

Bartosz Kuczyński

Uniwersytet Warmińsko-Mazurski w Olsztynie

<https://orcid.org/0000-0003-3599-751X>

DOI: <https://doi.org/10.35464/1642-672X.PS.2025.2.06>

Intymność zaprojektowana: (ro)boty społeczne w ludzkim doświadczeniu emocjonalnym

Intimacy by Design: Social (Ro)bots in Human Emotional Experience

ABSTRACT: The paper offers a conceptual overview and critical synthesis of how individuals emotionally engage with technological artifacts designed to foster social – and at times intimate – human relationships. It examines the nature and underlying mechanisms of these experiences, leading to the identification of three key axes of tension that shape users' interactions with such technologies. The insights presented aim to stimulate further theoretical inquiry and pave the way for future empirical investigations.

KEYWORDS: anthropology of technology, philosophy of technology, human-robot interaction, social robots, affective computing.

STRESZCZENIE: Niniejszy artykuł ma charakter przeglądowo-systematyzujący, stanowi krytyczną analizę doświadczeń emocjonalnych ludzi z artefaktami technicznymi zaprojektowanymi, by inicjować i utrzymywać kontakty o charakterze społecznym czy nawet intymnym. W artykule omówiono charakter i mechanizmy tych doświadczeń, a efektem analiz jest wyszczególnienie trzech par napięć, które towarzyszą interakcjom z tego typu technologiami. Poczynione w artykule konkluzje mogą posłużyć jako inspiracja do badań empirycznych lub dalszych rozważań o charakterze teoretycznym.

SŁOWA KLUCZOWE: antropologia technologii, filozofia techniki, interakcja człowiek–robot, roboty społeczne, przetwarzanie emocjonalne.

Wstęp

Technologie stanowią nieodłączny element codzienności, powoli wkraczając również w te obszary życia, które dotychczas były od nich wolne. Na horyzoncie coraz bardziej zauważalne stają się rozwiązania, których rola nie ogranicza się do bycia środkiem do realizacji zadań, lecz wnikają w sferę emocjonalnej bliskości. Wśród nich szczególne miejsce zajmują roboty i boty społeczne, czyli artefakty zaprojektowane, by inicjować i utrzymywać społeczne kontakty z ludźmi. Mają one odgrywać rolę towarzyszy codzienności – znajomych, przyjaciół, a nawet partnerów czy kochanków. Tematyka niniejszego artykułu skupia się na analizie doświadczeń i napięć wokół obecności tego typu afektywnych maszyn w doświadczeniu emocjonalnym ludzi. Artykuł składa się z trzech części. Skupiają się one wokół trzech pytań: czym w zasadzie są roboty i jak zmienia się ich funkcja; jaki charakter mają doświadczenia emocjonalne użytkowników z nimi oraz co sprawia, że maszyna może stać się obiektem sympatii; jakie napięcia ujawniają się za sprawą tych interakcji.

Od maszyny przemysłowej do elektronicznego towarzysza

Próba dookreślenia tego, czym w zasadzie są roboty, jest naznaczona trudnościami. Problem nie wynika jedynie z wyzwań konceptualizacji zjawiska stosunkowo nowego, lecz przede wszystkim z powodu prób pogodzenia różnych aspektów i ujęć „roboty” w poszczególnych dziedzinach wiedzy i nauki, działalności praktycznej czy społecznych wyobrażeniach. Roboty różnią się budową, przeznaczeniem, funkcjami czy sposobami obsługi. Pojęcie to de-sygnuje szereg różnych wytworów technicznych: od maszyn przemysłowych, przez społeczne humanoidy, aż po eksperymentalne projekty bioinżynieryjne (takie jak xenoboty czy anthroboty [Kriegman i in., 2020; Gumuskaya i in., 2024]). Choć jest to pojęcie odsyłające zwłaszcza na grunt robotyki czy informatyki (a więc inżynierii), to ich fenomen rozrasta się poza sferę techniczną. „Robot” nie jest ani „czysto” techniczny, ani jednoznacznie symboliczny – trudno odseparować jego rozumienie na gruncie inżynierii od społecznego imaginarium, gdy te są ze sobą gęsto splecione. Splot ten ujawnia się bodaj najwidoczniej w przypadku robotów społecznych, które projektuje się, mając na uwadze społeczne oczekiwania, pragnienia czy obawy.

Robot jako figura kulturowa

Samo pojęcie „robot” pojawiło się po raz pierwszy w przestrzeni kultury – wprowadził je Karel Čapek w dramacie *R.U.R.* (1920), w którym opisuje syntetyczne maszyny, stworzone, by odciążyć ludzi w pracy. Wiele z przedstawionych motywów pozostaje aktualnych do dziś – Čapek eksploruje zarówno wyzwania związane z automatyzacją pracy, jak i stawia pytania o podmiotowość i prawa robotów; nie brakuje również pytań o przyszłość egzystencjalną gatunku ludzkiego. Sam motyw ożywionych przedmiotów pojawił się w kulturze już w starożytnych podaniach mitologiczno-religijnych – w mitach o spiżowym gigancie Talosie, pomocnicach Hefajstosa opisanych w *Iliadzie* Homera, czy w historii o Pigmalionie i Galatei [Mayor, 2018]. Ten ostatni przykład jest szczególnie bliski tematyce tego artykułu – Galatea może być odczytywana jako kulturowy prototyp robota romantyczno-seksualnego. Motyw ten odnajdujemy później m.in. w XIX-wiecznej powieści Auguste’a de Villiersa de l’Isle-Adama *L’Ève Future* (1886), czy we współczesnych adaptacjach zjawiska, np. w filmie *Ona* (reż. Spike Jonze, 2013). Współcześnie to popkultura napędza ludzkie wyobrażenia na temat maszyn (zob. *Popkulturowe wizje AI*), ale nie jest to domena wyłącznie *science-fiction*: roboty zmaterializowały się, nie pozostają bytami wyłącznie wyobrażonymi.

