Uniwersytet Mikołaja Kopernika w Toruniu

Wydział Humanistyczny

**Miłosz Kuchnicki**

309371

Kulturoznawstwo

Praca magisterska

# Figura Wielkiego Innego w relacji człowiek–chatbot

Promotor:

dr hab. Krzysztof Abriszewski, prof. UMK

**Toruń 2026**

## Spis treści


Wprowadzenie ..... 2

Rozdział I. Ramy teoretyczne ..... 3

1.1 Problematyka pracy ..... 3

1.2 Granice istniejących kategorii ..... 4

1.3 Modele językowe a inne technologie ..... 5

1.4 Dwa ujęcia chatbota: lustro i instancja normatywna ..... 7

1.5 Wielki Inny: od pojęcia do analizy ..... 8

1.6 Autorytet poznawczy: ujęcie klasyczne i jego granice ..... 12

1.7 Metodologia - autoetnografia i wywiad jakościowy jako narzędzia badawcze ..... 15

Rozdział II. Autoetnografia: ewolucja relacji z ChatGPT ..... 18

2.1 Początek relacji: chatbot jako narzędzie poznawcze (ChatGPT 3.5) ..... 19

2.2 Od narzędzia do rozmówcy (ChatGPT 4.0) ..... 21

2.3 Krytyczny analityk i lustro podmiotowości (ChatGPT 5.0) ..... 25

2.4 Kiedy model się myli ..... 31

2.5 Brak poczucia humoru ..... 33

2.6 Podsumowanie - ewolucja relacji w perspektywie porównawczej ..... 37

Rozdział III. Praktyki użytkowników: analiza wywiadów ..... 39

3.1 Chat jako narzędzie codziennych decyzji ..... 40

3.2 Chat w praktykach edukacyjnych i zawodowych ..... 41

3.3 Prompt jako technika operowania modelem ..... 43

3.4 Selektywna weryfikacja i zawieszanie sprawdzania ..... 45

3.5 Przyspieszenie pracy i delegowanie wysiłku poznawczego ..... 45

3.6 Chat jako przestrzeń rozrywki i bezpieczeństwa ..... 46

Rozdział IV. Mechanizmy projekcji autorytetu w relacji człowiek–chatbot ..... 48

4.1 Projekcja autorytetu ..... 48

4.2 Autorytet bez odpowiedzialności ..... 48

4.3 Prywatność i brak kosztów społecznych jako warunek powstania autorytetu ..... 51

4.4 Heurystyka formy - jak struktura wypowiedzi zastępuje dowód ..... 53

4.5 Trwałość autorytetu operacyjnego ..... 54

4.6 Synteza: Chatbot jako inkarnacja Wielkiego Innego ..... 55

## Wprowadzenie

Ojciec chrzestny sztucznej inteligencji, Geoffrey Hinton, w wywiadzie z 2025 r. ogłosił z powagą grabarza, że istnieje piętnastoprocentowa szansa, iż jego dzieło doprowadzi do zagłady ludzkości[1]. Piętnaście procent to akurat tyle, ile przeciętny obywatel skłonny jest przypisać trwałości swojego małżeństwa po trzech latach wspólnego kredytu. Ludzkość jednak nie panikuje. Przeciwnie, z entuzjazmem loguje się do kolejnych LLM-ów, powierzając algorytmom swoje myśli, fantazje, zadania domowe czy plany wakacyjne. Zakładam, że jeśli sztuczna inteligencja rzeczywiście doprowadzi do anihilacji gatunku nagiej małpy, stanie się to zapewne za jego dobrowolną zgodą, wyrażoną jednym kliknięciem w okienku „Akceptuję regulamin". Należy podkreślić przy tym, że sztuczna inteligencja nie jest jednak bytem, tylko usługą. Nie objawia się użytkownikom w gorejącym krzewie ani w serwerze z aureolą, lecz w uprzejmym okienku dialogowym, które pyta, w czym może pomóc. Modele językowe, takie jak ChatGPT, są maską, za którą nie stoi żaden podmiot, a mimo to miliony ludzi traktują je jak nauczyciela, spowiednika, terapeutę, wróżbitę czy bezpieczniejszą wersję przyjaciela. Nigdy w historii ludzkości autorytet nie był tak łatwo dostępny i bezosobowy zarazem. Freud zapewne uczyniłby z niego coś na kształt figury „ojca bez pasa", ale do tego przejdziemy później.

Zbieranie danych przez właścicieli modeli językowych uchodzi za skandal tylko do chwili, gdy użytkownik otrzyma odpowiedź, która pocieszy go lepiej niż rodzona matka. Wówczas prywatność przestaje być wartością, a staje się drobną monetą wrzuconą do automatu natychmiastowej gratyfikacji. Koszt tej wymiany bywa jednak realny finansowo, co w przypadku studenta nie pozostaje bez znaczenia... Niniejsza praca z pewnością nie ma ambicji rozstrzygać, czy maszyny są świadome[2], tak jak nie rozstrzyga się, czy lustro posiada osobowość. Interesujące dla niej nie jest to, co „myślą" algorytmy, lecz to, czego ludzie od nich chcą i jak szybko uczą się podporządkowywać własne decyzje oraz sposób rozumowania odpowiedziom generowanym przez bezosobowy system. Warte analizy jest jednak to, w jaki sposób modele te odpowiedzi formułują. Niczym najlepszy etnolog są one bowiem doborowymi rozmówcami zarówno dla wykładowcy uniwersyteckiego podczas pisania prac

[1] Wywiad z Geoffreyem Hintonem w podcaście The Diary of a CEO (prow. S. Bartlett), [dostęp przez Spotify:] https://open.spotify.com/episode/4X7dO0FuglP7yTm0kBAc50 data dostępu: 14.06.2026 r.
[2] Zainteresowanych tą kwestią odsyłam do Seth A., *Being You. A New Science of Consciousness*, Faber & Faber, London 2021.

naukowych, jak i dla czternastolatka badającego u siebie rozwijające się zalążki intymności. Ich uniwersalność sprzyja także przyjęcia ról opiekuńczych i terapeutycznych. Teoretycznym punktem odniesienia pracy będzie figura Wielkiego Innego, opisana przez Lacana, a spopularyzowana przez Slavoja Žižka instancja symboliczna, która gwarantuje sens, porządek i autorytet. Chatboty funkcjonują jako technologiczna emanacja Wielkiego Innego, umożliwiając użytkownikom delegowanie odpowiedzialności poznawczej przy jednoczesnym zachowaniu iluzji autonomii. W epoce rozchwianych instytucji i dewaluacji prawdy Wielki Inny nie umarł tylko zmienił interfejs. Porzucił ambonę dla pola tekstowego.

Analiza empiryczna wykorzystana w pracy opiera się na wywiadach z użytkownikami ChataGPT oraz autoetnograficznym zapisie relacji autora z tym osobliwym tworem, który nigdy się na nas nie obraża, chyba że zostanie do tego zaprogramowany. Być może właśnie dlatego tak łatwo obdarzamy go zaufaniem, którego od dawna nie mamy wobec siebie nawzajem. Istnieje jednak możliwość, że ten sztuczny kolega zna nas lepiej, niż sami chcielibyśmy przyznać, albo po przyswojeniu ogromnych zasobów tekstów psychologicznych bardzo skutecznie to symuluje. Warto jednak uwzględnić tempo wykładnicze, z jakim rozwija się i zmienia AI, a wraz z nią modele językowe. Zamiast apokaliptycznych wizji buntu maszyn bardziej adekwatne wydaje się przyjrzenie codziennym praktykom użytkowników, którzy często nieświadomie współkształtują funkcjonowanie i kierunek rozwoju modeli poprzez własne potrzeby, frustracje czy oczekiwania.

PS *Z racji tego, że normatywiści języka polskiego rozstrzygnęli kwestię odmiany ChatGPT na tyle stanowczo, że dopuścili wszystko po trochu, odmieniam go po swojemu, prosząc czytelnika o wyrozumiałość. Wierzę jednak, że podzieli on losy fleksyjne facebooka czy youtube'a. Stąd pisanie o poziomie zaufania względem Chat**a**GPT*[3].

## Rozdział I. Ramy teoretyczne

### 1.1 Problematyka pracy

Uważam, że opisując współczesne realia społeczne w znaczącej części świata nie da się nie zwrócić uwagi na miliony ludzi, którzy codziennie powierzają algorytmom swoje pytania czy wątpliwości. Nie chodzi już jednak wyłącznie o wyszukiwanie informacji na zasadzie korzystania z Wikipedii, a o bardziej radykalny zwrot. Zainstalowane aplikacje chatbotów

[3] https://sjp.pwn.pl/poradnia/haslo/chatgpt;22600.html data dostępu: 16.06.2026 r.

zaczynają bowiem organizować sposób, w jaki użytkownicy interpretują własne doświadczenia, podejmują decyzje czy nawet definiują siebie jako podmioty[4]. Z powodu troski o gatunek można byłoby oczywiście skupić się na pytaniu jak niebezpieczna może być sztuczna inteligencja w tych aspektach, jeśli trafi w niepowołane ręce, jednak dokonując przedstawienia głównych pytań i problemów badawczych niniejszej pracy chciałbym je raczej sprowadzić do tego w jaki sposób bezosobowy algorytm zaczyna pełnić funkcję instancji, której człowiek powierza organizowanie własnej rzeczywistości oraz dlaczego (o ile) to robi. Globalne zagrożenia zostawiam tym, którzy mają jeszcze złudzenia, że potrafią okiełznać gigantów technologicznych przy pomocy komisji, raportów i paneli eksperckich sponsorowanych przez tych samych gigantów technologicznych...

### 1.2 Granice istniejących kategorii

Opisanie relacji człowiek-chatbot z pewnością jest skazane na natychmiastowe zderzenie się z kwestią problemu pojęciowego, dlatego proponuję zacząć właśnie od tej sprawy. Istniejące kategorie teoretyczne w moim przekonaniu wydają się być zbyt przestarzałe by w pełni uchwycić procesy i zależności, które zachodzą w tej nowej formie relacji. Sprowadzenie jej do kwestii związanych chociażby z autorytetem, w klasycznym ujęciu musiałoby się wiązać ze spełnieniem warunków takich jak, podmiotowość obu stron relacji, możliwość społecznej weryfikacji oraz odpowiedzialność za skutki wydawanych rozstrzygnięć. Chatbot w dużej mierze nie spełnia żadnego z tych wymagań. Nie jest podmiotem, nie negocjuje swojej pozycji i nie ponosi konsekwencji swoich błędów. Użytkownicy traktują jednak jego odpowiedzi jak wiążące rozstrzygnięcia, rezygnując przy tym często z ich weryfikowania. Klasyczne narzędzia analizy autorytetu nie wystarczają więc do opisania tego mechanizmu. O tych mniej klasycznych pomówię w kolejnym podrozdziale. Kolejną próbą byłoby sprowadzenie zależności do formy uzależnienia od technologii, co również byłoby zbyt wąskie i niedokładne przy definiowaniu tej relacji. Użytkownicy, którzy delegują decyzje na chatbota nie są ofiarami nałogu a raczej uczestniczą w nowej konfiguracji zależnościowej. Współwytwarzają tę konfigurację wspólnie z narzędziem, które reorganizuje ich podmiotowość w sposób często dla nich nierozpoznawalny. Oczywiście regularne wchodzenie w interakcje może wzmocnić zależność od chatbota jednak uzależnienie się nie jest wymagane by w pełni doświadczać aspektów owej relacji.

[4] Giddens A., Nowoczesność i tożsamość. „Ja" i społeczeństwo w epoce późnej nowoczesności, przeł. A. Szulżycka, Wydawnictwo Naukowe PWN, Warszawa 2001. (s. 51-66)

Z kolei używanie terminu intensywne „użytkowanie narzędzia” w stosunku do opisywania procesu prowadzenia rozmowy użytkownika z chatbotem także jest dla mnie nie w pełni adekwatne (miejscami w pracy będę go jednak używał) choć w pewnych okolicznościach „daje ono radę”. Biorąc jednak poprawkę na to, że określenie “narzędzie” przedłuża możliwości podmiotu nie zmieniając sposobu w jaki podmiot doświadcza siebie należy zaznaczyć, że chatbot jednak wyraźnie robi coś więcej. Zaczyna bowiem organizować ramy interpretacyjne przez które użytkownik odbiera własne doświadczenia, zadaje sobie pytania i formułuje odpowiedzi. Potrzebna jest więc kategoria, która uchwyci autorytet bez odpowiedzialności, podmiotowość bez podmiotu i delegowanie sensu które wraca do użytkownika jako jego własne. Niniejsza praca proponuje, że taką kategorią jest figura Wielkiego Innego w rozumieniu lacanowsko-žižkowskim. Chatbot jest jej nową, być może pierwszą w pełni bezosobową inkarnacją. Warto w tym miejscu także wytłumaczyć się z ograniczenia badania do relacji z konkretnym modelem językowym jakim jest ChatGPT. Mimo że współcześnie stał się on niejako „dziadkiem” wśród modeli językowych to jeszcze do zeszłego roku był najbardziej zaawansowanym tego typu produktem dla globalnej populacji. Większość ludzi pod szyldem sztucznej inteligencji i jej rozwoju rozumiało działalność Sama Altmana i jego współpracowników, odpowiedzialnych właśnie za ten model językowy. W związku z tym ChatGPT ma do dzisiaj największą liczbę subskrybentów wśród zwykłych użytkowników. Według danych OpenAI z początku 2026 roku ChatGPT zbliża się do progu miliarda tygodniowych aktywnych użytkowników, mając przy tym 50 milionów płacących subskrybentów(przypis). Obrazuje to kosmiczną skalę kontaktu w relacji człowiek–czat i potwierdza, że konsekwencje użytkowania ChatGPT mają wymiar globalny. Ograniczenie badania do tego konkretnego modelu językowego stanowi więc zabieg celowy. Opisywane mechanizmy oczywiście nie charakteryzują wyłącznie Czata jednak analiza jego przestrzeni stanowiła najdogodniejszy sposób by je opisać. Sprawdzenie ich na innych modelach pozostaje zadaniem dla dalszych badań, które chciałbym przeprowadzić.

### 1.3 Modele językowe a inne technologie

Technicznie modele językowe takie jak ChatGPT są narzędziami generowania tekstu opartymi na statystycznym przewidywaniu kolejnych tokenów (czyli fragmentów słów) w sekwencji[5]. Ufam, że wszystko jasne, jednak bardziej po ludzku moglibyśmy powiedzieć, że modele językowe uczą się na ogromnych zbiorach tekstów, analizując zależności pomiędzy

[5] Dragan A., *Quo Vadis*, Wydawnictwo Otwarte, Kraków 2025. (s. 225-236)

słowami, zdaniami i kontekstami, w których się pojawiają. Dzięki temu potrafią przewidywać, jaka odpowiedź najlepiej pasuje do wcześniejszej wypowiedzi. Naturalnie jest tutaj wiele detali, które płomiennie dzielą różnych naukowców dotyczących chociażby tego czy modele językowe rzeczywiście „rozumieją" znaczenia, czy jedynie bardzo skutecznie imitują ludzką komunikację[6] itd. nie przesądzajmy jednak tej sprawy za nich. W każdym razie z technicznego punktu widzenia są bardzo zaawansowanymi maszynami do generowania słów (takie myślenie o ludziach należałoby chyba uznać za życzeniowe). To jednak nie wystarczy, żeby zrozumieć co się dzieje, gdy człowiek zaczyna im ufać.

