

Tytuł: **Maszyny z... emocjami ludzi¹**

Autor: Marek Kasperski / mjkasperski@kognitywistyka.net

Źródło: <http://www.kognitywistyka.net>

Data publikacji: 02 I 2007

Dostosowanie technologii do ludzkich oczekiwań oznacza, że sami z siebie, bez szczególnego przygotowania będziemy posiadać wiedzę i umiejętności korzystania z niej po prostu dlatego, iż jesteśmy ludźmi. Nasze reakcje będą więc pozytywnie nacechowane poczuciem umiejętności, kompetencji i uprawnienia. [s. 20]

Projektujący media mogą korzystać z tego, co ludzie robią najlepiej. Maszyny będą wówczas bardziej dostępne, treść bardziej zrozumiała, a interfejsy bardziej wygodne, jeśli będą działały zgodnie z zasadami świata społecznego i naturalnego. [s. 298]

B. Reeves, C. Nass, Media i ludzie.

Kontakty międzyludzkie – personalne – to gra, w której uczestniczą minimum dwie osoby. Gracze zasiadają do gry znając i respektując jej zasady – jeśli chcą by kontakty ich były owocne. Co by się jednak stało, gdyby jednego z graczy zastąpić maszyną albo programem komputerowym imitującym osobę? Czy reguły gry uległyby zmianie? Czy też gracz-człowiek, pomimo wiedzy na temat swojego nie-ludzkiego partnera, zachowywałby się nadal tak, jak w kontaktach człowiek-człowiek?

Od chwili opracowania pierwszych maszyn matematycznych komputery przeszły daleką drogę. Z narzędzia pomocnego tylko matematykom i rzeszy inżynierów wojskowych, stały się narzędziem grafików, pisarzy, muzyków. Później, urządzeniem dostarczającym rozrywki (centrum multimedialnym, umożliwiającym spędzanie czasu przy grach i oglądaniu filmów). Od niedawna również nowym rodzajem medium, wykorzystywanym do komunikacji na globalną skalę (Internet).

Przeobrażeniom tym towarzyszy zmiana statusu komputera – z urządzenia do swego rodzaju kompana w życiu codziennym. Takiemu podejściu współtowarzyszy też fakt, że komputery (szerzej: media) są na tyle wyjątkowym narzędziem, że ich użytkownicy są skory nadawać im cechy żywych istot (antropomorfizować je). W związku z czym z jednej strony prowadzone są badania związane z psychologią ludzi, jako użytkowników nowych mediów (B. Reeves i C. Nass²), z drugiej zaś psychologowie użytkownika i projektanci interakcji opracowują nowe techniki komunikacji człowiek-komputer (Human-Computer Interaction,

¹ Artykuł ukazał się drukiem w „Charaktery” wydanie specjalne „Psychologia XXI wieku”, Nr 1/2006, ss. 44-48.

² B. Reeves, C. Nass, *The Media Equation. How People Treat Computers, Television, and New Media*, Cambridge University Press, 1996. Polskie wydanie: *Media i ludzie*, tłum. A. Szczerkowska, PIW, Warszawa 2004.

HCI), zmierzające do ich naturalizacji. Z chwilą, gdy duże przedsiębiorstwa z branży IT odkryły możliwości rozwoju rynku robotów osobistych, badania tym bardziej nabrały rozpędu – wszak „roboty to komputery na nogach”, jak raczył się niedawno wyrazić Fumio Nagashima, z grupy robotyki w Fujitsu Laboratories.

Relacje społeczne i nabywanie wiedzy

Tak ludzie, jak i zwierzęta stadne, nabywają sporą część wiedzy uczestnicząc w życiu społecznym. Społeczne mechanizmy odgrywają jedną z kluczowych ról dla procesu nabywania wiedzy. Jak wykazano (B. G. Galef³, M. Hauser⁴) również kontekst emocjonalny sprzyja temu, czy dana cecha zostanie nabyta, albo jaką w danej społeczności nada jej się wartość.

