

Streszczenie

Celem niniejszej pracy jest zbadanie uwarunkowań poznawczych i społecznych antropomorfizacji wirtualnego agenta konwersacyjnego podczas interakcji krótkotrwałej. Celem jest również pogłębienie rozumienia zjawiska antropomorfizacji chatbotów, które wciąż rzadko jest obiektem badań. Jako antropomorfizację rozumie się rodzaj wnioskowania o ludzkich właściwościach obiektu, które są nieobserwowalne (Epley i in., 2007). Obejmują one między innymi cechy określane powszechnie jako *typowo ludzkie* oraz cechy osobowości. Przypisywanie tych właściwości współwystępuje z nasiloną reakcją emocjonalną, którą również uznano za miarę zjawiska.

Syntetyzując dotychczasową wiedzę nad zagadnieniem, przeprowadza się przegląd literatury z zakresu psychologii, filozofii, a także te na pograniczu informatyki (interakcje człowiek - robot, sztuczna inteligencja).

Analizowany stan wiedzy odnosi się do teorii psychologicznych, w tym do *Trójczynnikowej Teorii Antropomorfizacji* Epleya, Waytza i Cacioppo (2007), a także do badań psychologicznych nad antropomorfizacją i zjawiskami pokrewnymi, uwzględniając jednocześnie specyfikę obiektu, jakim jest chatbot. Na podstawie dokonanej analizy, charakteryzuje się zjawisko antropomorfizacji chatbota i identyfikuje czynniki warunkujące antropomorfizację, w tym jej determinanty społeczne oraz poznawcze.

Chatbot, jako obiekt cechujący się wysokim podobieństwem do człowieka, jest antropomorfizowany automatycznie - na poziomie nieświadomym, a także na poziomie mniej lub bardziej uświadomionym. Badania własne dotyczą bardziej uświadomionych aspektów zjawiska, uwidaczniających się w deklaracjach. Koncentrują się one na uwarunkowaniach poznawczych antropomorfizacji, takich jak: dostępność informacji na temat chatbota przed interakcją oraz wartości niejawne (orientacje życiowe) badane kwestionariuszem *Rzeczywiste – Nierzeczywiste* (Mudyń, 2010). Jako uwarunkowania społeczne traktuje się natomiast: pozostawanie w izolacji podczas interakcji z chatbotem, pozostawanie w obecności innej osoby obcej podczas interakcji z chatbotem oraz nasilenie poczucia samotności mierzonego *Skalą Samotności de Jong Gierveld* (de Jong-Gierveld i Kamphuls, 1985; Grygiel i in., 2013). Przeprowadzone badanie przebiega w warunkach interakcji krótkotrwałej: 10 - minutowej rozmowy z wirtualnym agentem na dowolny temat, po której badani wypełniają kwestionariusze. Eksperyment obejmuje trzy manipulacje eksperymentalne: w warunku E1

przed interakcją badani otrzymują informacje o chatbocie, w warunku E2 badani podejmują interakcję bez poprzedzającej jej informacji i pod nieobecność innych osób, w warunku E3 – bez informacji i w obecności innej, obcej osoby w zasięgu wzroku. W warunku kontrolnym badani nie mają informacji o chatbocie i podejmują z nim kontakt w obecności eksperymentatora. Antropomorfizację mierzy się kwestionariuszowo z użyciem *Individual Differences anthropomorphism questionnaire* (IDAQ; Waytz, Cacioppo, i in., 2010), listy cech osobowości autorstwa zespołu Haslama (Haslam i in., 2005) oraz skonstruowanej na użytek pracy *Skalę Uczuć*.

