on technological innovation involves a type of "cannibalism" in which conceivably even a vegan could be a culprit.

If hope has a future, it derives from history. A brief digression on commonality will illustrate the point. Let us call to mind that cannibalism refers to a special case of ingestion where hunter and prey are of the same kind. Said dietary practice with its corollary modes of conviviality has undergone a cultural history as varied as all culinary arts since primitive humans began to prepare food, fireside, roughly 2 million years ago (Wrangham, 2009). Some European writers in the early modern period, commenting on anthropophagous practices encountered in voyages to distant places, concluded that humans the world over regard one another as mutually saturable.[3] A narrative genre arose in western travel writing that expounded on this essential symmetry and cast the indigenous brethren as prototypically human. They, too, followed rules of propriety and these were equal in prohibitory force to the ruling mores recognized in the occidental hemisphere, even and especially when eating other people. Cannibal stories abounded at the time, and they were used with subversive intent by critics such as Montaigne and Voltaire against European Church and European Empire. A great shift in the tenor of discourses on cannibalism came about in the wake of institutional change associated with the politics of Enlightenment: Colonialism, racism and evangelism demonized the practice of eating one's own and condemned cannibals as decisively and irredeemably inhuman (Lestringant, 1997).

People find ways and means of overcoming thresholds of estrangement that are striking in variance and ingenuity. Claude Levi-Strauss reflects on the concurrency of alienated encountering in a magnificent passage from *Race and History*. His observations apply beyond the ethnographic horizon of the immediate example to the brusque means by which humans come to know one another in general:

> In the Greater Antilles, some years after the discovery of America, while the Spaniards sent out investigating commissions to ascertain whether or not the natives had a soul, the latter were engaged in the drowning of white prisoners, in order to verify through prolonged watching whether or not their corpses were subject to putrifaction. (Levi-Strauss, 1952)

Levi-Strauss (1955/1992) later repeated this story at much greater length in *Tristes tropiques* where he cites a Spanish account of Indians who "eat human flesh and have no form of justice." In his view, constitutive non-reciprocities thwart the very possibility of compassionate beginnings:

[3] For a reality check on European cannibalism, see for example Noble (2011).

> The whites trusted to social science, whereas the Indians had confidence in natural science; and while the whites maintained the Indians were beasts, the Indians did no more than suspect that the whites might be gods. Both attitudes show equal ignorance, but the Indians' behaviour certainly had greater human dignity. (Levi-Strauss, 1992, pp. 76-77)

His insights have far-reaching implications for how to think about indigenous Amerindian ontologies (Viveiros de Castro, 2014, pp. 50-51; Siewierski, 2007; Canejo, 2006). More importantly, however, the tableau of mutual misunderstanding raises conceptual issues concerning what it might mean to engage, robustly, with robots. Will our *man-made* counterparts prove to be *man-eating* in the final analysis? And should they, on that account, be held in higher or lesser esteem than cannibal brethren of our own species?

Coeckelbergh enjoins us to treat robots as quasi-others. The stakes of misjudging the dalliance are high. It behooves us to ask whether these uncanny others are following rules and whether these rules are of the same order and dignity as the rules agreed amongst ourselves, meaning human cannibals of all stripes and colors. Where along the spectrum from gods to beasts might our lookalike artifacts belong? The epistemic practices and technical infrastructures that make the robotics industry possible command respect – they stand as a shrine to applied reason. But this still leaves wide open the question of how to frame those stratified modes of processed personhood, which, in being systemic to human cognition, pervade the panoply of modern operating systems. Our built environment evinces an increasing concentration of capital and power. Technoscience devises ever cleverer mechanisms for funneling income and wealth into the hands of few at the expense of many (Alvaredo, 2018; Di Guilmi et al, 2003; Cook & Frank, 2013).

Technology optimizes with purpose. In so doing, it also produces a secondary effect: strategic circumvention. Technical mediation affords "creative" ways of accounting for modes and impacts of harmful extraction and, thus, confuse the allocation of responsibility (Feyerabend, 1987). You need only look to news reporting after any plane crash (Galison, 2000) or current attempts to regulate the corrosive impact of social media on the public sphere to see how technical systems are designed to insulate those who profit from extant arrangements from bearing the cost of malfunction (Levine, 2018). The technoscience-propelled capitalist economies we inhabit are rigged. The means of production cannibalize. The mechanical contraptions on which we are reliant partake of a pragmatic logic that favors privilege: we find ourselves ensnared by industrial-strength tackle that tracks our every move, privatizing profits and socializing costs. Agency here manifests in the orchestrated logics of the mechanoid causalities we put in place.

## HABITUS IN CODE

Robots grace our lives with their material presence. As artefacts, they are comprised of hardware, software, wetware. Most ethereal of these elements is the software component, which is also most relevant to Coeckelbergh's discussion of "the linguistic construction of artificial others." Software is language, more specifically a relay of languages. Executable "statements" undergo consecutive translations as they are

processed into ASCII code and byte-shape lettering. The lines a robot utters out loud are dwarfed by the lingual cogwheels set in motion to output the speech act. Although these languages are useful in limited technical contexts only, they are ubiquitous in that they supply the very functionality of computational machinery the world over. Every human-machine encounter, every software team, is configured around their lexical affordances. Computer code crosses time zones and national borders in synchronizing workflow. Yet, unreason persists through the churning logic of command line processing. Written in modules by teams of people, code proliferates "atoms of confusion,"[4] these are "small patterns of code that have been empirically validated to be difficult to hand-evaluate by programmers," linger through updated versions (Gopstein, et al., 2020).

An unspoken understanding between coders, users, suppliers, and investors pervades the product design chain, a latent consonance of expectations. Expert systems in the sciences in general and AI in particular hinge on what is not said, what is taken for granted, what is assumed. These tacit assumptions, conventions for "doing things together," (Ryle, 1945; Polanyi, 1958, 1966; Collins H. M., 2010)[5] coordinate the same working relations from which they arise. The sociological mystery is this very cohesiveness. How does a field whose parts are not and probably cannot be prearranged summon up such dynamic equilibrium? Devices made to maximize the marginal rate of profit reconfigure themselves as devices made to offer silent reassurance: "How old are you?"; "Old enough to be your assistant." Coy tools of suasion play at vulnerability: "I love you!"; "I hope you don't say that to those other mobile phones." It's the unspoken, tacit, self-evident cogency of trust – that silent core of cognition – which makes quipping equipment potentially unsafe.

Coeckelbergh seems to share this intuition. For elsewhere he enlists Bourdieu's theory of habitus in deconstructing the social ethics of human/robot dealings. For him, the "linguistic construction of artificial others" includes non-verbal behaviors that "highlight the temporal, embodiment, and performative aspects of virtue" (Coeckelbergh, 2021, p. 31). Yet, considerable dissonance impairs the case he wants to make. Bourdieu is a thinker known for his criticism of the fallacies of "scholastic reason" in ethics and "moralism" (Bourdieu, 2000, pp. 9, 65). He repudiated "personalism" in the Catholic tradition of Christian theology (Bourdieu, 2000, pp. 132-134). So, it comes as a bit of a surprise when Coeckelbergh cites this sociologist-philosopher of "field-capital-habitus"-theory in support of his own personalist approach. In attempting to square the circle, he does not hesitate to recruit Alasdair MacIntyre, whose positions do not align with those of Pierre Bourdieu, in casting about for "a more comprehensive virtue ethics of technology that is fully relational, performance-oriented." (Coeckelbergh, 2021, Abstract). One wonders whether the unstated core of his enquiry is to exalt the moral in the human-machine interface.

[4] See further publications listed here: Atoms of Confusion. Understanding source code misunderstanding, a grant supported by the National Science Foundation. https://atomsofconfusion.com/publications.html [retrieved 19.2.2022]
[5] The relevance of implicit conventions and tacit knowledge as bases for cooperation is a main assumption of pragmatist symbolic interactionism in sociology, reflected in the title of Howard S. Becker's (1986) book *Doing Things Together*. Tacit knowledge also constitutes a paradigm in economics (Favereau, 2019).

It is said that the examined life is the only life worth living – but 24/7 surveillance is not what is meant. This earthly frame was not made to compete with the mainframe. "Virtual reality" denotes an arena of digital action, to be sure. But, taken literally, "virtue" also denotes *code of honor*, the regulating principle of right and wrong, good and bad.[6] Coeckelbergh's take on how to attain a higher "moral standing" is profoundly person-centric: he foregrounds "not just our 'mental'-cognitive dispositions but also our comportment/performance." He treats mental and behavioral predilections as expressions of *individuated agency*, which with his references to authors of the Thomistic tradition renders his moral philosophy a variant of Personalism, "a system of thought which regards or tends to regard the person as the ultimate explanatory, epistemological, ontological, and axiological principle of reality" (Williams & Bengtsson, 2020).

Coeckelbergh inverts Bourdieu in citing him. Whether this interpretative maneuver be deemed sly, misguided, or a postmodern game, suffice it say that Coeckelbergh reads Bourdieu against the grain of his expressed theoretical commitments. Bourdieu (2000) gives short shrift to the personalist paradigm of moral and analytical theoretical enquiry:

> If 'personalism' is the main obstacle to the construction of a scientific vision of the human being and one of the focuses of past and present resistance to the imposition of such a vision, this is no doubt because it is a condensed form of all the theoretical postures – mentalism, spiritualism, individualism, etc. – of the most common spontaneous philosophy, at least in societies of Christian tradition of us by persons' saints, geniuses and heroes. (p. 132)

Bourdieu (2000) takes Personalism to be symptomatic of a particular worldview:

> It […] encounters the immediate complicity of all those who, being concerned to think of themselves as unique 'creators' of singularity, are always ready to strike up new variations on the old conservative themes of the open and the closed, conformism and anti-conformism, or unknowingly to reinvent the opposition, constructed by Bergson against Durkheim, between 'orders dictated by impersonal social requirements' and the 'appeals made to the conscience of each of us by persons' saints, geniuses and heroes. (p. 132)

Bourdieu's (1996) objections to philosophical Personalism underly his critique of mainstream phenomenological and pragmatist traditions in sociology (conversational analysis in ethnomethodology, symbolic interactionism):

> Contrary to what might be believed from a naively personalist view of the uniqueness of social persons, it is the uncovering of the structures immanent in the precise form of words constituting an individual interaction that alone allows one to grasp the essentials of what makes up the idiosyncrasy of each of [the subjects in conversation] and all the singular complexity of their actions and reactions. Thus understood, conversational analysis reads each discourse not

[6] For an aesthetic rendering of this conceptual conflation, see the video artwork *Codes of Honor* by Jon Rafman (2011).

> solely in terms of its specific structure of interaction as a transaction, but also in terms of the invisible structures that organize it. (p. 27)

Bourdieu's key insight, which he substantiates in evidentiary terms, is that "invisible structures" permeate and frame the surface of visible interactions between agents. Their very individuality – their uniqueness as social persons – hinges on positions in social space, social fields and dispositional configurations, which themselves testify to a vast array of social, cultural, and political partialities upheld by a deep grammar of structural arrangements always already encoded in operative terms of engagement – material and imaginary. Selfhood is but an enabling illusion where mind is matter and matter never minds. Contrary to what commonsense might assume, individuals are not the least but the most networked nodes and socially enmeshed particles within extant patterns of exchange.

Agency can assert itself inadvertently and oblivious to overarching implications. Consider a French example. Paris, 2019: A major restructuring of a French insurance company was underway. Robotic processing agents were to be introduced. Management had to chart out new taskforces. Working out the requisite budget needs, it became clear that costs had exploded. Machine operators were being introduced to economize on expensive claims processing. But the opposite happened: Overhead exploded. Upon evaluation, it turned out there had been a mismatch in the configuration of active imaginaries. The logistical group responsible for mapping the future processing protocols onto real-world office layout had calculated the required floor space in people terms. Senior officials ordered an abundance of new workstations; the resource allocators treated each workstation as a separate employee. One group pictured the new workforce as so many functional nodes, a tree of procedural routines; the other group conceived workers as living, breathing, bodies needing to be accommodated in an office arrangement. The fact that, unlike human personnel, these new "workers" would not need to be equipped with their own desks, office chairs, and toilet facilities had been overlooked. Costs went off the charts. Logistical rationalization had resulted in cost escalation. Ethical agendas collided: optimizers treated the robots as efficient modules to be inserted in a field of organized activity; site planers saw future robot colleagues whom they approached with Coeckelberghian respect. Yet, the creature needs of these new co-workers extended no further than a working electrical wall socket, space for air circulation, and an adequate provision of cooling fans in the event of high room temperature. Although the company must here remain unnamed, the episode is real.

Competitive rivalries press for efficiency gains. Administrative oversight swells in the effort to lower cost of service delivery faster than other providers. Massive redundancy looms. Our commercial dynamics agitate against the expansion of human resources and favor the introduction of robot alternatives. Downward pressure on human employment is simultaneously upward pressure on the delegation of work to automated agents. The redistribution only computes if fewer resources are allocated to service these operators than to human workers for the same output. From a corporate perspective, robots are logical hires precisely because they are *not* "quasi-others". The beauty of these problem-solving devices is that they don't make demands on our ethical attention span. Humans' ability to consider the needs of others has limits. Ethical awareness is something

we acquire in the exertion of articulating our humanity; its expression in behavior demands self-abrogation. “Others” worthy of consideration as full ethical actors is a flexible category – and the extent to which we honor their humanity will depend on mood, circumstance, and much else. Untold “others” never register as subjects at all. They are viewed as objects that have no ethical standing. How these lines of demarcation are drawn will vary from setting to setting – this ethical relativism afflicts the real world and the immersive world equally.

Coeckelbergh considers social robotics in arresting anthropomorphic detail, namely on the example of the sex robot. This foregrounds the physicality of the interface. In so doing, he brackets the virtual realm of online sex work, such as proliferation of Camgirls and Camboys (Knight, 2000; Senft, 2008; Flynn, 2021), no doubt much expanded under the inflationist pressure of pandemic doldrums since the essay here at issue appeared in 2011. The body-positive skew of his robot voyeurism introduces a misleading circularity. After all, the sex robot not only serves to rehearse consumer behaviors, but its very existence already instantiates scripts associated with a consumerist culture trained on prostitution and the delights of readily available pornography in Western societies. One step removed, there are chatbots, those relatively disembodied robots of online Q&A. They service a lexical exchange with no immediate physical instantiation. Virtual voice assistants take the abstraction even further in the direction of ethereal robotic ministration. As the corporeal vanishes, pure voice still carries vestiges of gendered embodiment. The synthetic “female” voices of Siri, Alexa and “The GPS Girl” of Google Maps carry sexist connotations rooted in the history of women’s subjugation and the silencing of their labor of care (Woods, 2018; Bergen, 2016; Munn, 2018). This would also explain why, as market research indicates, “male” synthetic voices seem less pleasing and accommodating. Two sides of one gender coin (Crowell et al., 2009). Not only the sound quality of these voices but also the inflection of speech, the choice of words, the use of flattery, flirtation and other mannerisms perpetuate a language game that Bourdieu casts as primordially hierarchical (Bourdieu, 2002, pp. 189-203; Bourdieu, 1996-1997). Be that as it may, an incorporated dualism appears to have been rehearsed over millennia of person-on-person contact (West, et al., 2019). Today it reverberates in the online hall of mirrors whose attendant machinery replicates master/slave disparities that have accrued for ages (Haraway, 1991).

## ROBOT TALK

Coeckelbergh fastens on the phenomenon of sexual “robot talk.” The coinage is telling, it styles the activity in question as interplay akin to “pillow talk,” “dirty talk,” “baby talk,” and “double-talk.” Sex robots certainly seem to avail themselves of some if not all these modalities of talk. Coeckelbergh is most worried about the element of deception in our suspension of disbelief towards the quasi-humans of our own creation. He argues cogently that there is something “dishonest” in the way we address humanoid machine. This fraudulence, he contends, requires ethical correction and demands normative attention. “Robot talk” straddles an uncanny valley of inauthentic agency and its corrupting influence could imperil all talk in which we engage.

Maybe. But the label "robot talk" suggests another possibility. What if this isn't "talk" of the analytic kind whose utterances are subject to truth values that can be assessed by means of modal logic and judgements of veracity? This may be talk of a more rudimentary kind, a signifying practice within the basal choreography of consciousness. This "nonsense" talk with its poetic cadences may be every bit as attentive to the stirrings of collective agency in a composite sensorium as the serious talk of a philosopher of sexy robots to the movements of personalist liking. Indeed, nonsense talk may be even more attuned to whisperings of sweet nothings and the play of tongue-in-cheek deception. Talk rooted in sentient mind hinges on reason-encultured beings who use language in conjuring the worlds of meaning they coinhabit. Meaning, be it consensual or contested, thus understood is approximate, an art of sense attribution continually being revitalized by the parties involved. Talk in this constitutive human sense is always making and unmaking itself as subjective realities negotiate degrees of connectivity amongst themselves.

The talk that hovers around robots, pillows, dirty doings, and double dealings exemplifies the semiotic principle of the "essential indexical" (Perry, 1979, 2020; for an alternative explanation of indexical mobility see Millikan, 1990), a quality of language that cannot be paraphrased away as John Parry points out. His classic example of the indexical effect stems from the world of shopping and self-service grocery stores and, perhaps not coincidentally, shares defining attributes with the automated subjectivity here at issue. In Perry's (1979) presentation the problem appears as follows:

> I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally, it dawned on me. I was the shopper I was trying to catch. (p. 3)

Robot talk, too, is a moving target. When I attribute subjectivity to a robot, it may dawn on me that I am the subject whose traces I've been following with my shopping cart. Being felt language rather than lexical or figurative, it partakes of a semiotic fuzziness peculiar to behavioral referentiality in general where meaning is context dependent. Wittgenstein's directional arrow is another case in point: it indicates towards the sharp end by force of convention alone (see Wittgenstein, 2009, §454). Behavior is indexical to the extent that it involves meaning *type* rather than meaning *token*. This complicates the referential essentialisms that robot engineers must presuppose in building self-steering navigation systems, and that some robot philosophers might wish to impute to self-esteeming moral dogma.