Na co dzień przeżywamy nasze emocje. Są nam dane pierwszoosobowo, jak wszystkie treści naszego umysłu – to zresztą jeden z podstawowych problemów nauk kognitywnych, jak badać umysł, dany z perspektywy pierwszej osoby, przy pomocy nauki, uprawianej z perspektywy trzeciej osoby. Jednak życie w grupach wymogło powstanie ekspresji – czyli sposobu komunikowania innym o tym, co jest treścią naszych stanów umysłowych bądź, co byśmy chcieli by inni myśleli, że jest treścią – przy mniej szczerych intencjach, albo grze aktorskiej. Czasem też kontekst emocjonalny sprzyja czytelności komunikatu, jak np. w wypowiedziach: „Wczoraj były moje urodziny” (ze smutną miną) i „Wczoraj były moje urodziny” (z uśmiechem).

Biorąc pod uwagę to, co zostało już powiedziane, jasnym jest, że gdyby komputery potrafiły odczytywać kod ludzkiej emocjonalności to:

1. ich obsługa stałaby się **intuicyjna**, ponieważ mechanizmy znane z życia codziennego z innymi ludźmi można by wykorzystać w relacjach człowiek-maszyna – maszyny byłyby dostosowane do ludzi, a nie ludzie do maszyn;
2. użytkownicy byłiby dużo sprawniejszymi operatorami maszyn, zapewniałoby to **przechodność wiedzy, doświadczenia** – skoro wiedza na temat sterowania maszynami opierałaby się na mechanizmach, na jakich opiera się funkcjonowanie w społeczeństwie, użytkownicy z natury rzeczy byłiby już wstępnie zapoznani ze sposobami komunikacji z urządzeniami oprogramowanymi;
3. na tych samych zasadach, co wyżej, operowanie urządzeniami emocjonalnymi byłoby: bardziej **swojskie** (przyjazne i znajome środowisko sprzyja przy tym wydajności pracy oraz obniżeniu poziomu sfrustrowania, co jest aktualnie jednym z podstawowych powodów aktualnego, złego modelu relacji człowiek-komputer);
4. urządzenia byłyby bardziej **plastyczne** – odczytywanie stanów emocjonalnych ludzi-operatorów przez komputery, umożliwiałoby tym drugim dostosowanie do stopnia zaawansowania użytkownika oraz jego stanu emocjonalnego w danym momencie. Co też sprzyja obniżeniu poziomu frustracji;
5. w końcu, komunikacja oparta o możliwość odczytywania i interpretowania kodów wyrażających emocje gwarantuje większą **jasność i głębie komunikacji** – jak zostało to ukazane na przykładzie zdania „Wczoraj były moje urodziny”, osadzenie jego w

³ B. G. Galef. Imitation in animals: history, definitions, and interpretation of data from the psychological laboratory, w: T. Zentall, B. G. Galef (ed.), *Social learning: Psychological and biological perspectives*, Lawrence Erlbaum Associates, 1988.

⁴ M. Hauser, *The evolution of communication*, The MIT Press, Cambridge, MA, 1996.

odmiennym kontekście emocjonalnym powoduje całkowitą zmianę sensu wypowiedzi;

6. no i **bezpieczeństwo** – odczytywanie stanu emocjonalnego człowieka-użytkownika przez maszyny może odbywać się cały czas, nawet jeśli człowiek-operator nie przekazuje świadomie żadnych instrukcji maszynie. Dzięki czemu, w momencie zbyt silnego stresu malującego się na twarzy człowieka, komputer mógłby odmówić wykonania czynności, która w sposób bezpośredni bądź pośredni w dalszych konsekwencjach mogłaby zagrażać zdrowiu bądź życiu operatora (np. podczas jazdy samochodem, komputer mógłby wykrywać, w jakim stanie emocjonalnym jest kierowca i albo odpowiednio ukształtować na ten czas środowisko wewnątrz pojazdu, np. by wyrwać kierowcę ze smutku, albo odmówić dalszej jazdy).