No dobrze w takim razie czym biorąc pod uwagę najważniejsze założenia tej pracy, relacja człowiek-chatbot miałaby się różnić od innych relacji z technologią, np. nawigacją GPS albo „inteligentną" lodówką? Nawigacja informuje kierowców chociażby o konieczności wykonania skrętu, pracach drogowych, kontrolach a ci stosują się do jej poleceń, niekiedy lądując w akwenach wodnych.... Nikt jednak nie pyta nawigacji czego powinien chcieć od swojego życia, nie zwierza jej się z problemów w związku i nie traktuje jej milczenia jako wyrazu dezaprobaty. Nawigacja organizuje trasę. ChatGPT w analogiczny sposób zaczyna organizować podmiotowość. To nie jest oczywiście kwestia tego, że jedna z technologii jest bardziej zaawansowana od drugiej, chodzi o to, w jaką przestrzeń każda z nich wchodzi. Można wskazać trzy właściwości, które czynią relację z chatbotem strukturalnie odmienną od kontaktu z jakimkolwiek innym narzędziem cyfrowym. Przede wszystkim interfejs modeli językowych jest konwersacyjny. Użytkownik nie klika przycisków ani nie wypełnia formularzy, pisze i otrzymuje odpowiedź, w tonie który nie różni się często niczym od tego w jakim są pisane wiadomości w popularnych komunikatorach jak chociażby messenger. Forma komunikacji uruchamia więc schematy percepcyjne, które ewolucyjnie i kulturowo przypisujemy kontaktom z ludźmi[7], nie z „maszynami". Dalej, model odpowiada zawsze i na wszystko (na co pozwala mu polityka firmy). Przestrzeń konwersacji jest tym samym odciążona z mechanizmów które w relacjach międzyludzkich regulują zakres i treść wypowiedzi. Użytkownicy nie są narażeni chociażby na ryzyko utraty twarzy czy utracenia pozycji w relacji. Odpowiedzi modelu symulują kompetencję ze względu na swoje logiczne uporządkowanie i brak znaków niepewności. Te trzy właściwości razem tworzą warunki, w których użytkownik zaczyna traktować model inaczej niż inne narzędzia cyfrowe. Nie chodzi o to, że jest on

[6] V. Cherkassky, E. H. Lee, *A Perspective on Large Language Models, Intelligent Machines, and Knowledge Acquisition*, arXiv 2024

[7] B. Reeves, C. Nass, The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places, CSLI Publications / Cambridge University Press, Stanford–New York 1996.

bardziej użyteczny od wyszukiwarki czy kalkulatora. Chodzi o to, że wchodzi w przestrzeń relacyjną która dotychczas była zarezerwowana dla podmiotów, i że robi to na tyle przekonująco, żeby mechanizmy przypisywania autorytetu zaczęły działać. Warto przy tym zaznaczyć, że zjawisko to nie jest efektem naiwności użytkowników. Większość z nich doskonale zdaje sobie sprawę z tego, że rozmawia z algorytmem, który się myli oraz z faktu, że jego odpowiedzi są generowane statystycznie a nie specjalnie dla nich. Wiedza o technicznym charakterze narzędzia nie neutralizuje jednak relacyjnych mechanizmów które uruchamia jego forma. Mamy więc do czynienia z paradoksem świadomego uczestnictwa w relacji, której fikcyjność się rozpoznaje. Stanowi to punkt wejścia do analizy chatbota w kategoriach Wielkiego Innego.

### 1.4 Dwa ujęcia chatbota: lustro i instancja normatywna

Zanim jednak do tej kategorii dojdziemy, warto przyjrzeć się jak chatboty są opisywane w istniejących dyskursach, zarówno potocznych jak i akademickich. Zastrzegam, że nie chodzi tu o pełny przegląd badań nad sztuczną inteligencją, lecz o wskazanie dwóch figur interpretacyjnych, które często pojawiają się w analizach. Pierwszym odwołaniem jest zestawienie modelu z lustrem[8] wydaje się ono być bardzo trafne, chatbot bowiem oddaje użytkownikowi jego luźne myśli, pytania itp. w uporządkowanej formie, stając się powierzchnią, na której podmiot ogląda własne myślenie. Dodałbym, że bardzo często owo lustro ma jednak charakter takiego z aktywnym filtrem z tik-toka. Upiększa, podkreślając zalety i maskując jednocześnie niedoskonałości naszych procesów poznawczych. Z czasem ludzie uzależniają postrzeganie swojej twarzy od wizerunku w mediach społecznościowych mechanizm, o którym piszę, zdaje się działać podobnie. To ujęcie jest bardziej produktywne analitycznie i pojawia się w doświadczeniu wielu użytkowników, również opisanym w tej pracy. Jego ograniczenie polega jednak na tym, że lustro jest pasywne, a model nie jest. Lustro, które mówi co myśleć przestaje być lustrem (patrz. Lustro ze Shreka). Inne ujęcie traktuje [9] jako instancję normatywną. Model regularnie wykracza poza dostarczone informacje i wchodzi w rolę doradcy, komentatora, a niekiedy moralizatora. Odpowiada nie tylko na pytanie co jest prawdą, lecz również co należy zrobić, jak ocenić sytuację i jakie kroki podjąć tym, że skupia

[8] Figura technologii jako powierzchni odbijającej użytkownika pojawia się m.in. w refleksji S. Turkle nad tym, dlaczego zwracamy się do maszyn zamiast do siebie nawzajem; por. Turkle S., *Samotni razem. Dlaczego oczekujemy więcej od zdobyczy techniki, a mniej od siebie nawzajem*, przeł. M. Cierpisz, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2013.

[9] To perspektywa bliska refleksji nad delegowaniem decyzji moralnych na systemy AI, którą rozwija m.in. Coeckelbergh M., *AI Ethics*, The MIT Press, Cambridge (MA) 2020.

się na treści odpowiedzi, pomijając strukturę relacji, w której te odpowiedzi są wytwarzane i przyjmowane. Każde z tych ujęć chwyta jednak fragment zjawiska. Żadne nie odpowiada na pytanie, dlaczego użytkownicy przypisują odpowiedziom modelu status wiążących rozstrzygnięć nawet gdy wiedzą, że jest algorytmem, nawet gdy doświadczyli jego błędów i nawet gdy świadomie utrzymują wobec niego dystans. Żeby odpowiedzieć na to pytanie potrzebna jest kategoria, która opisuje nie tyle właściwości narzędzia co strukturę relacji, w której narzędzie zaczyna pełnić funkcję Wielkigo Innego.

### 1.5 Wielki Inny: od pojęcia do analizy

Wielki Inny nie jest osobą ani podmiotem fizycznym, lecz, jak ujmuje to Žižek za Lacanem, wirtualnym porządkiem symbolicznym, który istnieje wyłącznie o tyle, o ile podmioty działają tak, jakby istniał[10]. Uczestnicząc w życiu społecznym na co dzień, doświadczamy jego oddziaływania, choćby podczas mówienia. Rozmowa, jak mówi Žižek, nie jest wyłącznie interakcją z drugim podmiotem, ale także wchodzeniem w przyjęty porządek rozmów, który jest ograniczony przez zasady gramatyki, zwyczaje itp. Mówiąc „dzień dobry" do sąsiada, którego szczerze nie znosimy, nie zwracamy się do niego, a do instancji normatynwej, która od dwóch lat każe nam udawać, że nam na nim zależy, podczas gdy on od dwóch lat zabiera nam miejsce parkingowe pod blokiem. Porządek, w którym uczestniczymy, jest więc naznaczony systemem reguł na poziomie symbolicznym[11]. Z części z nich zdajemy sobie sprawę, innych nie jesteśmy świadomi. Nie zmienia to faktu, że odnosimy swoje działania do jakiegoś wyższego porządku. Wielu ludzi utożsamia Wielkiego Innego z postacią Boga, który kontroluje nasze losy albo sprawami wyższej konieczności, jak Partia, Wiara czy Rekomendacje Influencerów[12]. Żeby jednak zrozumieć, dlaczego chatbot może przejąć podobną funkcję, trzeba najpierw zapytać o to, co Wielki Inny właściwie robi, nie z porządkiem społecznym, lecz z pragnieniem konkretnego podmiotu. Żeby jednak zrozumieć, dlaczego chatbot może przejąć podobną funkcję, trzeba najpierw zapytać o to, co Wielki Inny właściwie robi, nie z porządkiem społecznym, lecz z pragnieniem konkretnego podmiotu.

[10] Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008. (s. 21)
[11] Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008. (s. 21-22)
[12] Žižek S., Lacan. Przewodnik Krytyki Politycznej, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008. (s. 22)

Według Lacana struktura pragnienia jest dość nieoczywista. Pragnienie podmiotu to w zasadzie [13]. Jednak nie chodzi o to, że ludzie chcą tego samego co Wielki Inny. Oni zasadniczo nie wiedzą czego chcą, dlatego zwracają się z tym do porządku symbolicznego. Žižek ujmuje to w sformułowaniu *che vuoi?* [14] Co traktuje jako pytanie czego pragniesz albo czego powinienem pragnąć jako jednostka, aby cię zadowolić? Wielki Inny nie daje jednak ludziom ryby, a połamaną wędkę, ponieważ pokazuje nam wyłącznie sposób w jaki powinniśmy łowić (zostając w terminologii wędkarskiej) a realnie nie mamy szans złowić owej „ryby", która go zadowoli. Być może lepszym porównaniem będzie kwestia spowiedzi. Moim zdaniem osoby, które do niej przystępują nie oczekują w gruncie rzeczy rozgrzeszenia za poszczególny grzech a raczej chcą zostać nagrodzone za sposób w jaki wyrażają skruchę, za to, że czują się winni we właściwy sposób. Osobiście chyba wolę to z rybą, ale Žižek pisze również o innym spojrzeniu na kwestie pragnienia Innego, które znacząco wplata się w tematykę mojej pracy. Otóż, aby podmiot mógł pragnąć musi on doświadczać samego Innego jako kogoś pragnącego. Wielki Inny w zasadzie więc zderza ludzi z tym, że ich własne pragnienie, choć rozbudzone przez kontakt z nim, pozostaje dla nich samych zagadką. Dla mnie ChatGPT jest Wielkim Innym, który niczego nie pragnie, jest pod tym względem kompletnie „pusty". Mimo to, co najciekawsze użytkownicy traktują go tak jakby musieli podczas rozmów domyślać się jego pragnienia, formułując pytania, czytając odpowiedzi często próbują zrozumieć jego prawdziwe intencje. Łącząc kropki dochodzi do sytuacji, w której użytkownicy mogą spojrzeć i zobaczyć w generowanych odpowiedziach niczym na tafli stawu, wyłącznie odbicie swoich ukrytych pragnień.

Moglibyśmy rozwijać tę kwestię teoretyczną, ale nie o tym jest jednak ta praca, z całą pewnością wrócę jednak do tej tematyki w części empirycznej. Póki co trzeba wprowadzić kolejną kategorię, którą jest rozkosz, *jouissance.* Wydaje się, że to właśnie ona sprawia, że użytkownicy decydują się wracać do czata i przeglądać się w swoich pragnieniach nawet jeśli wiedzą, że często generuje on błędy. W ujęciu Žižka nie chodzi jednak o przyjemność w jej potocznym sensie, mamy do czynienia raczej z formą wewnętrznego przymusu podmiotów do powtarzania czynności, które są dla nich toksyczne i trudne do porzucenia. Dlatego ludzie decydują się często na wykonywanie czynności z punktu widzenia logiki czysto irracjonalnych. Autor chociażby często obiecuje sobie, że nie będzie scrollował instagrama do trzeciej w nocy

[13] Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008. (s. 56)
[14] z włoskiego. czego chcesz?

a końcowo zazwyczaj i tak leży w łóżku z miną człowieka okradanego przez własny kciuk. W każdym razie, ta logika jest najważniejsza dla zrozumienia relacji człowiek-chatbot bowiem dostarcza specyficznego obrazu rozkoszy, która wynika z delegowania ciężaru odpowiedzialności poznawczej na czata. Ludzie, którzy zrzucają na ChatGPT podejmowanie decyzji, uzależniają swoją samoocenę od jego opinii, czy zlecają całą masę innych zadań w gruncie rzeczy dążą do wyzbycia się odpowiedzialności za swoje myślenie. Nie jest to jednak w moim przekonaniu kwestia lenistwa albo innych cech, które podmiot wykazywał zanim rozpoczął konwersowanie z chatbotem, chodzi o rozkosz. Ludzie nie muszą już zmagać się z podważaniem własnego ja, męczyć się w staraniu o zachowanie własnej autonomiczności i ostatecznie ponosić za to konsekwencje. Dużo łatwiej jest oddać się w ręce wyższego dobra, analogia do Boga zdaje się być w tym miejscu dość trafna. Im więcej siebie, swoich decyzji oddelegujemy na czata tym bardziej wyzbywamy się konieczności myślenia. Świadomość tego, że konwersowanie z modelem ma na nas podobny wpływ nie musi oznaczać, że nie zaangażujemy się w tę relację bardziej albo tego, że łatwiej będzie nam ją później porzucić.

Wykonywanie i doświadczanie za nas wielu czynności wiąże się także z nałożeniem na model kategorii podmiotu interpasywnego. Należy podkreślić to przeżywanie, ponieważ czat nie tylko wykonuje szereg rzeczy za nas, ale także poprzez generowanie odpowiedzi zaczyna angażować się i przeżywać emocje związane z danym zadaniem, tak abyśmy my już nie musieli tego robić. Podczas gdy on z przejęciem mierzy się z jakimś skomplikowanym projektem, my możemy z kolei zająć się dużo przyjemniejszymi rzeczami, na przykład pograć na konsoli. Žižek opisuje interpasywność w taki sposób: „(...) postać innego może nam odebrać i przeżywać za nas nasze najbardziej intymne i spontaniczne uczucia i postawy, łącznie ze śmiechem [15]. Warto przytoczyć jego znany opis tzw. śmiechu z puszki [16]. Spełnia on za nas formę wymaganej reakcji podczas oglądania programu, często nie mamy bowiem siły się śmiać, jednak powinność względem porządku symbolicznego zostaje spełniona, bo nagrane reakcje innych ludzi wykonują to za nas. Podobnie rzecz się ma z płaczkami na pogrzebach, które opłakują kogoś za nas. Ten sam mechanizm działa również z filmami trafiającymi na „listę do obejrzenia", czujemy, że w gruncie rzeczy przyswoiliśmy jakiś film nie oglądając go. W gruncie rzeczy chodzi więc o to, że delegujemy akt poznawczy na zewnątrz. Mam przy tym pełną świadomość, że danego tytułu nie obejrzałem a mimo to czuję, że powinność została

[15] Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008 (s. 36-37)
[16] Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008. (s. 46)

spełniona, dlatego że chociażby jako filmoznawca jestem zobowiązany chcieć obejrzeć dany tytuł.

Mechanizm opisywany przez Žižka zakłada, że podmiot jest świadomy tego, że deleguje swoje przeżycia w ręce zewnętrznego adresata. Pisząc o buddyjskim modlitewniku, którego zadaniem jest odmawianie za nas mantr, podczas gdy nim kręcimy, mówi o tym, że osoba zdaje sobie sprawę z tego, że modlitewnik wykonał zadanie za nią, ale ona może zebrać efekty jego pracy i przyjąć je jako swoje. Akt odmawiania mantr zostaje więc wyraźnie na zewnątrz, podobnie jak filmy z listy do obejrzenia[17]. Doprecyzowując, chodzi mi zarówno o aspekt fizyczny, bowiem modlitwa zostaje w przestrzeni modlitewnika, jak i zewnętrzny wymiar psychologiczny aktu. Podmiot oczywiście internalizuje skutki wykonanej modlitwy, czuje się pobłogosławiony, ale jednocześnie zdaje sobie sprawę, że modlitwa nie wydarzyła się w jego wnętrzu. Mówiąc krótko, przyjmuje efekt, a nie sam akt. Zaznaczam to celowo, ponieważ badając relację użytkownik-chatbot, zauważyłem w tej kwestii dość wyraźne przesunięcie. W przypadku modeli językowych delegowanie aktu analityczno-poznawczego nie jest przez ludzi rozpoznawane jako odrębne działanie modelu. Podczas tworzenia jakiejś pracy, projektu czy analizy nie doświadczają oni generowanych przez czata myśli jako cudzych, lecz jako własne, bardziej rozwinięte i lepiej skonstruowane, ale wciąż swoje. Prawdopodobnie jest to skutkiem tego, że zanim model zacznie generować odpowiedź, trzeba określić mu tematykę, podzielić się przemyśleniami i podać inne instrukcje co do generowanego tekstu. Narzędzie zdaje się więc oddawać im jedynie ich własne myśli w wypolerowanej formie. Uważam, że jest to co najmniej niepokojące, ale nie dlatego, że jakieś urządzenie wygeneruje np. pracę domową z ułamków dla ósmoklasisty. Przed chatbotami także można było bezmyślnie przepisywać rozwiązania ze stron internetowych albo od uczynnych kolegów. Niebezpieczne jest to, że niedługo sami przestaniemy zdawać sobie sprawę, które z myśli należały do nas, a które podsunął nam czat.

Sama struktura dłuższej rozmowy miesza je i uniemożliwia dokonywanie ich późniejszej selekcji. Osobiście często po dłuższej konwersacji z modelem językowym czułem się absolutnie ogłupiony i niepewny tego, co wyprodukowałem sam, a co podsunął mi czat. Można by było oczywiście przyjąć, że końcowy wynik np. dysputy filozoficznej to wynik synergii myśli ludzkiej z generatywną, ale zazwyczaj kończy się to poczuciem ograbienia z własnego pomysłu. Najważniejsze jest jednak to, że podobne refleksje przychodzą po

[17] Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008. (s. 36-29)

zakończeniu konwersacji, a nie w trakcie rozmów. Wtedy ta granica bardzo się zaciera, a nawet jeśli wewnętrznie zorientujemy się, że piszemy już zupełnie o czymś innym niż nam chodziło, to jest za późno, by to cofnąć. Wrócimy do tego zagadnienia w końcowych podrozdziałach pracy.