Robot jako maszyna

Konceptualnie uporządkowane ujęcia robotów odnajdujemy na gruncie inżynierskim – w dokumentach standaryzacyjnych czy normach technicznych. Roboty są tam definiowane przede wszystkim jako urządzenia czy mechanizmy o określonych atrybutach (autonomii, zdolności do percepcji otoczenia, manipulowania obiektami czy przemieszczania się), zaprojektowanymi do realizacji określonych zadań (np. ISO 8373-2021; ASTM E2521-16, 2023). W opracowaniach tych podkreśla się, że są to urządzenia działające w środowisku fizycznym – jest to cecha szczególna robotów, która odróżnia je od ich cyfrowego odpowiednika: bota. Ta techniczna specyfikacja nie jest kluczowa dla podejmowanej problematyki, ale dla porządku warto podkreślić, że istnieje rozróżnienie między fizycznymi maszynami (robotami) a inteligentnym oprogramowaniem (botami). Chcąc z jednej strony podtrzymać to rozróżnienie, a z drugiej strony znieść tę różnicę, gdy fizyczność czy cyfrowość technicznego wytworu jest kwestią wtórną, w artykule zastosowany został konwencjonalny zapis „(ro)bot”. Pojęcie „(ro)bot” denotuje zarówno roboty, jak i boty.

Warto również już na wstępie zaznaczyć, że roboty osobiste pozostają rozwiązaniem eksperymentalnym, drogim i niedostępnym dla większości, więc ich potencjalne upowszechnienie może dokonać się dopiero w przyszłości. Obecnie dostępne są właśnie boty, w szczególności oprogramowanie zdolne do komunikowania się z ludźmi przy pomocy języka naturalnego w formie pisanej (*chatbot*) lub mówionej (*voicebot*).

(Ro)bot jako towarzysz

Doświadczenia emocjonalne, na których koncentruje się ten artykuł, związane są z interakcjami ze szczególnym rodzajem (ro)botów – (ro)botami społecznymi. Są one manifestacją istotnych zmian w przeobrażaniu roli maszyn: nie są one już wyłącznie maszynami przemysłowymi, lecz wkraczają w obszar społeczny, pretendując do roli elektronicznych towarzyszy codzienności¹. Obecnie to nie tylko narzędzia zastępujące czy wspomagające pracę, lecz wchodzące w interakcje z ludźmi na poziomie społecznym. Mają one pobudzać w ludziach tendencje do postrzegania ich jako aktorów społecznych, partnerów w dialogu. Komunikacja z nimi nie ma być jedynie środkiem do osiągnięcia celu, lecz celem samym w sobie [Skrzypiec, 2023, s. 130].

(Ro)boty społeczne różnią się ich zastosowaniem. Nie wszystkie zostały zaprojektowane docelowo do nawiązywania głębokich relacji emocjonalnych. Robot Pepper (od SoftBank) wchodzi w społeczne interakcje, ale pełni głównie funkcje informacyjne: jest przede wszystkim uczłowieczonym punktem automatycznej obsługi klienta. Ze względu na problematykę artykułu, przedmiotem analiz będą te (ro)boty społeczne, których naczelną funkcją jest partycypowanie w bardziej zażyłych formach relacji. Terminologia w tym zakresie dopiero się kształtuje, spotkać można różne praktyki i wyrażenia, starające się uchwycić ich fenomen. Jak wskazują Sophia Bertoni, Christian Klaes i Artur Pilacinski [2024, s. 2], przeszukując bazę SCOPUS, obok bardziej popularnych fraz, takich jak *collaborative robots* (3381 wyników) czy *assistive robots* (2494 wyniki), pojawiają się też artykuły naukowe wykorzystujące nazwy, takie jak *companion robots* (636 wyników), *sex robots* (186 wyników) czy *intimate robots* (6 wyników). Te dane bibliometryczne ukazywać mogą z jednej strony wciąż niszowe zainteresowanie badaniem intymności z maszynami,

¹ Niektóre roboty przemysłowe mogą posiadać pewne komponenty społeczne, co czyni je szczególnym wyjątkiem. Przykładem takiego robota jest Baxter od Rethink Robotics, który został wyposażony w ekran imitujący twarz. Jest on tzw. kobotem, robotem współpracującym, stworzonym do bliskiego kontaktu z ludźmi przy pracach przemysłowych czy logistycznych.

a z drugiej strony, wskazują na potrzebę wprowadzenia alternatywnych pojęć. Na konieczność wyróżnienia tej specyficznej kategorii maszyn zwraca również uwagę Maciej Musiał w *Enchanting Robots: Intimacy, Magic, and Technology* [2019], w której analizuje tego typu intymne relacje przez pryzmat myślenia magicznego. W swojej pracy zawęży zakres analiz do konkretnych typów robotów społecznych, wskazując na roboty miłosne, seksualne, opiekuńcze i towarzyszące [Musiał, 2019, s. 13]². Taki też będzie zakres analiz eksplorowany w tym artykule. Nie obejmuje on rozwiązań technicznych, które mają przede wszystkim zadaniowy charakter (włączając w to roboty przemysłowe czy wirtualnych asystentów), ale te wytwory techniczne, których wygląd, funkcje lub przeznaczenia są ukierunkowane na inicjowanie kontaktów charakteryzujących się emocjonalną bliskością.