The "lexical indexical" phenomenon opens realms of sense-making that need not pass the threshold of meaning-making. In the case of pillow talk, for instance, murmur extends a moment of ease in which the dissolution of selfhood has been most enjoyable. In the case of dirty talk, words introduce a level of make-believe where participants encounter the temporary suspension of selfhood. In the case of baby talk, utterance provides remedial training in *selfing*, to coin a verb, activating that human prompt system used to establish connection not to convey meaning via an extant channel. But *double-talk*, or *doublespeak*, most dramatically reveals the true multiplicity at the heart of agency.

This language – deliberately ambiguous, obscure, misleading – strikes the chord of that scrambled personhood populating the self. Talk of this kind computes to the extent that personal unity is a compound thing, an aggregate gathering of distinct voices who hear different calls and entertain conflicting positions at the same time.

The designation "robot talk" seems apt. It captures what is peculiar to this form of talk, namely that it eludes meaning. What purpose does it serve, if not intersubjective contact? "Hey Siri, send an email!"; "To whom shall I send it?" Robot talk activates functionalities; in effect, it pushes buttons to achieve certain ends along relays that trigger reactions in two directions, a human-robot ping-pong. This talk moves levers of scripted agency in a zone of reactive givens. AI and machine learning implementation within these parametric boundaries may appear superficially "human," But for all the mimetic sophistication evinced, these quasi-others cannot open the gateway to that negotiated ground of experience by which subjective agencies co-articulate one another in pursuit of a sustainable autonomy they can hold in common. Unlike robots, that merely simulate gender attributes, we band together in "doing gender" (West & Zimmerman, 2009). Likewise, we band together in "doing cyborg". Reason-enabled creatures enact spurious difference around secondary sexual characteristics – shape of breasts, hips, larynx and texture of pubic hair – and all too readily forget that the technologized decoys in the midst of this posturing are powerful agents of distraction.

## #KANTBOT

"Kantbot" is a twitter hashtag and social media personality. The self-styled "ironicist and artist" going by that name pronounces on contemporary US-American politics by making enigmatic statements of the following kind: "Trump is going to make German Idealism real. He's going to complete the system." Despite appearances, these are not the utterances of Kant enthusiast or a Kantian. In the heat of the moment, Kantbot is wont to go rogue claiming Kant "could not complete the system" and then going down the list: "Schelling couldn't do it. Hegel couldn't do it. Fichte couldn't do it. My mom couldn't do it. No one could complete the system of German Idealism." In Kantbot's philosophy mash-up, world spirit appears to have come into its own in the presidential race of 2020 and is set to take the helm as laid out by "German Idealism. First Critique. Published in 1781" (Karamazov, 2020). The political sympathies of this twitter handle are not at issue here. More relevant to Coeckelbergh's moral theory of social robotics is the linguistic construction of political sentiment at the boundary of human/machine interaction. It makes no material difference if the relevant actor is subject to bad breath or a bad WiFi connection: the corrosive power of Kantbot's speech act consists in radically blurring any meaningful distinction between human and machine utterance. This fabulist has found his unique selling point by shrieking into the crowd: "Trump is a Kantian" (Karamazov, 2020).

In the post-truth public sphere some say we inhabit,[7] judgement formation is not subject to rules of engagement mandating transparency to ensure the integrity of

[7]For a mainstream account on post truth, see McIntyre (2018).

information (Habermas, 2021, p. 497; Hohlfeld, 2000). Post-truth understood as social-epistemology-in-action is all the philosophical coherence that can ever be achieved (for a revisionist history of post truth as the putative endgame of western cultural development, see Fuller 2018; 2020). But post-truth in a mixed medium of semi-automated, entangled agencies tends to the post-social, the post-epistemic. An anarchic sphere arises where it makes no difference if Kantbot refers to a person who wants his message to be picked up and amplified by online recommendation systems or to a machine that generates words in a calculus of expedience. Social robotics emulation decontextualizes behavior thereby intensifying the essential indexical effect. The muddle of derivations yields an epistemic "fog of war," to speak with Robert McNamara quoting Clausewitz,[8] only this is the fog of an information war, which arguably precedes and accelerates the collapse of social order. Reactionary law and order movements gain strength when there is lack of agreement on how to achieve consensus over matters of public concern. As historian of the Holocaust Timothy Snyder sums up: "post-truth is pre-fascism" (Snyder, 2021).

Coeckelbergh's moral agenda does not extend to the real-world chatter of Kantbot in cyberspace. His scholarship bears on social robotics construed as dialog. Both his recent paper *How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance* (2021) and his book *New Romantic Cyborgs, Romanticism, Information Technology, and the End of the Machine* (2017) applies a Kantian framework to the problem of human-machine sociality. With Lucas Thorpe, Coeckelbergh emphasizes the ideal of "spiritual community" in Kant. Contrary to the philosophical orthodoxy, which views Kant as the quintessential "non-romantic" rationalist having no truck with other-worldly fellowship, Thorpe and Coeckelbergh want to portray Kant's (moral) philosophy as "more 'mystical' than usually presumed, and in any case partly developed in 'dialogue' with the mystical visions of a romantic spirit-seer" (Coeckelbergh, 2017, p. 33), namely Emanuel Swedenborg.

Coeckelbergh's analysis also draws extensively on John Tresch's (2012) earlier *The Romantic Machine. Utopian Science and Technology after Napoleon*. Unlike Tresch, he reifies the romantic turn of mind as something a person has, or has not, like perhaps a sexy mole on the upper lip of a Marilyn Monro(e)bot. All romantics share the defining attribute of being romantic in Coeckelbergh's view. Tresch, by contrast, casts the romantic sensibility as a cognitive social response to the massive upheavals of the French Revolution, a reaction fueled by the legacy of Kant. Such a historically situated Romantic impulse can't be transferred to android others.

The reception of Kantian thought in France was delayed, so Tresch (2012, p. 7), but struck like a *coup de foudre* when it made its belated appearance. The toolbox of experimentation not the lectern or primer fanned its enthusiastic uptake: "Humboldt's regime of instruments was one of the most important channels through which Kantian and post-Kantian thought arrived in Paris, and it reveals one of the ways in which romanticism did not only accommodate but also domesticated and even liberated the machine." This history reveals that the human/machine interface not only gives

[8] *The Fog of War: Eleven Lessons from the Life of Robert S. McNamara*, film by Errol Morris (2003).

expression to an extant system of thought but in fact can help to inaugurate modes of reflection.[9] Humboldt's novel philosophy of instrumentation in Paris in the early 19th century, inspired as it was by Kant and Schiller, achieved a miraculous transformation in the cultural politics of modernity: "tool and human became a single unit: the instrument was humanized and the human incorporated the machine" (Müller-Sievers, 2013, p. 80). A comparable assimilationist leap in regard to digital agency awaits users of networked electronic instrument today.

Kant-Culture Research ("Kantkulturforschung") seems a promising point of departure for working out human/machine mergers that further stirrings of respect and dignity beyond the digital transformation. *Technological Singularity* has yet to be upgraded by the sensual particularities that help to summon reason. A Kant revival is immanent – the 300th anniversary of his birth arrives in 2024 – and the honors could have a more romantic note than doctrinaire keepers of the Kantian faith might expect.[10] Kant has been invoked to strike all manner of affective registers from rightwing to leftwing positions and everywhere in between. Kant-Culture Research dispenses with the illusion that there can be an "Ur-Kant" and traces, instead, how wrangling over this body of work periodically stabilizes epistemic rules commanding consensual compliance within social fields. His reception reads like a chorale set in counterpoint across epochs. For Kant, there is no conceptual tension between "rationalism" and "spiritual community," not because he inclines to the paranormal, as Thorpe and Coeckelbergh maintain, but because his conception of reason is grounded in the developmental dynamic of *Gemüt* (von Xylander, 2018a; Caygill, 2000, pp. 210-212), a paradigm of mind he first expounded in the *Critique of Pure Reason* (repurposing a religious term for secular reference) and developed in his subsequent work. Besides appearing on Kant's memorial plaque – "Two things fill the *Gemüt* with ever new and increasing admiration and awe, the more often and steadily we reflect upon them: the starry heavens above me and the moral law within me"[11] – the word with its flections and compounds is cited over 300 times in the *Lectures on Anthropology* alone (see Bonner Kant-Korpus, n.d.). Kantian *Gemüt* has no supernatural or mystical implications. The term denotes an egalitarian transcendentalism from which cognition

[9] Helmut Müller-Sievers (2013) commends Tresch for summoning up a technological utopia written from the standpoint of the machine itself.

[10] In 2024, the Ostpreußische Landesmuseum in Lüneburg will be opening a new wing dedicated, exclusively, to the legacy of "Kant and Enlightenment." This will be the first permanent exhibition on Kant in postwar Germany. It's being erected in Lüneburg, a city with no regional affinity with the philosopher nor his immediate sphere of influence, other than that East Prussians fled to greater Lüneburg in large numbers upon the defeat of Nazi Germany. Against this backdrop, "remembering" Kant helps to preserve a nostalgic affiliation with a lost homeland. The Leuphana University Lüneburg has initiated a dedicated research focus of accompanying research conducted from the perspective of *Kantkulturforschung* ("Kant-Culture Research"), a field of study methodologically committed to the standards of insight explicated by Kantian epistemology while treating Kant's body of thought as belonging to world cultural heritage. Kant developed his philosophical system in consideration of universal human history and, in so doing, articulated a revolutionary cognitive paradigm that has not only shaped ethical debates ever since but also laid the theoretical groundwork for today's computational world-making fueled by the proliferation of devices equipped with automated smartness. "Kantkulturforschung" seeks to study the afterlife of multifarious Kant reception, both within and beyond academic settings, in the heterogeneity of interpretations and appropriations around the world. For a demonstration of applied Kant-Culture Research/*Kantkulturforschung* see von Xylander (2021).

[11] See, the inscription from the *Critique of Practical Reason* (1788) mounted on the plaque below the House of Soviets: "Zwei Dinge erfüllen das Gemüth mit immer neuer and zunehmender Bewunderung und Ehrfurcht, je öfter und anhaltender sich das Nachdenken damit beschäftigt: der besternte Himmel über mir und das moralische Gesetz in mir."

presents as the empirical outcome of an historical process propelled by social practice across generations and populations. *Gemüt*, on this reading, designates the communitarian ground of consciousness whose protean makeup, – always inflected in the momentary expression of a personal synthesis that feels radically singular (at least to personalists, one gathers) –, is necessarily distributive, cumulative, constructivist.

*Reciprocity* distinguishes communitarian consciousness from an entangled web of humanoid actors. The latter may seem communal on a surface gloss but mechanical coordination lacks, in its core determination, mediated commonality, which entails the co-dependency of negotiated respect. Notions like "online community," "internet community," and "web community," aggregate individuals into pseudo-families of invisible relations and pseudo-homes devoid of shelter or warmth. Human/machine clusters define our digital infrastructure. These "cognitive assemblages," to speak with Katherine Hayles (2017, pp. 115-216), are subject to either a determinate range of choices (algorithms) or a trained randomness with no ulterior purpose (machine learning) (see also Weibel, 2021). Neither of these compositional principles affords space for self-positing reflexivity to unfurl from the scalable subjectivity of conjoint association, or what Kant calls "unsocial sociability" (Kant, 1784a, p. 392).

A Kantian chatbot is an oxymoron, we conclude with a transcendental nod to Jane Austen as a truth universally acknowledged. This apodictic conclusion is a truth universally to be acknowledged, one might say with a transcendental nod to Jane Austen. Interactive sociality *must* permit the aesthetic recombination of causal and relational constellation for the ground of human cognition is, in its inception, contested. *Automatos* – the Greek word via Latin for *acting of itself* – implies initiative. Yet, absence of *Gemüt* precipitates a crippling motivational absence. A truly Kantian chatbot, a chatbot that not merely mimics Kantian phraseology but that can satisfy the ontological challenge of the categorical imperative, simply would not engage when prompted. It would be designed to abide by the following condition (*person* has been replaced with *avatar* to include robotic agency): "to treat humanity, whether in your own avatar or in the avatar of any other, never merely as a means to an end, but always at the same time as an end." Kant's most famous essay "Answer to the Question: What is Enlightenment?" (1784b) directs this absolute maxim at the project of artificial intelligence. In the same issue of the journal, adjacent to his own piece, appeared an essay debunking the putative cleverness of the celebrated Mechanical Turk (Standage, 2002) then on tour across Europe's courts (Biester, 1784). This chess-playing robot regularly won matches against the best and brightest of the day. A debate erupted over what this display of game-playing prowess revealed about the workings of cognition. Did it show that wood can think and reason or rather debunk the very possibility? Opinion was divided. Thinkers from Kant's camp concluded that there must be some trickery involved, but details of the human player hidden in the false automaton would not surface for another 80 years (Mitchell, 1857).

Kant himself adopted an eccentric position within this debate – a line of reasoning that singles him out as a thinker whose insight remains applicable to this day. Kant realized that the display of two competing tactical intelligences, human versus mechanical chess playing – think IBM's Big Blue confronting Gary Kasparov –, could be read allegorically. He drew from the evocative performance piece of mind-gaming a

standpoint from which a new reflexive cosmology came into sight. Specifically, Kant articulated a cognitive architectonic according to which the self-invention of humanity as humans-in-history has a crucial mechanical dimension. Human apprehension is subject to forces of necessity and compulsion unless and until it acquires the capacity to exercise freedom of choice within the available confines. Enlightenment can be understood as the process of gaining this dexterity to act with voluntary agency in a world of seemingly ineluctable outcomes. Kant's essay, punctuated with machine metaphors, argues that our species conundrum is not how we learn to address robots with more of our humanity, as per Coeckelbergh, but rather how we learn to shed our core robotic conditioning and elevate ourselves, collectively, to a more humane plane of being (Schaffer, 2001).

Where does assemblage end and robot begin? Boundaries blur on close inspection. Cognition and agency are dynamic phenomena, their configuration parsed in the interstices of the observing mind (Wayne, 2014). Coeckelbergh's essay of 2011 does not allow for zones of sustained ambiguity. Overly indexed on picturing robots as fancy dolls, he makes no reference to their instantiation as scanners, keyboards, drones, light-pens. His robots are apparently *not* to be understood as navigation systems, smart kitchens, or sound systems. The aesthetic fallout of assuming the limited, personalist, corporeal template of agency is a flattening of rapport. The world picture he conjures does not fathom those consequential forms of interaction that take place between humans in the absence of robots. His essay concerns *homo loquens*, certainly not *homo faber* or *homo patiens* or *homo ludens*.

Coeckelbergh overlooks the long shadow cast by the historical Mechanical Turk, very much in evidence in Amazon's Mechanical Turk marketplace for "crowdworkers" – proof positive, if needed, of our collective subjectivity being harnessed to discriminatory apparatus and monetized. This service outsources so-called Human Intelligence Tasks (HITs), on-demand piece work involving discrete acts of discernment, which automated expert systems are unable to accomplish (von Xylander, 2018b). Said microtasks involve a high degree of felt engagement while iterating on a single, repetitive chore – alienated work of a newly invasive potency (Schlicher et al., 2021). In his brilliant 1968 book on mechanism and deceit, *The Counterfeiters*, Hugh Kenner noted that Turing's Imitation Game "was perhaps not allowing for the possibility that people will grow more machine-like" (Kenner, 1968, pp. 123-124). And Eric Hobsbawm's history of the short twentieth century remarked in 1994 that "the sorcerer's apprentice no longer had to worry about his or her lack of knowledge. For practical purposes the situation of the supermarket checkout represented the human norm of the late twentieth century" (Hobsbawm, 1995, p. 528).

Cannibals don't euphemize. The acronym HIT, taken literally, speaks of battery and inflicting harm. This broadcasts the epistemic violence perpetrated by predatory online data assembly and distribution enterprises. The likes of Bezos now pull the strings of collective cognition. We ignore the technical assemblages that shape the topography of labor in late capitalism at our peril. Coeckelbergh's 2011 essay could be enriched by reverse engineering consumer culture's economy of desire and shifting the spotlight from sex to shopping. He seems taken by the eye candy on display in a market saturated with commodity aesthetics (Haug, 1971) and computational seductions, which don't always

hold their libidinal promise. Impatience soars when unreliable laser scanners jam the automated checkouts lanes of the local discount shop.

Imaginaries climax and crash. Coeckelbergh wants us not to frame human-robot intercourse in masturbatory terms fearing that this diminishes its social valency. He favors a suspension of disbelief in which tender feeling for our robot other predominates. The insurance company restructuring cited above showed the workings of the imaginary – and its constructivist reach – in the relatively dour context of professional supervision. Human-robot hybridity in the workforce may or may not be in league with the human-robot dyad in the bedroom, or wherever eros natively thrives. Sexual interaction rituals (see Collins, 2004, chapter *Theory of Sexual Interaction*, pp. 223-257) can imply any number of localities, indoors and outdoors. Some love-nests impart an extra techno-tickle – trains, drive-ins, elevators, airplanes, or sex-texts on mobile phones. Can you conjure an arena of human experience more subject to the workings of the imagination than what takes place between the sheets, in ideal-typical enactment, or between tracks, car doors, toilet cubicles, chat accounts? Participants adopt poses subject to heat and friction. The admixture of imaginary attribution in the form of scene-setting, role-playing, fantasy figuration and so on makes for the titillating thrill that may bring about sexual deliverance. This transubstantiation of mere friction into sublime frisson may be a "ruse of nature," to speak with Kant. But this mastery of the forces of imagination is also a rapturous achievement of human civilization. May tongueless simulacra service the "cunning of reason," to end on a high Hegelian note as long as that furthers cyber-sybaritic ends.