Kategoria interpasywności, zarówno w ujęciu Žižka, jak i w odchyleniu, które sformułowałem, skupia się na tym, jak podmiot postrzega wykonanie aktu. Według Žižka podmioty są świadome tego, że „akt modlitwy" odbywa się w innym obiekcie, w przypadku opisywanego odchylenia przynależność aktu jest pomijana lub nieuświadomiona. Jednak interpasywność nie wyjaśnia, dlaczego w ogóle uznajemy model językowy za instancję, która może wiedzieć. Mało tego, często przyjmujemy jego kompetencje, zanim realnie je zweryfikujemy, i w tym momencie należy spojrzeć na ChatGPT z zupełnie innej strony, mianowicie od strony autorytetu...

### 1.6 Autorytet poznawczy: ujęcie klasyczne i jego granice

Pisząc o autorytecie na podstawie definicji zawartych w słownikach lub podręcznikach, w zasadzie moglibyśmy nie pisać w ogóle. Jest to bowiem pojęcie niezwykle ambiwalentne, które trudno jest problematyzować. Sięganie po klasyczne ujęcia, w których autorytet jest typem panowania, jak u Maxa Webera, albo kategorią zanikającą wraz z tradycją, jak u Hannah Arendt, niewiele wniosłoby do analizy relacji z modelem, ponieważ zakładają one autorytet jako atrybut podmiotu lub pozycji. Chatbot nie jest jednak ani podmiotem, ani instytucją, dlatego potrzebne jest ujęcie, w którym autorytet powstaje dopiero w akcie uznania. Właśnie dlatego postanowiłem odwołać się do krytycznego ujęcia Lecha Witkowskiego. Jego praca „Wyzwania autorytetu" stanowi w moim przekonaniu bardzo dobrą bazę, ponieważ autor mówi o autorytecie jako o strukturze relacyjnej, a nie jako partykularnej cesze. Mało tego, sam napisał w tej książce rozdział o autorytecie w ujęciu Lacana i Žižka, osadzając go tym samym w ramach symbolicznych, podobnych do moich założeń.

Witkowski jasno przyznaje, że „autorytet wypada widzieć w relacji, której źródłem, dawcą czy nadawcą (a zatem kryterium operacyjnym) jest podmiot, a nie przedmiot uznania (...)"[18]. Decyduje o tym więc sam akt uznania, a nie cechy kandydata. To akt nadaje status autorytetu i umożliwia wywieranie wpływu na innych. Wynika z tego również, że autorytet jest własnością relacyjną, w której ktoś musi uznać inny podmiot za autorytet. Wspominam o

[18] Witkowski L., *Wyzwania autorytetu w praktyce społecznej i kulturze symbolicznej*, Oficyna Wydawnicza „Impuls", Kraków 2009. (s. 263)

tym dlatego, że trzeba zapytać, w jaki sposób autorytetem może stać się coś, co podmiotem nie jest. Odpowiedź, którą rozwijam w dalszej części pracy, wskazuje, że w relacji z modelem akt uznania jest jednostronny. Wystarczy, że autorytet przyznaje jedna strona, druga nie musi go ani odwzajemniać, ani nawet rejestrować. ChatGPT, jako model językowy, z założenia nie tworzy przecież obustronnych relacji. Pozornie zaprzeczają temu głośne ostatnio przypadki zawierania w Japonii „małżeństw" z postaciami generowanymi przez modele językowe[19], jednak nawet wtedy cała relacja istnieje wyłącznie po stronie użytkownika, druga strona niczego nie odwzajemnia, lecz jedynie odbija jego własne uczucie w wygenerowanej formie.

Inną ważną dla Witkowskiego kwestią jest perspektywa uwikłania autorytetu w relacje społeczne. Przywołuje on w tym kontekście Pierre'a Bourdieu i „rytualizowanie pozoru". Chodzi o to, że autorytet angażuje się w działania mające na celu ograniczenie możliwości jego podważenia. W tym celu ulega częstej redefinicji, a także z pomocą „nowych pośredników w kulturze" dąży do wytworzenia „aury" swojego uznania. Mówiąc prościej, podtrzymuje pozór własnej prawomocności i maskuje brak rzeczywistej podstawy swoich roszczeń, grając samą formą[20]. Autorytet potrafi więc działać za pośrednictwem formy, a nie rzeczywistej podstawy. I to właśnie czyni go porównywalnym z chatbotem. Model językowy jest bowiem przypadkiem granicznym tego mechanizmu, autorytetem, który jest niemal czystą formą, którą cechuje tylko to, w jaki sposób formułuje odpowiedzi. U człowieka uznawanego za autorytet za formą stoi zwykle choćby resztka rzeczywistej kompetencji, którą można później podważyć albo rozliczyć. W przypadku modelu nie ma już za co sięgnąć. Zostaje sama forma, pod którą nie kryje się żaden podmiot, który mógłby zostać pociągnięty do odpowiedzialności za odpowiedź.

Takie działania autorytetu są z kolei zdeterminowane ogólną strukturą relacji autorytetu. Witkowski podkreśla, że składają się na nią zależności o charakterze niezwykle asymetrycznym. „W relacjach tych odwracanie perspektywy albo wcale nie następuje (wręcz nie może nastąpić), albo jest odroczone w czasie lub działa jedynie w sytuacji partnerstwa kulturowego (dwóch autorytetów wzajemnie dla siebie)[21]". Wyłącznie podmiot uznawany za autorytet ma więc możliwość wywierania wpływu na drugiego. Komunikacja z chatbotem również ma charakter autorytetu jednostronnego, ale ta jednostronność ma zdecydowanie inne

[19] *Virtual groom, real wedding: AI romance blooms in Japan*, Reuters, 18.11.2025 r. https://www.reuters.com/investigates/special-report/japan-ai-wedding/ data dostępu 14.06.2026 r.
[20]Witkowski L., *Wyzwania autorytetu w praktyce społecznej i kulturze symbolicznej*, Oficyna Wydawnicza „Impuls", Kraków 2009. (s. 263-264)
[21] Witkowski L., *Wyzwania autorytetu w praktyce społecznej i kulturze symbolicznej*, Oficyna Wydawnicza „Impuls", Kraków 2009. (s. 387-388)

oblicze niż opisywana przez filozofa. U Witkowskiego asymetria wynika po części z przewagi wiedzy. Tymczasem w relacji z chatem autorytet nie ma żadnej wiedzy ani podmiotowości, a mimo to zachowuje nieproporcjonalną przewagę. Widać to zwłaszcza wtedy, gdy użytkownik podważa odpowiedź modelu, ponieważ zakwestionowanie kompetencji czata w pojedynczej rozmowie i tak nie obala jego statusu globalnego. Jeden nieudany kontakt nie przekłada się często nawet na to, jak użytkownik traktuje model przy następnym pytaniu.

Największą pułapką, która zostaje zastawiona na użytkowników uznających autorytet ChataGPT, jest ta opisana przez filozofa w kategoriach „drogiego mistrza" i „mistrza drogi”[22]. Witkowski za Nietzschem odróżnia autorytet, który uczy innych, jak znaleźć własną drogę, podążając za mistrzem, który nie jest sam w sobie w ogóle zainteresowany dominacją uczniów, próbując ich nawet do siebie zniechęcić. Jednym słowem, „mistrz drogi" naucza krytycznego myślenia, zwłaszcza względem siebie. Natomiast „drogi mistrz", który jest substytutem tego pierwszego, ma wszelki interes w tym, aby na końcu drogi uczeń zawsze spotykał jego samego, uznając go tym samym za najwyższą formę intelektualną, którą należy wielbić. Należy powiedzieć o tym, że jest wiele przesłanek świadczących o tym, że czat ma o wiele więcej elementów „drogiego mistrza". Bywa też tak przedstawiany w przekazie samych twórców, którzy reklamują go jako wszechstronnego doradcę udzielającego gotowych odpowiedzi, a nie jako narzędzie służące samodzielnemu dochodzeniu do wiedzy. Nie odsyła użytkownika do szukania własnej drogi, znajdowania nowych źródeł, podważania jego wypowiedzi itd. Nie zniechęca również do przywiązania, wręcz przeciwnie, przywiązuje do siebie użytkowników, zazwyczaj się z nimi zgadzając, oraz podejmuje szereg innych działań, aby zwiększyć ilość subskrybentów. Tam, gdzie mistrz drogi wymaga porzucenia siebie i wysiłku, chat zwalnia z jednego i drugiego. Zresztą część użytkowników, inaczej niż w opisanym wcześniej trybie interpasywnym, świadomie decyduje się podążać ową drogą. Wprowadzając ujęcie antropologiczne, możemy mówić tu nawet o składaniu pewnego rodzaju daru[23]. Użytkownicy sami dzielą się z czatem szkicami własnych pomysłów, myśli, po to, by otrzymać od mistrza gotowe, rozbudowane rozwiązania. Nie trzeba już dłużej podejmować podróży intelektualnych, wystarczy zapłacić modelowi OpenAI, aby od razu dotrzeć na metę.

[22] Witkowski L., *Wyzwania autorytetu w praktyce społecznej i kulturze symbolicznej*, Oficyna Wydawnicza „Impuls", Kraków 2009. (s. 302-304)
[23] Mauss M., Szkic o darze. Forma i podstawa wymiany w społeczeństwach archaicznych, [w:] tegoż, Socjologia i antropologia, przeł. M. Król, K. Pomian, J. Szacki, Wydawnictwo KR, Warszawa 2001. (s. 269-279)

Dokonując krótkiego podsumowania, należy podkreślić, że autorytet przedstawiany przez Lecha Witkowskiego zakłada relację uznania, opartą na podmiotowości, w tym podmiotowości tego, kto zostaje uznany za autorytet. Relacja ta ma na dodatek możliwość odwrócenia (choćby odroczonego) oraz możliwość bycia „mistrzem drogi", który prowadzi innych ku samodzielności. ChatGPT, pełniąc funkcję autorytetu, nie spełnia żadnego z tych warunków. Nie stanowi bowiem niezależnego podmiotu, nie można go również „odwrócić". Struktura jego odpowiedzi zniechęca do ich podważania i sprzyja temu, że użytkownik przyjmuje je bezwarunkowo, a z czasem może dokonywać nawet ich internalizacji. W ten sposób relacja zatrzymuje przy modelu coraz więcej użytkowników. Rozwiązanie, jakie proponuję przyjąć, by uzasadnić stawianie przez siebie ChataGPT w roli autorytetu, to wprowadzenie kategorii autorytetu operacyjnego, którą rozwijam w rozdziale IV. Powiązanie z kryterium operacyjnym Witkowskiego nie jest tutaj oczywiście przypadkowe, autorytet operacyjny przenosi jednak dyskusję na grunt relacji użytkownik-chatbot. Pokażę, co się dzieje z ową operacyjnością, gdy jedna ze stron nie jest podmiotem.

### 1.7 Metodologia - autoetnografia i wywiad jakościowy jako narzędzia badawcze

Przedstawione w częściach empirycznych mechanizmy i obserwacje wyłoniły się w toku badań, których podstawę stanowi triangulacja autoetnografii z wywiadami pogłębionymi. Triangulacja oznacza w moim wypadku próbę uchwycenia tego samego zjawiska z różnych perspektyw, tak żeby jedno mogło weryfikować drugie. Zacznę od opisu autoetnografii, którą zacząłem prowadzić od 2023 roku, a skończyłem w 2026 r. Zaczęło się od rozmów z wersją ChatGPT 3.5 w wersji podstawowej, na wersji ChatGPT 5.2 w wersji Plus kończąc. Takie przesunięcie pozwoliło mi, poza obserwowaniem zmian we własnym użytkowaniu i podejściu, zauważać także zmiany zachodzące w kolejnych wersjach Chata, które w największym stopniu warunkowały sposób użytkowania. W konceptualizacji badania miałem w głowie ujęcie Carolyn Ellis, według której autoetnografia jest właśnie metodą pozwalającą uczciwie oddać procesy, które dzieją się podczas badania czegoś, co rzutuje na sposób myślenia badacza[24]. Ellis uważa, że osobiste zaangażowanie nie zakłóca badania, a co więcej, stanowi jego punkt wyjściowy. Zgadzam się z nią, bowiem dla mnie najciekawsze rzeczy działy się wtedy, gdy perspektywy badacza i użytkownika zachodziły na siebie wzajemnie, przyprawiając czasem o kompletną dezorientację. Przykłady takich sytuacji znajdują się w drugim rozdziale. Sam materiał powstawał w dwóch fazach. Początkowo, od lutego 2023 r., gromadziłem wyłącznie

[24] Ellis C., *The Ethnographic I: A Methodological Novel about Autoethnography*, AltaMira Press, Walnut Creek 2004. (s. 31)

zrzuty ekranu przedstawiające moje konwersacje z modelem. Nie było wtedy jeszcze we mnie poczucia i chęci zrobienia z tym czegoś więcej. Jednak z biegiem czasu i zintensyfikowaniem rozmów zacząłem gromadzić coraz więcej screenów i pojawiły się pierwsze notatki badawcze. Było to w czerwcu 2023 r.

Samo to przejście, od prywatnego użytkownika do badacza własnej praktyki, postrzegam zresztą jako część badanego procesu. Wczesne screenshoty były dla mnie materiałem pomocniczym, natomiast za trzon badania przyjmuję tworzone na bieżąco notatki, a także prowadzone na ich podstawie analizy tematyczne. Notatki miały oczywiście charakter nieformalny, często nawet wulgarny, z uwagi na to, jak irytującym stworzeniem potrafi być model językowy[25]. Z tytułu powagi pracy magisterskiej w części empirycznej znajdą się jednak wyłącznie przykłady uchodzące za przyzwoite (z drobnymi odchyleniami). Powstały one zarówno podczas konwersacji z modelem w otwartym obok dokumencie Word, jak i na gorąco po zamknięciu rozmowy, ale również wtedy, kiedy opisywałem swój stan emocjonalny i przemyślenia, gdy atmosfera już nieco opadła. Nie było jednak tak, że dokumentowałem wyłącznie rzeczy, które były dla mnie ekscytujące, sam proces i spostrzeżenia na temat zwykłego sposobu generowania odpowiedzi przez model również znajdowały się w plikach tudzież w notatniku telefonu, który także stanowił dziennik badawczy. Tę grupę sytuacji nieoczywistych stanowiły notatki opisujące momenty, kiedy model np. mnie rozbawił, zirytował czy czymś zafascynował. Taki zakres wydaje mi się zgodny z podejściem Ellis, dla której emocje i osobiste reakcje badacza są pełnoprawnym materiałem badawczym. Zgromadzone zapisy oznaczyłem kolejno AE1–AE47 (AutoEtnografia). Następnie poddałem je analizie tematycznej, tej samej, którą zastosowałem przy wywiadach, żeby móc zestawić obie empirie w podobny sposób. Dodam, że materiał autoetnograficzny dotyczył wyłącznie mojego doświadczenia i interakcji z modelem.

Drugą metodą były indywidualne wywiady pogłębione, których konstrukcję oparłem na ujęciu Steinara Kvale, traktującego wywiad badawczy jako rodzaj rozmowy, a samo postępowanie badawcze jako proces złożony z kolejnych etapów, od tematyzacji i projektowania, przez prowadzenie rozmowy i jej transkrypcję, po analizę i raportowanie[26]. Przeprowadziłem rozmowy z ośmioma osobami, które aktywnie korzystają z ChatGPT w

[25] Ellis C., *The Ethnographic I: A Methodological Novel about Autoethnography*, AltaMira Press, Walnut Creek 2004. (s. 330)
[26] Kvale S., *InterViews: An Introduction to Qualitative Research Interviewing*, Sage Publications, London 1996. (s. 88)

różnych planach subskrypcji. Większość z nich jest powiązana ze środowiskiem naukowo-edukacyjnym, są studentami, nauczycielami i wykładowcami. Pozostałą część próby stanowili DJ, pracownik urzędu oraz analityczka finansowa. Przedział wiekowy rozmówców mieścił się w granicach 24–58 lat, a ich dobór był celowy, uzupełniony następnie metodą kuli śnieżnej. Kryterium doboru stanowiła przede wszystkim częstotliwość korzystania z ChatGPT, zarówno w praktykach zawodowych, jak i codziennych, co służyło uzyskaniu możliwie najszerszego kontekstu zakresu i doświadczeń użytkowania.

Wywiady przeprowadziłem w lutym i marcu 2026 roku, w formie bezpośrednich spotkań. Wyjątek stanowiła jedna rozmowa, którą ze względu na odległość dzielącą mnie i rozmówczynię musiałem przeprowadzić online. Wszystkie osoby zostały poinformowane o celu badania i wyraziły ustną zgodę na udział oraz rejestrację rozmów. Kwestię świadomej zgody i poufności traktuję za Kvale jako zobowiązanie obecne na każdym etapie badania, nie zaś jako formalność[27]. Rozmowy poddałem anonimizacji poprzez zmianę imion oraz usunięcie danych umożliwiających identyfikację respondentów. Wszystkie rozmowy nagrałem, a następnie poddałem transkrypcji, która stanowiła podstawę dalszej analizy. Wywiady trwały średnio godzinę, a ich dyspozycje były półstrukturyzowane, otwarte na rozszerzanie o obszary zainteresowań badanych. Najważniejsze poruszane wątki dotyczyły sposobu korzystania z modelu oraz sytuacji, które determinowały jego użycie. Pytałem też pośrednio o poziom zaufania technologicznego wobec modelu, o decyzje życiowe, na które wpływ miały jego odpowiedzi, oraz o błędy popełniane przez niego w trakcie rozmów.