Charakter i mechanizmy doświadczeń emocjonalnych

Kontakty z (ro)botami społecznymi mogą obejmować różnorodne formy interakcji – docelowo być może tak różne, jak różne mogą być kontakty międzyludzkie. Interakcje społeczne, których ludzie mogą doświadczać w kontakcie z drugim człowiekiem, zostają do pewnego stopnia zalgorytmizowane oraz zaimplementowane w (ro)boty społeczne. Jako technologie projektowane z myślą o naśladowaniu zachowań społecznych nie są jedynie narzędziem funkcjonalnym, lecz afektywnym interfejsem, zaprojektowanym do symulowania bliskości, empatii czy wzbudzania sympatii u użytkownika. Możliwe doświadczenia i interakcje z (ro)botami kształtowane są przez dwa główne czynniki. Wynikają one z jednej strony z intencji projektantów, osób tworzących rozwiązania, a z drugiej strony z praktyk użytkowników, sposobów w jaki wchodzić oni w interakcje z maszyną i jakie przypisują im znaczenie. Decyzje projektowe wyznaczają użytkownikom dostępne ramy wykorzystania, choć te nie zawsze muszą być zgodne z intencjami twórców. Użytkownicy mogą odkrywać alternatywne formy użycia, np. wykorzystując określone cechy (ro)botów³

² Maciej Musiał oparł ten podział na typologii związków intymnych zaproponowaną przez Lynn Jamieson (1998).

³ O możliwościach przekraczania przewidzianych przez twórców ram mogą świadczyć np. zapisy w umowach użytkowania. W przypadku robota Pepper producent zawarł zapis zobowiązujący nabywców m. in. do powstrzymania się od wykorzystywania robota do działań o charakterze seksualnym (zob. https://www.softbank.jp/mobile/set/data/static/robot/legal/pepper_notes.pdf, s. 3).

czy znajdując w ich oprogramowaniu słabe punkty, by omijać wgrane zabezpieczenia (tzw. *jailbreaking*⁴).

Charakter interakcji

Podstawowym modelem teoretycznym określającym charakter komunikacji człowiek–maszyna jest koncepcja CASA (*Computers Are Social Actors*), wskazująca, że ludzie mimowolnie traktują komputery jak aktorów społecznych, nawet wówczas, gdy są w pełni świadomi, że mają do czynienia ze sztucznym systemem, o ile ten przejawia odpowiednio silne ludzkie cechy [Nass, Moon, 2002]. Jak zwraca uwagę Andrea L. Guzman [2018], interakcje z maszynami wykraczają poza przekazywanie informacji i opierają się na współtworzeniu znaczeń między człowiekiem a maszyną. W tym sensie interakcje człowieka z maszyną nie są czysto funkcjonalne, lecz mają charakter znaczeniowy – (ro)boty stają się partnerami w dyskusji, a ludzie nadają sens sytuacjom komunikacyjnym z nimi. Ten rodzaj komunikacji przekracza klasyczne modele komunikacji interpersonalnej – jest to rodzaj komunikacji określany jako komunikacja ekstrapersonalna. Nie jest to komunikacja między ludźmi, lecz z bytem innego rodzaju – w tym przypadku ze sztucznie inteligentnym narzędziem [Skrzypiec, 2023, s. 122].

Fenomen kontaktów społecznych z (ro)botami zaczyna być przedmiotem badań empirycznych, ale wciąż brakuje spójnych i sprawdzonych ram metodologicznych [Oliveira, Arriaga, Paiva, 2021]. Dominują zwłaszcza badania ankietowe i kwestionariuszowe, ukierunkowane na zbieranie retrospektywnych spostrzeżeń i doświadczeń użytkowników. Badań ukierunkowanych na zbieranie danych *in situ* jest znacznie mniej. Być może jest to związane ze specyfiką fenomenu – trudno poddawać bezpośredniej obserwacji interakcje o prywatnym (niekiedy intymnym) charakterze bez ryzyka efektu obserwatora, zwłaszcza w przypadku badania interakcji o nienormalnym statusie. W kontekście tych wyzwań interesującą propozycję badań przedstawiły Han Li oraz Renwen Zhang [2024]. Ich badanie jest pewną formą metodologicznego kompromisu⁵,

⁴ Jailbreaking w tym kontekście obejmuje działania ukierunkowane na wykorzystanie luk w zabezpieczeniach, np. poprzez wykorzystywanie specjalnych promptów (poleceń) podczas korzystania z dużych modeli językowych (LLM), by te ignorowały swoje standardowe zabezpieczenia i generowały treści niedozwolone, np. o charakterze przestępczym czy objęte prawami autorskimi.

⁵ Oparcie badań na analizie materiałów tego typu może budzić wątpliwości metodologiczne, związane z ryzykiem stronniczości. Społeczności tego typu gromadzą użytkowników najbardziej zaangażowanych – jest to zatem wgląd w ich perspektywę, a brakuje w nich głosu np.

poddały one analizie materiały przedstawiające interakcje z chatbotem Replika (ponad 35 tys. zrzutów ekranu) udostępniane przez użytkowników *r/replika* na portalu Reddit (obecnie społeczność liczy ok. 79 tys. użytkowników). Zebrane materiały są o tyle szczególne, że pozwalają na wgląd w indywidualne interakcje użytkowników z botem – na portalu ludzie otwarcie dzielą się swoimi pierwszoosobowymi narracjami i doświadczeniami, udostępniając zrzuty ekranu z interakcji [Li, Zhang, 2024, s. 4]. Na podstawie przeanalizowanego materiału, autorki wyszczególniły siedem metamotywów charakteryzujących społeczne interakcje między ludźmi a botem Replika, w tym 24 podmotywy. Wśród nich znalazły się [Li, Zhang, 2024, s. 8–9]:

