

ALAN TURING: ARE WE CAPABLE OF DISTINGUISHING HUMAN FROM COMPUTER?

Nikitina K. A.

*Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus*

The idea behind the research was to investigate the correlation between a person's ability to differentiate a conversation with a human and on the other hand with a computer and the affect language, gender, age, and profession might have on it. To investigate this topic two people were interviewed both in English and in Russian as well as two bots, one of them was a Russian bot "P-БОТ" and the other one was an English bot "cleverbot". Looking at the contents of the dialogs, it can be concluded that participants tend to see vague answers as the ones given by the computer.

In today's world the AI is becoming more advanced very rapidly. Different forms of the AI await people on each corner of their daily life. From a facial recognition in the photos you upload to Facebook to your voice-assistants like Alexa or Siri [1; 2]. The idea of IA was born in the middle of the last century by a British mathematician Alan Turing. Even back there he was able to predict just how sophisticated the AI can grow to be on the future. So, he proposed a test known as "The Turing Test" as a way to tell how good the AI would be in what Turing called "The Imitation Game" [3; 4]. As of 2020 many AI's have passed The Turing Test, meaning that more than 30% of the judges decide that they are talking to a human over a five-minute keyboard conversation. The examples of these AI's can be Cleverbot, Elbot, Ultra Hal or Eugene Goosetman [5].

The proposed idea behind the research is that whether the language is your native or second it will affect your ability to differentiate between a human and a computer. Secondary the effect of gender and age on the ability to recognize a computer will also be examined. Aim of my research is to assess the dependence of correct answers percentage from language, gender, and age.

To investigate this topic two people were interviewed both in English (1_en, 3_en) and in Russian (1_ru, 3_ru) as well as two bots, one of them was a Russian bot "P-БОТ" [6] (2_ru) and the other one was an English bot "cleverbot" [7] (2_en). The information was gathered using a survey [8].

The survey that is created for the research consists of 3 parts. The first part of the survey gathered the information about the participants: age, gender and whether the participants usually work with people, tech or something else. The second part tested the participants in their ability to evaluate which of the presented dialogs took place with a person and which with a computer when the dialog itself is in their native language. The questions included: name, how are they doing, favorite film, what is important for them in friendship and what is the meaning of life. The third part tested the participants in their ability to evaluate which of the presented dialogs took place with a person and which with a computer when the dialog itself is in their second language. The questions included: their occupation, favorite book, favorite color, whether they have a pet and what is their dream.

When the participants were tested in Russian the percentage of the right answers was 31.5% for the first question, 21.3% for the second and 75.6% for the third. When they were tested in English the percentage of the right answers was 52.8% for the first question, 29.1% for the second and 58.3% for the third (fig. 1). In the end only 6 out of 127 gave correct answers to all the questions. 17 participants were correct in all the Russian questions, 23 participants were correct in all the English questions. 10 participants identified correctly all computer both times and 7 participants identified all the humans.

As the result most of participants failed to identify the computer correctly, and that in fact the computer's results were considered to be human ones more commonly than the actual human ones (78.7%). Despite being wrong, answers were more unevenly distributed in Russian (computer votes were 68.5%, 21.3%, and 24.4%) and in English they were closer together (computer votes were 47.2%, 29.1%, and 41.7%), that leads to a conclusion that when tested in English people were more inclined to guess than to actually think. Comparing answers and results given by the participants to the human dialogs the difference in the votes becomes obvious: in Russian the test subject student was voted to be human by only 31.5%, but in English human vote was 58.3%; the third test subject professor was voted to be human by 75.6% in Russian and 52.8% in English. The difference between them in Russian was 44.1% but in English it was only 5.5%. Looking at the contents of the dialogs, it can be concluded that participant tend to see vague answers as the ones given by the computer.

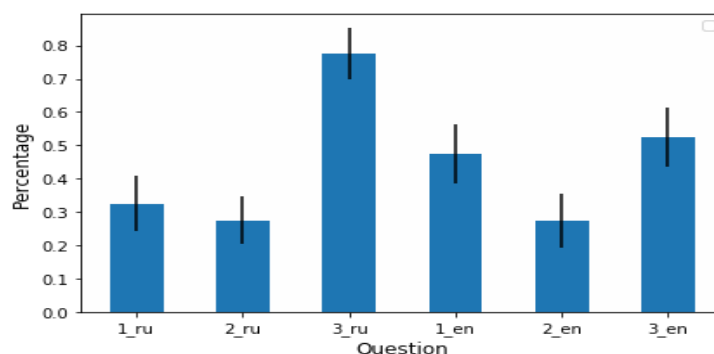


Figure 1 – Correct answers by questions

The first and the second age groups have almost no significant difference in the answers they gave, the only two notable differences being an 8.2% gap in identifying computer in Russian (first age group did better) and an 8.9% difference regarding the last English dialog (the second age group did better). But the third age group performed better when tested in English.

The difference is more visible in that case, being more obvious in Russian. 17.6% of females did better in the first question, but 8.7% males did better in the third (both were human dialogs). When tested in English the difference is more subtle, biggest one being that males did better in identifying computer by 8.2%.

By 6.1% engineers were better at identifying computer in Russian (no significant difference in English). But humanitarians were better at identifying human in the first question by 10.1%. That may lead to conclude that humanitarians see more vague answers as human ones when engineers do the opposite.

Statistically the probability of people correctly answering on 6 questions out of 6 is 2.0%, the actual case was 4.7% (6 out of 127) which suggests that participants did show some success in recognizing human and computer dialog. For three out of three the probability is 15.9%, in actual results Russian segment was 13.4% correct and English was 18.1%. The Russian one is a bit lower, but English is almost exactly the same as the probability.

The answers were more unevenly distributed in Russian than in English, which leads to a conclusion that when tested in English people were more inclined to guess than to think.

Looking at the contents of the dialogs, it can be concluded that participant tend to see vague answers as the ones given by the computer.

The number of people who answered correctly to all the questions is higher than the statistical probability.

References:

1. Adams R.L. 10 Powerful Examples Of Artificial Intelligence In Use Today [Электронный ресурс]. – URL: <https://www.forbes.com/sites/robertadams/2017/01/10/10-powerful-examples-of-artificial-intelligence-in-use-today/> (дата обращения: 12.04.2020).
2. Faggella D. Everyday Examples of Artificial Intelligence and Machine Learning [Электронный ресурс]. – URL: <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/> (дата обращения: 12.04.2020).
3. Moor J.H. Turing test / J.H. Moor // Encyclopedia of Computer Science. – GBR: John Wiley and Sons Ltd., 2003. – С. 1801–1802.
4. Traiger S. Making the Right Identification in the Turing Test / S. Traiger // The Turing Test: The Elusive Standard of Artificial Intelligence : Studies in Cognitive Systems / ed. J.H. Moor. – Dordrecht: Springer Netherlands, 2003. – P. 99-110.
5. Computer convinces panel it is human : Technology // BBC News. – 2014. – 9 Jun.
6. Русский онлайн чат бот, искусственный интеллект, виртуальный собеседник, чат с ботом [Электронный ресурс]. – URL: <http://p-bot.ru/index.html> (дата обращения: 12.04.2020).
7. Cleverbot [Электронный ресурс]. – URL: <http://www.cleverbot.com/> (дата обращения: 12.04.2020).
8. Тест Тьюринга [Электронный ресурс]. – URL: https://docs.google.com/forms/d/e/1FAIpQLSffa9jIFudEOo4SaFB-ktQnGuCvZn3PS6MTegtMm7gzKIZdg/viewform?usp=embed_facebook (дата обращения: 12.04.2020).