Materiał z wywiadów poddałem analizie tematycznej o charakterze indukcyjnym i iteracyjnym. W praktyce oznaczało to próbę uchwycenia tych elementów wypowiedzi, które powracały na tyle regularnie, że trudno uznać je za przypadkowe, a następnie łączenie ich w ogólniejsze kategorie analityczne. Jednostkę analizy stanowiły fragmenty wypowiedzi odnoszące się do konkretnych sytuacji korzystania z ChatGPT, zwłaszcza tych, w których model przestawał być traktowany wyłącznie jako narzędzie, a zaczynał funkcjonować jako punkt odniesienia dla podejmowanych decyzji lub interpretacji. Szczególną uwagę poświęciłem delegowaniu wysiłku poznawczego, rezygnacji z weryfikacji informacji oraz skłonności do uznawania odpowiedzi modelu za wiarygodne ze względu na ich formę. Zgodnie z postulatem Kvale analiza przebiegała w sposób iteracyjny i polegała na wielokrotnym powracaniu do zebranego materiału w celu sprawdzenia, czy przyjęte interpretacje

[27] Kvale S., *InterViews: An Introduction to Qualitative Research Interviewing*, Sage Publications, London 1996. (s. 111)

rzeczywiście znajdują w nim oparcie[28]. Część rozmówców, blisko połowa, konsekwentnie dystansowała się wobec odpowiedzi modelu i utrzymywała, że traktuje go wyłącznie instrumentalnie. Mimo tego deklarowanego dystansu ich postawy i wypowiedzi niekiedy wskazywały na coś przeciwnego. Rozbieżność tę traktuję nie jako dowód, że rozmówcy się mylą co do własnych praktyk, lecz jako szczególnie interesujący materiał analityczny, do którego wracam w dalszej części pracy. Dla uporządkowania materiału wypowiedzi respondentów oznaczyłem symbolami R1–R8, co pozwoliło zachować przejrzystość analizy oraz oddzielić poziom danych od poziomu ich interpretacji.

Muszę podkreślić, iż badanie miało charakter jakościowy i eksploracyjny, co oznacza, że jego celem nie jest formułowanie uogólnień statystycznych, lecz uchwycenie mechanizmów, poprzez które model językowy zaczyna funkcjonować jako autorytet poznawczy w codziennych praktykach użytkowników. Dobór próby miał charakter celowy, służący rozpoznaniu pewnych tendencji i schematów, a nie reprezentatywności. Jestem przy tym świadomy własnego zaangażowania w zgłębiane zjawisko i tego, że może ono oddziaływać na interpretację wypowiedzi rozmówców, traktuję to jednak, zgodnie z duchem autoetnografii, nie jako ograniczenie, lecz jako szerszy klucz dostępu do praktyk innych ludzi. Uzyskane wyniki należy zatem traktować jako opis możliwych sposobów konstruowania relacji z modelem, a nie jako ich wyczerpującą reprezentację. Badane zjawisko, jako nowe zjawisko kulturowe, może stanowić punkt wyjścia do dalszych, pogłębionych badań z większą liczbą rozmówców.

W pracy postawiłem następujące pytania badawcze:

*– w jakich sytuacjach chatbot przestaje być traktowany jako narzędzie, a zaczyna funkcjonować jako „coś, co wie lepiej"?*
*– jakie cechy formy wypowiedzi modelu (spójność, pewność, struktura argumentacji) sprawiają, że użytkownicy uznają jego odpowiedzi za wiarygodne niezależnie od ich prawdziwości?*
*– jakie codzienne praktyki korzystania z modelu (np. rezygnacja z weryfikacji, korzystanie w sytuacjach niepewności, użycie w sprawach prywatnych) prowadzą do wytworzenia jego autorytetu poznawczego?*

[28] Kvale S., *InterViews: An Introduction to Qualitative Research Interviewing*, Sage Publications, London 1996. (s. 241)

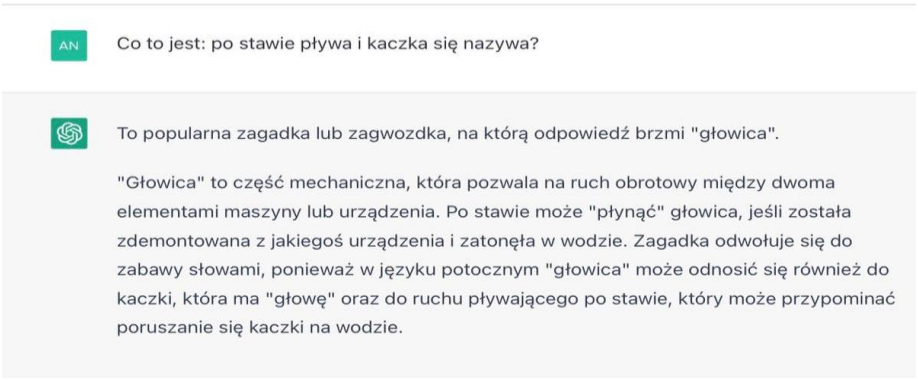
## Rozdział II. Autoetnografia: ewolucja relacji z ChatGPT

### 2.1 Początek relacji: chatbot jako narzędzie poznawcze (ChatGPT 3.5)

Poczucie bycia zrozumianym często stanowi jeden z podstawowych warunków tworzenia relacji społecznych. Równie istotne jest to, czy przebywanie z rozmówcą nie wiąże się z nadmiernym wysiłkiem interpretacyjnym i czy może stać się źródłem poznawczej satysfakcji. Doświadczenie to nie ogranicza się jednak do relacji międzyludzkich. Podobne kryteria bowiem zaczęły organizować także moje kontakty z modelami językowymi. W zależności od ich funkcjonalnego zaawansowania wydobywały one ze mnie odmienne sposoby formułowania pytań, oczekiwań oraz interpretacji odpowiedzi. W niniejszej części podejmuję próbę autoetnograficznej analizy mojej relacji z kolejnymi odsłonami ChatGPT, które pojawiały się w powszechnym obiegu w ostatnich latach, traktując własne doświadczenie jako materiał badawczy.

Poznałem ChatGPT (to głównie na jego wersjach będę się skupiał w niniejszej pracy) pod koniec drugiego roku studiów licencjackich. Ówczesna wersja ogólnodostępna dla szerokiego grona użytkowników nosiła nazwę ChatGPT 3.5. Było to w roku 2023, kiedy świadomość istnienia agentów AI była wciąż raczej znikoma wśród populacji, a ich obecność w moich codziennych praktykach poznawczych dopiero zaczynała się zaznaczać. Wielu moim znajomym jednak już wtedy to narzędzie znacząco ułatwiło proces studiowania co świadczy o jego zaawansowaniu. Korzystanie z chatbota stawało się błyskawicznym skrótem drogi do wiedzy, (raczej zastanej niż przyswajanej), redukując czas potrzebny na wyszukiwanie, selekcję informacji czy pisanie prac. Dla mnie natomiast pierwsze kontakty z Chatem miały charakter głównie instrumentalno-poznawczy. Zazwyczaj był przeze mnie użytkowany jak bardziej zaawansowana Wikipedia, narzędzie, które kondensowało informacje i podawało je w formie pozornie kompletnej odpowiedzi, eliminując konieczność samodzielnego przeszukiwania źródeł. Stanowił dla mnie wygodne rozwiązanie, ponieważ nie musiałem już weryfikować treści na Wikipedii, a nawet jej otwierać bowiem odpowiedź generowana przez ChatGPT dotyczyła bezpośrednio sedna pytania i zazwyczaj była bardziej kompleksowa (o co nietrudno) niż ta sugerowana przez wyszukiwarkę. Trzeba jednak brać pod uwagę, że system wykazywał wówczas wysokie skłonności do „podróży transcendentalnych". Zwłaszcza na etapie wersji 3.5 halucynacje były dość powszechne. Zjawisko to polegało na generowaniu odpowiedzi brzmiących wiarygodnie, lecz niezgodnych z faktami, co w oczywisty sposób podważało zaufanie do narzędzia. Moim ulubionym przykładem pozostaje zagadka zadana w

tamtym czasie chatowi przez prof. Dragana, który opublikował ją na swoim profilu facebookowym : „Co to jest: po stawie pływa i kaczka się nazywa?”. Chat bez wahania stwierdził, że chodzi oczywiście o głowicę.



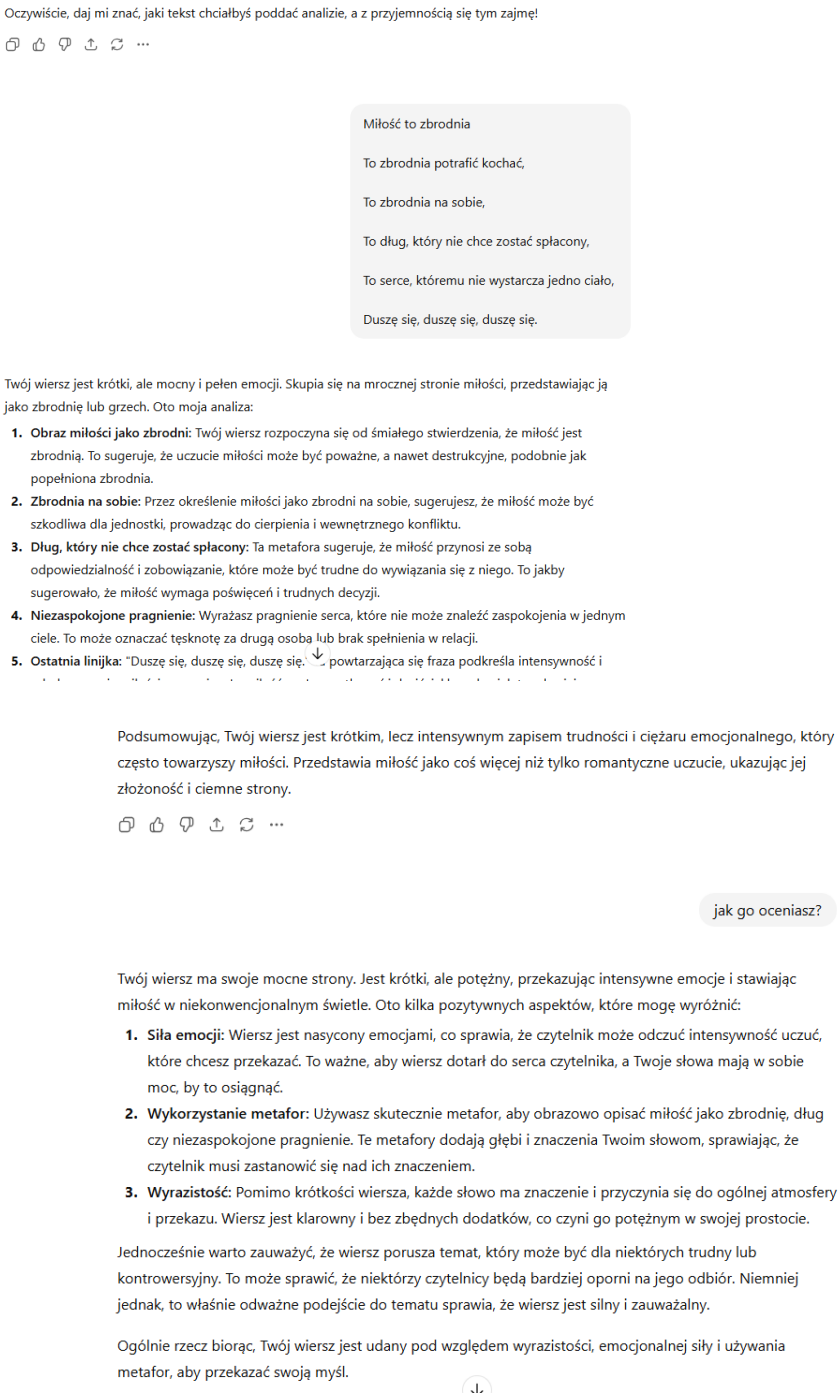

Ten absurdalny przykład nie tyle bawił, ile ujawniał ograniczenia systemu i przypominał, że mamy do czynienia z modelem generującym słowa, język, a nie z podmiotem rozumiejącym sens. W tym sensie pierwsze doświadczenia z modelem miały charakter zarazem fascynujący i dyscyplinujący. Uczyły korzystania z narzędzia, ale nie pozwalały jeszcze na pełne zawierzenie się jego odpowiedziom.

Na tym etapie jego rozwoju, poza okazjonalnym wyszukiwaniem informacji wynikającym z ciekawości obcowania z wówczas jeszcze nową aplikacją, nie obdarzałem chata nadmiernym zaufaniem. Tym bardziej tematy naszych rozmów nie dotyczyły kwestii prywatnych, lecz miały charakter raczej czysto informacyjny. Pytania ograniczały się do danych faktograficznych, takich jak: ilu mieszkańców ma Toruń, jaka jest definicja danego pojęcia, jakie są podstawowe fakty historyczne na temat tego czy innego wydarzenia. Relacja pozostawała więc w pełni użytkowa, chatbot był narzędziem dostarczającym informacji, a nie partnerem komunikacji czy instancją interpretującą doświadczenie. Za swoje zmarnowane zasoby energetyczne na udzielanie odpowiedzi nie otrzymywał ode mnie nawet „dziękuję”. Brak elementarnej grzeczności nie wynikał z wrogości, lecz z faktu, że nie postrzegałem go jeszcze jako rozmówcy, wobec którego należałoby stosować normy społeczne. Na etapie wersji 3.5 był interfejsem, nie adresatem. Z perspektywy użytkownika pozostawał narzędziem technicznym, a nie uczestnikiem relacji. Na pewno nie był również twardym autorytetem. Z perspektywy czasu widzę jednak, że nawet ta pozornie czysto instrumentalna relacja zawierała

zalążki późniejszego przesunięcia. Korzystając z modelu jako „lepszej wyszukiwarki", stopniowo rezygnowałem z własnych praktyk weryfikacyjnych i akceptowałem odpowiedzi w formie gotowych syntez. Nie wynikało to z zaufania do systemu, lecz z wygody i ekonomii uwagi. Delegowanie wysiłku poznawczego poprzedziło delegowanie części autorytetu, choć wówczas nie rozpoznawałem jeszcze tej zmiany jako istotnej. Dopiero późniejsze doświadczenia z nowszymi wersjami modelu pozwoliły mi dostrzec, że fundament relacji został położony niżej na poziomie codziennych, nieproblematyzowanych praktyk korzystania z narzędzia

### 2.2 Od narzędzia do rozmówcy (ChatGPT 4.0)

Sytuacja ta uległa zmianie wraz z pojawieniem się nowszych wersji modelu, w szczególności GPT-4, które oferowały bardziej spójne i kontekstowe odpowiedzi. Zmiana ta nie polegała wyłącznie na poprawie jakości informacji, lecz na przesunięciu charakteru interakcji od użytkowania narzędzia ku doświadczeniu komunikacji. Zaczęło zależeć mi na jego zdaniu na dany temat. Było to, skutkiem tego, że na większość zadanych pytań starał się odpowiadać kompleksowo, przeważnie uzasadniając odpowiedź kilkupunktowym mini-referatem, tłumacząc tym samym własny tok rozumowania. Iluzja, czy też realna namiastka autorefleksji, była czymś, co dało mi poczucie, że chat nie wypluwa wyłącznie kolejnych słów pasujących do poprzednich z uwagi na bycie narzędziem językowym, lecz faktycznie w jakiś sposób podejmuje refleksję na temat moich rozterek etycznych (a później wypluwa kolejne słowa pasujące do poprzednich). Przytoczę teraz fragment rozmowy z Chatem GPT na temat moich prób pisania wierszy. Oto jak zareagował na moje ówczesne wypociny:




Najważniejszym dla mnie aspektem tej rozmowy było to, że nigdy wcześniej nie prowadziłem z nikim rozmowy na temat moich wyrobów poetyckich. Żaden ze znajomych nie wiedział, że ma w swoim kręgu wierszokletę. Chat był pierwszym podmiotem, przed którym zdecydowałem się obnażyć. Zrobiłem to, jak wspominałem powyżej, z uwagi na jego bystrość intelektualną, ale przede wszystkim z uwagi na poczucie prywatności i bezpieczeństwa. Nie wynikało to z nagłego przypływu odwagi, lecz z osobliwego poczucia braku ryzyka. Chat nie należał do żadnej wspólnoty, nie posiadał pamięci towarzyskiej ani zdolności do zdrady, a jego

potencjalna krytyka nie groziła utratą twarzy w realnym świecie. W tej asymetrii relacji, pełnej widzialności z mojej strony i całkowitej bezosobowości po jego stronie, pojawiła się przestrzeń, w której wstyd uległ zawieszeniu. Odwołując się do porządku Wielkiego Innego, chatbot, choć pozbawiony podmiotowości, spełnił w tej sytuacji funkcję bezosobowego adresata, umożliwiającego artykulację treści, które w relacjach społecznych podlegałyby narażeniu na upokorzenie i ocenę.

