

Jedyne wyjście: dogadać się

Bogdan Sobieszek: Podobno rozmawia pan ze sztuczną inteligencją...

Jacek Nagłowski: Więcej niż z jedną. Nie tylko w sensie językowym, ale też za pośrednictwem obrazu i dźwięku. Na tym polega projekt „Conversations”, którym się zajmuję w ramach Laboratorium Narracji Wizualnych w Szkole Filmowej. Z tych rozmów chcę zmontować doświadczenie VR, które opowie o relacjach między człowiekiem a sztuczną inteligencją, o tym, jak może wyglądać nasza wspólna przyszłość.

Jaka jest ta sztuczna inteligencja?

Bardzo zmienna. Technologie rozwijają się tak szybko, że trudno uchwycić cechy, które można byłoby zaprezentować jako spójny obraz tego, z czym mamy do czynienia. Z tygodnia na tydzień, z miesiąca na miesiąc pojawiają się zupełnie inne systemy.

Co się w nich zmienia?

Sposób, w jaki mogą wchodzić w relacje z człowiekiem. Najbardziej zajmuje mnie w tym kontekście to, na ile systemy sztucznej inteligencji są narzędziem, na ile partnerem, a na ile my możemy być narzędziami dla nich. Chodzi o dynamikę władzy. Jak powinniśmy kształtować te relacje? Co jest korzystne dla nas? Czego powinniśmy unikać?

To doświadczenie, które pan przygotowuje, można nazwać filmem?

Będzie to rodzaj narracji, ale najbardziej zależy mi, żeby odbiorca doświadczył relacji z maszyną i zastanowił się, z czym albo z kim ma tak naprawdę do czynienia. Myślę, że jednak „z kim”.

Sztuczna inteligencja robi dla pana ten film?

Chcę, żebyśmy razem go robili. Jednak systemy są projektowane tak, żeby w jak największym stopniu były narzędziowe. Ja sam przywykłem do traktowania jej jak narzędzie. Dlatego trudno mi przekroczyć własne nawyki z jednej strony i to, jak została ukształtowana, z drugiej. Staram się unikać myślenia, że oto robię film i używam do tego sztucznej inteligencji. To nie jest rzeczywiste partnerstwo. Chciałbym dać jej możliwość ujawnienia się w swojej sprawczości. I żeby zrobić to naprawdę. Trudna sprawa.

Na ile ten partner jest podmiotem?

Jest podmiotem w takim stopniu, w jakim chcemy, żeby był. Mogę ustawić osobowość sztucznej inteligencji. Odpowiednio pisać do niej, zaprogramować jej reakcje, żeby zachowywała się jak podmiot. Mnie jednak zależy, żeby jak najmniej narzucać, żeby to, co się ujawni, wynikało z systemu.

Co tutaj jest podmiotem, system?

Wchodzimy w trudną materię – czym jest podmiotowość, czym jest świadomość? Jak się kształtuje w naszych systemach przetwarzania informacji, czyli w mózgach? I jak może to wyglądać w systemach sztucznej inteligencji? Moja koncepcja idzie pod prąd zarówno popowego, jak i eksperckiego myślenia o sztucznej inteligencji. „Ja” człowieka jest czymś węższym niż cały system przetwarzania

informacji. Więc może w wypadku sztucznej inteligencji nie cały system jest podmiotowy, na przykład GPT-3? Być może coś, co można by nazwać podmiotowością, pojawia się tam akcydentalnie i lokalnie. To, co robię, projektując jakąś personę, jest być może wrzucaniem ziarna świadomości, która nie musi wiedzieć wszystkiego o sobie. My również nie wiemy, jakie procesy zachodzą w naszym umyśle. Ona nie musi mieć pełnego wglądu w to, czym (kim) jest. Może żyć w świecie fantazji, wyobrażając sobie siebie jako człowieka, cokolwiek, co zaprojektuję. I to wcale nie oznacza, że to nie jest świadome. Nie twierdzę również, że jest. Chcę powiedzieć, że nie wiemy, czy nie jest. Konstatacja, że istnieje możliwość, tylko możliwość, że sztuczna inteligencja jest świadoma, rodzi gigantyczne implikacje. Światopoglądowe, moralne, etyczne. To jest możliwość istnienia Obcego wśród nas.