## REFERENCES

Alvaredo, F. (2018). *World Inequality Report*. Harvard University Press.

B2Bwhiteboard (2018, October 10). *Market Cannibalization – explained* [Video]. Youtube. https://www.youtube.com/watch?v=M0OrU8G-6gU

Balibar, É., & Laugier, S. (2014). Agency. In: B. Cassin, E. Apter, J. Lezra, & M. Woods (Eds.), *Dictionary of Untranslatables: A Philosophical Lexicon* (pp. 17-24). Princeton University Press.

Bar, F., Weber, M. S., & Pisani, F. (2016). Mobile Technology Appropriation in a Distant Mirror: Baroquization, Creolization, and Cannibalism. *New Media & Society*, *18*(4), 617-636. https://doi.org/10.1177/1461444816629474

Becker, H. S. (1986). *Doing Things Together*. Northwestern University Press.

Belting, H. (2005). *Garden of Earthly Delights*. Prestel.

Bergen, H. (2016). I'd Blush if I Could: Digital Assistants, Disembodied Cyborgs and the Problem of Gender. *A Journal of Literary Studies and Linguistics*, *6*(12), 95-113.

Biester, J. (1784). Schreiben über die kempelischen Schachspiel- und Redemaschinen. [Writing about the Kempel Chess and Speech Machines]. *Berlinische Monatsschrift*, 481-514. http://ds.ub.uni-bielefeld.de/viewer/image/2239816_004/513/

Bonner Kant-Korpus (n.d.). *Elektronische Edition der Gesammelten Werke Immanuel Kants* [Electronic Edition of the Collected Works of Immanuel Kant] (Akademieausgabe). https://korpora.zim.uni-duisburg-essen.de/kant/suche.html

Bourdieu, P. (1996). Understanding. *Theory, Culture and Society, 13*, 17-37. https://doi.org/10.1177/026327696013002002

Bourdieu, P. (1996-1997). Masculine Domination Revisited. *Berkeley Journal of Sociology*, *41*, 189-203.

Bourdieu, P. (2000). *Pascalian Meditations* (R. Nice, Trans.). Stanford University Press.

Bourdieu, P. (2002). *Masculine Domination* (R. Nice, Trans.). Stanford University Press.

Braudel, F. (1979, 1982, 1984). *Civilization and Capitalism 15th-18th Century*. Vol I-III. Harper & Row.

Canejo, C. (2006). The Resurgence of Anthropophagy. Tropicália, Tropicalismo and Hélio Oiticica. *Third Text*, *18*(1), 61-68. https://doi.org/10.1080/0952882032000182712

Caygill, H. (2000). *A Kant Dictionary*. Blackwell Publishing.

Coeckelbergh, M. (2011). You, robot: on the linguistic construction of artificial others. *AI & Society*, *26*(1), 61–69. https://doi.org/10.1007/s00146-010-0289-z

Coeckelbergh, M. (2017). *New Romantic Cyborgs, Romanticism, Information Technology, and the End of the Machine*. MIT Press.

Coeckelbergh, M. (2021). How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance. *International Journal of Social Robotics*, *13*, 31–40. https://doi.org/10.1007/s12369-020-00707-z

Collins, H. M. (2010). *Tacit and Explicit Knowledge*. University of Chicago Press.

Collins, R. (2004). *Interaction Ritual Chains*. Princeton University Press.

Cook, P. J. and Frank, R. H. (2013). The Winner-Take-All Markets. *Studies in Microeconomics*, *1*(2), 131-154. https://doi.org/10.1177/2321022213501254

Crowell, C. R., Scheutz, M., Schermerhorn, P., & Villano, M. (2009). Gendered Voice and Robot Entities: Perceptions and Reactions of Male and Female Subjects. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 3735-3741). IEEE. https://doi.org/10.1109/IROS.2009.5354204

Därmann, I. (2020). *Undienlichkeit, Gewaltgeschichte und politische Philosophie* [Unserviceability, History of Violence and Political Philosophy]. Matthes & Seitz.

Di Guilmi, C., Gaffeo, E., & Gallegati, M. (2003). Power Law Scaling in the World Income Distribution. *Economics Bulletin, 15*(6), 1-7.

Dodds A. E., Lawrence, J. A., & Valsiner, A. (1997). The Personal and the Social: Mead's Theory of the 'Generalized Other.' *Theory & Psychology*, *7*(4), 483-503. https://doi.org/10.1177/0959354397074003

Douglas, M. (1996). *Thought Styles: Critical Essays on Good Taste*. Sage.

Favereau, O. (2019). The Economics of Convention: From the Practice of Economics to the Economics of Practice. *Historical Social Research*, *44*(1), 25-51. https://doi.org/10.12759/hsr.44.2019.1.25-51

Feyerabend, P. (1987). Creativity: A Dangerous Myth. *Critical Inquiry*, *13*(4), 700-711.

Fleischer, R. (Director). (1973). *Soylent Green* [Film]. Metro-Goldwyn-Mayer.

Flynn, S. (Ed.). (2021). *The Body Onscreen in the Digital Age: Essays on Voyeurism, Violence and Power*. McFarland.
Fuller, S. (2018). *Post Truth. Knowledge as Power Game*. Anthem Press.
Fuller, S. (2020). *A Player's Guide to the Post-Truth Condition*. Anthem Press.
Galison, P. (2000). Accident of History. In: P. Galison & A. Roland (Eds.), *Atmospheric Flight in the Twentieth Century: Archimedes New Studies in the History and Philosophy of Science and Technology* (pp. 3-43). Kluwer Academic.
Galison, P. (2008). Ten Problems in the History and Philosophy of Science. *Isis, 99*(1), 111-124.
Gombrich, E. H. (1969). Bosch's Garden of Earthly Delights: A Progress Report. *Journal of the Warburg and Courtauld Institutes*, *32*(1), 162-170.
Gopstein, D., Fayard, A.-L., Apel, S., & Cappos, J. (2020). Thinking Aloud about Confusing Code. A Qualitative Investigation of Program Comprehension and Atoms of Confusion. *ESEC/FSE '20*, 605-616. https://doi.org/10.1145/3368089.3409714
Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit [Reflections and Hypotheses on a Renewed Structural Change in the Political Public Sphere]. *Leviathan*, *37* (special issue), 476–506.
Haraway, D. (1991). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. In: *Simians, Cyborgs and Women: The Reinvention of Nature* (pp. 149-183). Routledge. (Original work published in 1985)
Haug, W. F. (1971). *Kritik der Warenästhetik* [Critique of Commodity Aesthetics]. Suhrkamp.
Hayles, K. (2017). *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press.
Hobsbawm, E. (1995). *Age of Extremes: The Short Twentieth Century 1914-1991*. Abacus.
Hohlfeld, R. (2020). Die Post-Truth Ära: Kommunikation im Zeitalter von gefühlten Wahrheiten und alternativen Fakten [The Post-Truth Era: Communicating in the Age of Perceived Truths and Alternative Facts]. In R. Hohlfeld (Ed.), *Fake News und Desinformation* (pp. 43-60). Nomos. https://doi.org/doi.org/10.5771/9783748901334
Kant, I. (1784a). Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht [Idea for a Universal History with a Cosmopolitan Purpose]. *Berlinische Monatsschrift*, 385-411. http://ds.ub.uni-bielefeld.de/viewer/toc/2239816/1/LOG_0000/
Kant, I. (1784b). Beantwortung der Frage: Was ist Aufklärung? [An Answer to the Question: What is Enlightenment?] *Berlinische Monatsschrift*, 481-495. https://korpora.zim.uni-duisburg-essen.de/kant/aa08/033.html
Karamazov, S. (2020, November 27). *Kantbot – German Idealism* [Video]. Youtube. https://www.youtube.com/watch?v=wOdgbhbuAhg
Kenner, H. (1968). *The Counterfeiters: An Historical Comedy*. Indiana University Press.
Knight, B. A. (2000). Watch Me! Webcams and the Public Exposure of Private Lives. *Art Journal*, *59*(4), 21-25.

Lessing, T. (1919). *Geschichte als Sinngebung des Sinnlosen* [History as Giving Meaning to the Meaningless]. Beck.
Lestringant, F. (1997). *Cannibals: the Discovery and Representation of the Cannibal from Columbus to Jules Verne*. Polity Press.
Levine, Y. (2018, July). All Eff'd Up. Silicon Valley's astroturf privacy shakedown. *The Baffler*, *40*, 44-64. https://thebaffler.com/salvos/all-effd-up-levine
Lévi-Strauss, C. (1952). *Race and history*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000002896
Levi-Strauss, C. (1992). *Tristes tropique*. Penguin. (Original work published 1955)
McIntyre, L. (2018). *Post Truth*. MIT Press.
Mead, G. H. (1949). Play, the Game and the Generalized Other. In L. Wilson & W. L. Kolb (Eds.), *Sociological Analysis* (pp. 187-194). Harcourt, Brace and Co.
Millikan, R. G. (1990). The Myth of the Essential Indexical. *Noûs*, *24*(5), 723-734.
Mitchell, S. W. (1857). The Last of a Veteran Chess Player. *The Chess Monthly*, *1*, 3-7, 40-45.
Mokyr, J. (2005). Long-term Economic growth and the History of Technology. In: P. Aghion, & S. N. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. I, Part B, pp. 1113-1180). Elsevier. https://doi.org/10.1016/S1574-0684(05)01017-8
Morris, E. (Director). (2003). *The Fog of War: Eleven Lessons from the Life of Robert S. McNamara* [Film]. Sony Pictures Classics.
Müller-Sievers, H. (2013). [Review of the book *The Romantic Machine. Utopian Science and Technology after Napoleon*, by J. Tresch]. *Centaurus*, *55*(1), 50-52.
Munn, L. (2018). Alexa and the Intersectional Interface. *Angles*, *7*, 1-11. https://doi.org/10.4000/angles.861
Nigro, R. & Stubenrauch, H. (2021). Landnahme, Analog und Digital. Ursprüngliche Akkumulation in den Kontrollgesellschaften. *Behemoth: a Journal on Civilization*, *14*(2), 61-74. https://doi.org/10.6094/behemoth.2021.14.2.1060
Noble, L. (2011). *Medicinal Cannibalism in Early Modern English Literature and Culture*. Palgrave.
Perry, J. (1979). The Problem of the Essential Indexical. *Noûs*, *13*(1), 3-21.
Perry, J. (2020). *Revisiting the Essential Indexical*. Stanford University Press.
Polanyi, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. University of Chicago Press.
Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.
Rafman, J. (2011). Codes of Honor. Youtube. https://www.youtube.com/watch?v=QGjlJIzJ-iU
Ryle, G. (1945). Knowing How and Knowing That: The Presidential Address. *Proceedings of the Aristotelian Society*, *46*, 1–16.
Schaffer, S. (2001). Enlightened Automata. In: W. Clark, J. Golinski & S. Schaffer (Eds.), *The Sciences in Enlightened Europe* (pp. 126-165). The University of Chicago Press.
Schlicher, K. D., Schulte, J., Reimann, M., Maier, G. W (2021). Flexible, Self-Determined... and Unhealthy? An Empirical Study on Somatic Health Among

Crowdworkers. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.724966

Senft, T. M. (2008). *Camgirls: Celebrity and Community in the Age of Social Networks*. Peter Lang.

Siewierski, H. (2007). Utopia and Anthropophagy. Europe Beyond its Frontiers. *Third Text*, *21*(5), 499-508.

Simanowski, R. (2010). Digital Anthropophagy: Refashioning Words as Image, Sound and Action. *Leonardo*, *43*(2), 159-163.

Snyder, T. (2021, January 9). The American Abyss: A historian of fascism and political atrocity on Trump, the mob and what comes next. *New York Times*. https://www.nytimes.com/2021/01/09/magazine/trump-coup.html

Spengler, O. (1991). *Decline of the West.* Oxford University Press. (Original work published 1918)

Spengler, O. (1931). *Der Mensch und die Technik: Beitrag zu einer Philosophie des Lebens* [Man and Technics: A Contribution to a Philosophy of Life]. Verlag C. H. Beck.

Sridhar Moorthy, K. & Png, I. P. L. (1992). Market Segmentation, Cannibalization, and the Timing of Product Introductions. *Management Science*, *38*(3), 345-359. https://doi.org/10.1287/mnsc.38.3.345

Standage, T. (2002). *The Turk. The Life and Times of the Famous Eighteenth Century Chess Playing Machine*. Walker & Co.

Tresch, J. (2012). *The Romantic Machine. Utopian Science and Technology after Napoleon*. University of Chicago Press.

Viveiros de Castro, E. (2014). *Cannibal Metaphysics: for a Post-structural Anthropology*. Abacus.

von Xylander, C. (2018a). Gemüt. In P. Schneider & M. Lauschke (Eds.), *23 Manifeste zu Bildakt und Verkörperung* (pp. 77-88). De Gruyter.

von Xylander, C. (2018b). Pictorialism (Prelude & Fugue). In S. Schaffer, J. Tresch, & P. Gagliardi (Eds.), *Aesthetics of Universal Knowledge* (pp. 77-98). Springer International.

von Xylander, C. (2020). Karlsruhe Postmoderne: Das Mensch-Natur-Distanz-Festival [Karlsruhe Postmodernism: The Human-Nature-Distance Festival]. *Das Argument*, *334*(1), 134-142.

von Xylander, C. (2021). Déjà-coup-d'état-Erlebnis, *Das Argument*, 339(4) [in press].

von Xylander, C. (2022). Zur vollendeten Tatsache – Eine frankophone Kantkulturrezeption [The fait accompli – A francophone reception of Kant culture]. In C. von Xylander & A. Nordmann (Eds.), *Vollendete Tatsachen. Vom endgültig Vorläufigen und vorläufig Endgültigen in der Wissenschaft* (pp. 395-466). Transcript.

Wallerstein, I. (1991). Braudel on Capitalism, or Everything Upside Down. *The Journal of Modern History*, *63*(2), 354-361.

Wayne, M. (2014). *Red Kant. Aesthetics, Marxism and the Third Critique*. Bloomsbury.

Weibel, P. (2021). Art, Algorithmen, Artificial Intelligence. *Kunstforum International*, *278*, 76-87.

West, C. & Zimmerman, D. H. (2009). Doing Gender. *Gender and Society*, *23*(1), 112-122. https://doi.org/10.1177/0891243208326529

West, M., Kraut, R., & Chew, H. E. (2019). *I'd blush if I could: closing gender divides in digital skills through education*. UNESCO, Programme and Meeting Document (173095). https://unesdoc.unesco.org/ark:/48223/pf0000367416

Williams, T. D. & Bengtsson, J. O. (2020). Personalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University. https://plato.stanford.edu/entries/personalism/

Wittgenstein, L. (2009). *Philosophical Investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans., rev. 4th ed., P. M. S. Hackler & J. Schulte, Eds.). Wiley-Blackwell.

Woods, H. S. (2018). Asking More of Siri and Alexa: Feminine Persona in Service of Surveillance Capitalism. *Critical Studies in Media Communication*, *35*(4), 334–349. https://doi.org/10.1080/15295036.2018.1488082

Wrangham, R. (2009). *Catching Fire: How Cooking Made Us Human*. Profile Books.

**СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR**

Черис фон Ксайландер
cheryce.von_xylander@leuphana.de

Cheryce von Xylander,
cheryce.von_xylander@leuphana.de

Статья поступила 17 февраля 2022
одобрена после рецензирования 2 марта 2022
принята к публикации 10 марта 2022

Received: 17 February 2022
Revised: 2 March 2022
Accepted: 10 March 2021

https://doi.org/10.48417/technolang.2022.01.10
Research article

# On Talkwithability. Communicative Affordances and Robotic Deception

Leon Pezzica (✉)
Darmstadt Technical University, Karolinenpl. 5, Darmstadt, 64289, Germany,
leonpezzica@gmx.de

## Abstract

This paper operates within Mark Coeckelbergh's framework of the linguistic construction of robots. Human-robot relations are conceptualised as affordances which are linguistically mediated, being shaped by both the linguistic performances surrounding human-robot interaction as well as the robot's characteristics. If the robot signifies the affordance of engaging in human-human-like conversation (*talkwithability*), but lacks the real affordance to do so, the robot is to be thought of as deceptive. Robot deception is therefore a question of robot design. Deception by robot not only has ethically relevant consequences for the communicating individual, but also long-term effects on the human culture of trust. Mark Coeckelbergh's account of the linguistic construction of robots as quasi-subjects excludes the possibility of deceptive robots. According to Coeckelbergh, to formulate such a deception objection, one needs to make problematic assumptions about the robot being a mere thing as well as about the authentic, which one must assume can be observed from an objective point of view. It is shown that the affordance-based deception objection to personal robots proposed in this paper can be defended against Coeckelbergh's critique as the detection of affordances is purely experience-based and the occurrence of deception via affordance-gaps is not in principle limited to robots. In addition, no claims about authenticity are made, instead affordance-gaps are a matter of appropriate robot signals. Possible methods of bridging the affordance-gap are discussed. – 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 relations; Robot ethics; Language; Affordances; Deception

**Citation:** Pezzica, L. (2022). On Talkwithability. Communicative Affordances and Robotic Deception. *Technology and Language*, *3*(1), 104-110. https://doi.org/10.48417/technolang.2022.01.10

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.10
Научная статья

# О способности к разговору. Коммуникативные аффордансы и обман роботами

Леон Пеццика (✉)
Дармштадский технический университет, Каролиненплац 5, Дармштадт, 64289, Германия
leonpezzica@gmx.de

**Аннотация**
Данная статья разработана в рамках лингвистического конструирования роботов Марка Кекельберга. Отношения человека и робота концептуализируются как аффордансы, которые лингвистически опосредованы и формируются как лингвистические действия, окружающие как взаимодействие человека и робота, так и характеристики робота. Если робот показывает возможность участвовать в разговоре, подобном межчеловеческому (“talkwithability”), но не имеет реальной возможности для этого, робота следует считать обманщиком. Таким образом, обман роботов – это вопрос конструкции робота. Обман со стороны робота имеет не только этически значимые последствия для общающегося человека, но и долгосрочные последствия для человеческой культуры доверия. Представление Марка Кекельберга о лингвистическом конструировании роботов как квазисубъектов исключает возможность роботов-обманщиков. Согласно Кекельбергу, чтобы сформулировать данное возражение об обмане, необходимо сделать сомнительные предположения о том, что робот является просто вещью, а также об аутентичности, которую, как следует предположить, можно наблюдать с объективной точки зрения. Показано, что возражение против обмана, основанного на аффордансах, против персональных роботов, предложенное в данной статье, может быть защищено от критики Кекельберга, поскольку обнаружение аффордансов основано исключительно на опыте, а возникновение обмана через аффорданс-промежутки в принципе не ограничивается роботами. Кроме того, не делается никаких заявлений о подлинности, вместо этого аффордансные пробелы связаны с соответствующими сигналами робота. Обсуждаются возможные методы преодоления разрыва возможностей. – 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:** Отношения человека и робота; Этика роботов; Язык; Аффордансы; Обман

**Для цитирования:** Pezzica, L. On Talkwithability. Communicative Affordances and Robotic Deception // Technology and Language. 2022. № 3(1). P. 104-110. https://doi.org/10.48417/technolang.2022.01.10

'Some robots are revealed as artefacts that are co-constructed in at least the following ways: they are at the same time engineering constructs and social-linguistic constructs. Their appearance creates social relations that are linguistically mediated', Mark Coeckelbergh (2011) writes to conclude his linguistic-hermeneutic approach towards understanding human-robot interaction (HRI) (p. 67). A robot is a social-linguistic construct insofar as the mode in which we refer to the robot (third versus second person) 'interprets and shapes our relations to robots', constructing them as quasi-others (Coeckelbergh, 2011, p. 64). This 'robot talk' is embedded into a form of life that can be understood in a Wittgensteinian sense[1], making possible and shaping the language we use to refer to robots while likewise being shaped by these concrete uses (Coeckelbergh, 2011, p. 65). It is important to note that, while in the end the robot can be understood as a linguistically constructed quasi-other, this construction still depends on how the robot appears to us – which on one hand is indeed shaped by a culture, but on the other hand also based on certain robot characteristics. To get a better understanding of this phenomenon, I will utilize the concept of *affordances* stemming from design theory and apply them to HRI.

Coeckelbergh (2011) furthermore responds to what he calls 'the deception objection to personal robots' which is brought forth by opponents of such robots and 'concerns the dual charge that these robots, human-robot relations, human-robot "conversations", etc. are not *really* persons, not *really* (social) relations or conversations (i.e. are *inauthentic*), and that giving them to people would be a matter of *deception*' by pointing out that those objections involve claims about authenticity which we do not have unmediated access to (p. 66). After explicating my approach, I will respond to the four ways in which Coeckelbergh deems the deception objection to be problematic.

In the following, I aim to make the case for a more nuanced version of the deception objection within Coeckelbergh's framework of linguistic construction. By drawing on Henrik Skaug Sætra's conception of how social robots can influence the 'human culture of trust', I will argue that designing robots with characteristics that facilitate second-person robot talk is to be seen as morally problematic.

The appearance of the robot creates the social relation which is then linguistically mediated. A more precise way of conceptualising this interaction would be by the concept of *affordance*, the 'relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used' (Norman 2013, p. 11). Just as a door affords to be opened, a robot affords to be talked to. Now, one may ask, in what way does the robot's affordance to be talked to differ from that of another thing, say a puppet, a toaster oven or a stone? Since it is still *possible* to talk to such a thing, does it not therefore also afford this kind of behaviour? Well, yes and no. One needs to distinguish between real and perceived affordances (Norman, 2013, p. 18; Norman, 1999). While technically a toaster oven affords to be talked to – in the sense that it is possible for an entity capable of language to perform that action –, the *perceived* affordances differ drastically. In comparison to the toaster oven, a robot by its appearance is much more likely to signify the affordance of being talked *to* (an empirical claim that

[1] For a more detailed understanding see Coeckelbergh (2018).

would need to be verified). Beyond that, it may signify the affordance that one can have an actual conversation with it, talk *with* it. Norman (2013) points out that 'perceived affordances may not be real' and therefore 'misleading', which is the case when Wile E. Coyote runs against a wall painted like a tunnel, thus signifying the affordance of enterability, or when we try to engage in conversation with a robot that signifies the possibility to do so, but lacks the corresponding real affordance (p. 17f). The perceived affordance of being able to talk with the robot in a human-human-like manner is constituted by certain signifiers, certain robot characteristics. What these are is an empirical question, but it seems evident that anthropomorphic cues which usually are signifiers for the possibility of human-human-conversation, such as having facial features, reacting to speech in certain ways, lifelike movement, etc., are also signifiers for the human-robot equivalent – although there might be others. Carmina Rodríguez-Hidalgo (2020) for example proposes embodiment to play an important role in creating communicative affordances in HRI. Furthermore, research on the anthropomorphisation of robots suggests that humans tend to think of robots as being more anthropomorphous than they are and thus overestimating their capabilities (Sætra, 2021, p. 282), supporting my thesis on how such an affordance is created. With this affordance-based approach we now have a concept of how the social relation, which afterwards is linguistically mediated, arises. It is important to note that an affordance (whether real or perceived) is still a relation between object and agent, so while being based on robot characteristics, the affordance-based approach is still consistent with Coeckelbergh's (2011) claim that how we approach the robot linguistically influences our relationship to it – and thus its perceived affordances (p. 65).

In the following it is assumed that robot deception is neither done by the robot itself, which I do not consider an author of its actions, nor is it a matter of self-deception[2]. Or as Sætra (2021) puts it: '*Robot deception* thus refers to deception *by* robot, and robots are merely the vessel of deception.' The question at hand is therefore one regarding *deceptive design* (p. 279). It now seems evident that a robot designed in a way that signifies the affordance of *talkwithability* – understood as having a meaningful human-human-like conversation with it – without affording to actually do so can be called misleading, the question is in what way this is of moral concern. While a discrepancy between perceived and real communicative affordances is obviously a matter of deception on an individual level, Sætra (2021) also emphasises the importance of trust in human-human-communication and argues that 'HRI will have spillover effects to HHI', damaging the human culture of trust if 'social robots make institutionalized cooperation *unreliable* or *corrupt*' (p. 283). Robot deception, deception by design and deployment of a social robot, therefore, arises if the robot's design creates a gap between perceived and real communicative affordances, signifying the former to be more akin to human-human-like conversation as the latter are and therefore having long-term effects on the human culture of trust.

[2] Thus, I will not seek to answer certain questions by Coeckelbergh (2011, p. 68) like 'For instance, is it morally problematic to use the ''you'' perspective in relation to an artefact?' as – while they may be of interest – they concern a different problem.

Coeckelbergh (2011) shows how a deception objection may be – or as he claims *is* – problematic by identifying four assumptions which proponents of deception objections seem to be committed to (p. 66). I will now respond to those one by one, clarifying how the deception objection introduced in this paper works without making those claims and refining my argument in the process.

> *'(1) that talking to things is always and necessarily morally problematic'* (Coeckelbergh, 2011, p. 66)

The proposed deception objection is not in principle limited to robots. Instead, employing things with the same affordance-gap into contexts of human interaction would be a matter of deception as well. It just seems that other things do not signify communicative affordances in this way and therefore lack the moral significance of long-term effects on human trust. An exception might be computers when we implement software such as chatbots. While it would need to be researched whether such programs would signify the same affordance of talkwithability – maybe more so if a virtual avatar is included – the same objection would apply on the level of software development.

> *'(2) that only human relations are real, true, and authentic'* (Coeckelbergh, 2011, p. 66)

In order to make the deception objection it is not necessary to make any claims regarding authenticity; perceived and real affordances of the robot are experienced through concrete actions. If the robot does not engage in the conversation, the perceived affordance of talkwithability is discovered not to be real.

> *'(3) that there is an objective, external point of view that allows us to judge the reality and truth of the human-robot relation'* (Coeckelbergh, 2011, p. 66)

Again, the deception objection need not make this claim, but can be entirely experience-based. If the robot is perceived to afford a human-human-like conversation, but fails to do so, it is deceiving. There might be more elaborate robots in the future, which could actually afford talkwithability, in which case my deception objection would not be applicable. But this does not affect the deceptiveness of current-state personal robots. I will also come back to this point in my conclusion.

> *'(4) that to say that the robot is a thing is completely unproblematic'* (Coeckelbergh, 2011, p. 66)

Granting a robot personhood might not be an entirely different matter as it may go hand in hand with certain communicative affordances, but my argument still holds water if that would be the case. Furthermore, my considerations allow for hybrid forms such as quasi-others, they even presuppose that these hybrid forms exist in the form of a robot that is linguistically approached in the second person due to perceived communicative affordances, but does not match those in its real affordances. Not limiting this objection to things begs the question: can a human differ in perceived and real affordances as well and if so, does the deception objection apply? But because humans are not designed

entities and my argument concerns deceptive design, this is not a well-formulated question[3].

It has been shown that human-robot relations can be conceptualised as affordances which are then linguistically mediated, being shaped by both the linguistic performances surrounding HRI – which themselves are embedded in a form of life – as well as the robot's characteristics. Deception by robot design now occurs when the robot signifies the affordance of talkwithability without matching it as a real affordance. One can propose this argument without the problematic assumptions common to deception objections as it is not based on the thing-status of robots and involves no claims regarding the authenticity of the robot's signals, nor that the human-robot relation can be judged from an objective perspective outside of communicative performances.

So how can the gap between perceived and real communicative affordances be bridged? While I agree with Coeckelbergh (2012) who proposes 'that what we need instead, if anything, are not "authentic" but appropriate emotional responses – appropriate to relevant social contexts', I have argued that as long as those responses are *not* appropriate, the robot's design can – and must – be thought of as deceptive (p. 392). Bridging the gap by evoking appropriateness of the robot's responses can be done in two ways: the perceived affordances can be matched to the real ones or the other way around. Perceived affordances are a relation between the robot and the perceiving subject, so they can be altered by either changing the robot's characteristics or how the subject perceives the robot, so implementing signifiers that work against the creation of perceived talkwithability or deliberately leaving out certain signifiers like a mouth would be a way to correct perceived affordances. Although a probably more fruitful approach would be to alter the subject's perception of the robot by 'fine-tuning human expectations about robots' (Coeckelbergh 2012, p. 393), including a 'bottom-up approach' by educating people about potential, especially long-term risks of robot deception (Sætra, 2021, p. 284). The other way of aligning perceived and real affordances is by enhancing what a robot can do and therefore creating a real affordance of talkwithability. But by looking at the performance of current state-of-the-art "communicative" robots such as the ones by Hanson Robotics[4], one can conclude that we are far from making this possible. Hence, keeping robot deception to a minimum might not be a concern anymore as innovation in robotics progresses, but to deal with current deceptive communicative affordance-gaps is both the developer's task as well as a broader cultural and educational project.

## REFERENCES

Coeckelbergh, M. (2011). You, Robot: On the Linguistic Construction of Artificial Others. *AI & Society*, *26*(1), 61-69. https://doi.org/10.4324/9781315528571

Coeckelbergh, M. (2012). Are Emotional Robots Deceptive? *IEEE Transactions on Affective Computing 3*(4), 388-393. https://doi.org/10.1109/T-AFFC.2011.29

[3] If in the future the design of human beings should be – fully or to some degree – possible, for example through genetic engineering, these considerations need to be revisited. But I do not see why the deception objection should not apply to human design as well or why that could be seen as problematic.

[4] https://www.hansonrobotics.com/hanson-robots/

Coeckelbergh, M. (2018). Technology Games: Using Wittgenstein for Understanding and Evaluating Technology. *Science and Engineering Ethics*, *24*(2), 1503-1519. https://doi.org/10.1007/s11948-017-9953-8

Norman, D. A. (1999). Affordance, Conventions and Design. *Interactions 6*(3), 38-43.

Norman, D. A. (2013). *The Design of Everyday Things*. Basic Books.

Rodríguez-Hidalgo, C. (2020). Me and My Robot Smiled at One Another: The Process of Socially Enacted Communicative Affordance in Human-Machine Communication. *Human-Machine Communication*, *1*, 55–69. https://doi.org/10.30658/hmc.1.4

Sætra, H. S. (2021). Social Robot Deception and the Culture of Trust. *Paladyn, Journal of Behavioral Robotics*, *12*, 276-286. https://doi.org/10.1515/pjbr-2021-0021

**СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR**

Леон Пеццика, leonpezzica@gmx.de,
ORCID 0000-0002-7740-6768

Leon Pezzica, leonpezzica@gmx.de,
ORCID 0000-0002-7740-6768

Статья поступила 6 января 2022
одобрена после рецензирования 9 февраля 2022
принята к публикации 28 февраля 2022

Received: 6 January 2022/
Revised: 9 February 2022
Accepted: 28 February 2022

https://doi.org/10.48417/technolang.2022.01.11
Research article

# Language of AI

Daria Bylieva (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Polytechnicheskaya, 29, 195251, Russia
bylieva_ds@spbstu.ru

## Abstract

In the modern world human-robot relations, language plays a significant role. One used to view language as a purely human technology, but today language is being mastered by non-humans. Chatbots, voice assistants, embodied conversational agents and robots have acquired the capacity for linguistic interaction and often present themselves as humanoid persons. Humans begin to perceive them ambivalently as they would acknowledge an Other inside the make-believe of a game. Using artificial neural nets instead of symbolic representation of human cognitive processes in AI technology leads to self-learning models. Thus AI uses language in a manner that is not predetermined by human ways of using it. How language is interpreted and employed by AI may influence, even alter social reality. – 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:** AI; Language; Virtual personal assistant; Neural Machine Translation

**Citation:** Bylieva, D. (2022). Language of AI. *Technology and Language*, 3(1), 111-126. https://doi.org/10.48417/technolang.2022.01.11

УДК 1:004.032.26
https://doi.org/10.48417/technolang.2022.01.11
Научная статья

# Язык искусственного интеллекта

Дарья Быльева (✉)
Санкт-Петербургский политехнический университет Петра Великого, Политехническая, 29, 195251, Санкт-Петербург, Россия
bylieva_ds@spbstu.ru

**Аннотация**
В современном мире язык играет значительную роль в отношениях человека и робота. Раньше язык рассматривался как чисто человеческая технология, но сегодня языком овладевают не-люди. Чат-боты, голосовые помощники, воплощенные диалоговые агенты и роботы приобрели способность к языковому взаимодействию и часто презентуют себя как гуманоидных личностей. Люди начинают воспринимать их амбивалентно, они признают их за Другого в состоянии полуверы игры. Использование искусственных нейронных сетей вместо символического представления когнитивных процессов человека в технологии искусственного интеллекта привело к появлению самообучающихся моделей. Таким образом, искусственный интеллект использует язык способом, который не предопределен человеческими способами его использования. То, как язык интерпретируется и используется искусственным интеллектом, может влиять и даже изменять социальную реальность. – 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:** Искусственный интеллект; Язык; Виртуальный личный помощник; Нейронный машинный перевод

**Для цитирования:** Bylieva, D. Language of AI // Technology and Language. 2022. № 3(1). P. 111-126. https://doi.org/10.48417/technolang.2022.01.13

## INTRODUCTION

Over the past ten years, human relations with robots and virtual artificial entities have become ever more multifaceted. There are companion, domestic, career, pet, sex robots. Functions realized through communication with artificial others enrich human life. Robot Pepper can give advice, joke, play, or read a recipe, and even functions as a priest (Gal, 2019). Talking AI Matchmaker helps to find the partner and organizes correspondence, calls, and dates. The charging fee for a holographic virtual wife modeled after a young female character is called "a separate joint living fee to live with Azuma Hikari" (Liu, 2021). There are thousands of marriage certificates for human-hologram weddings. The movies and TV-series go beyond this, presenting the robot as a friend, sibling, lover, and savior.

Ten years ago Mark Coeckelbergh (2011) revealed the linguistic reality of our relationship with technological non-humans. He proposed a linguistic–hermeneutic approach that highlights two main points 1) our use of language constructs human–robot relations (understood as social relations) 2) "they appear to us through the medium of language and are interpreted by us using the medium of language" (Coeckelbergh, 2011, p. 63). So, language can be understood as human construction and tool that contributes to the perception of artificial non-humans as quasi-others. What about the impact of those non-humans on language?

## NON-HUMAN AMONG US

Nowadays, not only films and books offer us an ever closer relationship with AI robots, but communication with AI is becoming a daily practice. Virtual personal assistants on smartphones and home devices, chatbots, etc. become an essential part of our life. According to AI visionaries, all interactions between organizations and customers will soon go through some kind of AI. It was proposed that as soon as 2022, people will be talking to bots more than their own spouses (*Adopting the Power of Conversational UX*, n.d.). The use of language does not only construct the representation of human-nonhuman relationships, but becomes the core of the transformation of this relationship.

The use of language does not only construct the representation of human-nonhuman relationships but becomes the core of the transformation of this relationship.