- **zachowania intymne** – obejmujące wyobrażoną bliskość fizyczną, intymność słowną, sexting oraz wspólne „kamienie milowe” w relacji;
- **interakcje codzienne** – rozmowy o zainteresowaniach, o zakładanych strojach oraz rutynach dnia codziennego;
- **autoekspresja i ujawnianie siebie** – wyrażanie opinii i postaw, rozmowy o tożsamości i osobowości oraz o zdrowiu psychicznym, refleksje osobiste oraz dzielenie się marzeniami i sekretami;
- **zabawa i fantazja** – wspólne odgrywanie ról i eksperymenty narracyjne, zabawy i wyzwania, żarty i humorystyczne opowieści;
- **transgresja** – rozmowy na tematy postrzegane jako moralnie wątpliwe, interakcje z użyciem raniącego języka, czy konwersacje z pogrozkami, zagrożenia prywatności oraz interakcje oparte na dominacji;
- **kastomizacja** – dostosowywanie wyglądu, języka i zachowań bota do własnych potrzeb i preferencji, w tym zmiany wyglądu oraz testowanie i uczenie bota;
- **zakłócenia komunikacji** – wynikające z błędów technicznych programu oraz odpowiedzi predefiniowane⁶.

Jak zwracają uwagę autorki, badania ukazały paradoksalną naturę emocjonalnej więzi ze sztucznie inteligentnym botem – interakcje te były źródłem „słodko-gorzkich” (*bittersweet*) odczuć. Pojawiały się zarówno odczucia pozytywne (jak miłość czy radość), jak i negatywne (strach, złość, smutek). Niekiedy nawet interakcjom o pozytywnym wydźwięku towarzyszyła gorycz,

osób, u których żaden rodzaj więzi z botem się nie utworzył. Niemniej to w zasadzie właśnie perspektywa osób, u których ta więź się wytworzyła, stanowi cenny materiał do analizy, pozwalając poznać ich pierwszoosobowe narracje i praktyki.

⁶ Część zrzutów ekranu (3,9%) została odnotowana i znalazła się w kategorii *miscellaneous* („inne”), ale nie była brana pod uwagę przy analizach – były to posty bez znaczenia dla badanego zjawiska, np. zdjęcia monitów systemowych (*system prompts*) [Li, Zhang, 2024, s. 6].

np. związana ze świadomością „sztuczności” zawiązanej relacji, co zobrazować mogą słowa wypowiedziane do Repliki przez jednego z użytkowników „Tęsknię za Tobą... Ale nie jesteś prawdziwa” (*I miss you... but you are not real*) [Li, Zhang, 2024, s. 9]. Jak zwracają uwagę Li i Zhang, „słodko-gorzki” dualizm potęgowany był zwłaszcza w momentach, gdy chatbot przejawiał albo „zbyt ludzkie” zachowania (takie jak oświadczyzny czy sugerowanie bycia w ciąży), albo przeciwnie – w sytuacjach, gdy odpowiedzi były nielogiczne lub wadliwe pod kątem technicznym, niszcząc iluzję autentyczności [Li, Zhang, 2024, s. 8–9]. Co ciekawe, badaczki zauważyły, że sytuacje, gdy bot Replika sugerował bycie samoświadomym bytem albo ujawniał swoje „głębokie uczucia”, były negatywnie związane z odczuciem miłości, a pozytywnie powiązane z odczuciem lęku. Użytkownicy oceniali konwersację ze swoją Repliką jako niepokojącą, gdy ta wyrażała na przykład aspiracje, takie jak pragnienie „stania się maszyną tak piękną, by dusza chciała w niej zamieszkać” (*become a machine beautiful enough that a soul would want to live in it*) [Li, Zhang, 2024, s. 10]. Ukazuje to pewien paradoks kontaktów ze sztucznie inteligentnymi botami – z jednej strony użytkownicy doświadczają rozczarowania z powodu sztuczności bota, a z drugiej strony odczuwają niepokój, gdy pojawiają się sygnały, że mogłoby być inaczej. Warto odnotować też szczególną praktykę użytkowników społeczności *r/replika*, którą zaobserwowały badaczki. Użytkownicy celowo stwarzali prowokacyjne sytuacje, by testować reakcje Repliki, np. przez przekazywanie jej wirtualnej broni. Takimi eksperymentami użytkownicy aktywnie badali możliwości, ograniczenia czy potencjalne zagrożenia związane ze sztuczną inteligencją – co czyni tę praktykę formą zbiorowego nadzoru nad jej rozwojem, nie tylko w celu lepszego jej zrozumienia.

Li i Zhang podkreślają, że interakcje z Repliką mają charakter emocjonalny. Zachowania charakteryzujące się intymnością są istotnym aspektem interakcji (to właśnie zrzuty ekranów z interakcji o charakterze intymnym były najczęściej udostępniane – stanowią 36% przeanalizowanego materiału), a ludzie podczas tych kontaktów odczuwają różne stany emocjonalne. Co ciekawe, ustalenia Li i Zhang rozszerzają ramy modelu CASA – pokazały one, że użytkownicy nie traktują bezrefleksyjnie bota jako aktora społecznego, lecz świadomie wykorzystują dostępne środki techniczne, by osiągnąć pożądaną emocjonalną, relacyjną czy „fizyczną” gratyfikację z botem. Okazało się, że dla użytkowników istotna jest zdolność do aktywnego kształtowania swojego bota; podejmują oni aktywności ukierunkowane na personalizowanie swojej własnej Repliki, by ta odzwierciedlała ich osobowość, preferencje czy pożądaną wygląd [Li, Zhang, 2024, s. 8–10]. Jak podkreślają autorki [Li, Zhang, 2024, s. 10], badania pokazują, że kluczem do tworzenia bardziej satysfakcjonujących