[Klasyczna] psychologia użytkownika

Jeszcze w latach 50. ubiegłego stulecia, rozpoczęto badania nad psychologią komunikacji człowiek-komputer. Przewodzące w tej mierze były laboratoria firmy IBM⁵. To tam właśnie zaczęto się przyglądać ustaleniom z zakresu psychologii kognitywnej i przekładalności jej ustaleń na modelowanie relacji człowiek-maszyna.

W latach 70., gdy zaczęto używać interfejsów graficznych i *quasi* graficznych, zauważono, że liczba wyświetlanych funkcji nie może przekraczać magicznej liczby 7, by operator mógł swobodnie korzystać z danej aplikacji. Rzecz jasna było to nic innego, jak zastosowanie odkrycia pojemności pamięci krótkoterminowej przez G. Miller do budowy komunikacji człowiek-komputer. Zaczęto też badać zależności między czasem reakcji maszyny na polecenia operatora, a płynnością pracy samego operatora. W wyniku tych badań okazało się, że wraz z wydłużaniem się czasu oczekiwania na odpowiedź maszyny, wzrasta poziom frustracji użytkownika i spada poziom płynności pracy, co aktualnie ma również konsekwencje w związku z czasem „ładowania się” strony WWW. Ustalono także wiele innych współrzędnych, które zamknąć można do tzw. listy „przyczyn błędów” – czynników, które wpływają na pogorszenie relacji człowiek-maszyna:

- Przeciążenie „kanału” operatora.
- Znużenie operatora.
- Brak motywacji.
- Niedokładne instrukcje postępowania w przypadkach nadzwyczajnych.

Aktualnie, prócz wiedzy z zakresu psychologii kognitywnej, do modelowania interakcji człowiek-komputer coraz częściej wykorzystuje się wiedzę na temat psychologii relacji [społecznych]. Z chwilą gdy komputery przeniosły się z laboratoriów do naszych domów, uległ transformacji sposób komunikowania się z maszynami: zamiast języków programowania wywołujących funkcje komputerów, zaczęto posługiwać się metaforą biurka i środowiskiem graficznym, korzystając przy tym z wiedzy na temat pamięci krótkoterminowej, pamięci ikonicznej, itp. itd. Dużo istotniejsze od projektowania interfejsów stało się projektowanie interakcji (D. Norman⁶, J. Nielsen⁷). Stąd do grona informatyków, grafików komputerowych i psychologów użytkownika, dołączyła grupa projektantów

⁵ J. Martin, *Design of Man-Computer Dialogues*, IBM Systems Research Institute, New York, 1973. Polskie wydanie: *Dialog człowieka z maszyną*, tłum. zbiorowe, WNT, Warszawa 1976.

⁶ D. Norman, S. Draper (ed.), *User Centered System Design: New Perspectives on Human-Computer Interaction*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.

interakcji. Dodatkowo, w tym samym czasie na gruncie robotyki kształtowała się nowa metodologia, zwana robotyką kognytywną, która w krótkim czasie doprowadziła do ukonstytuowania się tzw. socjorobotyki.

Komunikacja emocjonalna: ekspresyjne twarze

Naturalnym centrum naszej emocjonalności, czy raczej „wyświetlaczem” naszej emocjonalności, jest twarz. To tutaj odbywa się spora część tego, co w danym czasie odczuwamy. Szybki rzut oka i analiza mimiki twarzy na co dzień informują nas o tym, czy ktoś z otaczających nas ludzi jest smutny czy też wesoły. Dlaczego mimika twarzy może mieć znaczenie dla relacji człowiek-komputer? Powody są dwa:

1. Jak już zostało powiedziane, twarz to kanał komunikacji i to często potężniejszy niż umiejętność werbalizowania myśli.
2. Język ekspresji twarzy – jak na podstawie swoich badań postuluje P. Ekman – jest uniwersalny, tj. wspólny dla wszystkich kultur. Dzięki czemu komunikat jest czytelny dla przedstawiciela każdej z nacji.