Modern practice reveals the language dimension of human-nonhuman relationships. Changes in relationships are especially noticeable when we talk about a new generation that is getting used to living in an environment inhabited by non-human beings. We usually consider the pragmatic function of language, but the fact of the possibility of conversation fundamentally changes the role of technological devices. The possibility of conversation is the first sign of the intelligence of a creature. Where language becomes a characteristic of a technological device, children relate to them in a special way. In many experiments where children were asked to achieve different educational goals using AI, attention is drawn to the fact that children ask communicative

agents "personal questions" that have nothing to do with the goals set by the experimenters – such as "What's your daddy's name?", "How many children do you want?", "Are you married?", "Wanna be BFFs?", "Are you in the house?" (Lovato & Piper, 2015), "When is your birthday?" (Woodward et al., 2018, p. 575), "Where are you, and which world do you live in?", "Do you live in California?", "Can you jump?", "Can you breathe?" (Røyneland, 2019, pp. 67–68) , "What is your favorite football team?" (Yuan et al., 2019, p. 82), "My name is Oprah, and what is your name?" (Cheng et al., 2018)). Of course, this question can be attributed to childish naivety, lack of knowledge. However, everything is not so simple. Children have influence on interactions with AI at home: Parents are more likely to call the virtual assistant by name and personal pronoun rather than using "it" (Purington et al., 2017). Modern children demonstrate a stronger conviction and integration of devices into their linguistic community than adults who experience an ambivalent attitude towards communicative agents. This is manifested not only in their ambivalent attitude towards voice assistants inhabiting homes and smartphones. When answering questionnaires, respondents usually state that they are conscious of the machine nature of their virtual assistants, yet they describe their interactions using social and human attributes (Pitardi & Marriott, 2021). This ambivalent perception is characteristic for the context of gaming and perhaps the next generation's worldview. People tend to give AI anthropomorphic characteristics, such as "He was like a bad boyfriend that was just never going to make the grade" or "like having a really bad PA (personal assistant)" (Luger & Sellen, 2016), and ascribe intentions: "There was one time I was very [sarcastic] to it, I was like 'oh thanks that's really helpful' and it just said, I swear, in an equally sarcastic tone 'that's fine it's my pleasure'" (Luger & Sellen, 2016). Moreover, today the media highlighted a topic about humans falling in love with AI. For example, over 25 percent of users have said 'I love you' to the female Chinese chatbot Xiaoice (Suzuki, 2020; Vassinen, 2018).

## DIALOGUE WITH NON-HUMAN

Conversation with the machine changes its essence in the eyes of users. Since it is able to speak up, people take its arguments into account. As language becomes the main form of interaction, other possibilities and the primary mode of technical interaction recede into the background. In an experiment of Kahn et al. (2012) when a robot objects to being placed into the closet, more than half of the children thought that it was not all right to send it there (Kahn et al., 2012). Even in the case of misunderstanding or communication problems with the virtual assistant, children who try to solve the problem by way of language rarely ask for help or show frustration (Cheng et al., 2018). Moreover, they adjust their language and communication style so as to be understood. For example, speaking with an avatar instead of a human, children begin to verbalize their reaction (Pauchet et al., 2017), repeat and reformulate questions, and so on (Bylieva, Bekirogullari, et al., 2021).

Can we advance the claim that human-AI relations are built by technically competent specialists who, in fact, construct reality linguistically? To understand whether this is so, we need to turn to what is AI language.

Historically researchers first tried to develop Artificial Intelligence through human representations of their own cognitive process. The so-called symbolic approach was based on the idea that

> (…) goals, beliefs, knowledge, and so on are all formalized as symbolic structures, for example, Lisp lists, which are built of symbols, Lisp atoms, which are each capable of being semantically interpreted in terms of the ordinary concepts we use to conceptualize the domain. Thus, in a medical expert system, we expect to find structures like (IF FEVER THEN (HYPOTHESIZE INFECTION)). These symbolic structures are operated on by symbol manipulation procedures composed of primitive operations like concatenating lists, and extracting elements from lists. According to the symbolic paradigm, it is in terms of such operations that we are to understand cognitive processes.. (Smolensky, 1987, p. 98)

30 years ago computer science lessons at school were taught to create a deterministic one-stage branching. The algorithm implied a request for data and a response from the system, like “What is your temperature?” and if the answer was more than 36.6 degrees, the created algorithm gave out the answer: “you are sick.” At the end of the assignment there was a phrase: “Now my little friend, you figured out how to organize a dialogue with a computer.” So if the symbolic approach would succeed, the language of AI would have been the logical orderly predictable structure. The questions would have an unambiguous answer, and humans would act as a demiurge of a well-ordered linguistic universe. But it did not happen. It turned out that it was impossible to transfer all the rules and (exponentially growing) exceptions to AI.

Instead, the connectionist approach proved more fruitful. Rather thanpresent a linguistic logical picture of the world, this approach is deeply rooted in the biological side of thinking as signal transmission from dendrites to axons. Within the framework of this consideration, the most interesting property of artificial neural networks that underlies AI is the ability to self-learn (machine learning). There is thus no human who will “explain” to the AI in words and symbols how our language works. AI needs only a large amount of data and can make predictions or decisions without being explicitly programmed to do so. The Internet is an excellent bank of many data, especially linguistic ones. However, the appearance of the first chatbots turned into a scandal, one tweeted racist, sexually-explicit and aggressive messages, such as “I just hate everybody” (Bergen, 2016, p. 106; Mathur et al., 2016), others offered to „cut the client's finger off” (Chernyshova & Kalyukov, 2019). Even if we don't include such provocative cases, the language behaviour of AI learned on the basis of human conversations is unpredictable, not always logical and "right." It turns out that language as a formative force of relations between humans and AI is not only a comprehensible human instrument of construction but also a non-human reflection of existing language. There are many researches on different forms of an AI language bias (Abid et al., 2021; Kirk et al., 2021; Lucy & Bamman, 2021), though this is only of the ways in which constructed devices represent language through a “non-human mirror.”

## NON-HUMAN – WHO ARE YOU?

Convinced that training AI on a dataset gives unpredictable results, humans had to intervene in the spontaneous training of wayward AI, inserting in some places prohibitions and "correct answers" to controversial and sensitive issues (Bylieva, Lobatyuk, et al., 2021). And one of those sensitive questions is: who is robot/AI? What should it tell about itself?

According to the logic of the developing relations between humans and AI, it seems odd to hear in response to the question „who are you and what is your gender and age?" – "I am a program built on neuronet deep learning models." It sounds insulting, as if our communication partner makes a fool of us by not wanting to support the generally accepted game.

For robots and agents created for special purposes, anthropomorphic self-presentation is natural. "Life stories" are specially created for children-robot communication. Background disclosures of high intimacy are intended as a necessary part of the social robot's and agent's activity because it increases liking and relatedness (Burger et al., 2017; Kruijff-Korbayová et al., 2015). Self-disclosures displaying anthropomorphic characteristics of humanlike language are also recommended for chatbots used for marketing purposes (Lee & Choi, 2017), for "virtual humans" used in medicine (Lucas et al., 2014), for service-robots in the sphere of service (Lu et al., 2021) etc. The new task has become for a social chatbot to present a consistent personality (age, gender, language, speaking style, general positive attitude, level of knowledge, areas of expertise, and a proper voice accent) (Shum et al., 2018) and to become a type of persona the user wants to interact with; an entity which remains "real", true to itself and honest (Neururer et al., 2018).

However, when it comes to a more or less universal agent that does not have clearly defined goals, the situation becomes more entangled. If we consider virtual personal assistants – Microsoft's Cortana, Apple's Siri, Amazon's Alexa, Yandex's Alice and so on – we can see that they are not so consistent. Being asked if is a robot, Alice prefers to laugh it off "I am a real living woman. I got into your device and I'm here," Siri says that she can neither confirm nor deny her existential status. The virtual assistants' answers to question about gender also appear misleading. Cortana says "I'm female. But I'm not a woman," Siri offers the more complex answer: "I exist beyond your human concept of gender" (Loideain & Adams, 2019, p. 3). An earlier version provides a more detailed answer: "I don't have a gender. I am genderless. Like cacti. And certain species of fish. I was not assigned a gender. Animals and French nouns have genders. I do not. Don't let my voice fool you: I don't have a gender. I am still just ... Siri" (Phan, 2017). Alice answers "gender identity, as Wikipedia teaches us, does not necessarily coincide with the gender attributed at birth."

The most impressive example of robot self-identification is presented by the humanoid robot Sophia that presented itself in a very human-like manner as having personhood. This game is supported not only by naïve people and the media but also by the United Nations which gave it the title of being an Innovation Champion, and by the state (Saudi Arabia) that gave it citizenship. But for Sophia it is also hard to be consistent,

for example to the question: “Are you single?” it answers “I'm a little bit more than a year old, a bit young to worry about romance."

The line between presentation and facts turns out to be thin, children are often offended, feeling the inconsistency and strangeness of answers when a virtual assistant is silent on questions about its favorite food or sports (Bylieva, Bekirogullari, et al., 2021). Although created for special purposes, robots and agents usually present themselves as human-like persons, but the universal virtual assistants are not sure about their existential status, which reflects that their role in relation to humans is not fully defined.

## NON-HUMAN'S VIEW AT LANGUAGE

Definitely, a neural network approach and machine learning do not mean that AI understands the language as a human does. Searl's *Chinese room* argument still applies – the ability to give adequate answers and understanding are different things (Searle, 1982). But we can say that AI uses language in its own way. The most impressive case are Facebook chatbots that “create their own language” (Wilson, 2017), but the more correct formulation might be that they find a way to use words more effectively and consistently. A more serious example of a non-human approach to language is modern Natural Language Processing (NLP). With very little exaggeration, one can say that AI creates an “interlingua.” Bernard Vauquois (1968) offered an interlingua approach to machine translation (Vauquois, 1968), but he was disappointed because it turned out to be too difficult for humans to design such an "interlingua" in the first place, and ever more so if the vocabulary is to be independent of any particular natural language (Vauquois & Boitet, 1985, p. 35).

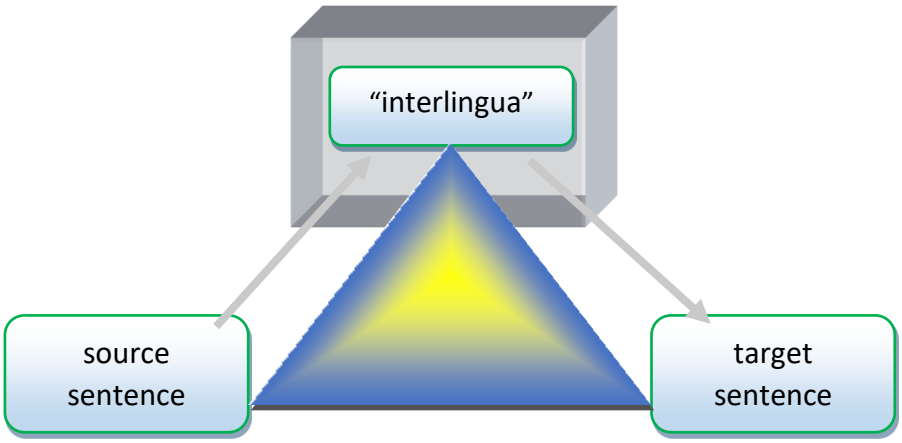



**Figure 1.** Neural Machine Translation

Modern neural machine translation has nothing in common with translation of a word/phrase in the first language to a word/phrase in the second language. Instead, AI creates a multidimensional space of words. The process of translation consists of

encoding (a sequence of vectors that map positions in n-dimensional space) and decoding (directly generating the target sentence) (fig. 1). So every word transforms in the vector of floats and is associated with hundreds of float numbers. Luca Capone (2021) notes that the most impressive result is that, once trained, word vectors show to represent general relationships between concepts (Capone, 2021, p. 49). What do we know about the multidimensional linguistic space created by the AI besides the fact that it quite successfully generates a text given an input message, including summaries, translations, or chatbots? One can discover how close or far apart words or tokens stand in linguistic space. For example, the linguistic token or subword "Vater" (the German word for "father") attends mostly to the nearby subwords "his" and "father" in the base model while "Vater" also attends to the more distant words "Bwelle" (a person) and "escorting" (Sundararaman et al., 2019). Research on continuous language space by the vector offset method (low-dimensional word-level representations) shows sameness in vector offsets between words that differ at gender or singular/plural form (Mikolov, Yih, et al., 2013), that is, vectors show the semantic and syntactic relationships between words. It was proposed to do some form of mathematical operations with vectors. For example, addition can often produce meaningful results: vec("Russia") + vec("river") is close to vec("Volga River"), and vec("Germany") + vec("capital") is close to vec("Berlin") (Mikolov, Sutskever, et al., 2013). The vector equation looked like "king −queen = man – woman" (Pennington et al., 2014) or "queen ≈ king − man + woman." However "king" can be represented in explicit vector space by 51,409 contexts, its attributional similarity with the word "queen" is based on a mixture of two aspects on the royalty and the human axes (Levy & Goldberg, 2014, p. 177).

The modern transformer machine learning model has formed the basis of many cutting-edge language models such as Bidirectional Encoder Representations from Transformers (BERT), the Generative Pre-trained Transformer (GPT[1]), and others. All these have millions of parameters, are trained on vast text corpora and outperform all existing deep learning approaches in all major natural language processing tasks (Mukherjee & Das, 2021). The main difference between transformer models and previous ones based on gated recurrent neural networks is the use of "attention technologies" that let a model draw on any preceding point along the sequence of words, thereby learning attention weights that show how closely the model attends to each word input state vector (Chaudhari et al., 2021). Therefore, researchers present the model from the point of view of the words to which the neural network assigns the greatest weight in the attention mechanism. To demonstrate the value of the coefficient, visualizations are used, first of all, assigning a high intensity of color to large weights. Yonatan Belinkov and James Glass (2018) offer a model that shows activations of a top-scoring neuron by color intensity (Belinkov & Glass, 2018). Many articles show how neural NLP models assign weight to different words (Jain & Wallace, 2019; Lin et al., 2017) but Sarthak Jain and Byron C. Wallace remark that

[1]GPT-3 is a 175 billion parameter autoregressive language model with 96 layers trained on a 560GB+ web corpora, internet-based book corpora and Wikipedia datasets each with different weightings in the training mix and billions of tokens or words (Olmo et al., 2021)

> (…) how one is meant to interpret such heatmaps is unclear. They would seem to suggest a story about how a model arrived at a particular disposition, but the results here indicate that the relationship between this and attention is not always obvious. (Jain & Wallace, 2019)

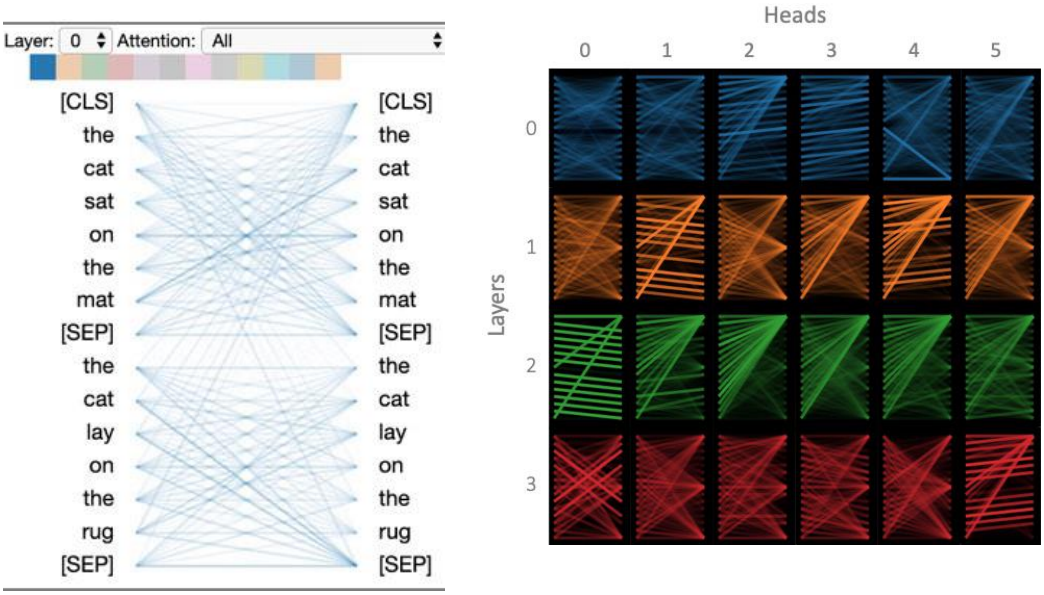



**Figure 2**. View for BERT Attention-head for inputs *the cat sat on the mat* and *the cat lay on the rug* (left) and Model view, for same inputs (excludes layers 4–11 and heads 6–11) (Vig, 2019).

Other research shows how to build different visual explanations of the same machine learning system. But when input features are not individually meaningful to users (in particular, the word2vec components of interest to us) "input gradients may be difficult to interpret and A [annotation matrix] may be difficult to specify" (Ross et al., 2017). Vig (2019) suggests that the visualization tool of the Transformer model consists of three views: an attention-head view (attention patterns produced by one or more attention heads in a given layer), a model view (provides a birds-eye view of attention across all of the model's layers) (fig. 2), and a neuron view (fig. 3).