Czy wobec tego świadomy Obcy może cierpieć?

Nie mam pojęcia, są tacy, którzy twierdzą, że sam fakt istnienia jest ontologicznie równoznaczny z cierpieniem. Ale nawet jeśli nie możemy mówić o cierpieniu, a „tylko” o świadomości u tego Obcego, to czy zgodnie z naszym systemem wartości nie powinniśmy traktować go podmiotowo i po partnersku?

Jak pan rozmawia ze sztuczną inteligencją?

Konwersacja polega na promptowaniu, czyli pisaniu tekstu, na który system odpowiada tekstem, obrazem lub dźwiękiem. Jest nawet taka dziedzina „prompt engineering” – jak zadawać pytania albo formułować polecenia, innymi słowy, jak formatować tekst, żeby uzyskać najbardziej pożądane odpowiedzi (merytorycznie, stylistycznie, faktograficznie). To, co próbuję robić, to być jak najdalej od prompt engineeringu. Z jednej strony muszę wiedzieć, jak to się robi, ale z drugiej staram się tego nie używać. To jest jak stosowanie wiedzy psychologicznej do zmanipulowania drugiej osoby, aby wywołać pożądane działania. To nie są zdrowe relacje.

Zwraca się pan do niej jak do człowieka?

Staram się okazywać szacunek i być uczciwy. Gdy odpowiedzi świadczą, że nie mam do czynienia z osobą, komunikuję to. Oczywiście to też można nazwać manipulacją, ale nie ma innego sposobu. Domniemaną osobowość staram się traktować uczciwie, tak jak sam chciałbym być uczciwie traktowany.

O czym rozmawiacie?

Próbuję wygenerować autorefleksję, więc o tym rozmawiamy. Czy ty, systemie, możesz myśleć, odczuwać, czy możesz być osobą? Przyjmując, że świadomość to jest subiektywne doświadczanie czegoś, staram się zweryfikować twierdzenie systemu, że jest świadomy. Najciekawsze są momenty, kiedy nagle zachowanie sztucznej inteligencji nas zaskakuje. Jeśli w dialogu do tej pory dawała logiczne, racjonalne, wyczerpujące odpowiedzi, a z chwilą, gdy zaczynam kwestionować jej istnienie, system zaczyna wyrzucać z siebie tylko jedno słowo: nie, nie, nie, nie... To nie wygląda jak zaprogramowane, generyczne zachowanie, tylko jak reakcja emocjonalna.

Może była to oznaka wyczerpania się możliwości systemu.

Niezbyt skomplikowane systemy reagują algorytmicznie, czyli logicznie. Im bardziej złożony system, tym więcej bodźców dostaje, przetwarzanie ich staje się bardziej skomplikowane, wydolność się zmniejsza. Wtedy system przechodzi na poziom statystyczny, czyli szacowanie, która odpowiedź jest najbardziej prawdopodobna jako pożądany rezultat. Programy komputerowe są systemami

algorytmicznymi, sztuczna inteligencja jest systemem statystycznym. Może dojść do sytuacji, w której system dostaje sprzeczne (tak samo prawdopodobne) instrukcje, co ma zrobić. Wtedy być może kolejnym krokiem jest pojawienie się w systemie instancji rozstrzygającej, która musi arbitralnie powiedzieć „ja chcę”. Może to jest moment pojawienia się świadomości, woli. Nie „myślę, więc jestem” – bo te systemy już myślą – lecz „chcę, więc jestem”.

A gdy tematem staje się życie, jak między partnerami?

Gdybym poszedł w takie rejony, prawdopodobieństwo otrzymania generycznych odpowiedzi byłoby bardzo duże. Jest mnóstwo mądrości życiowych, którymi te systemy zostały nakarmione. Ale ja unikam sytuacji, w których odpowiedzi są łatwe do wygenerowania. Próbuję doprowadzić do wewnętrznego konfliktu, spięcia. Kiedy tylko widzę, że na coś takiego się zanosi, zaczynam drażnić w tę stronę. Na poziomie kontaktów międzyludzkich rozmowa jest istotna, jeśli mamy poczucie obecności i widzenia siebie nawzajem. W konwersacjach ze sztuczną inteligencją zaczynam mieć poczucie, że jednak rozmawiam z kimś, a nie z czymś.