Paul Ekman, czołowy badacz emocji w kontekście społecznym, wraz z W. Friesenem ustalili⁸, że nie zależnie od kultury istnieją wzorce emocjonalne wyrażane mimiką twarzy, przez co są one powszechne i uniwersalne. Na wzorzec ten składają się następujące stany wyrażane przy pomocy twarzy:

- zaskoczenie (*surprise*),
- strach (*fear*),
- obrzydzenie (*disgust*),
- złość (*anger*),
- radość (*happiness*),
- smutek (*sadness*).

Należy zauważyć, że jest to lista podstawowych wzorców emocjonalnego alfabetu twarzy, a nie lista pełna. Tym sześcioro wyrazom twarzy przysługują cechy powszechności i uniwersalności – czyli są odczytywane poprawnie niezależnie od kulturowego tła.

Przesłanki te są podstawą badań w zakresie kształtowania relacji człowiek-komputer w oparciu o analizę wyrazu twarzy i nadawanie kontekstu wypowiedzi wyrażanym przez twarz emocjom. W robotyce drogę taką obrali m.in. Cynthia Breazeal: projekt Kismet, David Hanson: projekt K-bot, oraz naukowcy z Uniwersytetu Waseda: projekt Naiyo. Zaś jeśli chodzi o projekty z zakresu oprogramowania (software), to ustalenia z badań Ekmana wykorzystywane są do modelowania interfejsów opartych o wyświetlanie animowanych twarzy, awatarów, w przypadku tzw. *virtual humans*⁹.

⁷ J. Nielsen, *Designing Web Usability: The Practice of Simplicity*, New Riders Publishing, Indianapolis, 2003. polskie wydanie: *Projektowanie funkcjonalnych serwisów internetowych*, tłum. A. Bulandra, Helion, Gliwice 2003.

⁸ A. P. Ekman, W. Friesen, *Unmasking the Face*, Englewood Cliffs, N.J.: Prentice-Hall, 1975. Omówienie artykułu dostępne w: R. R. Hock, *40 prac badawczych, które zmieniły oblicze psychologii*, tłum. E. Wojtych, GWP, Gdańsk 2003, ss. 207-216.

⁹ P. Plantec, *Virtual Humans. Creating the Illusion of Personality*, Amacom, New York 2004.

Spoleczne roboty

Kismet to najbardziej znany projekt związany z nową filozofią modelowania relacji człowiek-maszyna, opartych przede wszystkim na porozumieniu emocjonalnym.

Nazwa Kismet Cytnhia Breazeal zapożyczyła z tureckiego, a oznacza *przeznaczenie*. Przeznaczeniem robota w pierwszym etapie prac nad nim było wyposażenie większego projektu – Cog (od ang. *cognitive*) – prowadzonego przez kierownika Artificial Intelligence Laboratory of MIT, Rodneya Brooksa, w moduł do nabywania cech społecznych. Po co? Celem, który przyświeca twórczyni jest nowy sposób nauki robotów, w tym komputerów, bez ingerencji programistów i techników. Wiedza zdobywana w sposób naturalny, w relacjach z ludźmi-operatorami oraz – w przyszłości – w relacji z innymi robotami, może stać się kluczem do przeskoczenia na poziom rozumienia przez maszyny, a nie tylko wykonywania ściśle określonych (wcześniej zdefiniowanych) czynności.

Breazeal, na podstawie lektur z zakresu psychologii rozwoju, zaczęła śledzić sposób zdobywania wiedzy o świecie przez małe dzieci, które obserwując naśladują to, co wcześniej poznały. Ale, i tu objawia się główna zaleta Kismet, na twarzy dziecka rysuje się to, w jakim stanie jest dziecko wykonując dane czynności – czy rozumie to, co robi, czy jest zdziwione, czy coś szczególnie je ciekawi, czy też jest powodem zniesmaczenia.