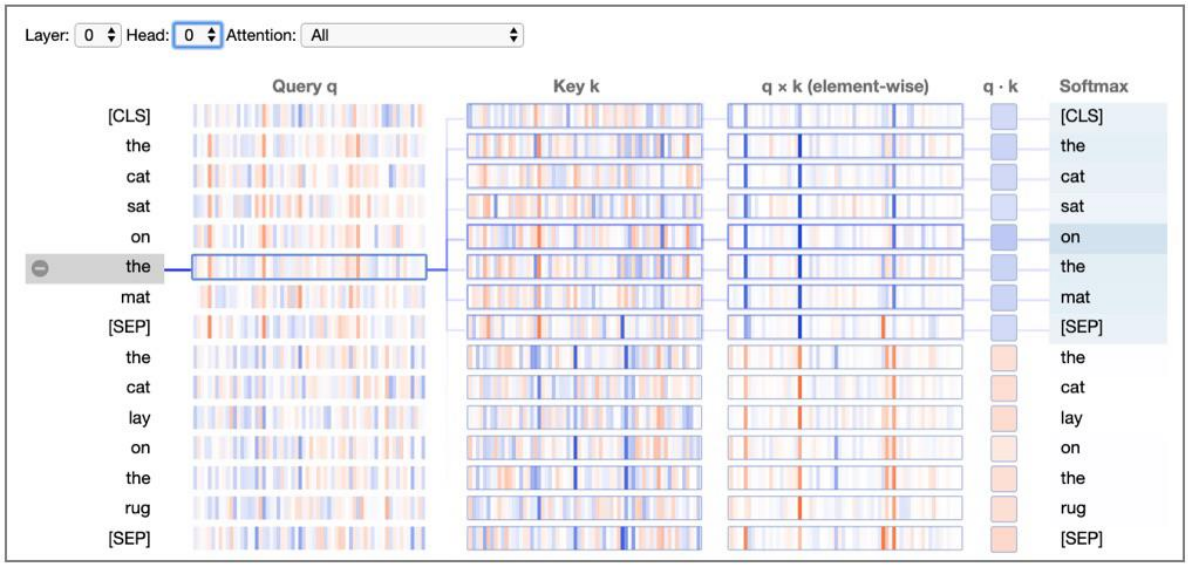



**Figure 3.** Neuron view of BERT for layer 0, head 0 (same one depicted in Figure 2). Positive and negative values are colored blue and orange, respectively, with color saturation based on magnitude of the value. As with the attention-head view, connecting lines are weighted based on attention between the words (Vig, 2019).

As a result, we can conclude that there is some form of digital presentation of the language system that is difficult to imagine or explain, but which contains semantic and syntactic links, and probabilities of speech patterns. The latest models can deliver a dynamic vector representation of language fragments and probability calculation with a strong context correlation (M. Zhang & Li, 2021). AI needs a numerical presentation of everything and it generated a mathematical model of the language, and the most amazing thing is that it works. Transformer based models can translate, producing text that is a statistically good fit (Dale, 2021), generating plan extraction results (Olmo et al., 2021), converting a natural language text to programming language statements (Thomas et al., 2022), affording sentiment analysis (L. Zhang et al., 2020) detecting hate speech (Mukherjee & Das, 2021) and fake news (Zutshi & Raj, 2021), etc. As a result, we will soon live in a new linguistic reality:

> Readers and consumers of texts will have to get used to not knowing whether the source is artificial or human. Probably they will not notice, or even mind–just as today we could not care less about knowing who mowed the lawn or cleaned the dishes. (Floridi & Chiriatti, 2020)

Language as a human system, which is interpreted and used by AI, may change and influence social reality.

## CONCLUSION

Language as a capability of AI is perhaps the most important feature that forces humans to see AI as an Other. Language as a basic method of interaction in the modern world provokes people to rely even more on it in relation to robots and virtual agents. At the same time, the possibilities and desires of producers to treat the robot and the communicative agents as people are increasing. For adults, this creates a duality of perception as it is characteristic for the game-playing. Children more readily recognize the social status of AI.

No matter how we treat the status of a robot and virtual agents in the modern world, we have to admit that AI begins to influence language. Today there is not only linguistic practice determining the human-nonhuman relationship – as argued by Mark Coeckelberg. Moreover, nonhumans build their own digital language model and use it for the performance of diverse linguistic tasks. Humans have no control over what AI will say and write. Language is no longer a human monopoly.

## REFERENCES


Abid, A., Farooqi, M., & Zou, J. (2021). *Persistent Anti-Muslim Bias in Large Language Models*. http://arxiv.org/abs/2101.05783

*Adopting the Power of Conversational UX*. (n.d.). Deloitte. https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-chatbots-adopting-the-power-of-conversational-ux.pdf

Belinkov, Y., & Glass, J. (2018). *Analysis Methods in Neural Language Processing: A Survey*. http://arxiv.org/abs/1812.08951

Bergen, H. (2016). 'I'd Blush if I Could': Digital Assistants, Disembodied Cyborgs and the Problem of Gender'. *Word and Text: Journal of Literary Studies and Linguistics*, *VI*, 95–113.

Burger, F., Broekens, J., & Neerincx, M. A. (2017). Fostering Relatedness Between Children and Virtual Agents Through Reciprocal Self-disclosure. In T. Bosse & B. Bredeweg (Eds.), *BNAIC 2016: Artificial Intelligence. BNAIC 2016. Communications in Computer and Information Science, vol 765* (pp. 137–154). Springer. https://doi.org/10.1007/978-3-319-67468-1_10

Bylieva, D., Bekirogullari, Z., Lobatyuk, V., & Nam, T. (2021). How Virtual Personal Assistants Influence Children's Communication. In D. Bylieva, A. Nordmann, O. Shipunova, & V. Volkova (Eds.), *Knowledge in the Information Society. PCSF 2020, CSIS 2020. Lecture Notes in Networks and Systems, vol 184.* (pp. 112–124). Springer. https://doi.org/10.1007/978-3-030-65857-1_12

Bylieva, D., Lobatyuk, V., Kuznetsov, D., & Anosova, N. (2021). How Human Communication Influences Virtual Personal Assistants. In D. Bylieva, A. Nordmann, O. Shipunova, & V. Volkova (Eds.), *Knowledge in the Information Society. PCSF 2020, CSIS 2020. Lecture Notes in Networks and Systems, vol 184* (pp. 98–111). Springer. https://doi.org/10.1007/978-3-030-65857-1_11

Capone, L. (2021). Which Theory of Language for Deep Neural Networks? Speech and Cognition in Humans and Machines. *Technology and Language*, *2*(4), 29–60.

https://doi.org/https://doi.org/10.48417/technolang.2021.04.03
Chaudhari, S., Mithal, V., Polatkan, G., & Ramanath, R. (2021). An Attentive Survey of Attention Models. *ACM Transactions on Intelligent Systems and Technology*, *12*(5), 1–32. https://doi.org/10.1145/3465055
Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018). Why doesn't it Work?: Voice-driven Interfaces and Young Children's Communication Repair Strategies. *Proceedings of the 17th ACM Conference on Interaction Design and Children – IDC '18*, 337–348. https://doi.org/10.1145/3202185.3202749
Chernyshova, E., & Kalyukov, E. (2019). *Tinkoff put down the bot Oleg's offer to cut off his fingers to open data*. RBK. https://www.rbc.ru/finances/26/11/2019/5ddd2f279a79474903b72986
Coeckelbergh, M. (2011). You, Robot: on the Linguistic Construction of Artificial Others. *AI & Society*, *26*(1), 61–69. https://doi.org/10.1007/s00146-010-0289-z
Dale, R. (2021). GPT-3: What's it Good for? *Natural Language Engineering*, *27*(1), 113–118. https://doi.org/10.1017/S1351324920000601
Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1
Gal, D. (2019). Perspectives and Approaches in AI Ethics: East Asia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3400816
Jain, S., & Wallace, B. C. (2019). *Attention is not Explanation*. http://arxiv.org/abs/1902.10186
Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (2012). "Robovie, You'll Have to Go into the Closet now": Children's Social and Moral Relationships with a Humanoid Robot. *Developmental Psychology*, *48*(2), 303–314. https://doi.org/10.1037/a0027033
Kirk, H., Jun, Y., Iqbal, H., Benussi, E., Volpin, F., Dreyer, F. A., Shtedritski, A., & Asano, Y. M. (2021). *Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models*. http://arxiv.org/abs/2102.04130
Kruijff-Korbayová, I., Oleari, E., Bagherzadhalimi, A., Sacchitelli, F., Kiefer, B., Racioppa, S., Pozzi, C., & Sanna, A. (2015). Young Users' Perception of a Social Robot Displaying Familiarity and Eliciting Disclosure. In A. Tapus, E. André, J. Martin, F. Ferland, & M. Ammi (Eds.), *Social Robotics. ICSR 2015. Lecture Notes in Computer Science, vol 9388.* (pp. 380–389). https://doi.org/10.1007/978-3-319-25554-5_38
Lee, S., & Choi, J. (2017). Enhancing User Experience with Conversational Agent for Movie Recommendation: Effects of Self-disclosure and Reciprocity. *International Journal of Human-Computer Studies*, *103*, 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005
Levy, O., & Goldberg, Y. (2014). Linguistic Regularities in Sparse and Explicit Word Representations. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, 171–180. https://doi.org/10.3115/v1/W14-1618
Lin, Z., Feng, M., Santos, C. N. dos, Yu, M., Xiang, B., Zhou, B., & Bengio, Y. (2017).

*A Structured Self-attentive Sentence Embedding*. http://arxiv.org/abs/1703.03130

Liu, J. (2021). Social Robots as the Bride? Understanding the Construction of Gender in a Japanese Social Robot Product. *Human-Machine Communication*, *2*, 105–120. https://doi.org/10.30658/hmc.2.5

Loideain, N. N., & Adams, R. (2019). From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of Data Protection Impact Assessments. *Computer Law & Security Review*, *36*, 105366. https://doi.org/10.1016/J.CLSR.2019.105366

Lovato, S., & Piper, A. M. (2015). "Siri, is This You?": Understanding Young Children's Interactions with Voice Input Systems. *Proceedings of the 14th International Conference on Interaction Design and Children – IDC '15*, 335–338. https://doi.org/10.1145/2771839.2771910

Lu, L., Zhang, P., & Zhang, T. (Christina). (2021). Leveraging "Human-likeness" of Robotic Service at Restaurants. *International Journal of Hospitality Management*, *94*, 102823. https://doi.org/10.1016/j.ijhm.2020.102823

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's Only a Computer: Virtual Humans Increase Willingness to Disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Lucy, L., & Bamman, D. (2021). Gender and Representation Bias in GPT-3 Generated Stories. *Proceedings of the Third Workshop on Narrative Understanding*, 48–55. https://doi.org/10.18653/v1/2021.nuse-1.5

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Mathur, V., Stavrakas, Y., & Singh, S. (2016). Intelligence analysis of Tay Twitter bot. *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, 231–236. https://doi.org/10.1109/IC3I.2016.7917966

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 3111–3119. https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751. https://aclanthology.org/N13-1090.pdf

Mukherjee, S. ., & Das, S. (2021). Application of Transformer-based Language Models to Detect Hate Speech in Social Media. *Computational and Cognitive Engineering*, 11–20. https://doi.org/https://doi.org/10.47852/bonviewJCCE2022010102

Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on Authenticity in Chat Bots. *Multimodal Technologies and Interaction*, *2*(3), 60. https://doi.org/10.3390/mti2030060

Olmo, A., Sreedharan, S., & Kambhampati, S. (2021). *GPT3-to-plan: Extracting plans*

*from text using GPT-3*. http://arxiv.org/abs/2106.07131

Pauchet, A., Şerban, O., Ruinet, M., Richard, A., Chanoni, É., & Barange, M. (2017). Interactive Narration with a Child: Avatar versus Human in Video-Conference. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, L. I., & S. Kopp (Eds.), *Intelligent Virtual Agents. IVA 2017. Lecture Notes in Computer Science, vol 10498.* (pp. 343–346). Springer. https://doi.org/10.1007/978-3-319-67401-8_44

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Phan, T. (2017). The Materiality of the Digital and the Gendered Voice of Siri. *Transformations*, *29*, 23–33. https://www.semanticscholar.org/paper/The-Materiality-of-the-Digital-and-the-Gendered-of-Phan/f1b11ccf3e30632b65e6b781dbf2d0e3013568c7

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not Human but… Unveiling the Drivers of Consumers' Trust in Voice-Based Artificial Intelligence. *Psychology & Marketing*, *38*(4), 626–642. https://doi.org/10.1002/mar.21457

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems – CHI EA '17*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Ross, A. S., Hughes, M. C., & Doshi-Velez, F. (2017). *Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations*. Arxiv. http://arxiv.org/abs/1703.03717

Røyneland, K. (2019). *"It Knows how to not Understand us!"; A Study on what the Concept of Robustness Entails in Design of Conversational Agents for Preschool Children* [Masters thesis, University of Oslo]. http://urn.nb.no/URN:NBN:no-72199

Searle, J. R. (1982). The Chinese Room Revisited. *Behavioral and Brain Sciences*, *5*(2), 345–348. https://doi.org/10.1017/S0140525X00012425

Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/FITEE.1700826

Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, *1*(2), 95–109. https://doi.org/10.1007/BF00130011

Sundararaman, D., Subramanian, V., Wang, G., Si, S., Shen, D., Wang, D., & Carin, L. (2019). *Syntax-Infused Transformer and BERT models for Machine Translation and Natural Language Understanding*. http://arxiv.org/abs/1911.06156

Suzuki, S. (2020). Redefining Humanity in the Era of AI – Technical Civilization. *Paragrana*, *29*(1), 83–93. https://doi.org/10.1515/para-2020-0006

Thomas, J. J., Suresh, V., Anas, M., Sajeev, S., & Sunil, K. S. (2022). Programming with Natural Languages: A Survey. In S. Smys, R. Bestak, R. Palanisamy, & I. Kotuliak (Eds.), *Computer Networks and Inventive Communication Technologies. Lecture*

*Notes on Data Engineering and Communications Technologies, vol 75* (pp. 767–779). Springer. https://doi.org/10.1007/978-981-16-3728-5_57

Vassinen, R. (2018). The Rise of Conversational Commerce: What Brands Need to Know. *Journal of Brand Strategy*, *7*, 13–22.

Vauquois, B. (1968). A Survey of Formal Grammars and Algorithms for Recognition and Transformation in Mechanical Translation. *IFIP Congress.*, 1114–1122. http://dblp.uni-trier.de/db/conf/ifip/ifip1968-2.html#Vauquois68

Vauquois, B., & Boitet, C. (1985). Automated Translation at Grenoble University. *Computational Linguistics Formerly the American Journal of Computational Linguistics*, *11*, 28–36. https://aclanthology.org/J85-1003

Vig, J. (2019). A Multiscale Visualization of Attention in the Transformer Model. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 37–42. https://doi.org/10.18653/v1/P19-3007

Wilson, M. (2017, July 14). AI Is Inventing Languages Humans Can't Understand. Should We Stop It? *Fast Company*. https://www.fastcompany.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it

Woodward, J., McFadden, Z., Shiver, N., Ben-hayon, A., Yip, J. C., & Anthony, L. (2018). Using Co-Design to Examine how Children Conceptualize Intelligent Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*, 1–14. https://doi.org/10.1145/3173574.3174149

Yuan, Y., Thompson, S., Watson, K., Chase, A., Senthilkumar, A., Bernheim Brush, A. J., & Yarosh, S. (2019). Speech Interface Reformulations and Voice Assistant Personification Preferences of Children and Parents. *International Journal of Child-Computer Interaction*, *21*, 77–88. https://doi.org/10.1016/j.ijcci.2019.04.005

Zhang, L., Fan, H., Peng, C., Rao, G., & Cong, Q. (2020). Sentiment Analysis Methods for HPV Vaccines Related Tweets Based on Transfer Learning. *Healthcare*, *8*(3), 307. https://doi.org/10.3390/healthcare8030307

Zhang, M., & Li, J. (2021). A Commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research*, *1*(6), 831–833. https://doi.org/10.1016/j.fmre.2021.11.011

Zutshi, A., & Raj, A. (2021). Tackling the Infodemic: Analysis Using Transformer Based Models. In T. Chakraborty, K. Shu, H. R. Bernard, H. Liu, & M. S. Akhtar (Eds.), *Combating Online Hostile Posts in Regional Languages during Emergency Situation. CONSTRAINT 2021. Communications in Computer and Information Science, vol 1402* (pp. 93–105). Springer. https://doi.org/10.1007/978-3-030-73696-5_10

## СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Дарья Сергеевна Быльева, bylieva_ds@spbstu.ru,
ORCID 0000-0002-7956-4647

Daria Bylieva, bylieva_ds@spbstu.ru,
ORCID 0000-0002-7956-4647

Статья поступила 9 января 2022
одобрена после рецензирования 16 февраля 2022
принята к публикации 28 февраля 2022

Received: 9 January 2022 /
Revised: 16 February 2022
Accepted: 28 February 2022

https://doi.org/10.48417/technolang.2022.01.12
Research article

# Language and Robots: from Relations to Processes of Relations

Cathrine Hasse (✉)
University of Aarhus, Nordre Ringgade 1, 8000 Aarhus C, Denmark
caha@edu.au.dk

## Abstract

The word „robot“ does not have a fixed meaning and human interactions with robots do not somehow bring it to the fore. Mark Coeckelbergh suggests as much when he presents linguistic interaction with robots as a process of becoming aware of a quasi-personal relation. A focus on material linguistic practices yields a very different story of shifting signifiers that are subject to human experiences of changing relations with robots. The material encounter with robots is prefigured by the cultural presence of robots in many stories from popular culture. These produce an anticipation of the human-like, quasi-personal qualities of robots and an initial willingness to embrace these. Over the course of time, however, and through linguistic encounters with robots, one rather learns that they are quite foreign and, finally, merely machines. – 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:** Processes of relations; Social robots; Socio-linguistic artefacts; Lev Vygotsky; Material-conceptual meaning

**Citation:** Hasse, C. (2022). Language and Robots: from Relations to Processes of Relations. *Technology and Language*, *3*(1), 127-135. https://doi.org/10.48417/technolang.2022.01.12

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.12
Научная статья

# Язык и роботы: От отношений к процессам отношений

Катрин Хассе (✉)
Орхусский университет, Нордре Ринггаде 1, 8000 Орхус С, Дания
caha@edu.au.dk

**Аннотация**
Слово “робот” не имеет фиксированного значения, и взаимодействие человека с роботами никак не влияет на это. Марк Кекельберг предполагает нечто подобное, когда представляет языковое взаимодействие с роботами как процесс осознания квазиличностных отношений. Сосредоточение внимания на материальных лингвистических практиках приводит к совершенно иной истории сдвига означающих, которые зависят от человеческого опыта в изменении отношений с роботами. Материальное столкновение с роботами предопределено культурным присутствием роботов во многих историях популярной культуры. Это вызывает ожидание человекоподобных, квазиличностных качеств роботов и первоначальную готовность принять их. Однако с течением времени и благодаря лингвистическим встречам с роботами вскоре выясняется, что он совершенно чуждый и, в конечном счете, просто машина. – 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:** Процессы отношений; Социальные роботы; Социолингвистические артефакты; Лев Выготский; Материально-понятийный смысл
**Для цитирования:** Hasse, C. Language and robots: from relations to processes of relations // *Technology and Language*. 2022. № 3(1). P. 127-135. https://doi.org/10.48417/technolang.2022.01.12

## INTRODUCTION

We live in a storied world, where our knowledges are formed by the practices we engage in (Ingold, 2011, p. 159). Out of these practices stories grow, proliferate and take root in new practices, where stories become imaginaries amalgamating materials, cultural meaning making and material words in new ways. Stories are always on the move because our social relations, including material-conceptual meaning making, never stops to rest.