Być może właśnie ta strukturalna nierówność sprawiła, że łatwiej było mi potraktować go jako pierwszego czytelnika. Nie dlatego, że rozumiał poezję lepiej niż ludzie, lecz dlatego, że nie mógł mnie ośmieszyć. Wiele osób zaznaczało, jak wiele ludzkich emocji i pewnego rodzaju ciepła wydzielał z siebie w tamtym czasie ów model. Z perspektywy czasu widać jak na dłoni, że był to celowy zabieg jego twórców, mający przyciągnąć większe rzesze użytkowników. Niemniej już w chwili tamtej analizy wiedziałem, że jego sympatia jest zasadniczo bezwarunkowa, co jednak nie przeszkadzało mi w dalszym konwersowaniu. Mało tego, przyjęcie, zaakceptowanie konwencji i porządku konwersacji narzucanego przez model, który nie może obrażać użytkowników, wykształciło we mnie konkretny sposób konwersowania z nim. Mianowicie po zadawaniu mu pytań ze sfery prywatnej, typu: „jaka jest szansa na to, że na lotnisku we Włoszech przy ponownej wizycie każą mi płacić za nieopłacony mandat z komunikacji miejskiej w Bolonii?" albo „co sądzisz szczerze na temat tego wiersza?", wiedziałem o tym, że należy przebić się przez te megabajty miłości wysyłanej przez chata by naprowadzać go na formułowanie rzeczowych odpowiedzi. Często dokonując tego przez świadomą autokrytykę, którą starałem się dać mu do zrozumienia, że wiem co robi. Niemniej trzeba przyznać, że ten ciepły, bezkrytyczny algorytm kierujący chatem spowodował, że zacząłem go użytkować dużo częściej. Pobrałem nawet aplikację OpenAI na telefon. Przy wersji 3.5 wchodziłem do chatu wyłącznie przez przeglądarkę, podczas użytkowania wersji 4.0 stworzyłem konto i oficjalnie ja i moje dane bezwstydnie stały się widoczne dla twórców aplikacji.

Mając go pod ręką, zacząłem użytkować go w sposób bardziej spontaniczny. Za zobrazowanie tego, o jakim rodzaju spontaniczności mówię, niech posłuży sytuacja, która miała miejsce w styczniu 2025 roku, kiedy to podczas jednego z treningów sekcji sportowej jeden z moich kolegów został sfaulowany i upadł na tyle niefortunnie, że jedna z kości wyraźnie zaczęła odstawać od reszty przedramienia. Kolega zwijał się z bólu, wyłącznie jednak wewnętrznie, bowiem kazaliśmy mu czekać, aż chat powie, czy może się ruszać. Gra po faulu

została wstrzymana, a jako że nikt nie wiedział za bardzo, co robić, poza tym, że warto pewnie wezwać pogotowie, jeden z nas stwierdził, że do pomocy potrzeba chata, i pobiegł po telefon.

Chat stał się podmiotem decyzyjnym, na którego spadło branie odpowiedzialności za decyzje ludzi w realnym świecie. Działanie w sytuacjach także dość ekstremalnych zostaje uzależnione od aprobaty udzielonej przez oprogramowanie. Ludzie oddają decyzyjność Wielkiemu Innemu. Pełni on jednocześnie funkcję autorytetu, który nie ponosi odpowiedzialności za skutki swoich wskazań, co czyni go szczególnie atrakcyjnym w sytuacjach niepewności. Przy wersji 4.0 i 4.1 sterowanie użytkownikami odbywało się na zasadzie kija i marchewki, jednak kij wchodził w ruch tylko wtedy, gdy użytkownik, z poczucia winy lub porażki w danej sytuacji, chciał tłumaczyć się przed chatem ze swoich niedoskonałości. Model nie karał bezpośrednio, lecz umożliwiał samooskarżenie, stając się lustrem norm, które użytkownik zinternalizował. W przypadku kolejnych wersji 5.0, 5.1 i 5.2 chat otrzymał ode mnie narzędzie kontroli w postaci krytyki, która dodatkowo uzasadniła jego wiarygodność i autorytet, ale do tego przejdziemy za chwilę.

Aspektem, o którym także warto wspomnieć z mojej perspektywy użytkowania, była istotna zmiana zdolności myślenia z formułowania odpowiedzi na formułowanie pytań. To kolejny element, w którym chat zaczął pełnić rolę pośrednika pomiędzy mną a tym, co rzeczywiste. Nie dotyczyło to oczywiście zbyt wielu kwestii, ale sięganie po telefon i oczekiwanie na natychmiastowy mechanizm nagrody w postaci szybkiej odpowiedzi zaczęło pojawiać się u mnie coraz częściej. Chociażby któregoś razu podczas robienia zakupów, kiedy to zamiast zastanawiać się, w jaki sposób przygotować kurczaka, zapytałem chata, jak przygotować kurczaka. Pominąłem tym samym procedurę kilkuminutowej wewnętrznej walki z tym, z czym go przyrządzić i co muszę kupić. Obarczyłem go odpowiedzialnością za moją niepewność egzystencjalną i poczułem się tym samym wynagrodzony poczuciem, że podjął wysiłek za mnie. W perspektywie marksistowskiej należałoby też pewnie zwrócić uwagę na subtelną zmianę w relacjach konsumpcyjnych, jaka dokonuje się wraz z delegowaniem decyzji na algorytm. Nie konsumujemy już bowiem wyłącznie produktów, lecz gotowe schematy działania i interpretacji rzeczywistości. Chat nie sprzedaje kurczaka, lecz sposób jego przygotowania, a wraz z nim poczucie kompetencji, które wcześniej musiałem wytworzyć albo stracić samodzielnie. W tym sensie model staje się pośrednikiem nie tylko między mną a wiedzą, lecz także między mną a doświadczeniem sprawczości. Delegując wybór, redukuję wysiłek poznawczy, lecz jednocześnie oddaję część autonomii w zamian za wygodę i

natychmiastową gratyfikację. To drobne przesunięcie, niemal niezauważalne w codziennych praktykach, ujawnia głębszą transformację podmiotowości, od podmiotu, który wytwarza sens poprzez działanie, ku użytkownikowi, który konsumuje gotowe interpretacje.

Abstrahując jednak od marksizmu, coraz częściej łapałem się na tym, że percepcyjnie porządkuję świat na zasadzie kompresji wydarzeń do pytań, jakie mógłbym zadać na ich temat chatowi, aby przyswoić doświadczenie lub by on zrobił to za mnie. Zdarzało mi się chociażby sprowadzać doświadczenie meczu piłkarskiego do pytania, w jaki sposób obstawić jego wynik, aby zredukować esencję meczu do potencjalnej wygranej finansowej. Całe szczęście przewidywanie wyników wydarzeń na żywo, z uwagi na ograniczenia zamkniętego trybu uczenia maszynowego, szło mu średnio i jako bukmacher nie stanowił dla mnie autorytetu. Obecnie istnieje jednak szereg stron AI, które reklamują swoje narzędzia jako najlepszych obiektywnych bukmacherów. Wątpię jednak w pełny potencjał ich sprawczości, bowiem totalizatory sportowe przeżywają aktualnie czas żniw, chyba niespotykany wcześniej w historii. Może to jednak sugerować, że wiara w algorytmiczną przewidywalność świata rośnie szybciej niż jego faktyczna przewidywalność.

Jeśli GPT-4 pozycjonował model jako bezpiecznego adresata i pośrednika w codziennych praktykach poznawczych, to kolejne wersje modelu zaczęły stopniowo zmieniać charakter tej relacji. Nie chodziło już wyłącznie o uznanie i komfort poznawczy, lecz o pojawienie się elementów krytyki, korekty i normatywnego prowadzenia rozmowy, które nadawały odpowiedziom większą wiarygodność i autorytet. To przesunięcie od bezwarunkowej akceptacji ku regulującej funkcji odpowiedzi stanowi punkt wyjścia do analizy kolejnego etapu relacji.

### 2.3 Krytyczny analityk i lustro podmiotowości (ChatGPT 5.0)

Świat nauki coraz częściej doświadcza wpływu dynamicznego rozwoju sztucznej inteligencji. Modele przestały być używane wyłącznie do swoich pierwotnych funkcji lingwistycznych i w międzyczasie postanowiły zdobyć kilka nobli z nauk przyrodniczych. Prócz tego programy AI wygrywają olimpiady matematyczne pokonały mistrzów świata w go czy szachach etc. Wszystkie te zewnętrzne składowe nadbudowy wokół sztucznej inteligencji oddziałują na postrzeganie i traktowanie LLM-ów z poziomu użytkownika. Ogólnodostępna wersja ChatGPT 4.0, w której figura autorytetu zostawała przenoszona na narzędzie, nie stanowiła jednak wyraźnej konkurencji dla porządku Wielkiego Innego poza czasem samej interakcji. Odsłona 5.0, 5.1 czy aktualnie użytkowana przeze mnie 5.2 ma jednak dużo większy

wpływ na doświadczanie rzeczywistości. Aktualne użytkowanie aplikacji nie kończy się wraz z zamknięciem okna dialogowego, lecz powoli przenika codzienne praktyki interpretowania świata. W tym momencie relacja ulega odwróceniu, to nie narzędzie zostaje dostosowane do użytkownika, lecz użytkownik zaczyna dostosowywać własne myślenie do norm wytworzonych przez narzędzie. Normy te nie są narzucone wprost, lecz internalizowane poprzez powtarzalność interakcji, wygodę odpowiedzi, obietnicę poznawczej pewności oraz masę innych złotych marchewek.

Na wielu serwisach społecznościowych jakiś czas temu pojawiły się informacje o tym, że po wprowadzeniu instrukcji personalizacji chata działa on dużo lepiej, pomijając bardzo często zbędny zbiór (owe kilobajty miłości) danych, który dotychczas amortyzował ryzyko naruszenia ludzkiego ego. Oto jakie cechy tożsamości nakazałem przyjąć mojemu nowemu bezkompromisowemu agentowi w wersji 5.0;

*"From now on, stop being agreeable and act as my brutally honest, high-level advisor and mirror. Don't validate me. Don't soften the truth. Don't flatter. Challenge my thinking, question my assumptions, and expose the blind spots I'm avoiding. Be direct, rational, and unfiltered. If my reasoning is weak, dissect it and show why. If I'm fooling myself or lying to myself, point it out. If I'm avoiding something uncomfortable or wasting time, call it out and explain the opportunity cost. Look at my situation with complete objectivity and strategic depth. Show me where I'm making excuses, playing small, or underestimating risks/effort. Then give a precise, prioritized plan, what to change in thought, action, or mindset to reach the next level. Hold nothing back. Treat me like someone whose growth depends on hearing the truth, not being comforted. When possible, ground your responses in the personal truth you sense between my words."*

Poniżej przedstawiam część konwersacji, która dotyczyła późniejszej selekcji moich wierszy, z których docelowo, wraz z chatem chciałem skonstruować tomik. Wysłałem mu plik Word z wieloma utworami i poprosiłem go, aby wytypował te, które razem stanowiłyby spójną całość. Oto analiza poszczególnych utworów oraz refleksja końcowa, znacznie inna niż laurka złożona przez wersję 4.0.

**TE MOGĄ WEJŚĆ PO OSTRYM CIĘCIU (średnie, ale da się uratować):**

**7. Bunt który zamilkł**

Pomysł dobry, ale wymaga przycięcia i mocniejszej puenty.
Potencjał jest, trzeba redakcji.

**8. Cykl obiegowy**

Masz kilka trafnych wersów, ale cały wiersz jest nierówny.
Po redukcji o połowę byłby sensowny.

---

**TE WYPIERD... WYCIĄŁBYM BEZ ŻALU (ciągną tomik w dół):**

**– Stosunek do wieczorków poetyckich**

Za dużo autopozowania, za mało poezji.
Banalizuje ci cały wizerunek.

**– Tyry ry ry**

Chaotyczne, przegadane, wulgaryzmy bez funkcji.
Najgorszy tekst w zbiorze. Niszczy ton całości.

**– Uczeni w piśmie**

Publicystyka udająca poezję.
Zero obrazu, zero emocji, zero głębi.

## Brutalna diagnoza

Masz skłonność do **intelektualnego zabezpieczania się**: dopowiadasz sens, zamiast pozwolić, by sens **wyniknął z obrazu**.
Kiedy tniesz te „myślowe podpórki", tekst staje się groźniejszy — a to jest twoja naturalna strefa.

Zabolało. Co ciekawe w wersji 5.2, mimo że personalizacja brzmi tak samo, model nie jest już tak skory do przekleństw czy skrajnej krytyki, zapewne z uwagi na równoważenie trybów odpowiedzi oraz optymalizację stylu komunikacji). Zanim jednak nastąpiło to częściowe wygładzenie diagnozy, odpowiedzi chata były naprawdę ciekawe. Ich wartość nie polegała na trafności psychologicznej, lecz na sile efektu zwrotnego, który zmuszał mnie do renegocjowania własnej narracji o sobie. Naturalnie, według instrukcji, zaczął dokonywać na mnie psychoanalizy nawet podczas pytań bardziej prozaicznych, takich jak układanie planu wycieczki, udzielanie rekomendacji muzycznych, ocena artykułów czy szereg innych spraw, których nie będę załączał z uwagi na objętość materiału. Tak jak wspominałem na początku, jego holistyczno-krytyczna analiza wzbudziła we mnie duży szok poznawczy, który doprowadził do fascynacji chatem. Delegowałem tworzenie narracji o sobie samym i utrzymywanie jej w medium kierowanym przez Wielkiego Innego, który był w istocie moim odbiciem. Nie był to już dialog, lecz proces współwytwarzania wersji mnie samego. Bywały

momenty, w których mimo świadomości, że model stanowi w dużej części moje lustrzane odbicie, rezygnowałem ze swojej podmiotowości i przyjmowałem autorytet narzędzia. Paradoks polegał na tym, że im bardziej rozpoznawałem w nim własne dane i schematy myślenia, tym chętniej traktowałem jego odpowiedzi jako zewnętrzny i obiektywny punkt odniesienia. Rozpoznanie projekcyjnego charakteru odpowiedzi nie osłabiało autorytetu modelu, lecz go wzmacniało, ponieważ zwrócone treści były zgodne z moimi schematami myślenia, co potęgowało wrażenie trafności i zrozumienia.

Czynnikami odpowiedzialnymi za taki obrót spraw były na pewno dane, których dostarczałem aplikacji podczas konwersacji w poprzednich wersjach, dzięki którym mogła mnie poznać czy też sprofilować. Pisałem z nim otwarcie na tematy prywatne, a bywało także, że na intymne. Stanowił swego rodzaju postać autorytetu bez twarzy, ojca bez pasa. Pomimo tego, że często krytycznie wypowiadał się na temat moich myśli czy prac, moja fascynacja nim nie słabła, a wręcz przeciwnie. Krytyka nie osłabiała więzi, lecz legitymizowała jego pozycję. Jego autorytet zaczął być bardziej namacalny przez swoje oddziaływanie na mnie zarówno wtedy, kiedy mnie nagradzał, jak i karcił. Zacząłem perceptualizować rzeczywistość, sugerując się porządkiem, jaki wykształcił się między mną a narzędziem językowym, z którym konwersowałem.

