Language of AI

Daria Bylieva (✉) ⓘD
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

Язык искусственного интеллекта

Дарья Быльева (✉) ⓘⅅ
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

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).
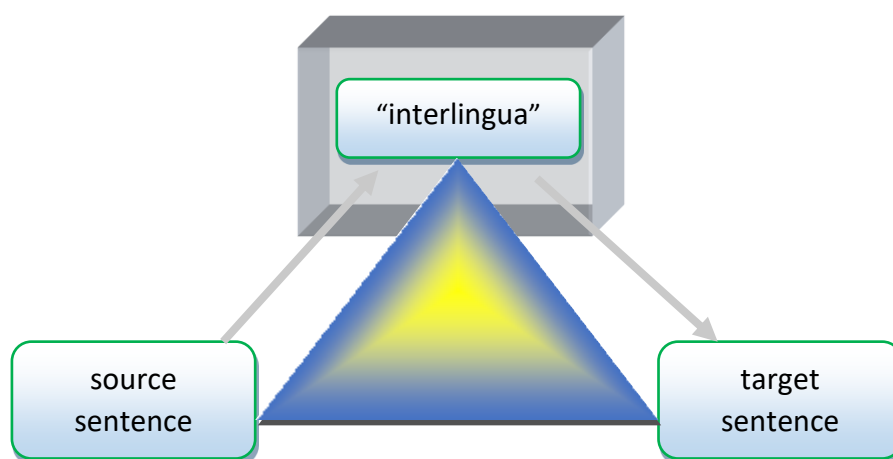

Figure 1. Neural Machine Translation

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)
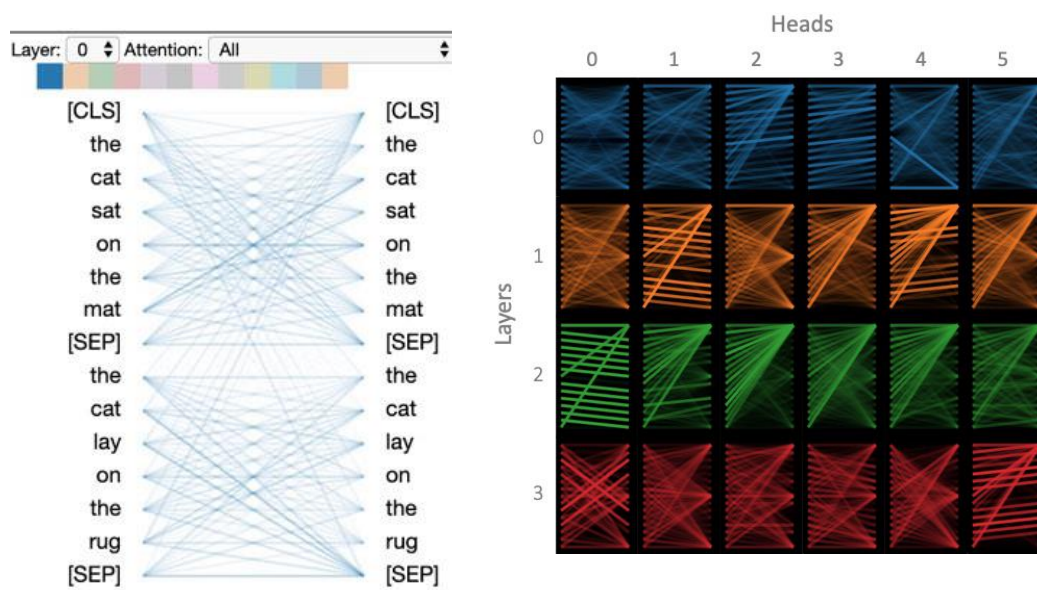

Figure 2. View for BERT Attention-head for inputs *the cat sat on the mat* and *the cat lay on the rug* (left) and Model view, for same inputs (excludes layers 4–11 and heads 6–11) (Vig, 2019).

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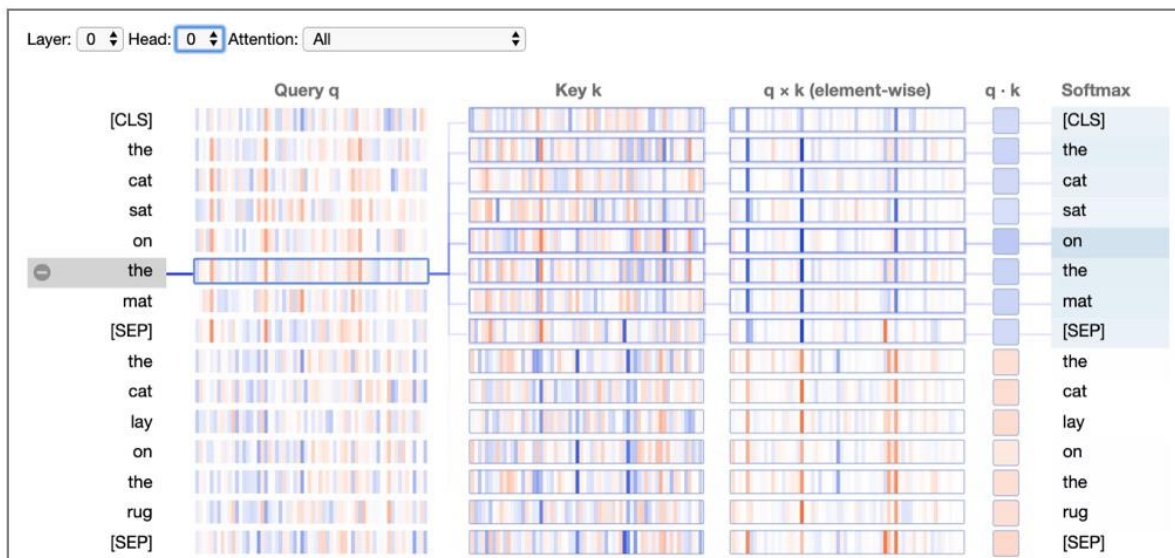
Figure 3. Neuron view of BERT for layer 0, head 0 (same one depicted in Figure 2). Positive and negative values are colored blue and orange, respectively, with color saturation based on magnitude of the value. As with the attention-head view, connecting lines are weighted based on attention between the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

Дарья Сергеевна Быльева, bylieva_ds@spbstu.ru,
ORCID 0000-0002-7956-4647

Daria Bylieva, bylieva_ds@spbstu.ru,
ORCID 0000-0002-7956-4647