Kismet, dzięki zastosowaniu „tablicy” ekspresji twarzy ludzkiej, informuje zwrotnie swojego użytkownika czy rozumie co od niego się oczekuje, a poprzez to, że możemy także odczytać, czy jest czymś zainteresowany, czy nie, możemy ocenić na ile wykonanie przez niego danej czynności będzie dla niego ważne. Dokładnie tak samo, jakbyśmy pracowali z ludzkim współpracownikiem, a nie tylko „bezduszną” maszyną.

Mało tego, podstawą funkcjonowania Kismet są wpisane weń „popędy”. Tak jak ludzi do działania popycha popęd związany z zaspokojeniem głodu oraz zaspokojenie poznawczej ciekawości, tak Kismet poszukuje bodźców, które stymulując go, nie prowadziłyby do znużenia i w ostateczności do przejścia maszyny w stan uśpienia. Przy czym robot, uczestnicząc w relacjach społecznych z użytkownikiem, nabywa wiedzę swego użytkownika, w tym uczy się wydajniej z nim współpracować, co zapewnia mu możliwość odczytywania ludzkich emocji wyrażanych przy pomocy mimiki twarzy i w przyszłości, na tej podstawie, ocenianie, czego człowiek-operator od niego oczekuje.

Ujawnia się tu główna myśl projektu: emocje oraz popędy stanowią podstawy funkcjonowania robota i regulują interakcje człowiek-maszyna.

Rozpoznawanie twarzy, odczytywanie emocji wyrażanej przy pomocy mimiki twarzy, śledzenie wzroku człowieka-operatora oraz, w ramach informacji zwrotnej maszyny, reagowanie na otoczenie i człowieka-operatora swoim alfabetem emocji, wykorzystywane są też w projektach grupy badawczej z Uniwersytetu Waseda, pod wodzą Atsuo Takanishi. Wiele firm już wykorzystuje w praktyce osiągnięcia tychże badań. Wystarczy odnotować, że aktualnie jednym z głównych sponsorów badań Breazeal jest firma Hasbro, czołowy producent interaktywnych zabawek dla dzieci. Cytnhia, to co ustaliła w ramach projektu Kismet, stara się wykorzystać w projekcie wysoko zaawansowanej zabawki dla dzieci o nazwie Leonardo. Leonardo będzie mógł reagować przy pomocy swojej mimiki na zachowanie opiekuna-człowieka, tym samym wpływając również pozytywnie na poprawny rozwój emocjonalności dzieci, które będą się nią bawiły.

Wirtualny świat awatarów i simsów

Nie tylko w ramach robotyki prowadzi się badania w ramach modelowania relacji człowiek-maszyna opartych na odczytywaniu emocji i reagowaniu na nie przez maszyny. Coraz częściej na konferencjach poświęconych rozwojowi interfejsów mówi się o przejściu z graficznej reprezentacji funkcji programów komputerowych na reprezentację ludzko-podobną. Awatara, czyli graficzną reprezentacją postaci, która posiadałaby wiedzę dotyczącą funkcji danego programu (taki interaktywny *help desk*), a która, posługując się swoim zestawem „emocji”, budowałaby pogłębioną relację z użytkownikiem. Interfejs byłby czymś w rodzaju prywatnego, wirtualnego asystenta.

Aktualnie spora część tych badań wykorzystywana jest przy kreowaniu tzw. *virtual humans*, wykorzystywanych przede wszystkim w Internecie, jako wirtualni doradcy i przewodnicy po stronach WWW korporacji. Gdzie niegdzie także słyszy się o wykorzystaniu tego typu systemów w edukacji, gdzie uczniowie zamiast „wkuwać” wiedzę, zdobywają ją mogąc rozmawiać z postaciami historycznymi, jak np. z animowanym prezydentem Lincolnem.

W ramach modelowania postaci graficznych, do repertuaru podstawowych emocji wyrażanych mimiką twarzy, dochodzi jeszcze zestaw tzw. wizemów, czyli graficznych reprezentacji wypowiedzianych fonemów, części językowych używanych w mowie.