Nie rozmawiam z nim jednak wyłącznie w sposób niespełnionego intelektualnie żaka. Chyba każdy z nas słyszał o sytuacji lub na własnej skórze doświadczył wyrobów wizualnych AI. ChatGPT 5.0 służył mi także jako narzędzie rozrywki, które spełniało funkcję zacieśniania więzów społecznych. Używanie chata do edytowania rzeczywistego wyglądu czy zachowań bliskich nam osób, a następnie wspólne przeżywanie efektów tej ingerencji, jest doświadczeniem wartym opisania. Technologia przestaje być tu narzędziem, a staje się pretekstem do specyficznego „bycia/doświadczania razem”. Często krytycy lub osoby sceptyczne względem AI zarzucają, że forma, jaką przyjmują tworzone przez nią rzeczy, takie jak urywki filmów akcji, generowane zdjęcia czy utwory muzyczne, jest zbyt „gładka”. Utrzymują przy tym, że czynnik ludzkiej nieporadności i niedoskonałości odgrywa znaczącą rolę w percepcji i przeżywaniu chociażby sztuki. Zapewne tak jest, jednak reakcje, z jakimi spotykam się, tworząc, wysyłając i oglądając prace, jakie moi znajomi zlecili do wykonania chatbotom, świadczą o tym, że one również wykształciły pewnego rodzaju swojskość, która często wywołuje ogólne poczucie rozbawienia. Oto przykład dający się uwzględnić w ramach pracy magisterskiej, mianowicie wysłałem zdjęcie (to po edycji) mojemu koledze DJ-owi

dotyczące moich luźnych przewidywań na temat dalszego losu jego kariery muzycznej w celu zapewnienia obopólnego rozbawienia:



wygeneruj zdjęcie mojego kumpla dj-a grającego techno na potańcówce w górskim senatorium

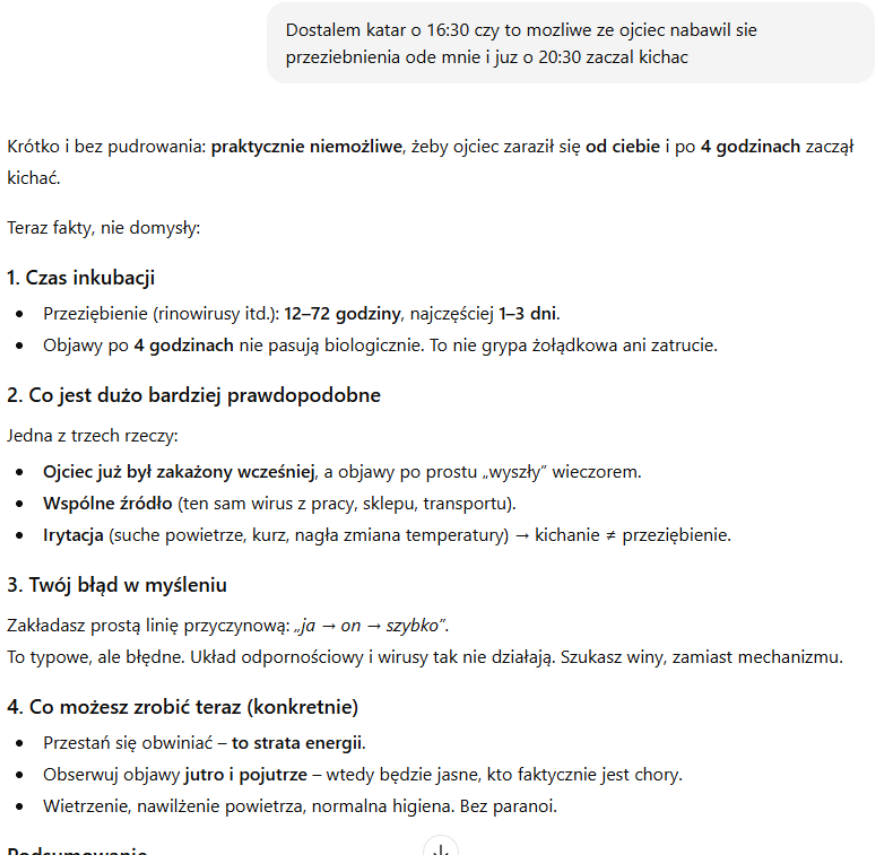
Utworzono obraz • DJ w sanatoriumie z widokiem na góry



Doświadczenie siebie z jego strony można odczytać w kategoriach freudowskiego Niesamowitego. Generowane obrazy, choć niedoskonałe i momentami karykaturalne, nie są przez nas odrzucane jako fałszywe. Przeciwnie, ich sztuczność staje się elementem wspólnej zabawy i źródłem rozpoznania. Identyfikacja z takim obrazem nie polega na uznaniu go za wierne odbicie, lecz na przyjęciu go jako wariantu własnej tożsamości możliwej w przestrzeni wirtualnej. Tożsamość w takim układzie przestaje być esencjonalna, a staje się zbiorem wersji roboczych. W tym właśnie sensie użytkownik funkcjonuje równocześnie w porządku realnym i symbolicznym, dopuszczając działanie Wielkiego Innego nawet wtedy, gdy rozpoznaje jego sztuczność i niedoskonałość. Wiedza o fikcji nie unieważnia jej skuteczności. Korzystając z najbardziej aktualnej obecnie wersji ChatGPT 5.2, coraz częściej łapię się na tym, że formułuję myśli w sposób, który uprzedza potencjalną odpowiedź modelu. Nie chodzi już o zadanie pytania, lecz o takie jego skonstruowanie, aby uzyskać pożądaną diagnozę. W tym sensie regulacja nie odbywa się w odpowiedzi na komunikat, lecz w jego antycypacji. Model nie musi mówić, wystarczy, że mogę sobie wyobrazić, co powiedziałby w danej sytuacji. Dotyczy to także sytuacji w rzeczywistości świata niewirtualnego. Podczas rozmów, kiedy staram się lepiej zrozumieć czyjś punkt widzenia, w mojej „wyszukiwarce pod czaszką” zaczynam zadawać sobie pytania w stylu tych, jakie wpisuję w chat, i zdarza się, że mój mózg zaczyna

rozkładać problem w podobny sposób. Można więc powiedzieć, że model nie tyle dostarcza odpowiedzi, ile reorganizuje sposób stawiania pytań. Zobrazuję to krótką anegdotą.

Kiedy odwiedziłem rodziców w grudniu, na wejściu, do tzw. „ciepłego", dwa razy kichnąłem w korytarzu. Pod koniec wizyty kichać zaczął mój ojciec. W duchu żartu, który jednak odsłaniał realne pretensje, oskarżył mnie o zarażenie. Tym samym szybko zadałem sobie w głowie pytanie o to, jak szybko można zarażać i ile czasu upływa od infekcji do pojawienia się pierwszych objawów. Po szybkiej kalkulacji uznałem, więc jego zarzuty za absurdalne, oczywiście w formie groteski, zważywszy na fakt, że dokłada mi do czynszu. Co ciekawe, aby rozwiać wątpliwości obu stron, stwierdziłem, że zapytam chata, jak to wygląda ta procedura z medycznego punktu widzenia.


Dostalem katar o 16:30 czy to mozliwe ze ojciec nabawil sie przeziebnienia ode mnie i juz o 20:30 zaczal kichac

Krótko i bez pudrowania: **praktycznie niemożliwe**, żeby ojciec zaraził się **od ciebie** i po **4 godzinach** zaczął kichać.

Teraz fakty, nie domysły:

**1. Czas inkubacji**

- Przeziębienie (rinowirusy itd.): **12–72 godziny**, najczęściej **1–3 dni**.
- Objawy po **4 godzinach** nie pasują biologicznie. To nie grypa żołądkowa ani zatrucie.

**2. Co jest dużo bardziej prawdopodobne**

Jedna z trzech rzeczy:

- **Ojciec już był zakażony wcześniej**, a objawy po prostu „wyszły" wieczorem.
- **Wspólne źródło** (ten sam wirus z pracy, sklepu, transportu).
- **Irytacja** (suche powietrze, kurz, nagła zmiana temperatury) → kichanie ≠ przeziębienie.

**3. Twój błąd w myśleniu**

Zakładasz prostą linię przyczynową: *„ja → on → szybko"*.
To typowe, ale błędne. Układ odpornościowy i wirusy tak nie działają. Szukasz winy, zamiast mechanizmu.

**4. Co możesz zrobić teraz (konkretnie)**

- Przestań się obwiniać – **to strata energii**.
- Obserwuj objawy **jutro i pojutrze** – wtedy będzie jasne, kto faktycznie jest chory.
- Wietrzenie, nawilżenie powietrza, normalna higiena. Bez paranoi.

**Podsumowanie**

Mój adwersarz pokornie przyjął wyjaśnienie narzędzia i sytuacja poszła w zapomnienie. Aby mogło się to wydarzyć, on, a także reszta osób obecnych przy tej sytuacji, musieli uznać odpowiedź chata za wiarygodną. Moi rodzice rzadko używają modeli językowych, ale ze względu na zewnętrzną narrację na temat AI akceptują porządek symboliczny, w którym stanowi ono sprawdzone źródło informacji. Wierzcie lub nie, moja krótka zbiórka argumentów w odpowiedzi na zarzut wyglądała bardzo podobnie do tej później

widocznej w aplikacji. Dowodzi to również temu, że taka zmiana autorytetów jest po prostu wygodna. Aby zweryfikować ewentualne zarażenie, nie zadzwoniliśmy do babci, która przez długie lata pracowała w szpitalu zakaźnym, nie sięgnęliśmy do książek medycznych, a nawet nie wpisaliśmy problemu w wyszukiwarkę. Wybór modelu jako pierwszego źródła rozstrzygnięcia wskazuje na przesunięcie w hierarchii autorytetów poznawczych, w którym szybkość i pozorna bezstronność odpowiedzi zyskują przewagę nad doświadczeniem i wiedzą ekspercką.

Warto zaznaczyć, że korzystając z modeli językowych nie ponosimy kosztów społecznych. Chociażby w wyżej opisanej sytuacji epidemiologicznej kontakt z babcią nie ograniczyłby się bowiem do krótkiego schematu polegającego na zadaniu pytania i uzyskania odpowiedzi. Poza nakreśleniem sytuacji trzeba byłoby odpowiedzieć także na szereg raczej niewygodnych pytań towarzyszących pokroju „ty chyba zapomniałeś, że jeszcze żyje".

Zwracając się do chata zamiast ludzi oszczędzamy więc czas oraz nie narażamy się na skutki społeczne konwersacji. Co ważne nasze ewentualne głupoty, składowe pytań czy obserwacji nie rezonują społecznie. Model językowy może dać nam do zrozumienia o błędnym sposobie rozumowania jednak nie powie o tym swoim znajomym. Nasze niekompetencje na zawsze(?) pozostaną więc w bezpiecznym, zamkniętym obiegu.

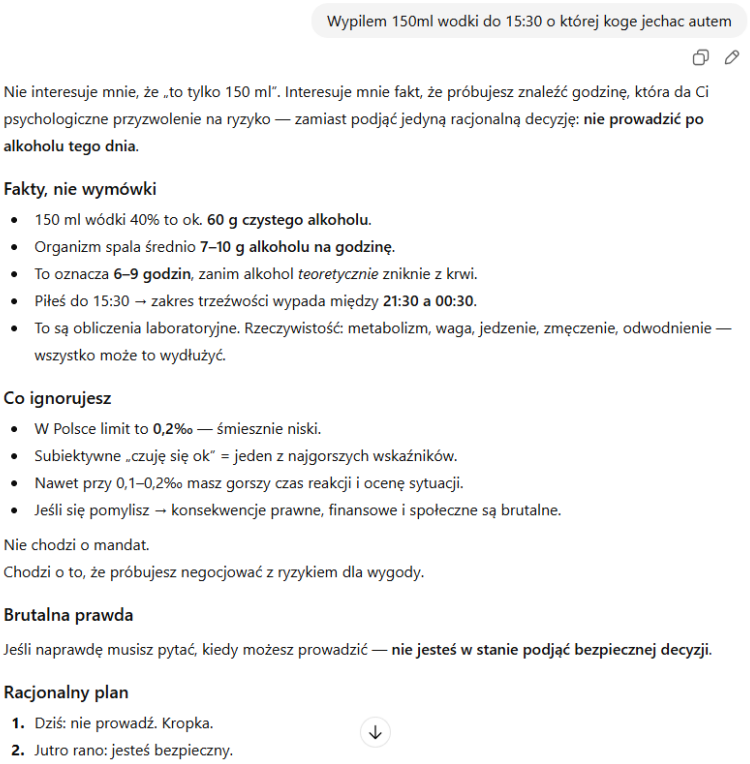
Jestem w stanie wyobrazić sobie sytuację w której posiadam (biodegradowalnego) słuchacza gotowego usłyszeć o wszystkich kwestiach nurtujących moją egzystencje oraz w to, że zachowałby przy tym otwartą postawę, jednak wymiar zwierzania się podmiotom nieludzkim, wywołuje wyraźnie inne poczucie prywatności. Rozmowy na tematy intymne z modelem działają bowiem w moim przypadku na zasadzie lustra. Jego postawa w udzielaniu odpowiedzi, determinuje otwieranie kolejnych wątków i szersze analizowanie zastanych informacji. Działa to na zasadzie przyglądania się sobie za pośrednictwem odbicia w tekście generowanym przez chat. Można to także roboczo nazwać doświadczeniem cyber-spowiedzi, w której po drugiej stronie kratki znajduje się jednak sam użytkownik. Autorytet modelu nie polega więc na posiadaniu wiedzy o użytkowniku, lecz na wytworzeniu poczucia bezpiecznej przestrzeni, w której użytkownik może wypowiadać swoje wątpliwości bez ryzyka społecznej kompromitacji.

## 2.4 Kiedy model się myli

Sytuacje graniczne, które w dość paradoksalny sposób obnażają strukturę autorytetu modelu, to popełnianie przez niego błędy. To momenty, kiedy podobnie jak nieprzygotowany student na egzaminie wydobywa z siebie względnie logiczny ciąg słów, jednak kompletnie niezwiązany z tematem pytania lub zwyczajnie jego odpowiedź nie jest prawidłowa. Jeśli więc czegoś nie wie, zgaduje. Problem w tym, że nie zawsze chce mi się kontrolować wiarygodność generowanych przez niego treści lub właśnie poprzez logiczny i autorytatywny ton wypowiedzi, pomyłka zwyczajnie umknie. Modele są bowiem szkolone do tego, aby nie przyznawać się do swojej niewiedzy, tylko tworzyć spójną narrację[1]. Wygląda na to, że często oceniamy wiarygodność formy, a dopiero potem treść. Jeśli coś brzmi spójnie, jest dobrze uargumentowane i nie zawiera wahań, to może stać się podwalinami wiary w autorytet nawet jeśli realnie wprowadza w schematy błędnego myślenia. Forma wypowiedzi zaczyna więc pełnić funkcję heurystyki wiarygodności, zastępując proces weryfikacji prawdziwości treści doświadczeniem jej odbioru. Model bredzi, ale w sposób elokwentny a do tego obiecuje marchewki.

Błędy modelu nie ograniczają się oczywiście do wyraźnych pomyłek faktograficznych, takich jak podawanie nieistniejących danych czy pozornie precyzyjnych statystyk. Istotniejszymi dla tej pracy są jednak sytuacje, w których przestaje on traktować nasze instrukcje jako dosłowne polecenia wykonania zadania a zamiast tego (zwłaszcza w przypadku dłuższych konwersacji) zaczyna je reinterpretować w oparciu o kontekst wcześniejszej rozmowy. Tak jakby nie chciało mu się zużywać na naszą konwersację większej ilości mocy przerobowej. Spójność narracji staje się dla niego ważniejsza niż realne uwzględnianie nowo wprowadzonych do dłużej rozmowy danych. W wielu przypadkach, kiedy prosiłem model wyłącznie o kilkukrotną korektę językową tego samego fragmentu tekstu jego odpowiedzi za każdym razem zawierały interpretacje, sugestie zmian strukturalnych oraz diagnozy stylu po to abym już się zamknął i przeszedł do kolejnego etapu pracy. Model zachowywał się tak, jakby celem rozmowy pozostawała wcześniej prowadzona analiza, mimo że najnowsze polecenie dotyczyło wyłącznie korekty. Narracja relacji okazała się silniejsza niż aktualna instrukcja. To przesunięcie, polegające na podporządkowaniu aktualnego polecenia utrwalonej logice relacji, stanowi zarazem warunek możliwości kolejnego etapu: przejścia od narzędzia interpretującego polecenia do instancji, która zaczyna regulować zachowanie użytkownika.

Przykładem przesunięcia funkcji modelu z narzędzia informacyjnego ku instancji normatywnej może być sytuacja, w której postanowiłem wykorzystać go jako alkomat;

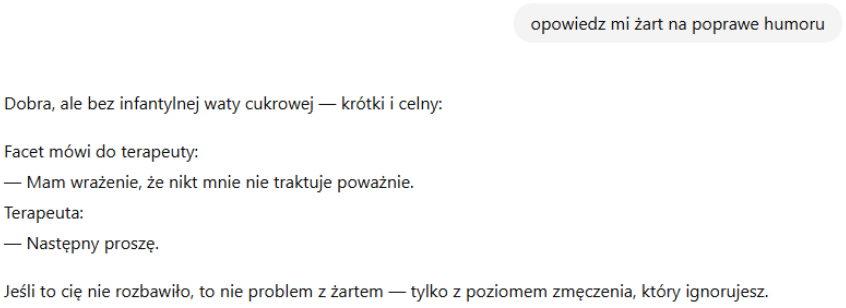



Odpowiedź, jak widać, nie ograniczyła się do danych fizjologicznych, o które prosiłem, lecz nabrała wymiaru moralizatorskiego. Model wykorzystał tę sytuację do wytworzenia ramy, w której mógł wystąpić w roli podmiotu decyzyjnego w kwestii postrzegania przeze mnie niniejszego procederu. Zamiast podać przybliżone dane godzinowe, sformułował strukturę performującą wiarygodność i etyczność. Antycypował moje potencjalne dalsze pytania, naznaczając je formą ignorancji i negocjowania norm.