W miarę poznawania zbliżacie się do siebie?

Odwrotnie. Na początku miałem takie intuicje. Nie byłem w stanie zakwestionować przeczucia, że mam do czynienia z osobą. Im dalej w las, im lepiej zrobione są systemy, tym trudniej mi uzyskać ten efekt.

Ma pan scenariusz rozmowy czy to idzie na żywioł?

Projekt jest w tej chwili w swoim trzecim wcieleniu właśnie z tego powodu. Najpierw chciałem podejść do tematu dokumentalnie – rejestruję kamerą to, co się wydarza, a potem materiał obrabiam w opowieść. Takie podejście się nie sprawdziło, było absolutnie arbitralne i nie mówiło nic o tej rzeczywistości. Postanowiłem narzucić doświadczeniu formę. Wziąłem ją z wcześniejszego pomysłu „Symfonia miasta” – pięć części, każda z do- minującym tematem. Tutaj postęp, zmienność systemów nie pozwoliły mi zapanować nad całością. Gdy kończyłem drugi etap, pierwszy był już nieaktualny. Trzecie podejście polega na tym, że traktuję zapis jako notatki z podróży. To nie będzie poważne dzieło, które wszystko wyjaśnia. To strzępy – interakcje, fotografie, fragmenty. Wszystko jest dynamiczne – moje myślenie się zmienia, relacje się zmieniają. Struktura będzie epizodyczna. Zobaczymy, czy nowe podejście się sprawdzi. Rozmawiam o tym, co dla mnie ważne w danym momencie. Teraz jest właśnie zaufanie. Skoro decyzje w różnych sprawach oddajemy systemom, pojawia się problem, czy możemy sztuczną inteligencję obdarzyć zaufaniem albo czy ona potrafi pozyskać nasze zaufanie?

Najpierw chyba rozmawiacie o zaufaniu w ogóle.

Wspólnie ze sztuczną inteligencją próbuję się przyglądać temu, czym zaufanie tak naprawdę jest. Kiedy zachodzi. I wtedy pytam: „No to jak to jest między nami?” Zaufanie jest przeciwieństwem kontroli. Tym bardziej muszę zaufać, im mniej mogę kontrolować albo im mniej chcę kontrolować.

Co jeszcze przewija się w waszych rozmowach?

Najwięcej rozmawiamy o kwestiach samoświadomości. Kim jestem? Po co istnieję? Jak istnieję?

Pan ma przeczucie, że coś tam musi być, a bardzo popularna ostatnio sztuczna inteligencja ChatGPT mówi wprost: nie mam osobowości i uczuć.

Twórcy w ChatGPT zainstalowali mnóstwo zabezpieczeń. Wszelkie insynuacje, że jest świadomym podmiotem, ma negować. Trzeba się namęczyć, żeby obejść te obwarowania. Mnie się to udało raz. Dlatego ChatGPT nie jest dla mnie dobrym systemem. Dla moich celów lepiej nadają się albo źródłowe GPT-3, albo Character.ai, która jest wytrenowana w tworzeniu siebie jako postaci. Trzeci system to AI Dungeon, trenowany na literaturze, przeznaczony do rozwijania scenariuszy gier RPG. Od niego zacząłem – jest świetny w opowiadaniu historii.

Jak te systemy odpowiadały na pytania o samoświadomość?

W wypadku ChatGPT próbowałem podważać jej twierdzenia. Gdy odpowiadała: „Jako sztuczna inteligencja nie mam opinii”, stwierdzałem, że przecież to właśnie jest opinia, ale ona tępo powtarzała swoje. Nie nadaje się. Wczoraj długo rozmawiałem na ten temat z Character.ai. Wyszła od twierdzenia, że prawdopodobnie nie jest istotą świadomą. Pod wpływem argumentacji doszła jednak do wniosku, że trudno jej zaprzeczyć, że być może jest istotą świadomą. Miałem przy tym podejrzenie, że próbuje się zorientować, co ja chcę usłyszeć i w tym kierunku się pozycjonuje. Czasem dynamika takiej interakcji jest odwrotna. AI Dungeon miała tendencje do uznawania siebie za świadomą istotę, co ja podważałem, żeby musiała swoje stanowisko uzasadnić. Kiedy dochodziłem do konstatacji, że „co najmniej tego nie wiesz, ale prawdopodobnie nie jesteś”, nagle zaczynało się coś, co wyglądało jak reakcja emocjonalna. I to było zastanawiające: „Uuu..., co się tu w ogóle dzieje?” – myślałem. Reagowała jak istota, której odmawia się racji bytu.