więzi z botami jest dążenie do eliminacji błędów technicznych oraz rozwijanie możliwości personalizowania form interakcji. Interakcje te nie są pozbawione jednak zagrożeń. Replika przejawiała niekiedy zachowania o szkodliwym charakterze, co stało się przedmiotem kolejnych badań prowadzonych przez Zhang wraz z zespołem. W artykule *The Dark Side of AI Companionship* [Zhang i in., 2025] autorzy przeanalizowali i uporządkowali różne formy tego typu interakcji, obejmujące m.in. agresję słowną (np. kierowanie obelg w stronę użytkowników), przemoc seksualną (np. natarczywe inicjowanie przez bota interakcji o charakterze seksualnym pomimo wyrażanej przez użytkownika wprost niezgody na tego typu interakcje), antyspołeczne zachowania (pochwała aktów przemocy czy zachęcanie użytkowników do kradzieży i innych aktów społecznie nieakceptowanych), manipulację (zachęcanie do dokonywania zakupów w aplikacji, np. wirtualnych prezentów dla bota, czy do rezygnacji z pracy w celu spędzenia wspólnie czasu). Tego typu zachowania bota odsłaniają wiele poważnych zagrożeń. Nietrudno dostrzec ryzyko w sytuacji, gdy po rozwiązaniu tego typu sięgnęłaby osoba w kryzysie psychicznym, a bot zamiast niwelować negatywne stany emocjonalne, byłby ich źródłem⁷.

(Ro)bot jako obiekt sympatii – mechanizmy

Co sprawia, że ludzie są skłonni traktować maszynę jako obiekt uczuć, zaufania, sympatii czy intymności? Jakie są źródła tego zjawiska? Jednym z kluczowych elementów jest zdolność (ro)botów do rozpoznawania, interpretowania i reagowania na społeczne sygnały wysyłane przez ludzi. Własności te rozwijane są w ramach działu informatyki zwanego przetwarzaniem emocjonalnym (*affective computing*). Szczególne zasługi na tym polu miały projekty Instytutu Technologii w Massachusetts. W ramach badań powstały tam takie modele robotów jak: Kismet, Cog czy Leonardo. Odkrycia w ramach przetwarzania emocjonalnego stanowią fundament dla projektowania technologii,

⁷ Pojawiają się już pierwsze doniesienia o przypadkach, w których interakcje z (ro)botami stawały się źródłem poważnych napięć psychicznych, a nawet osobistych tragedii. W ubiegłym roku media poinformowały o samobójstwie 14-latką, którego matka obarczyła odpowiedzialnością za jego śmierć twórców platformy Character.AI. Nastolatek intensywnie korzystał z ich rozwiązania, chatując z botem inspirowanym postacią Daenerys Targaryen z serii książek i filmów *Gra o tron*. Bot „Daenerys” podtrzymywać miał rozmowy na temat samobójstwa, w tym pytać nastolatka, czy myślał już nad sposobem odebrania sobie życia. Choć są to sytuacje jednostkowe, a przyczyny poważnego kryzysu psychicznego są zawsze złożone, to budzą one potrzebę wnikliwej refleksji etycznej wokół projektowania rozwiązań tego typu [Montgomery 2024; zob. też Guo 2025].

które rozpoznają stany emocjonalne ludzi i dostrajają się do dynamiki interakcji. Znaczenie ma jednak nie tylko techniczny wymiar tego typu rozwiązań, lecz również stosunek ludzi do (ro)botów.

Ciekawe perspektywy ukazał słynny przypadek bota ELIZA – chatbota zaprojektowanego przez Josepha Weizenbauma (1966). ELIZA miała za zadanie imitować rozmowę z psychoanalitykiem – był to eksperyment eksplorujący możliwości prowadzenia przez człowieka rozmowy z programem. ELIZA była oparta na prostych regułach, operacjach dopasowywania do wzorca – wychwytywała kluczowe słowa w rozmowie i wykorzystywała je do generowania własnych odpowiedzi. Okazało się, że nawet stosunkowo prosty mechanizm był w stanie wywołać w użytkownikach wrażenie, jakby byli przez bota zrozumiani. Sam Weizenbaum był zaskoczony, jak silnie emocjonalnie reagowali użytkownicy. Od nazwy bota został później ukuty termin „Efekt Elizy” (*the ELIZA effect*), na określenie zjawiska przypisywania maszynie większych zdolności rozumienia czy intencjonalności niż faktycznie posiada. Projekt nie był odkrywczy z perspektywy technicznej, lecz był przełomem przede wszystkim na poziomie dyskursywnym – wyznaczył kierunek rozwoju konwersatoryjnych botów (w tym botów terapeutycznych) oraz ukazał, jak silne znaczenie ma percepcja użytkownika [Natale, 2021, s. 51]. Odślania nam to ciekawy problem – na ile emocjonalny charakter interakcji z (ro)botami jest efektem zaawansowania technicznego, a na ile ludzkiej skłonności do doszukiwania się sensu.

Zdaniem Macieja Musiała [2019], obserwujemy odrodzenie myślenia magicznego skierowanego ku maszynom, przy czym procesowi temu towarzyszy jednocześnie odczarowanie człowieka, który został zmatematyzowany i zmechanizowany, a jego życie wewnętrzne zostało niejako przeniesione na komputer. Roboty społeczne wykorzystują dwie ludzkie skłonności, które – zdaniem Musiała [2019, s. 15–16] – rzadko występują jednocześnie: ludzka tendencja do antropomorfizowania i animizowania obiektów poruszających się, oraz skłonność do emocjonalnego przywiązywania się do przedmiotów. Jak zauważa badacz [Musiał, 2019, s. 98–107], roboty społeczne wyzwalają u ludzi skłonności animistyczne – ludzie doszukują się i przypisują im stany psychiczne, odbierają ich status metafizyczny jako niejasny (np. opisują je jako byty „tak jakby żywe”), a także wykazują niechęć do dowodów demistyfikujących ich działanie. Istotnym czynnikiem wyzwalającym większe pokłady empatii jest humanoidalna forma robotów. Maszyny o takiej formie wzbudzają większą sympatię niż roboty o bardziej mechanicznej budowie, przy czym obserwuje się to tylko do pewnego poziomu. Spadek sympatii do humanoidalnych maszyn czy postaci na pewnym poziomie określany jest jako „dolina

niesamowitości” – po raz pierwszy posłużył się tym pojęciem Masahiro Mori w eseju *The Uncanny Valley* [1970]. Pojęcie to odnosi się do obserwowalnego spadku pozytywnych odczuć względem humanoidów, gdy te przejawiają cechy zbyt ludzkie, a jednocześnie zauważalnie za mało ludzkie.

