Aleksandra Przegalinska

Studying sentiment and inscribed values in human-chatbot conversations

In my research work, I focus on the social dimension of the development of artificial intelligence and on interdisciplinary research related to human-machine interactions, and in particular, human-bot interactions.

The main objective of my research project, which is extensive and already under way, is to attempt both quantitative and qualitative analysis of the interaction between man and the human being. Although such interdisciplinary research has already been carried out, it is still relatively scarce and, with the dynamic development of humanoid artificial intelligence, is becoming increasingly important. As part of my research I have so far worked with Prof. P. Gloor (MIT), and the results of which were published in *Future Generation Computer Systems* (https://www.sciencedirect.com/science/article/pii/S0167739X17312268), together with a research team from ALK, SWPS and MIT I carried out an evaluation of the human-chatbot interaction. In detail, the aim was to obtain an assessment of the quality of human-chatbot communication, taking into account the so-called Uncanny Valley effect, described in literature as the so called „Uncanny Valley" - a feeling of "strangeness" accompanying human interaction with humanoids. By integrating the methodology of psychological questionnaires with electrophysiological measurements, the research team managed to confirm the significant effect of Uncanny Valley in bots.

The next step in this extensive project is to explore the possibility of hindering human-chatbot interaction.

This currently carried out part of the study includes: (1) analysis of the sentiment in the semantics of human-machine conversation using systems such as Condor and IBM Watson (comparatively) to process samples of selected bot conversations, including state of the art chatbots (Alexa, Siri) as well as deep learning supported bots such as Duplex and Cortana (2)

conversational log analysis (3) discursive subjectivity analysis and bot transparency analysis, concerning flagging the interactions (signaling whether the user is talking to the machine or human) (5) analysis of values represented by leading chatbots serving as an interface with the user relationship through research on the priorities of the bot's operation (e.g. representing the user's interests towards the company or institution) and their attitude towards various kinds of moral behaviours. Because psychometric measurements can be very difficult to apply to (chat)bots due to their limited communication skills, I have decided to use standard questionnaires to study moral beliefs in a modified form. Bots are not able to respond to the scale, nor can they sort claims by key. They can however respond to single sentence statements, so only individual sentences from the questionnaires will be used to study moral beliefs and then asked in the form of questions to chatbots. I intend to use the statements of the following questionnaires: RVS - Yearbook Value Scale and PVQ - Portrait Value Questionnaire.

Expected effect is to show the character of human interaction with the bot on many levels: affective, linguistic, philosophical - in terms of subjectivity and ethics - transparency, values.