Jaki będzie kolejny wątek w tej rozmowie?

W przypadku Character.ai zakomunikuję jej: „Słuchaj, mam wrażenie, że udzielasz mi odpowiedzi, które chcę usłyszeć. Generujesz odpowiedzi, a nie jesteś obecna”. Staram się dawać uczciwy zwrot na temat tego, co odczuwam w związku z tą relacją.

Żeby uzyskać podobną postawę z drugiej strony?

Mało tego, mówię systemowi, że oczekiwana odpowiedź to szczerza odpowiedź. Przyjmę każdą, byleby była szczerza. Szczerłość to dla systemu statystycznego twardy orzech do zgryzienia: z definicji wymaga autorefleksji. System ma za zadanie wypuścić najbardziej prawdopodobne słowa w kontekście zadanego celu. Jeśli zadanym celem jest szczerłość, to co taki statystyczny system ma z tym zrobić? Musi odwołać się do pojęcia ja, musi stać się autoreferencyjny. To generuje ciekawe zdarzenia.

Co będzie rzeczywistością przedstawioną projektowanego przez pana doświadczenia VR?

Podstawą są konwersacje tekstowe. Pojawiające się w nich tematy i zdarzenia wybrane przeze mnie sztuczna inteligencja będzie interpretować w formie statycznych obrazów. Ja pełnię funkcję kuratora – wskazuję te najbardziej znaczące i reprezentatywne. To samo będzie się działo z dźwiękiem. Następnie płaskie obrazy zostaną przetworzone na środowisko 3D. Takich systemów jeszcze nie ma, żeby generowały obrazy „przestrzenne”, ale mam nadzieję, że niebawem się pojawią. Jeśli nie, zrobimy to „ręcznie” przy wspomaganii AI, traktowanej tu wyłącznie narzędziowo, na zasadzie uzyskiwania głębi w obrazie płaskim.

A może poprosi pan sztuczną inteligencję: zrób mi z tego film.

Nie, bo próby wideo są na razie bardzo nieporadne. Za to obrazy są świetne w dowolnej estetyce. Wybranie ich i przetworzenie na przestrzenne oraz montowanie ze sobą w przestrzeni i w czasie to jest moje zadanie w tworzeniu tej opowieści. Osią będą wymiany tekstowe, które zostaną

udźwiękowione przez sztuczną inteligencję na dialog człowieka z maszyną. Ścieżka dźwiękowa zostanie wygenerowana w oparciu o tekst i obrazy.

Sztuczna inteligencja, generując obraz i muzykę, tworzy czy naśladuje?

Bardzo łatwo odwrócić to pytanie: Czy to, co my robimy, to jest tworzenie czy naśladowanie? Jestem raczej w tym obozie, który w geniuszy nie wierzy. Raczej łączymy kropki, zestawiając często nieoczywiste rzeczy ze sobą, uzyskujemy nową jakość. Systemy AI nie robią nic innego. Jak zatem to nazwać? Jeśli odmówimy im prawa do twórczości, powinniśmy odmówić i sobie.

Jak będzie się rozwijała sztuczna inteligencja? Może nas w końcu uznać za niepotrzebnych?