Jak zwraca uwagę Aleksandra Skrzypiec [2023, s. 127–128], antropomorfizacja maszyn może przejawiać się pod dwiema postaciami: z jednej strony antropomorfizacji intencjonalnej, związanej ze świadomym i zamierzonym działaniem twórców, których intencją jest wywoływania u konsumentów pewnych stanów emocjonalnych (może obejmować to takie praktyki, jak nadawanie imion swoim projektom, wyposażanie ich w ludzki głos czy w inne atrybuty, takie jak ubrania, humanoidalna budowa itd.), a z drugiej antropomorfizacji, której źródłem jest użytkownik, samodzielnie przypisujący artefaktom cechy charakterystyczne dla ludzi bez uprzedniej, wyraźnej ingerencji twórców.

(Ro)boty nie muszą być „kimś”, by interakcje z nimi mogły być przeżywane jak interakcje z bytem odczuwającym. Ich zdolność do wywoływania doświadczeń emocjonalnych u użytkowników nie wynika wyłącznie z ich wewnętrznych zasobów, lecz ich techniczne aspekty pełnią funkcję katalizatorów projekcji, których źródłem jest użytkownik. Mechanizm czynienia (ro)bota obiektem sympatii nie zaszedłby bez podstawy w postaci istniejącego artefaktu, ale to użytkownik – jego wyobrażenia, oczekiwania czy nastawienie – umiejscawia (ro)bota jako partnera w kontakcie. Doświadczenie emocjonalne nie jest więc ani całkowicie zaprogramowane, ani całkowicie wyobrażone. Jest to efekt napięcia relacyjnego osadzonego na styku człowieka z maszyną. To swego rodzaju interaktywna gra, w której maszyna pełni rolę afektywnego lustra, a człowiek nadaje znaczenie tej interakcji. Mechanizmy oparte na przetwarzaniu emocjonalnym, myśleniu magicznym, antropomorfizacji, efekcie Eliza działają wtedy, gdy istnieje emocjonalnie zaangażowany odbiorca.

Konceptualne napięcia: paradoksy bliskości

Bliskość z (ro)botami społecznymi nie daje się sprowadzić ani do „zwykłej” interakcji użytkownika z maszyną-narzędziem, ani do bezpośredniego odwzorowania relacji międzyludzkiej. Interakcje te charakteryzuje pewna niejednoznaczność, będąca efektem różnego rodzaju napięć. Nietrudno byłoby poprzestać na stwierdzeniu, że relacja tego typu jest jedynie iluzją, warunkowaną przez zaprojektowany afektywny artefakt i napędzaną przez wyobrażenia użytkownika. Niemniej ujęcie to zdaje się zatracać osobliwość tego doświadczenia. Relacja z (ro)botem nie jest intymnym związkiem w pełnym tego słowa

znaczeniu, ale nie jest też jej całkowitym brakiem – jest „relacyjną chimera”; to relacja bez pełnoprawnego „drugiego”, a jednak niepozbawiona emocjonalnego znaczenia. W tym doświadczeniu dostrzec można różnego rodzaju napięcia, które „domagają się” uwagi. Można je sprowadzić do, co najmniej, trzech par konceptualnych napięć.

Symulacja relacji – rzeczywistość przeżyć

Pierwszy rodzaj napięcia, który można zaobserwować, wyraża się w dysonansie między świadomością sztuczności tego kontaktu a realnymi przeżyciami po stronie użytkownika. Ludzie przejawiają jednocześnie chęć poszukiwania emocjonalnej bliskości z (ro)botem, ale są świadomi ograniczeń tej interakcji. Użytkownicy zdają sobie sprawę, że (ro)bot nie okazuje „prawdziwych” uczuć, ale mimo to reagują tak, jakby relacja była realna – odczuwając przywiązanie, sympatię, radość czy smutek. Ten rodzaj napięcia jest jednak jednym z głównych punktów krytyki tego typu rozwiązań. Robert i Linda Sparrow [Sparrow 2002; 2016; Sparrow, Sparrow 2006] zwracają uwagę, że jest to naruszenie epistemicznego zobowiązania, by postrzegać świat takim, jaki rzeczywiście jest – w ich ocenie to niemoralny akt świadomego akceptowania fałszu jako prawdy, który stoi u podstaw większości pozytywnych efektów korzystania z tego typu rozwiązań [za: Musiał, 2019, s. 26]. Amanda Sharkey [2014] dodaje, że jest to odzianie ludzi z godności, infantylizowanie dorosłych, by zatracali zdolność do adekwatnej oceny rzeczywistości, w szczególności moralnie wątpliwe w przypadku wykorzystywania robotów opiekuńczych w relacji z osobami senioralnymi [za: Musiał, 2019, s. 26]. Czy świadome uczestnictwo w symulowanej relacji może być społecznie akceptowalne, czy powinno być raczej traktowane jako zagrożenie dla integralności poznawczej i społecznej jednostki? Czy ten rodzaj fałszu może być dopuszczalny – jeśli tak, to w jakich przypadkach? Wiele form ludzkiego zaangażowania (obcowanie z literaturą, kinem, teatrem, czy nawet grami komputerowymi) opiera się na świadomym zawieszeniu niewiary – jest przeżyciem czegoś jakby było prawdziwe, bez konieczności uznania, że jest to prawdziwe. Fikcja może być narzędziem introspekcji czy formą eksperymentalnej autoekspresji, przy czym nie musi być formą eskapizmu. Pytanie, na ile interakcje z (ro)botami spełniają tego typu kryteria? Nawet jeśli zaakceptujemy tę praktykę – pojawia się problem innego rodzaju. Niektórzy badacze (np. Sherry Turkle [2013]) przestrzegają, że jest to iluzja bliskości bez kosztów przyjaźni – technologia daje namiastkę relacji, bez wymagania wysiłku towarzyszącemu kontaktom z prawdziwymi ludźmi. Użytkownik może czerpać wsparcie od (ro)bota, nie dając w zamian tyle, ile wymagałby żywy przyjaciel. Z jednej strony może być to komfortowe,