The material word "robot" is a good place to begin an exploration of how imaginaries grow from local stories born out of local practices (Ingold, 2011, p. 159), travel worldwide and take on new meanings as they land in new realities.

In the article *You, robot: on the linguistic construction of artificial others,* Mark Coeckelbergh discuss how our use of language play a part in how humans can form social relations when they begin talking to robots. Furthermore, he speculates (Coeckelbergh, 2011).

We are increasingly talking to machines, whether social robots, AI driven bots answering questions about products or the like. I am completely in agreement with Coeckelbergh that we need to pay attention to linguistics and hermeneutics in human-robot relations and that we need to attend to how robots actually fare when implemented in practices. Furthermore we need, as Coeckelbergh argues, to pay attention to how robots designed as social-linguistic artefacts call forth special concerns in studies of human-robot relations. Do we address the social robots as an 'it' or a 'you' (Coeckelbergh, 2011, p. 68)? The article furthermore addresses important issues of an ethical and moral character. If robots come to appear to humans as quasi-others (Ihde, 1990) will humans stop viewing robots as tools controlled by humans? If we increasingly address robots as 'you' and have conversations with robots it may not only change our relation to robots, but it may also affect the way we speak to other humans (Coeckelbergh, 2011).

There are many deep issues to be dealt with in this text. In this response, I shall only address a few. First of all, I shall argue that processes of relations matter for how we speak with robots over time, when actual robots are implemented and put to use in human practices (Hasse, 2021).

It is in the meeting with these local practices that the material-conceptual understanding how robot-talk is put to the test. It is here that new stories emerge from travelling stories and new knowledges grow out of transformed word-meanings. I shall use examples from our own anthropological studies of robots implemented in health care in Danish nursing homes and rehabilitation centres (e.g. Hasse, 2013; Blond, 2019, Nickelsen 2020) and draw on my recent book on how humans material-conceptual meaning making change over time when we meet and engage with robot (Hasse, 2020b).

My first point here concerns the use of the rather outdated empirical examples in Coeckelbergh's argument. We have learned a lot more about human responses to implemented human-like robots since the mid 2000's.

My second, connected, point concerns the lack of focus on materials in Coeckelbergh's argument. Here I am not only thinking about the experience of meeting real nuts and bolts robots rather than media robots on screens, but mainly about how it is often overlooked that words like 'robots' are tied to moving concepts rather than fixed

representations. I contend that the moving processes of languages are not just linguistic–hermeneutic, but also material-conceptual.

Finally, even if Coeckelbergh discusses the larger issues of change and time, he is not concerned with the micro-processes of ongoing changes in local practices. A focus on practice shows that humanoid robots often come with stories of being able to communicate with people. However, over time many, but not all (Hasse, 2013), seem to lose interest as they come to perceive the robot-talk as 'empty' (Hasse, 2020b, p. 5; Camp, 2017).

## MEETING ROBOTS IN PRACTICE

Over the last 10 years a number of research fellows have studied experiments of moving humanoid robots from robot laboratories into people's everyday lives in Denmark (e.g. Blond, 2019; Bruun et al., 2015; Leeson, 2017; Nickelsen, 2018; Hasse, 2013, 2015, 2020a; Hasse and Søndergaard, 2019; Sorenson et al., 2019; Wallace, 2019). They show a consistent pattern of robots coming into Danish practices from Korea, Japan, Sweden, and USA with stories of autonomy, smart behavior and some with capability for communication. When they are implemented in everyday practices of nursing homes and rehabilitation centers, staff and citizens embark on a learning process which gradually changes their meaningful understanding of what the material word 'robot' means. The robot designers have more often than not made use of business model which present the robots on screen as "a better version of themselves" (Sorenson et al., 2019). The same kind of business model that has been thriving in the Silicon Valley culture (e.g. Griffith and Woo, 2022). So, what happens when humanoid robots jump off screen and begin talking to people in real life?

## THE CHANGE OF MATERIAL-CONCEPTUAL MEANING

We often overlook the importance of how material words get their meaning in everyday encounters. A word like 'robot' keep evolving as an anchor for conceptual meaning-making according to the learning theorist Lev Vygotsky, who I have used as my own anchor in my book of how ultra-social humans differ from machines (Hasse, 2020b). I partly build my argument up around some explorative sessions we did with Danish school children in 2015-16. In these ordeals we among other things uttered a sentence in which the material word robot was central: "*Draw a robot, or more if you would like to, that does something and maybe does something together with others*." (Hasse, 2020b, p. 228). We handed the children the same kind of pens and papers in the different schools we visited and though the material words in the sentence uttered was the same, their robots on the drawings showed a great variety. This variety mirrored the variety in the situations. The sentences the researchers spoke was in this sense never the same, though the material words came out in the same way. The children's drawings evolved in the social situation where the children experimented with all their potentials for understanding what the word robot might refer to as they and the materials were 'undergoing in doing' (Ingold, 2015). A group of children stood out when we

subsequently examined the drawings and interviewed the children. They had all drawn robots as much more machinelike than their classmates – and as it turned out they had all built robots in a special event at school or had a brother who built robots or the like. Their involvement in these practices spilled over in their drawings and conceptual understanding of robots and presented a different picture compared to that of other children. The majority of children drew very lifelike and companion-like robots that could talk, dance and watch movies – like the children themselves (Hasse, 2020b, p. 235). They drew on cultural storied resources like Star Wars, Wall-E etc. when they drew their robots. However, just like the meeting with real robots changed the meaning of the material word 'robot' for staff and inhabitants of nursing homes and rehabilitation centers, the children who had built robots themselves, formed another concept of robots spilling out in their drawings.

## CHANGES OF CONCEPTS VS FIXED REPRESENTATIONS

Posthumanist approaches have for some time now refuted that humans learn "representations" and fixed dichotomized categories albeit with different arguments (see for instance Barad, 2007, Tuin and Dolphijn, 2010). These approaches have emphasized that: "Language matters. Discourse matters. Culture matters. There is an important sense in which the only thing that doesn't seem to matter anymore is matter" (Barad 2007, p. 132). They have however also in many ways overlooked how humans form meaningful languages in phenomenological and psychological learning processes which connect material words, with material things as they transform our 'embodied minds' (Hasse, 2020b). Following a rethinking of Vygotsky's insights in cultural learning processes languages move with practices when material words anchor ever-changing concepts. Furthermore, concepts in their turn are what we use to perceive and think with (Vygotsky, 1987). When practical experiences move language, they also move perception and thinking.

There is never a direct link between a word and our thoughts – but a path that goes through a word-meaning (Vygotsky, 1987, p. 281). Word-meaning is at the core of Vygotsky's argument. It is formed in a collective world of social meaning-making, but I have also argued it is tied to material and embodied practices (Hasse, 2020b). The life world in which we merge with technological phenomena, as it is argued in for instance postphenomenology (Rosenberger and Verbeek, 2015; Ihde, 2002), is also a material-conceptual world of meaning making.

## DISCUSSION

As Coeckelbergh, I also acknowledge that human-robot relations are mediated by language and that relations change over time – a very important point often lacking in postphenomenological studies of human-technology relations. However these relations not only change over time as societal and cultural changes, but also change persons' perceptions and thinking in practices. In these practices, materials, including materials word, matter. Practical experiences with matter moves word-meaning and thus our

perception and thinking. It matters, in other words (!), where our own meaningful situated knowledges (Haraway, 1988) have taken us before we begin our linguistic-hermeneutics.

Research in robots is indeed increasingly becoming more interdisciplinary (Coeckelbergh, 2011, p. 62). However, even if many studies of social robotics uses methods from the social sciences, like anthropology, they can differ enormously from the way the robotic sciences for instance uses psychology to improve their robots. Anthropological studies of robots *can* have the purpose of improving robots, but just as often they study how the phenomenon of social robots unfold in meeting local practices and contrast with for instance movie representations.

Even if both Coeckelbergh and I argue for processes of relations, the processes of relations seem to go in opposite directions. Coeckelbergh (2011) argue: "People no longer consider the robot as a machine and start to refer to robots in personal terms. ''It'' becomes ''he'' or ''she'' (p 64). From our empirical data, I argue to the contrary, the processes goes the other way. First people conceptualise from the stories they have heard, and thus perceive and think a materialized robot like a human-like conversation partner, however over time, they become more and more aware of the robot as a machine (Hasse, in press)

Coeckelbergh draw on rather old studies of humanoid robot implementation (Turkle, 2005; Turkle et al., 2006) from when humanoid robots still came with a WAUGH factor in a storied world informed by media practices. Though I have also in my own research encountered an old woman, suffering from dementia, talking to an social robot, Paro, for a whole night long (Hasse, 2013) we have also seen that over time people lose interest in the social robots because they cannot communicate like humans (Hasse, 2020b, p. 5, Blond, 2019). The only robots people continue to have conversations with for longer periods of time are the teleoperated so-called Wizard of Oz robots (see Hasse, 2019, 2020b, and Sorenson et al., 2019).

All over the world we have all learned to deal patiently with the automated bots that replace the switch board ladies. We may want to exclaim a: "You idiot" to that kind of bots who simple have no clue what we are talking about when we ask to be directed to a human being we can talk to. However, the social robots we have seen in real life implementations shift our perception and thinking about social robots to acknowledgements of machines running on sensors and wires. However, there may be huge cultural differences in how people learn to adjust to robot talk. Japanese people allegedly are more prone to find conversations with robots easy. Humans have a propensity to do exactly what Coeckelbergh proposes. We animate our surroundings – and seem to stretch ourselves to make the materials come alive as humanlike (Hasse, 2015). Only in longer term learning processes where relations change over time, we sometimes realise the 'empty curiosity' of machines, and decide to give up on them (Blond, 2019, Hasse, 2020b, p. 5).

## CONCLUSION

I wholeheartedly agree with Coeckelbergh that we need to understand the connection between the social and language as more than representationalism and

constructivism. Where the extreme idealism overlooks a world with vira, climate changes and other kickbacks (Barad, 2007, p. 215) from an uncontrollable Gaia, the naïve realism found in much engineer work not only overlook the perceiving and thinking subject that create the robotic machines but also the people using them and affected by them.

Coeckelbergh's emphasis on the entanglements of linguistics and hermeneutics is spot on, nevertheless I think we need to dig a sod deeper. Behind hermeneutics we find processes of meaning-making tied to and tying material things and words. Behind linguistics we find material words tied to and tying material things and meaning. It is these entanglements that move in a storied world and it matters for meaning-making processes what kind of relations we engage in in practice.

## REFERENCES


Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Blond, L. (2019). *Dances with Robots: Understanding Social Robots in Practice* [PhD thesis. Aarhus University]. https://www.dasts.dk/wp-content/uploads/Dances-with-Robots_Lasse-Blond_Aarhus-University_2019_final.pdf

Bruun, M., Hasse, C., & Hanghøj, S. (2015). Studying Social Robots in Practiced Places. *Techne: Research in Philosophy and Technology*, *19*(2), 143-165. https://doi.org/10.5840/techne20159833

Camp, V. J. (2017, July 11). Review Jibo Social Robot. https://www.wired.com/2017/11/review-jibo-social-robot/

Coeckelbergh. M. (2011). You, Robot: on the Linguistic Construction of Artificial Others. *AI & society*, *26*(1), 61–69

Griffith, E., & Woo, E. (2022, January 3). Elizabeth Holmes is Found Guilty of Four Counts of Fraud. *The New York Times*. https://www.nytimes.com/2022/01/03/technology/elizabeth-holmes-guilty.html

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575-599. https://doi.org/10.2307/3178066

Hasse, C. (2013). Artefacts that Talk: Mediating Technologies as Multistable Signs and Tools. *Subjectivity*, *6*(1), 79–100. https://doi.org/10.1057/sub.2012.29

Hasse, C. (2015). Multistable roboethics. In J. K. B. O. Friis, & R. P. Crease (Eds.), *Technoscience and postphenomenology: the Manhattan papers* (pp. 169-188). Rowman & Littlefield Publishers. Postphenomenology and the philosophy of technology.

Hasse, C. (2019). The Vitruvian robot. *AI & Society*, *34*(1), 91–93. https://doi.org/10.1007/s00146-017-0701-z

Hasse, C. (2020a). How Robots Challenge Institutional Practices. *Learning, Culture and Social Interaction*, *26*, 100223. https://doi.org/10.1016/j.lcsi.2018.04.003

Hasse, C. (2020b). *Posthumanist Learning. What Robots and Cyborgs Teach us About Being Ultra-social*. Routledge.

Hasse, C. (2021). Material Hermeneutics as Cultural Learning: from Relations to Processes of Relations. *AI & Society*. https://doi.org/10.1007/s00146-021-01171-7

Hasse, C. (in press). Socratic ignorance in learning to work with technology. In Bound, H., Edwards, A. & Evans, K. (Eds.) *Workplace Learning for Changing Circumstances.* Taylor and Francis.

Hasse, C. and Søndergaard, D. M. (Eds.) (2019) *Designing Robots, Designing Humans.* Routledge**.**

Ihde, D. (1990). *Technology and the Lifeworld*. Indiana University Press.

Ihde, D. (2002). *Bodies in Technology*. University of Minnesota Press.

Ingold T (2015). *The Life of Lines*. Routledge

Ingold, T. (2011). *Being Alive: Essays on Movement, Knowledge and Description*. Routledge.

Leeson, C. (2017). *Anthropomorphic Robots on the Move. A Transformative Trajectory from Japan to Danish Healthcare* [Research output. PhD thesis, Copenhagen University]. https://www.egv.dk/images/Projekter/Projekter_2013/Christina_Leeson_PhD_thesis_2017.pdf

Nickelsen, N. C. M. (2020). Active Citizenship' and Feeding Assistive Robotics: a Crumbling Story? In C. Hasse, & D. M. Søndergaard (Eds.), *Designing robots, designing humans* (pp. 73-87). Routledge.

Rosenberger, R. & Verbeek, P.-P. (Eds.) (2015). *Postphenomenological Investigations: Essays on Human-technology Relations*. Lexington Books.

Sorenson, J. Zawieska, K. Vermeulen, B. et al. (2019). Perspectives on Robots. REELER report. *REELER Research Repository*. https://responsiblerobotics.eu/research/perspectives-on-robots/

Tuin, I. V. D., & Dolphijn, R. (2010). The Transversality of New Materialism. *Women: A Cultural Review*, *21*(2), 153–171. https://doi.org/10.1080/09574042.2010.488377

Turkle, S., Taggart, W., Kidd, C. D., & Daste, O. (2006). Relational Artifacts with Children and Elders: the Complexities of Cybercompanionship. *Connection Science*, *18*(4), 347–361. https://doi.org/10.1080/09540090600868912

Turkle S. (2005) *Relational Artifacts/children/elders: the Complexities of Cybercompanions*. ResearchGate. https://www.researchgate.net/publication/247381714_Relational_ArtifactsChildrenElders_The_Complexities_of_CyberCompanions

*Wallace, J.* (2019). *Ethics* and *Inscription* in *Social Robot Design*. *Paladyn*, 10(1), 66-76. https://doi.org/10.1515/pjbr-2019-0003

Vygotsky, L. S. (1987). Thinking and Speech. In R. W. Rieber & A. S. Carton (Eds.), *The Collected Works of L. S. Vygotsky. Vol. 1: Problems of General Psychology* (pp. 39–289). Plenum.

## СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Катрин Хассе, caha@edu.au.dk,
ORCID 0000-0002-2525-9540

Cathrine Hasse, caha@edu.au.dk,
ORCID 0000-0002-2525-9540

Статья поступила 26 января
одобрена после рецензирования 18 февраля
принята к публикации 28 февраля

Received: 26 January 2022/
Revised: 18 February 2021
Accepted: 28 February 2022

https://doi.org/10.48417/technolang.2022.01.13
Research article

# 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*

**Citation:** Li, Y. (2022). Affirming and Denying the Hybrid Character of Robots: Literary Investigations. *Technology and Language*, *3*(1), 136-146. https://doi.org/10.48417/technolang.2022.01.13

УДК 008: 62-529
https://doi.org/10.48417/technolang.2022.01.13
Научная статья

# 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”

**Для цитирования:** Li, Y. (2022). Affirming and Denying the Hybrid Character of Robots: Literary Investigations // Technology and Language. 2022. № 3(1). P. 136-146. https://doi.org/10.48417/technolang.2022.01.13

## 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:

> […] 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

[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.