Wyniki dostarczają nowych informacji na temat struktury badanego zjawiska. Uzyskany układ korelacji wskazuje na jego niejednorodność. Wnioskuje się o obecności dwóch podtypów antropomorfizacji. Na podtyp pozytywny składają się pozytywne uczucia i odczuwanie uczuć ogółem oraz przypisywanie pozytywnych cech i cech ogółem, a także przypisywanie właściwości *typowo ludzkich*. Na podtyp negatywny składają się negatywne uczucia i przypisywanie negatywnych cech. Wyniki wskazują na ujemny wpływ dostarczonej informacji na aspekty podtypu negatywnego zjawiska i dodatni na aspekty podtypu pozytywnego – przypisywanie cech osobowości oraz uczucia. Odwrotne kierunki wpływu na badane zjawisko ma natomiast interakcja w obecności osoby obcej, a także nasilona samotność emocjonalna. Co więcej samotność, zarówno w aspekcie emocjonalnym jak i społecznym, koreluje dodatnio z podtypem negatywnym antropomorfizacji, ujemnie zaś z aspektami podtypu pozytywnego: uczuciami i przypisywanymi cechami osobowości. W grupie mężczyzn występuje dodatni związek między poczuciem samotności emocjonalnej, a przypisywaniem obiektowi cech *typowo ludzkich*. Wyniki nie wskazują na wpływ pozostawania w izolacji podczas interakcji krótkotrwałej na nasilenie antropomorfizacji. Wskazują natomiast, że na podstawie nasilenia społecznej orientacji życiowej można przewidywać zwiększone przypisywanie chatbotowi cech *typowo ludzkich*. Przypisywanie tych cech wiąże się także dodatnio ze skupieniem orientacji holistycznych. Natomiast polityczna orientacja życiowa współwystępuje z obniżeniem uczuć negatywnych podczas interakcji.

Na podstawie uzyskanych wyników wnioskuje się, że antropomorfizacja wirtualnego agenta występuje w dwóch podtypach, a nasilenie podtypu negatywnego współwystępuje z obniżonymi wskaźnikami podtypu pozytywnego. W analogii do procesów występujących w interakcjach międzyludzkich, wraz ze wzrostem podtypu negatywnego antropomorfizacji, może dochodzić do infrahumanizacji chatbota. W zakresie uwarunkowań poznawczych dostrzega się powiązanie wartości niejawnych – orientacji życiowych z antropomorfizacją. Identyfikuje się możliwą rolę, jaką odgrywają wartości niejawne w występowaniu zjawiska w

jego aspekcie poznawczym i bezpośrednim. Zauważa się, że wpływ dostarczonej informacji na antropomorfizację odbiega od początkowych założeń. Informacja może pełnić tu podwójną rolę: zwiększanie wiedzy o obiekcie, ale także zwiększenie komfortu interakcji. Wpływ obecności innej osoby podczas interakcji na aspekt negatywny antropomorfizacji może być związany z obniżeniem komfortu interakcji oraz *efektem interferencji społecznej* (Baron, 1986). Zauważa się też, że samotność, głównie w jej wymiarze emocjonalnym, wykazuje najwięcej powiązań z antropomorfizacją oraz stanowi jej predyktor. I co więcej, samotność wiąże się dodatnio z podtypem negatywnym zjawiska, co jest wynikiem nowym na tle dotychczasowych badań. Wydaje się, że osoby samotne przekładają na chatbota negatywne uczucia związane z innymi ludźmi, co z kolei może prowadzić do infrahumanizacji. Interesujący jest dodatni związek samotności emocjonalnej z przypisywaniem właściwości typowo ludzkich u mężczyzn – zależność ta może być wyjaśniana w kategoriach odmiennej niż u kobiet emocjonalności mężczyzn oraz odmiennych postaw wobec robotów. Wnioski z wyników uzupełnione są przez obserwacje wynikające z przeglądu treści rozmów z chatbotem, przeprowadzonych na potrzeby badania.

Wykryte zależności dostarczają wyników nowych, z racji uwzględnienia niestosowanych wcześniej narzędzi i manipulacji eksperymentalnych. W związku z tym otwierają drogę do dalszych badań nad zagadnieniem. Mogą pomóc także w projektowaniu interakcji z chatbotem tak, aby lepiej odpowiadać na oczekiwania użytkowników.

André de Torre

Abstract

The purpose of this paper is to investigate the cognitive and social determinants of anthropomorphism in short-term interactions with a virtual conversational agent. It also aims to deepen the understanding of the phenomenon of chatbots' anthropomorphism, which is still rarely studied. Anthropomorphism is defined as a form of inference about the human-like unobservable characteristics of an object (Epley et al., 2007). These includes, but is not limited to, characteristics commonly described as typically human and personality traits. Attribution of these properties co-occurs with heightened emotional response, which is also considered a measure of the phenomenon.

Synthesizing existing knowledge on the subject, a literature review is conducted in the fields of psychology, philosophy, as well as interdisciplinary research in human-robot interaction and artificial intelligence.

The analyzed state of knowledge refers to psychological theories, including the *Three Factor Theory of Anthropomorphism* by Epley, Waytz and Cacioppo (2007), as well as psychological studies on anthropomorphism and related phenomena, taking into account the specificity of the chatbot as the object of study. Based on the analysis, the phenomenon of chatbot anthropomorphism is characterized and the factors determining anthropomorphism are identified, including its social and cognitive determinants.