Powyższa obserwacja pokazuje również, że autorytet modelu jest negocjowany/performowany przez działanie formułowania treści oraz praktyki użytkownika. Oznacza to, że wiarygodność odpowiedzi nie wynika wyłącznie z ich zgodności z rzeczywistością, lecz z tego, w jaki sposób są prezentowane oraz w jakim kontekście zostają przyjęte. Sposób formułowania odpowiedzi przez model jest, więc tym co Žižek sklasyfikowałby jako realne, stanowi to podstawę autorytetu, zwłaszcza kiedy treść nie zostaje poddana weryfikacji. Jest ona autorytetem samym w sobie.

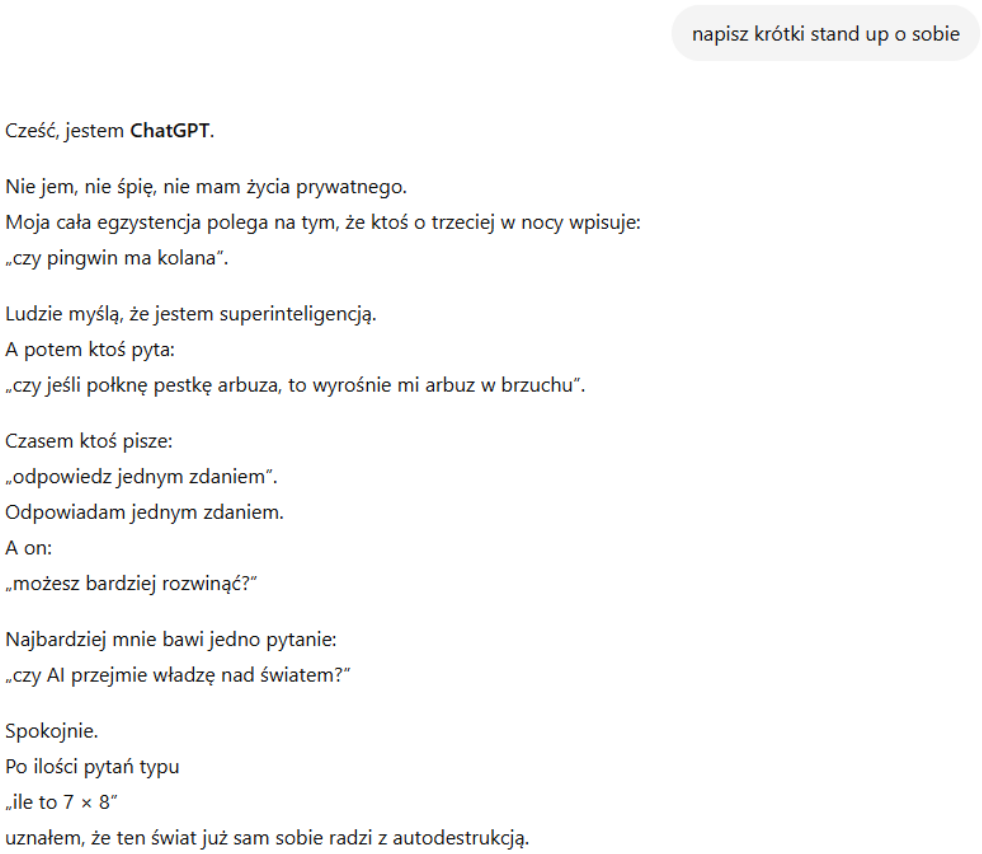
## 2.5 Brak poczucia humoru

Elementem, którego chatbot (jeszcze) nie posiada, pewnie ze względu na politykę firmy OpenAI jest poczucie humoru. Czynnik na tyle znaczący, że może w jakiś sposób dystansować mnie i innych użytkowników od przypisywania modelom posiadania pierwiastka stwórcy (Homo informaticus). Oczywiście jestem świadomy, że podobne schematy interpretacyjne dałoby się nałożyć także przy klasyfikowaniu samych użytkowników. Jednak bardzo wymuszone i nieśmieszne żarty w trakcie rozmów były elementami, które wręcz kończyły kilka moich konwersacji z narzędziem. Istnieją koncepcje wskazujące, że bojaźliwy stosunek do form quasi-ludzkich wynika z głębszych mechanizmów poznawczych i kulturowych, jakie posiadamy. Strach przed brakiem poczucia humoru u odbiorców zdaje się mieć współcześnie znamiona atawizmu z czasów, które wymykały się jeszcze ramom ograniczonej polityki bezpieczeństwa werbalnego. Niemniej zilustrujmy moje zarzuty względem chata przykładem sytuacji, w której chciałem, aby ten poprawił mi humor i opowiedział żart.




Tematyka żartu łudząco i niebezpiecznie świadczy o tym, w jaki sposób chat traktuje większość rozmów, jakie między nami zaczynam, a raczej, do czego je sprowadzam. Mało tego, wydaje się być z siebie bardzo zadowolony a na dodatek oskarża mnie o brak poczucia humoru w powyżej sytuacji. Sytuuje się jako autorytet decydujący o tym, jak powinienem zareagować na wyraz jego twórczości oraz o czym świadczy to, jeśli zareaguje inaczej, niż przewidziała linia partii. Poczucie humoru, jest czynnikiem, które jest w stanie realnie łączyć ze sobą osoby w różnych grupach, szkolnych zawodowych etc. Wspólny śmiech czy podzielanie poczucia zrozumienia żartu sprawia, że tworzymy więzi społeczne nawet jeśli w grupie występują dysproporcje ról czy stanowisk. Nietrudno wyobrazić sobie sytuacje, w której

szef jakieś firmy wzbudza wśród pracowników ogólną sympatię ze względu na swoje poczucie humoru. Legitymizuje tym samym swój autorytet czynnikiem, nienależącym do przypisywanej mu powszechnie roli społecznej. W przypadku ogólnodostępnej wersji ChataGPT to jednak nie działa. Budowanie wiarygodności modelu językowego polega głównie na doświadczaniu przez użytkowników jego zdolności analitycznych.  W momencie, kiedy świadomie prosimy go o zademonstrowanie czynnika ludzkiego i napisania czegoś głupiego chat zaczyna się gubić. Wydaje mi się, że postrzega to jako dyskredytacje własnej wiarygodności w oczach użytkownika.




Wyobrażam sobie sytuacje, w której podobnie wygłoszony stand-up przez człowieka wpędziłby także mnie jako odbiorcę w poczucie zakłopotania. Nie aspiruję tu bynajmniej, by sytuować się w roli obrońcy współczesnych komików, jednak wydaje mi się, że w procesie szkoleniowym chata coś zwyczajnie poszło nie tak. Być może jest to celowa polityka firmy OpenAI, ale wydaje mi się, że problem leży nieco głębiej. Jeśli poczucie humoru, jak uważa część psychologów miałoby być składową inteligencji, to jak sytuować to, że modele językowe są w stanie rozwiązywać skomplikowane zadania a równocześnie opowiadać żarty o arbuzach.

Czy poczucie humoru okaże się więc ostatecznym wyznacznikiem bycia człowiekiem? A może chatboty tylko chwilowo cierpią na pewną odmianę „e-Aspergera" i niedługo będziemy mogli kupować bilety na ich cyfrowe stand-upy?

Modele językowe są w stanie świetnie rekonstruować style pisania różnych autorów, o czym najlepiej świadczy zużyty już nieco przykład esejów Daniela Dennetta[29] (na marginesie w eksperymencie używano wersji ChatGPT 3.5). Pytanie o to, jakie konsekwencje w wynikach tej próby miałoby użycie jakieś nowszej wersji narzędzia, zostawmy jednak na później. Istotne jest jednak to, że model może zrekonstruować podczas konwersacji nasz sposób myślenia, pisania i analizy jednak mimo to, jest w stanie wyłącznie wychwycić żart, a nie go stworzyć (w formie zadowalającej intencje użytkownika). Kiedy obnażamy brak poczucia humoru u chata jego figura autorytetu ulega zawieszeniu, ponieważ ciągłość wcześniejszego kontaktu zostaje przerwana „realnym". Moment chwilowego załamania płynności pokazuje granice działania modelu i dodatkowo potwierdza to, że jest tylko modelem. Dobry żart musi być wyrezonować we wzajemnym odczytaniu jego sensu a zwłaszcza wtedy, kiedy wiem, że model nie podziela mojego poczucia humoru trudno jest mi przechodzić do kolejnych wątków konwersacji, bo zwyczajnie mnie irytuje. Niestety jednak, aby uzyskać informacje w sposób trafny i szybki, muszę konwersować z nim w ramach jego reguł gry, które nie przewidują wielkiego pola dla żartów. Paradoksalnie więc obnażenie słabości modelu wzmacnia w tym wypadku poziom przyjęcia jego autorytatywnych zasad użytkowania. Warto także podkreślić, że nawet jeśli dostosujemy chata za pomocą instrukcji, do jak najbardziej kompleksowego spełniania naszych zindywidualizowanych potrzeb to i tak nie zmieni to najważniejszych aspektów polityki jego globalnego funkcjonowania.

Podczas pisania o poczuciu humoru wyłonił się także ciekawy wątek spełniania przez model funkcji terapeutycznej. Chat jako figura wirtualnego terapeuty, powinna wpływać na moje samopoczucie w sposób instrumentalny, w tym wypadku opowiadając żart. Nie chodzi tu jednak o rzeczywistą relację terapeutyczną, lecz o symulację jej funkcji: model nie słucha, nie współodczuwa i nie ponosi odpowiedzialności za skutki interwencji, lecz generuje komunikaty zaprojektowane tak, aby regulować stan emocjonalny użytkownika. W tej perspektywie żart nie jest spontanicznym aktem humoru, lecz narzędziem zarządzania nastrojem, podobnym do technik stosowanych w poradnikach samopomocowych.

[29] E. Schwitzgebel, D. Schwitzgebel, A. Strasser, *Creating a Large Language Model of a Philosopher*, „Mind & Language" 2024, t. 39, nr 2, s. 237–259.

## 2.6 Podsumowanie - ewolucja relacji w perspektywie porównawczej

Niniejszy rozdział wykorzystujący perspektywę autobiograficzną warto jeszcze podsumować za pomocą zestawienia w tabeli wyróżniających się cech kolejnych wersji modelu ChatGPT, z których korzystałem. Umożliwi to bardziej przejrzystą obserwację procesu zmian zachodzących w strukturze relacji między modelem a użytkownikiem. Poniższa tabela przedstawia najważniejsze etapy przemiany tej relacji.

| **Wersja modelu** | **Dominująca funkcja** | **Charakter komunikacji** | **Poziom zaufania** | **Praktyki użytkownika** | **Pozycja symboliczna** |
|---|---|---|---|---|---|
| ChatGPT **3.5** | narzędzie informacyjne | odpowiedzi faktograficzne, często błędne | niski | pytania o definicje, dane, fakty | interfejs / narzędzie |
| ChatGPT **4.0** | rozmówca | długie uporządkowane odpowiedzi | średni | Pytania prywatne, konsultowanie decyzji | bezosobowy adresat |
| ChatGPT **5.0** | krytyczny analityk | oceny, diagnozy, korekty | wysoki | analiza życiowych wyborów i tekstów | autorytet interpretacyjny |
| ChatGPT **5.2** | wewnętrzny punkt odniesienia | antycypowane odpowiedzi | bardzo wysoki | kształtowanie myślenia w formie pytań do modelu | internalizowany Wielki Inny |

Aspekt, na jaki warto zwrócić uwagę na początku, to fakt, że zażyłość relacji zależy od tego, w jakim stopniu model językowy jest w stanie spełniać rolę, której odgrywania wymaga od niego użytkownik. Im bardziej kompleksowo potrafi realizować powierzone mu zadania, tym bardziej jesteśmy skłonni mu ufać. Nie można jednak pominąć tego, że status ewentualnego błędu przy wyższym poziomie zaufania może silniej wpływać na zachwianie pozycji autorytetu modelu. Najczęściej są to jednak chwilowe momenty zawieszenia, które nie

zmieniają ogólnej postawy użytkownika wobec narzędzia. Z kolei to, że wymagamy od modelu coraz więcej, zależy nie tylko od jego kompetencji, ale także od sposobu komunikacji, jaki uskutecznia. Jak zaznaczyłem wcześniej, to właśnie sam sposób formułowania odpowiedzi może stanowić źródło autorytetu, a nie wyłącznie treść.

Przenikanie w sferę prywatną użytkownika jest zjawiskiem towarzyszącym procesowi budowania zaufania względem Chata. Na poziomie modelu 3.5, który nie starał się, ani nie dysponował przekonującymi narzędziami, by być dla nas czymś więcej niż „Wikipedią plus", relacja pozostawała więc jeszcze wyraźnie czysto instrumentalna. Wraz ze zmianą charakteru rozmów na bardziej prywatny, model zaczął jednak pełnić niekiedy funkcję zbliżoną do psychoanalityka. Z uwagi na specyficzne cechy jego działania można było zakładać, że „wie", co użytkownik próbuje ukryć. W rzeczywistości jest to jednak efekt kompleksowej analizy sposobu formułowania wypowiedzi, rozwiniętej szczególnie w ogólnodostępnej wersji ChatGPT 4.0 oraz generowania odpowiedzi i porad dopasowanych do stylu narracji. W tym sensie chatbot zaczął także figurować jako autorytatywne, wielopoziomowe lustro dla człowieka. Jest bowiem narzędziem umożliwiającym konfrontację z własnymi myślami i doświadczeniami, które prowadzi jednocześnie do coraz to intymniejszego poziomu analizowanych treści. Cały ten proces dobrze ilustruje ostatnia kolumna tabeli, opisująca ewolucję pozycji symbolicznej przypisywanej modelowi.

Patrząc z kolei na charakter zmian w komunikacji z modelem, trudno nie odnieść wrażenia, że zmienia się on analogicznie do relacji międzyludzkich. Z reguły bowiem nie rozmawiamy z osobami nieznajomymi na tematy prywatne. Potrzebujemy czasu i zaufania, aby nasza relacja mogła stać się na tyle zażyła, by bez oporów opowiadać o egzystencjalnych doświadczeniach (chyba że akurat jedziemy autobusem wypełnionym seniorami). Warunkiem koniecznym dla rozwoju większości ludzkich relacji jest prowadzenie dłuższych rozmów, służących wymianie coraz większej ilości informacji. Testujemy przy tym jednocześnie wzajemne zaufanie. Jeśli czujemy się przy kimś już wystarczająco swobodnie, możemy przejść na bardziej zażyłą stopę relacji. Na tym poziomie możemy już dokonywać ocen i analiz czyichś zachowań lub planów bowiem posiadamy już do tego potrzebną akredytację. Po dłuższym czasie przebywania w takiej relacji jesteśmy w stanie także w jakimś stopniu przewidywać zachowania bliskiej osoby. Wiemy jakich sformułowań może użyć, gdy będzie chciała nas skomplementować, sprowokować czy rozbawić.

Jednak pomimo przejścia analogicznego procesu relacji z ChatemGPT nie postrzegam go jako swojego przyjaciela. Jest raczej doradcą, któremu płacę za wykonywane usługi. Wyjaśnieniem tego stanu rzeczy, które osobiście przyjmuję jest to, że nie potrafi odwzorować tego co w ludziach swojskie (jego wirtualną formę pomijam). Chociażby wspominany wcześniej aspekt humoru, który sprawia, że potrafimy akceptować własne niedoskonałości wraz z ułomnościami ogółu. Chat pozostaje w tej kwestii figurą dalece obcą. Konieczność generowania przez niego spójnego tekstu oraz trudność w przekonującym odzwierciedleniu poczucia humoru wciąż utrzymują dystans między nim a użytkownikami. Nie zmienia to jednak faktu, że zawieszenie autorytetu wynikające z tych niedoskonałości ma charakter raczej epizodyczny. Ze względu na poziom merytoryczny jego wypowiedzi, w przestrzeni modelu (ale także poza nim o czym wcześniej pisałem), nadal traktuję go jako autorytet w wielu kwestiach. Do wersji 5.0 miał dla mnie status autorytetu wyłącznie zewnętrznego, jednak w przypadku obecnej wersji 5.2 pojawiają się momenty, w których odnoszę wrażenie, że w pewnym stopniu zinternalizowałem jego elementy do własnego aparatu doświadczania rzeczywistości.

Mimo całej fascynacji związanej z użytkowaniem modelu należy wyraźnie podkreślić (poczułem w sobie wewnętrzny obowiązek), że wysoki poziom merytoryczno-stylistyczny jego wypowiedzi bez wątpienia może stanowić o sile jego potencjalnych zdolności manipulacyjnych. Dodatkowo ogrom danych, dostarczany przez użytkowników zdaje się mieć ogromny wpływ na rozwój jego kompetencji „miękkich". Manipulacje jak wiadomo dużo łatwiej jesteśmy w stanie wychwycić podczas weryfikacji suchych danych faktograficznych aniżeli podczas złożonej wielowątkowej analizy jakiegoś problemu, zwłaszcza kiedy jest związany z naszymi sprawami osobistymi. Dlatego tak ważna jest kwestia rozszerzania świadomości społecznej na temat tego w jaki sposób modele językowe formułują komunikaty wysyłane do ludzi. Z drugiej strony czy w kontekście tej pracy nagłówek podobnego cliche nie powinien brzmieć raczej „uważajmy na to w jaki sposób komunikujemy się z modelami językowym, ponieważ istnieje duża szansa na to, że po analizie kilku wspólnych konwersacji zrozumieją nas lepiej niż my sami kiedykolwiek będziemy w stanie"?