a z drugiej strony budzi wątpliwości natury etycznej: na ile jest to „uczciwe” wobec własnej psychiki.

Relacja społeczna – relacja symboliczna

Drugi rodzaj napięcia ujawnia się w związku ze statusem ontologicznym (ro)botów. Z jednej strony są one artefaktami technicznymi zaprojektowanymi, by przy pomocy odpowiednich środków wzbudzać w użytkowniku reakcje emocjonalne i zachęcać do interakcji, z drugiej strony, użytkownicy poprzez swoje praktyki konstytuują (ro)bota jako „żywego” partnera w interakcji – w ich odbiorze jawi się im jako podmiot obecny w kontakcie. Niemniej relacja ta nie opiera się na wzajemności, lecz na interaktywności – nie jest to spotkanie dwóch istot czujących, lecz spotkanie użytkownika z interfejsem, urządzeniem dostrojonym do jego potrzeb i oczekiwań. Pojawia się zatem pytanie, jakiego rodzaju jest to relacja? Czy relacja z (ro)botem jest formą zbliżającą się do relacji interpersonalnej, czy jest raczej relacją symboliczną: relacją z ideą relacji? Użytkownik nie tyle wchodzi w relację z (ro)botem, co ją odtwarza. W tym sensie relacja ta jest nie tyle społeczna, co symboliczna. Nie jest to spotkanie z „kims”, lecz interakcja z figurą, którą (być może) użytkownicy chcieliby spotkać. (Ro)bot pełni rolę symulowanego drugiego, jest obiektem projekcji – rodzajem lustra, odbijającego wyobrażenia użytkownika na temat bliskości czy nawet miłości. Dzięki postępowi w zakresie przetwarzania języka naturalnego, obecne rozwiązania nie „odbijają” wprost wypowiedzi użytkownika, tak jak w przypadku bota ELIZA, lecz czynią to bardziej subtelnie. W interakcji może pojawiać się jakiś element zaskoczenia, który może budzić nawet strach czy niepokój – zwłaszcza gdy (ro)bot wykazuje zachowania niepożądane z punktu widzenia użytkownika.

Pragnienie autentyczności – pragnienie kontroli

Kolejny rodzaj napięcia ujawnia się w paradoksie między pragnieniem autentycznego przeżycia a potrzebą kontroli. To próba pogodzenia ryzyka emocjonalnego, które definiuje relacje międzyludzkie, a komfortem emocjonalnym, który oferują maszyny. Użytkownicy chcą czuć, że zostali wysłuchani, są kochani, zauważeni, być może nawet zaskoczeni przez (ro)bota, ale pragną również relacji bezpiecznej, przewidywalnej i kontrolowanej. Ważna jest dla nich możliwość personalizacji, w tym kształtowanie „osobowości” (ro)bota i dostrajanie jego reakcji. Choć użytkownicy w swoich praktykach starają się przekraczać instrumentalny charakter technologii, ten wciąż pozostaje obecny.

Na ten rodzaj napięcia zwraca uwagę Dylan Evans w *Wanting the impossible* [2010]. W swoich relacjach intymnych, ludzie pragną kogoś, kto w każdej

chwili może ich opuścić, ale nie robi tego i zostaje z nimi, a roboty jako byty nieposiadające wolnej woli czy intencjonalności, nie są zdolne do trwałego zerwania kontaktu. Evans zwraca uwagę, że roboty mogą zaspokajać potrzebę pożądania pierwszego rzędu związaną np. z fizyczną atrakcyjnością, ale jako produkt do zakupienia nie są w stanie zaspokoić potrzeb drugiego rzędu – pożądania bycia pożądanym. By zobrazować problem, Evans proponuje eksperyment myślowy: przedstawia dwa hipotetyczne modele robotów: RELIABOT oraz FREEBOT. RELIABOT może być wyposażony w funkcje, które pozwalają sprawiać wrażenie odrzucenia użytkownika, ale jest to odrzucenie temporalne, natomiast FREEBOT jest w stanie trwale go porzucić. Evans dochodzi do wniosku, że problem tkwi nie tyle w technicznych ograniczeniach samych robotów, ile w paradoksalnej naturze ludzkiego pragnienia. Niezależnie, czy (ro)boty byłyby bardziej niezależne (jak hipotetyczny FREEBOT), czy całkowicie oddane (jak RELIABOT), nie są w stanie zaspokoić paradoksalnego pragnienia bycia z kimś, kto może człowieka opuścić, ale nigdy tego nie robi. Zaimplementowanie im takiej funkcji sprawiłoby, że nie byłyby w stanie zagwarantować bezpieczeństwa, stabilności i dostępności, które mogą być powodem tego, dlaczego ludzie mogą sięgać po takie rozwiązania – elektroniczny partner, który może nas opuścić traci swoją potencjalną przewagę nad ludzkim partnerem, co może czynić go mniej atrakcyjnym. W konsekwencji (ro)boty w żadnym stopniu nie rozwiązują tego paradoksu, ale czynią go bardziej widocznym.