The chatbot, as an object characterized by a high resemblance to a human, is anthropomorphized automatically - at the unconscious level, as well as at a more or less conscious level. The study focuses on the more conscious aspects of the phenomenon manifested in declarations. It is focused on the cognitive determinants of anthropomorphism, such as pre-interaction information about the chatbot and implicit values (life orientations) measured by the Real-Unreal questionnaire (Mudyń, 2010). The social determinants considered are isolation during interaction with the chatbot, presence of another unfamiliar person during interaction with the chatbot, and the intensity of loneliness measured by the de Jong-Gierveld Loneliness Scale (de Jong-Gierveld and Kamphuls, 1985; Grygiel et al., 2013). The conducted study takes place in the context of short-term interaction: a 10-minute conversation with a virtual agent on any topic, after which the subjects complete questionnaires. The experiment includes three experimental manipulations. In condition E1, before the interaction, the subjects receive information about the chatbot. In condition E2, participants engage in interaction without prior information and in the absence of other individuals. In

condition E3, participants interact with a chatbot without preceding information but in the presence of another unfamiliar person within sight. In the control condition, subjects have no information about the chatbot and interact with it in the presence of the experimenter. Anthropomorphism is measured through questionnaires using the Individual Differences Anthropomorphism Questionnaire (IDAQ; Waytz, Cacioppo, et al., 2010), a list of personality traits developed by the Haslam team (Haslam et al., 2005), and the Feelings Scale constructed for the purpose of this study.

The results provide new insights into the structure of the investigated phenomenon. The obtained correlation pattern indicates its heterogeneity. The presence of two subtypes of anthropomorphism is inferred. The positive subtype is characterized by positive feelings, overall emotional experience, attribution of positive traits, traits in general, attribution of typically human characteristics. The negative subtype is characterized by negative feelings and attribution of negative traits. The results indicate a negative impact of provided information on aspects of the negative subtype of the phenomenon and a positive impact on aspects of the positive subtype, specifically the attribution of personality traits and emotions. Conversely, interaction in the presence of another person and heightened emotional loneliness have opposite effects on the studied phenomenon. Furthermore, emotional and social loneliness positively correlate with the negative subtype of anthropomorphism and negatively correlate with the positive aspects, such as feelings and attributed personality traits. In the male group, there is a positive association between emotional loneliness and the attribution of typically human characteristics to the object. The results do not indicate an influence of isolation during short-term interaction on the intensity of anthropomorphism. Instead, it is observed that increased attribution of typically human characteristics to the chatbot can be predicted based on the intensity of social orientation. Attribution of these traits is also positively associated with holistic orientations. Political life orientation, on the other hand, co-occurs with reduced negative feelings during interaction.

Based on the obtained results, it is concluded that anthropomorphism of a virtual agent occurs in two subtypes, and the intensity of the negative subtype is negatively related to the positive subtype. In analogy to the processes that occur in human interactions, as the negative subtype of anthropomorphism increases, chatbot infrahumanization may occur. Regarding cognitive determinants, a connection is observed between implicit values (life orientations) and anthropomorphism. The study identifies a possible role of implicit values in the cognitive and immediate aspects of anthropomorphism. It is noted that the influence of the provided information on anthropomorphism deviates from the initial assumptions. Information can play

a dual role here: increasing knowledge of the object, but also increasing the comfort of the interaction. The impact of the presence of another person during interaction on the negative aspect of anthropomorphism may be related to decreased interaction comfort and the social interference effect (Baron, 1986). It is also noted that loneliness, mainly in its emotional dimension, shows the most connections with anthropomorphism and is a predictor of it. Moreover, loneliness is positively related to the negative subtype of the phenomenon, which is a novel finding compared to previous research. It appears that lonely individuals transfer negative feelings associated with other people onto the chatbot, which, in turn, can lead to infrahumanization. Of interest is the positive association of emotional loneliness with the attribution of typically human characteristics in men - this correlation can be explained in terms of men's different emotionality than women and different attitudes towards robots. The conclusions of the results are supplemented by observations derived from a review of the content of the chatbot conversations conducted for the study.

The presented findings are new due to the inclusion of previously unused tools and experimental manipulations. Therefore, they open the way for further research on the topic. They can also help in designing interactions with chatbots that better meet user expectations.

Anne de Torre