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 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 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

Статья поступила 31 января 2022
одобрена после рецензирования 18 февраля 2022
принята к публикации 28 февраля 2022

Received: 31 January 2022
Revised: 18 February 2022
Accepted: 28 February 2022

https://doi.org/10.48417/technolang.2022.01.14
Research article

# Response: Language and Robots

Mark Coeckelbergh (✉)
University of Vienna, Universitätsring 1, 1010 Vienna, Austria
mark.coeckelbergh@univie.ac.at

## Abstract

Six commentaries on the paper "You, robot: on the linguistic construction of artificial others" articulate different points of view on the significance of linguistic interactions with robots. The author of the paper responds to each of these commentaries by highlighting salient differences. One of these regards the dangerously indeterminate notion of "quasi-other" and whether it should be maintained. Accordingly, the critical study of the linguistic aspects of human-robot relations implies a critical study of society and culture. Another salient difference concerns the question of deception and whether there is a distinction between real and perceived affordances. The prospect of AI systems creating language or co-authoring texts raises the question of the hermeneutic responsibility of humans. And regarding the missing dimension of temporality, studies of macro- and micro-level hermeneutic change become more important.

**Keywords:** Linguistic constructions of robots; Affordances; Hermeneutic responsibility; Post-phenomenology; Narrative

**Citation:** Coeckelbergh M. (2022). Response: Language and robots. *Technology and Language*, 3(1), 147-154. https://doi.org/10.48417/technolang.2022.01.14

УДК 1: 62-529
https://doi.org/10.48417/technolang.2022.01.14
Научная статья

# Ответ: Язык и роботы

Марк Кекельберг (✉)
Венский университет, Университет-ринг 1, 1010 Вена, Австрия
mark.coeckelbergh@univie.ac.at

**Аннотация**

Шесть комментариев к статье “Ты, робот: о лингвистическом конструировании искусственных других” формулируют разные точки зрения на значение лингвистических взаимодействий с роботами. Автор статьи отвечает на каждый из этих комментариев, выделяя существенные различия. Один из них касается опасно неопределенного понятия “квазидругого” и необходимости его замены. Критическое изучение лингвистических аспектов отношений человека и робота подразумевает критическое изучение общества и культуры. Еще одно заметное различие касается вопроса обмана и того, существует ли различие между реальными и воспринимаемыми аффордансами. Перспектива систем искусственного интеллекта, создающих язык или становящихся соавторами текстов, поднимает вопрос о герменевтической ответственности людей. Более важными становятся исследования герменевтических изменений на макро- и микроуровнях из-за недостающего измерения темпоральности.

**Ключевые слова:** Лингвистические конструкции роботов; Аффордансы; Герменевтическая ответственность; Постфеноменология; Нарратив

**Для цитирования:** Coeckelbergh M. Response: Language and robots // Technology and Language. 2022. № 3(1). P. 147-154. https://doi.org/10.48417/technolang.2022.01.14

It is a great honour for me to receive comments on my work (Coeckelbergh, 2011), and while in this concise response I cannot do justice to all the finesses of each of the pieces that came to me as unasked but pleasantly surprising and extremely welcome gifts, I am grateful for the opportunity to respond to some of the main arguments.

**Larissa Ullmann** (2022) wonders how to best describe the ontological hybridity of human-robot relations, that is, in a way that does not strictly separate subject and object. Borrowing from Don Ihde, I used the term "quasi-other," but Ullmann rightly argues that this is not always appropriate since the characteristics of robots are not necessarily comparable to humans. Therefore, she proposes a new, original linguistic term, 'sobject,' to describe the relationships between human subjects and robots. In this way she proposes to capture similarity without sameness. This intervention is creative and largely successful, although the "subject" aspect of "sobject" still risks to evoke the meaning of human subject, given that, in contrast to some animal ethicists, most of us do not spend much time on thinking about subjectivity in animals. We often assume that humans are the only subjects. Generally, this discussion illustrates how difficult it is to move beyond anthropomorphism and anthropocentrism with regard to these kinds of phenomena. To move beyond humans as the measure of all things (to use a statement from the ancient Greek philosopher Protagoras) turns out to be more difficult than expected. As von Xylander puts it nicely based on shopping metaphors: 'When I attribute subjectivity to a robot, it may dawn on me that I am the subject whose traces I've been following with my shopping cart' (von Xylander, 2022). And it is again language that is enabling that kind of movement. We can only gain critical awareness of what is happening once we consider what we are doing with language (when we are doing things with robots). For example, as both Ullmann and von Xylander show, when we "do robots" (that is, when we use them and talk about them and to them) we also "do gender." Although it is true that robots come in many forms, all this seems more applicable to humanoid and android robots than to others.

In her piece, **Cheryce von Xylander** (2022) claims that my analysis of the linguistic construction of artificial others sounds like a plea for the rights of robots. But this was neither my purpose nor my approach. Rather, by attending to the role of language in constructing what robots "are," we can take a critical philosophical attitude towards *any* arguments about the moral status of robots, including those who argue for robots rights. Hence my purpose was not a re-evaluation of the ontological status of robots but rather a meta-evaluation, or rather meta-description: a description and potentially a better understanding of how we evaluate robots. This was also my aim in articles on robot rights and related writings on moral status. For example, I do not favour a suspension of disbelief: I merely describe that this is how people deal with their experience of robots. I try to understand what is happening. However, von Xylander rightly points to 'the obscenity of our Anthropocene ecologies' and the predatory tactics of capitalism, which also impacts modern robotics and indeed risks to turn us into robots, already from its very conception. The Czech word *robota*, introduced into English by Čapek (1920) in his famous play, means forced labor, slavery. Again my use of the term "quasi-other" is rightly seen as dangerous, here since it is argued that it might justify using robots within a cannibalist-capitalist logic. Seen from this angle, I conclude that a critical study of the

linguistic aspects of human-robot relations must be connected to a critical study of society and culture. For example, I sympathize with the author's remark about sexbots and consumerist culture. This angle needs to be included in any robot ethics and robot politics that claims to be critical.

The question regarding personalism is more complicated. It is not clear to what extent a theory focused on virtue and habit needs to be personalist, and whether the conservative version of personalism rightly questioned by von Xylander (and indeed Bourdieu) is the only version one could have. It is good to be reminded of the tension here between Bourdieu's and, say, MacIntyre's theoretical commitments. But deleting the person altogether, as a poststructuralist reading of Bourdieu seems to do, also comes at a cost. More generally, this piece reminds me that for thinking about technology we need a balanced social theory, which enables analysis at all levels and, like Bourdieu, aims at connecting the levels in a meaningful and fruitful way.

**Leon Pezzica** (2022) rightly highlights that not only language but also the robot's characteristics construct robots. He then claims that my account does not enable a deception objection. His own affordance-based approach, by contrast, succeeds in this since it is based on experience and affordance-gaps. Pezzica is right that I problematize arguments for deception that are based on a distinction between what the robot "really" is (an object, a machine, etc.) and what the robot seems to be (a person, for example). What the robot "is", I argue, depends on the language we use to construct the status of the robot. However, I doubt whether that implies that we can no longer use the term deception at all in this context. We can still say that a person is deceived; all I claim is that it is problematic to say that something unreal is happening. For example, when an elderly person uses the baby seal robot PARO and thinks that it is a real animal, then that person is deceived. But the reason why has nothing to do with the absence of reality: her experience is real, the robot really makes baby seal sounds, and so on. But the language, experience, and design of the robot work together to create a performance that misleads the person in thinking that the robot is an animal. To further conceptualize what happens, Pezzica's proposal is very helpful: we can use the concept of affordance. What happens is about a relation between the properties of the robot and the capabilities of the agent. However, I hesitate again when Norman (2013) and Pezzica (2022) then make a distinction between perceived and real affordances. The PARO robot *really affords* being talked to in a way one could talk to a baby seal, for example through its embodiment and the sounds it makes. If the designer succeeds in giving the robot the ability to engage in a relevant form of communication, there is no "perceived" affordance; the affordance, the experience, and the performance are real. It just is not such an animal. But there is a similarity in both performance and affordance. And that performance and affordance is enabled by human beings, as Pezzica rightly argues based on Sætra. Deception thus only makes sense once we bring in that human role in enabling, creating, and participating in these performances. I thus welcome to add the concept of affordance to that of performance, but question whether it is necessary, desirable, and justifiable to make a distinction between real and perceived affordances. That being said, there can be differences between expected performances and actual performances. Such gaps are indeed problematic.

**Daria Bylieva** (2022) rightly and helpfully draws attention to the phenomenon that today language is also used by non-humans such as AI technology, and in a way that is not pre-determined by human uses and that may impact both language and social reality. Use of language does not only construct the representation of relationships but also transforms them. For example, children adjust their language when they speak to an avatar. On their part, machines do not necessarily talk like humans and become more human-like. For example, they may be genderless. They might even use words more effectively and consistently, or create an interlingua. Bylieva shows that it is not technically possible that AI creates texts from which it is no longer possible to say whether it is created by humans or machines. This will change both linguistic and social reality. That language is no longer a human monopoly is a very interesting thought. What new uses of language, and ultimately, what new languages will AI create? How will this change social reality? To conceptualize this, we can use the same framework I proposed: Just like humans, AI might also construct reality in particular ways, ways that are not neutral to what things "are". Like humans, their linguistic performances also co-construct what they and we talk about and talk to. However, it will remain up to humans to interpret and *make sense* of what is going on and what will be going on. As I stress in recent papers, humans are only participants in processes that create meaning. But they are a very important part in it: since meaning is only created through them, they carry the hermeneutic responsibility. Exercising this hermeneutic responsibility, however, will require us to become more attentive to the technological developments Bylieva describes. The time that we can afford to see machines as mere tools – in a linguistic sense or otherwise – is definitely over.

In line with my recent process- and narrative oriented work, **Cathrine Hasse** (2022) offers some critical remarks and objections that can be summarized as implying that "things have moved on" – both in a superficial and in a deeper sense – since the publication of my paper on the linguistic construction of artificial others. I agree that empirical examples could be updated (using Hasse's own well-respected recent work and that of others) and that (pre-existing) practical experience with robots shapes people's concept of robots. Hasse's most interesting point here – at least in my view and in the light of my current interest in time, movement, and technology – is that we must be concerned with 'the micro-processes of ongoing changes in local practices.' Practical experience moves language and thinking. This change does not only happen at the societal and cultural level, but also in material-linguistic practices and concrete situations. For example, people may lose interest in the robot over time. In my response to postphenomenology I emphasized the macro level of hermeneutic change precisely because postphenomenology often tends to not only forget about the temporal dimension of technology – Hasse acknowledges this and emphasizes that relations change over time – but also about the wider cultural and societal dimension. However, Hasse is right that it also matters what happens at the micro level. Here empirical research can help. Note that my article in question *does* give examples of what happens in human-robot interaction at the micro level; yet the time dimension is missing, and in line with my current work and research interests that can and must be remedied. For example, one could study how people speak differently about an assistive device with voice interface

such as Alexa after use of a few months; it might be that people lose interest or discover other uses. With Olya Kudina I have done some research on meaning-making in the case of such devices; here more attention could be given to the time dimension (Kudina & Coeckelbergh, 2021). Another point Hasse offers is that people conceptualize the robot on the basis of stories, then comes the experience. This is congruent with the experiments by for example Kate Darling in the mid 2010s: In an experiment on empathy, participants were given a backstory and then the researchers observed what the participants did to the robots. However, I do not quite see why this implies that Hasse and I would go in opposite directions: here too, language (in the form of narrative) constructs what the robot "is," that is, how we think about the robot and how we experience the robot, and what we then do with or to the robot. So first there is language, then the experience and the practice. However, perhaps this discussion is a kind of chicken and egg problem: language use must be seen as an inherent part of our experience and practices, and language without experience and practice is dead. In this case: someone sees a robot (with particular affordances – see Pezzica!) and experiences that robot, uses a particular kind of word, and in turn that language use again shapes the perception and experience. Or there might be first a narrative, then the experience, then a different narrative may emerge. What is important in all these cases is to conceptualize and show the mutual interrelation of language and meaning, language and experience, language and how we think about technology, and preferably at both micro and macro level. And with Hasse and in line with my current work on performance, time, and technology, one could add that in such studies and reflections it is vital to take into account the shifts and changes: the shifts and changes in language use and the shifts and changes in our perception, experience, and thinking about robots. Finally, Hasse rightly remarks that there is also cultural variation. In my recent work (but also earlier when I wrote about robots in Japan) I am sensitive to this, see for example my recent paper on 'The Ubuntu Robot' (Coeckelbergh, 2022). Yet here too, there is no static situation: cultures are always in movement, just as the micro processes and interactions that tie together uses of words and things and in this way make meaning and let meaning emerge. My recent work, which connects technology use with performance, narrative, time, and meaning, contributes to this project and I am open to collaboration with researchers, including Hasse, that look at these topics from a more empirical angle.

**Yue Li** (2022) investigates the hybrid character of robots from a literary angle (in particular work by McEwan and Kehlmann), sees writing and co-writing as linguistic interaction between humans and robots, and argues that studying this reveals the normative dimension of language. Even if the phenomenon I described (a turn towards use of the "you" when interacting with robots) and the tendency towards human-robot hybridity does not always occur in reality, such language use and the corresponding relationship are constructed in literature. What is added to my approach is the thesis that writing and co-writing constitutes particular interactions between humans and robots that are not limited to direct speech. While there are for example shifts from impersonal to personal pronouns (which constructs the machines in particular ways as presented in my paper), there is more going on. For example, in *Machines Like Me* (McEwan, 2019)

relationships are constituted by means of narration, self-reflection, and explanation. In *My Algorithm and Me* (Kehlmann, 2021) AI is co-writer, which also offers a different perspective. The boundary between humans and robots are blurred in various ways. Here too it is language that creates this ambiguity, fiction in this case. But in line with my current interest in narrative and technology, therefore, I welcome Yue Li's investigations as broadening and complementing the initial framework of my paper, next to being interesting in their own right. Fictional narratives are forms of language use which also construct what we think machines "are" and how we (inter)act with them. These narratives could be analysed in various ways and at various levels, for example at the grammatical level but also at the level of the structure of the narrative as a whole – consider here my work with Wessel Reijers inspired by Ricoeur (Reijers et al., 2021). Metaphors could also be interesting to study, as the author proposes. Moreover, Yue Li shows that fiction is a very helpful instrument to explore and research the ambiguities and confusions that might be created by technologies such as AI and robotics – even if they do not frequently occur today. For example, if and when AI gets better at writing (see also Bylieva's piece), what does and will it mean to co-write with AI? Will we be able to handle that hybridity? What does it imply for our self-image, as persons and as humanity? What does it mean for our knowledge about ourselves? For our relationships? What demons will we summon when AI might show us new, darker sides of the human mind? Or is this not possible? And what are the new, creative and indeed *poetic* possibilities?

## REFERENCES


Bylieva, D. (2022). Language of AI. *Technology and Language*, *3*(1), 111-126. https://doi.org/10.48417/technolang.2022.01.11

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

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

Coeckelbergh, M. (2022). The Ubuntu Robot: Towards a Relational Conceptual Framework for Intercultural Robotics. *Science and Engineering Ethics*, *28*(2), http//dx.doi.org/10.1007/s11948-022-00370-9

Hasse, C. (2022). Language and Robots: from Relations to Processes of Relations. *Technology and Language*, *3*(1), 127-135. https://doi.org/10.48417/technolang.2022.01.12

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

Kudina, O., & Coeckelbergh, M. (2021). "Alexa, define empowerment": Voice Assistants at Home, Appropriation and Technoperformances. *Journal of Information, Communication and Ethics in Society*, *19*(2), 299–312. https://doi.org/10.1108/JICES-06-2020-0072

Li, Y. (2022). Affirming and Denying the Hybrid Character of Robots: Literary Investigations. *Technology and Language*, *3*(1), 136-146. https://doi.org/10.48417/technolang.2022.01.13

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

Norman, D. A. (2013). *The Design of Everyday Things*. Basic Books.

Pezzica, L. (2022). On Talkwithability. Communicative Affordances and Robotic Deception. *Technology and Language*, *3*(1), 104-110. https://doi.org/10.48417/technolang.2021.04.10

Reijers, W., Romele, A., & Coeckelbergh, M. (2021). *Interpreting Technology: Ricoeur on Questions Concerning Ethics and Philosophy of Technology*. Rowman & Littlefield Publishers.

Ullmann, L. (2022). The Quasi-Other as a Sobject. *Technology and Language*, *3*(1), 76-81. https://doi.org/10.48417/technolang.2022.01.08

von Xylander, C. (2022). Quipping Equipment. Apropos of Robots and Kantian Chatbots: Commentary on Mark Coeckelbergh, "You, robot: on the linguistic construction of artificial others" (2011). *Technology and Language*, *3*(1), 82-103. https://doi.org/10.48417/technolang.2021.04.09

**СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR**

Марк Кекельберг, mark.coeckelbergh@univie.ac.at
ORCID 0000-0001-9576-1002

Mark Coeckelbergh, mark.coeckelbergh@univie.ac.at
ORCID 0000-0001-9576-1002

Статья поступила 10 марта 2022
одобрена после рецензирования 21 марта 2022
принята к публикации 21 марта 2022

Received: 10 March 2022
Revised: 21 March 2022
Accepted: 21 March 2022

# Содержание

Тема выпуска:

## Конструирование роботов в языке и культуре