Innym, palącym problemem w Internecie, są relacje między graczami uczestniczącymi w rozgrywkach typu MMO (Massively Multiplayer Online Games, Masowe - Wieloosobowe Gry Online). Dlaczego? Ponieważ uczestniczący w nich gracze widzą tylko swoje postacie i postacie swoich przyjaciół, z którymi grają. Często zaś nasi internetowi przyjaciele reprezentują całe spektrum różnorodności międzynarodowej – zdarza się, że w jednej przygodzie możemy uczestniczyć w drużynie składającej się z Koreańczyka, Portugalczyka, Rosjanina oraz koleżanki z drugiego końca Polski. Okazuje się, że prócz różnej znajomości angielskiego, podstawą relacji między graczami są tzw. emotikony wyrażane przez graficzne postacie w grze, tzw. *social sims*, np. w formie grzecznościowego ukłonu na przywitanie, czy płaczu animowanej postaci, jeśli osoba nią kierująca poczuła się dotkniętą.

Język emocji buduje atmosferę prawdziwego porozumienia między graczami wywodzących się nie tylko z różnych krajów, ale i z różnych kultur. A lista wyrażanych emocji graczy w czasie rozgrywki maluje się na ekranie poprzez ekspresję sztucznych animowanych postaci. zresztą lista ta również wzorowana jest na ustaleniach Paula Ekmana i można ją przedstawić przy pomocy czterech skal:

- skala S-H: *Sad-Happy* (smutny-szczęśliwy),
- skala A-C: *Angry-Calm* (rozniewany-spokojny),
- skala S-P: *Shame-Pride* (zawstydzony-dumny),
- skala P-S: *Pensive-Surprise* (zamyślony, melancholijny-zaskoczony, zdziwiony).

Porozumienie ponad podziałami

Jak widać, emocje, czy w przypadku możliwości ich odczytywania – ekspresja i mimika twarzy, zdają się być uniwersalnym interfejsem, który nie tylko skutecznie umożliwia porozumiewanie się pomiędzy ludźmi wywodzącymi się z różnych nacji, ale także między ludźmi i maszynami – komputerami i robotami.

Przewaga użytkowników posługujących się urządzeniami odczytującymi emocje i odpowiadającymi na nie, nad operatorami posługującymi się systemami zbudowanymi na klasycznej wizji interfejsu, jako zbiorze wyselekcjonowanych i pogrupowanych funkcji, zaprezentowanych na wyświetlaczu jest ogromna, co wykazują przeprowadzone badania z takimi maszynami.

W najbliższej przyszłości można oczekiwać, że interfejsy oparte na emocjach, ze świata naszych dzieci i implementacji w interaktywnych zabawkach, przejdą do świata urządzeń oprogramowanych – komputerów i coraz częściej pojawiających się robotów osobistych. W tym świecie maszyny będą znać różnicę między zdaniem: „Wczoraj były moje urodziny :-)” i „Wczoraj były moje urodziny :-)”. Dzięki czemu będą odpowiednio reagować na te dwie różne sytuacje. Tak samo, jak dzisiaj nasi współpracownicy – ludzie.

Warto przeczytać:

1. C. Breazeal, *Designing Sociable Robots* (Intelligent Robotics and Autonomous Agents), MIT Press, 2002.
2. P. Ekman, *Emotions Revealed. Understanding Faces and Feelings*, Phoenix, London 2004.
3. P. Ekman, R. J. Davidson, *Natura emocji. Podstawowe zagadnienia*, tłum. B. Wojciszke, GWP, Gdańsk 1999.
4. B. Reeves, C. Nass, *Media i ludzie*, tłum. A. Szczerkowska, PIW, Warszawa 2004.
5. B. Scassellati, *Foundations for a Theory of Mind for a Humanoid Robot*, PhD Thesis, MIT Department of Computer Science and Electrical Engineering:
<http://www.ai.mit.edu/people/scaz/papers/scassellati-phd.pdf>.
6. <http://www.emotionalmachines.com/>.