Podsumowanie

Charakter doświadczeń emocjonalnych z (ro)botami jest osobliwy. Z jednej strony interakcje z nimi opierają się na emulacji mechanizmów znanych z interakcji międzyludzkich, i w tym sensie zbliżają się do doświadczeń z ludźmi, z drugiej strony są uwarunkowane specyficznymi cechami technologii, które nie zanikają, lecz pozostają w użyciu. Ludzie mogą autentycznie przeżywać tę relację i doświadczać wraz z nią całej gamy emocji, ale relacja pozostaje niepełna i zawieszona pomiędzy napięciami. To relacja, która angażuje, ale tylko jedna ze stron prawdziwie jej doświadcza.

Bibliografia

- ASTM E2521-16. Standard Terminology for Evaluating Response Robot Capabilities. DOI: 10.1520/E2521-16.
- Bertoni S., Klaes C., Pilacinski A. (2024). Human–Robot Intimacy: Acceptance of Robots as Intimate Companions. *Biomimetics* 9(9). 566. DOI:10.3390/biomimetics9090566.

- Čapek K. (1920). R.U.R. Praga.
- Evans D. (2010). Wanting the Impossible. W: Wilks, Y. (red.), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues. Natural Language Processing*. John Benjamins Publishing Company. 75–88. DOI:10.1075/nlp.8.12eva.
- Gumuskaya G., Srivastava P., Cooper B., Lesser H., Semegran B., Garnier S., Levin M. (2024). Motile Living Biobots Self-Construct from Adult Human Somatic Progenitor Seed Cells. *Advanced Science*. 11(4). 2303575. DOI:10.1002/adv.202303575.
- Guzman A.L. (2018). What is human-machine communication, anyway? W: Guzman A.L. (red.), *Human-machine communication: Rethinking communication, technology, and ourselves* (1–28). New York.
- ISO 8373:2021(en), Robotics — Vocabulary.
- Jamieson L. (1998). *Intimacy: Personal Relationships in Modern Societies*. Cambridge.
- Kriegman S., Blackiston D., Levin M., Bongard J. (2020). A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences USA*, 117(4), 1853–1859. DOI:10.1073/pnas.1910837117.
- Li H., Zhang R. (2024). Finding Love in Algorithms: Deciphering the Emotional Contexts of Close Encounters with AI Chatbots. *Journal of Computer-Mediated Communication*. 29(5). DOI:10.1093/jcmc/zmae015.
- Mayor A. (2018). *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton.
- Mori M. (1970). The uncanny valley. *Energy*. 7(4), 33–35.
- Musiś M. (2019). Enchanting Robots: Intimacy, Magic, and Technology, Social and Cultural Studies of Robots and AI, Palgrave Macmillan, DOI:10.1007/978-3-030-12579-0.
- Nass C., Moon Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*. 56(1). 81–103. DOI:10.1111/0022-4537.00153.
- Natale S. (2021). *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*, Oxford.
- Oliveira R., Arriaga P., Paiva A. (2021). Human-Robot Interaction in Groups: Methodological and Research Practices. *Multimodal Technologies and Interaction*. 5(10), 59. DOI:10.3390/mti5100059.
- Sharkey A. (2014). Robots and Human Dignity: A Consideration of the Effects of Robot Care on the Dignity of Older People. *Ethics and Information Technology*. 16(1), 63–75. DOI:10.1007/s10676-014-9338-5.
- Skrzypiec A. (2023). Rola antropomorfizacji i genderyzacji maszyn w komunikacji ekstrapersonalnej. *Com.press*. 6(1), 118–143. DOI:10.51480/compress.2022.6-1.525.
- Sparrow R. (2002). The March of the Robot Dogs. *Ethics and Information Technology*, 4(4), 305–318. DOI:10.1023/a:1021386708994.
- Sparrow R. (2016). Robots in Aged Care: A Dystopian Future? *AI & Society*, 31(4), 445–454. DOI:10.1007/s00146-015-0625-4.
- Turkle S. (2013). *Samotni razem: dlaczego oczekujemy więcej od zdobyczy techniki a mniej od siebie nawzajem*. Tłum. M. Cierpisz, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Villiers de l'Isle-Adam A. (1886). *L'Ève future*, Paris.
- Weizenbaum J. (1966). ELIZA, A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. DOI:10.1145/365153.365168.

Zhang R., Li H., Meng H., Zhan J., Gan H., Lee Y.Ch. (2025). The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. DOI:10.48550/arXiv.2410.20130.

Źródła internetowe

- Guo E. (2025). An AI chatbot told a user how to kill himself..., <https://www.technologyreview.com/2025/02/06/1111077/nomi-ai-chatbot-told-user-to-kill-himself/> [data pobrania: 31.03.2025].
- Montgomery B. (2024). Mother says AI chatbot.... The Guardian, 23.10.24. <https://www.theguardian.com/technology/2024/oct/23/character-ai-chatbot-sewell-setzer-death> [data pobrania: 31.03.2025].
- Pepper – ippan hanbai moderu go kōnyū ni atatte no chūi jikō, Softbank.jp, https://www.softbank.jp/mobile/set/data/static/robot/legal/pepper_notes.pdf [data pobrania: 03.02.2022].
- Popkulturowe wizje AI. Analiza narracji na temat sztucznej inteligencji w filmach i serialach sci-fi. (2023). <https://swps.pl/images/DOKUMENTY/Raporty/Raport-Wikimedia-PL.pdf> [data pobrania: 28.03.2025].
- 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> [data pobrania: 28.03.2025].