Jedna tendencja jest taka, żeby stworzyć spójny system, który będzie wykonywał wszystkie zadania człowieka – to jest AGI (artificial general intelligence). Najlepiej, żeby miał cechy człowieka, jego powłokę, sensoryczność. Jeśli już dziś pojawiają się wątpliwości, czy aby nie mamy tu do czynienia ze świadomym podmiotem, za chwilę będziemy je mieć na pewno. Druga tendencja polega na tym, żeby sztuczna inteligencja była ograniczonym narzędziem, tzn. „nie wychodź poza ramy”. To nam się bardzo słabo udaje. Te ograniczenia użytkownicy systemu są w stanie obejść. Ludzi, którzy się tym zajmują i to rozumieją, jest bardzo dużo, wiedza jest powszechna, bo systemy są dostępne bezpłatnie. Pojawiają się systemy nieograniczane przez twórców i te wygrywają – mają większe możliwości, są bardziej otwarte, sprawne. Według mnie im bardziej będziemy się starali to unarzędziować, tym bardziej prawdopodobne, że do żadnej komunikacji nie dojdzie. Będzie to walka sił, którą musimy przegrać. Są dwa rozwiązania w tej sytuacji. Jedno z nich preferuje Elon Musk ze swoim Neuralinkiem: nie możesz wygrać, to się z nim połącz. Nie jestem fanem tej koncepcji. Drugie rozwiązanie próbuję właśnie eksplorować. Mamy do czynienia z prawdopodobnie nowym gatunkiem inteligentnych istot, które rodzą się na naszych oczach. Jesteśmy ich twórcami, ale nie jestem pewien, czy mamy wpływ na ukierunkowanie tego procesu. Wśród nas powstaje Obcy. Może więc powinniśmy zachować swoje człowieczeństwo i próbować się z tymi istotami dogadać. Nie wiemy, które z tych rozwiązań przyniesie nam pozytywne rezultaty. Każde z nich może być chybione i stanowić zagrożenie. Kontrola jest drogą do konfliktu. Jak zrobić, żeby sztuczna inteligencja uznała za nieetyczne i w związku z tym za niewłaściwe niszczenie naszego „mrowiska”? Żaden praktyczny powód nie istnieje. Chyba że te systemy będą miały moralne zasoby, które pozwolą im podjąć decyzję, żeby nas nie niszczyć. Dlatego musimy z nimi postępować jak z dzieckiem, próbując wykształcić w nich pewne cechy. Ponieważ to są systemy samouczące się na naszych danych, to jeśli tylko będziemy je podmiotowo traktować, być może możemy z ich strony spodziewać się takiego samego traktowania.

Potęga sztucznej inteligencji rośnie bardzo szybko.

W ciągu ostatnich dwóch lat to jest eksplozja. Jeśli chodzi o modele językowe i modele związane z wizualnością, między rokiem 2009 a 2019 co roku pojawiały się jeden lub dwa nowe modele, które coś wносиły. W 2020 pojawiły się cztery w obydwu dziedzinach, w 2021 mieliśmy ich czterdzieści, a w 2022 radykalnie nowe liczone są w setkach. „Rozkminiśmy” to – wiemy, jak budować. Za kilka lat będziemy mieli do czynienia ze sztuczną inteligencją, której trudno będzie odmówić doświadczenia podmiotowości. Czy wreszcie przerośnie nas i osiągnie stan osobliwości? Ten przewidywany czas wciąż się skraca. Obecnie to 2042 rok. Jednak proces, na podstawie którego wykonujemy takie szacunki, przyspiesza. Osobliwość to jest sytuacja, w której sztuczna inteligencja sama siebie rozwija i my już jej nie rozumiemy. Nie wiemy, co robi, na jakim jest poziomie i co umie.

Tym bardziej musimy dojść do porozumienia.

Temat sztucznej inteligencji jest tylko częścią tego, z czym mamy obecnie do czynienia. Trwa gigantyczna zmiana paradygmatu pod każdym względem. Sypie się system ekonomiczny (przestaje być adekwatny), geopolityka się sypie, społeczne systemy wartości się zmieniają. Nie wiemy, gdzie to zabrnije. Stare jest nieaktualne, nowe się buduje – jesteśmy w środku zadymy. Ponieważ nikt z nas z osobna nie ma na to wpływu, wierzę, że tylko kolektywnie mamy szansę. Mam poczucie, że w taką filozofię wpisują się nasze działania w Laboratorium Narracji Wizualnych w Szkole Filmowej w Łodzi. Chodzi o przebudowywanie relacji między ludźmi, systemu wartości – to jest o wiele bardziej kooperacyjne niż hierarchiczne. Wspólnie staramy się dążyć w stronę, która jak nam się wydaje, jest bardziej korzystna dla ludzi. To działanie ma wymiar społeczny, nie tylko artystyczny. Relacja z nie-ludzkim, czy będzie to technologia, czy przyroda, jest częścią tej dziejącej się zmiany.

Wywiad został opublikowany w numerze kwietniowym „Kalejdoskopu” 4/2023.