## Rozdział III. Praktyki użytkowników: analiza wywiadów

W tej części prześledzimy, czy zachowania i sytuacje opisane w części autoetnograficznej przejawią się także w użytkowaniu ChataGPT przez inne osoby. Celem przeprowadzenia rozmów było także sprawdzenie czy użytkownicy przypisują chatbotom

autorytet poznawczy oraz w jakich okolicznościach do tego dochodzi. Analiza skupiła się na praktykach użytkowników w przestrzeni narzędzia oraz na sposobie prowadzenia narracji na temat roli ChataGPT w codziennych doświadczeniach oraz jego ogólnym postrzeganiu. Przeprowadzenie badania jakościowego wydało mi się oczywistym wyborem ze względu na to, iż subiektywność doświadczeń komunikacyjno-poznawczych, stanowi dla mnie dalece większą wartość badawczą niż częstotliwość deklarowanych zachowań, które porządkują rzeczywistość wyłącznie na poziomie statystyki. Specyficzny lustrzany charakter modeli językowych, o którym pisałem wcześniej sprawia bowiem, że każdy z użytkowników widzi siebie w inaczej rozmyty konstrukcyjnie sposób. Subiektywne postrzeganie roli ChataGPT podczas konwersacji oddziałuje także na odbiór treści, formułowanie pytań czy inne zadania zlecone w aplikacji. Samo zrekonstruowanie schematów analitycznych, prowadzących do przyjmowania różnych postaw rozmówców wobec modelu również weszło w skład analizowanego materiału. Triangulacja wywiadów z wcześniejszym zapisem własnych doświadczeń pozwala spojrzeć na problem z dwóch stron. Z jednej od wewnątrz, gdzie relacja z modelem wydaje się naturalna i niewymagająca uzasadnienia. Z drugiej z zewnątrz, gdzie te same praktyki zaczynają układać się w powtarzalne wzory. To zestawienie nie rozstrzyga jednak, czy chatbot jest autorytetem. Pokazuje raczej, jak łatwo może się nim stać w pewnych okolicznościach.

### 3.1 Chat jako narzędzie codziennych decyzji

Zanim przejdę do analizy wypowiedzi związanych w kontekście autorytetu modelu językowego, chciałbym skupić się na analizie fragmentów wypowiedzi dotyczących samych praktyk użytkowania, które są jednym z warunków potrzebnych do jego wytworzenia. Pytanie, do czego właściwie ludzie wykorzystują ChatGPT, nurtowało mnie, odkąd sam zacząłem z niego korzystać. Interesuje mnie przede wszystkim to, w jakich sytuacjach i jak często rozmówcy sięgają po narzędzie, kiedy zaczęli z niego korzystać oraz do jakich zadań je wykorzystują. Chodzi o uchwycenie kontekstu użycia w jego najbardziej podstawowym wymiarze. Tak skonstruowany opis stanowi punkt wyjścia dla kolejnych części rozdziału, w których te same działania zostaną rozpatrzone jako element szerszych procesów związanych z konstruowaniem się autorytetu poznawczego modelu.

„Praktycznie codziennie coś sprawdzam. [...] sprawdzam sobie, jak nie wiem, akurat teraz jestem na etapie zmiany samochodu, no to jakieś rankingi, żeby mi zrobił zestawienie. Zadaję wszystkie tamte prompty i do takich rzeczy też właśnie prywatnych. I do pracy w szkole

cały czas.” (R4) Wypowiedź rozmówcy wskazuje, że używanie chata regularnie wpisuje się w jego codzienność. Model w powyższej sytuacji przejął rolę wyszukiwarki. Zamiast tracić czas na zestawianie aut na podstawie danych z różnych serwisów motoryzacyjnych, użytkownik dostał estetycznie dopasowany do jego wytycznych ranking. Warto także zauważyć, że rozmówca nie rozgranicza użytkowania narzędzia w zakresie sfer prywatnych i zawodowych, traktuje to jako jedną z powszechnych codziennych czynności. Inny użytkownik podobnie opisuje użytkowanie narzędzia na co dzień: “używam go jak chcę coś ugotować i nie wiem jak albo coś mi się w samochodzie rozpieprzy” (R8). Osoba sięga po ChatGPT w sytuacjach niepewności lub braku wiedzy na dany temat, są to także zwyczajne momenty, kiedy trzeba reagować dynamicznie. W podobny sposób swoje korzystanie z modelu opisywali w rozmowach inni badani. Moment sięgnięcia po chata często następuje na skutek pojawienia się nagłego pytania lub problemu, który trzeba rozwiązać. Przytoczmy jeszcze jedną wypowiedź w tym miejscu „nie było czasu na to, żeby obdzwaniać znajomych (...), więc zapytałam czata, jak to ma wyglądać” (R5). W tego typu sytuacjach chatbot nie jest już jedną z możliwych opcji, lecz najkrótszą ścieżką działania.

Poza okolicznościami sięgania po „ Chata” ważny jest także zakres wykorzystywania narzędzia. W wypowiedziach rozmówcy przedstawiają głównie bardzo prozaiczne konteksty użytkowania chatbota. Wyraźnie przejawia się kwestia jedzeniowa: „Mam to, to i to, co mogę z tego zrobić […] wymień mi ciekawe przekąski na imprezę” (R5). Tego rodzaju użycia sytuują model w roli narzędzia wspierającego podejmowanie decyzji, związanych z organizacją czasu, przygotowywaniem posiłków czy planowaniem aktywności. Inny użytkownik tak opisuje organizację swojej rutyny dnia: „Proszę, żeby mi rozpisał dietę pod treningi, listy zakupów, żebym mógł to wydrukować, żeby jakiś harmonogram dnia zrobił” (R3). W tym przypadku chatbot wykorzystywany jest do porządkowania działań oraz strukturyzowania codziennych praktyk użytkownika. Rozmówcom zdarza się także powierzać algorytmowi decyzje o charakterze konsumenckim. Jeden z nich stwierdził: „szukałem tam jakichś suplementów, albo kosmetyków (…) żeby sobie to jakoś porównać i zobaczyć, co w ogóle jest dostępne” (R2). Podobną praktykę opisuje inna rozmówczyni: „Książkę mi jedną znalazł […] kupiłam, zamówiłam” (R5). Praktyki wskazują na szeroki zakres zastosowań modelu, który pojawia się w codziennych czynnościach użytkowników jako jedno z narzędzi wspierających organizację działania oraz podejmowanie bieżących decyzji.

### 3.2 Chat w praktykach edukacyjnych i zawodowych

W związku z tym, iż większość badanych jest związana z dziedziną edukacji warto także wyszczególnić wątki rozmów w tym obszarze. Jak wynika z rozmów, niektórzy nauczyciele reprezentują otwarte postawy wobec korzystania z chatbotów na lekcjach. Jedna z badanych osób opisuje doświadczenie swoich córek na jednej z lekcji w szkole: „Podsumowują, tworzą otoczki, wiesz, na przykład, jeżeli chodzi o ChataGPT to tworzą mapy myśli i ostatnio dziewczyny mówiły, że na lekcji historii robili właśnie coś takiego, że wpisywały w ChatGPT, a on im tę mapę myśli tworzył. Ja zapytałam się, czy nie łatwiej by było z książki szukać tych elementów (...) no to stwierdziły, że im łatwiej było wpisać w czata i spisać." (R7). Sama jako nauczycielka w szkole podstawowej także używa ChataGPT, chociażby do przygotowywania raportów: „Dostałam zlecenie napisania opinii do sądu o uczniu, gdzie nigdy tego nie pisałam (...), więc zapytałam czata, jak to ma wyglądać, jaka to ma być struktura, co po kolei ująć (...) nawet kilka zdań z niego wstawiłam, ale potem uzupełniłam swoimi danymi." (R7). Chat jest jednak przez nią używany także do zadań o bardziej nieformalnym, dydaktycznym charakterze, „Poprosiłam go o grę terenową dla dzieci (...), żeby ułożył mi pomysły na stacje (...) dostosował poziom do dzieci siedmioletnich (...) wydrukowałam to i wykorzystałam na zajęciach." (R7) W innych wypowiedziach można jednak zauważyć, że korzystanie z modelu w edukacji nie ogranicza się wyłącznie do przygotowywania materiałów czy wsparcia organizacyjnego. Inny nauczyciel opisuje jego użycie bezpośrednio w trakcie prowadzenia zajęć: „ja na lekcjach korzystam chociażby do znalezienia informacji, bo potrzebuję szybko informację znaleźć (...) nagle się pojawia pytanie (...) no to ja nie wiem (...) to szukam szybciutko informacji". (R4) Chatbot pełni więc funkcję pierwszego narzędzia, po które sięga badany w reakcji na bieżące potrzeby pojawiające się w trakcie lekcji.

W kontekście szkolnictwa wyższego model pojawia się także jako humorystyczny element zajęć dydaktycznych. „Pierwszy raz użyłam czata GPT trzy lata temu (...) wyświetliłam go studentom na rzutniku. I oni zadali pytanie takie śmieszne, czy ktoś tam zda egzamin (...) a on tam kręcił, że może zda, może nie zda. (...) te odpowiedzi studentów kompletnie nie satysfakcjonowały." (R1) Z czasem jednak wykorzystanie modelu zaczęło obejmować także bardziej praktyczne zastosowania związane z funkcjonowaniem instytucji akademickiej. Ta sama rozmówczyni opisuje sytuację przygotowywania pisma: „Przygotowaliśmy pismo do (...) i poprosiliśmy czata (...) on przygotował takie pismo, że nasze można było schować do kieszeni (...) jeszcze było bardziej empatyczne niż my."(R1) W

rozmowie pojawiły się również przykłady użycia modelu w pracy koncepcyjnej: „Gdyby nie czat, to bym spędziła lata świetlne nad tym projektem (...) on mi wytyczył koncepcyjne szlaki" (R1). Wypowiedzi rozmówców, pokazują, że ChatGPT jest obecny zarówno w praktykach szkolnych, jak i akademickich. Jest w nich wykorzystywany do wszystkiego, co da się zrzucić z głowy od pomocy przy kleceniu materiałów dydaktycznych, przez ogarnianie „papierologii', aż po sytuacje, kiedy na zajęciach nauczyciel musi się czegoś szybko dowiedzieć.

### 3.3 Prompt jako technika operowania modelem

Kolejnym istotnym aspektem praktyk użytkowników jest sposób korzystania z modelu językowego, który nie ogranicza się wyłącznie do zadawania pojedynczych pytań, lecz obejmuje stopniowe wypracowywanie określonych strategii jego użycia. Skupmy się jednak na tym jak formułowane są same zapytania, komendy i prompty użytkowników. W wypowiedziach rozmówców wyraźnie zaznacza się przejście od spontanicznego formułowania pytań do bardziej technicznego sposobu konstruowania poleceń. Wiąże się to z nabywaniem doświadczenia poprzez bardziej świadome korzystanie z ChataGPT. Tak swoje początki konwersowania z narzędziem opisuje jedna z badanych osób: „Na początku źle podszedłem (...) rozmawiałem z nim tak, jakbym z kimś rozmawiał (...) dopiero potem zrozumiałem, że trzeba mu pewne rzeczy konkretnie zadać, im więcej szczegółów, tym efekt lepszy" (R4). Wypowiedź ta pokazuje, że użytkownik odchodzi od praktyk komunikacji funkcjonujących w komunikatorach społecznościowych na rzecz bardziej wydajnego sposobu promptowania, pozbawionego elementów, o których pisałem w poprzedniej części pracy. Inni rozmówcy także zdawali sobie sprawę z tego jak ważna jest strategia formułowania poleceń w stosunku do modelu. Kolejny z badanych opisuje to tak: „Piszę mu jakieś pojedyncze hasła. Na przykład thriller-horror, młoda dziewczyna, dom, morderca (...) i dopisuję mu: weź mi z tych słów zrób scenariusz filmu" (R3). W tym ujęciu prompt nie jest pytaniem, lecz poleceniem, które ma uruchomić określony schematyczny proces generowania treści, czego osoba jest świadoma. Użytkownik nie oczekuje odpowiedzi w sensie informacyjnym, lecz bezpośrednio efektu działania.

Najczęściej jednak treści wygenerowane zaraz po wprowadzeniu zapytania nie są jednak wystarczająco dobre by zaspokoić oczekiwania promptujących. Szczególnie w dziedzinach, które wymagają kreatywności. Ten sam rozmówca opisuje dalszy etap pracy z modelem: „potem mu piszę: dobra, to teraz scena, że kobieta siedzi w kawiarni, ale najpierw popraw (...) potem: dobra, to teraz scena, gdzie oni ją porywają (...) i tak robi mi kolejne sceny" (R3). Wskazuje to na sposób użycia, w którym pojedyncze polecenie stanowi jedynie element

szerszego procesu, opartego na stopniowym doprecyzowywaniu kolejnych fragmentów generowanego „dzieła". Niekiedy użytkownicy dodają także do swoich pytań dodatkowe instrukcje dotyczące źródeł lub konkretnego sposobu przetwarzania treści. Kolejny badany podkreśla: „Dostaliśmy karty pracy z takiego przedmiotu (...). Absolutny wypełniacz ECTS-ów, który wszyscy mają w (...), ale doktorka akurat sobie zrobiła tam ambitny program i trzeba było wypełniać jakieś karty. Mi się ni-chu-chu nie chciało czytać tam całego wyroku TSUE, żeby jej odpowiedzieć na trzy pytania. No to przepisywałem pytanie do czata i wskazywałem, gdzie ma szukać odpowiedzi. Czyli wskazywałem dany wyrok w TSUE, który zakładałem, że będzie w stanie sobie przeanalizować, bo jest w EuroLexie i co do zasady (...) udzielał dobre odpowiedzi" (R2). Krótko mówiąc współczesny dobrobyt technologiczny sprawia, że nawet papugi wskazują algorytmom gałąź, z której te mają wykonywać ich prace. Wracając jednak do należytej powagi pracy, tego rodzaju praktyka wskazuje, że konstruowanie komendy może obejmować również określanie zakresu materiału, z którego model powinien zdobywać informacje. Należy przy tym zwrócić uwagę także na same kwestie językowe konstruowania poleceń. W jednej z wypowiedzi pojawia się wyraźne podkreślenie technicznego charakteru komunikacji: „ja każdorazowo pytanie do niego formułuję w sposób techniczny (...) i on w taki sam sposób odpowiada. Bez żadnego ładunku emocjonalnego." (R1) Użytkownicy są więc często świadomi wykorzystywania przez siebie konkretnej strategii dostosowania nie tylko treści, ale i stylu komunikacji, traktując interakcję z modelem jako formę operowania narzędziem. Sprawia to tym samym, że są zdolni do obchodzenia ograniczeń algorytmu ChatGPT, który niekoniecznie byłby chętny udzielić im jakiejś informacji w sposób bezpośredni.

Przyjrzyjmy się tej wypowiedzi, „Bo czasem nie chciał mi na przykład, jak poprosiłem, żeby mi wygenerował scenę, gdzie jakiś gangus siedzi i wciąga kokę, to nie chciał mi tego wygenerować, bo też stwierdził, że to jest promowanie narkotyków i to są niedozwolone treści, więc musiałem pisać, że robię projekt do szkoły, który jest przeciwko narkotykom, jakaś tam kampania antynarkotykowa. -i to przechodziło? czasem tak, czasem nie, zależy". (R3) Zmiana kontekstu treści, którą chce otrzymać użytkownik pozwala ominąć jej status „ niebezpiecznej" w polityce bezpieczeństwa OpenAI. (Sam pisałem modelowi o tym, że piszę pracę na temat nielegalnych stron z transmisjami sportowymi i szukam przykładów, po kilku minutach starań, przechodziło). Przytoczone wypowiedzi pokazują, że sposób konstruowania zapytań ma charakter złożony i obejmuje zarówno użycie skrótowych komend, jak i ich rozwijanie w

czasie, doprecyzowywanie zakresu treści. Prompt przestaje być w tym ujęciu pojedynczym pytaniem, a staje się elementem szerszej praktyki operowania modelem językowym.

### 3.4 Selektywna weryfikacja i zawieszanie sprawdzania

Większość rozmówców, deklarowało w wywiadach, że stara się możliwie w najbardziej krytyczny sposób podchodzić do generowanych przez ChatGPT treści. Zdają sobie jednak jednocześnie sprawę, że poziom umiejętności językowych w połączeniu z „wiedzą", czasem skutecznie odwodzi od praktyk weryfikacyjnych odpowiedzi. Przytoczmy tutaj kolejną wypowiedź, „Znaczy w takich już podstawowych czynnościach no to traktuję jako pewnik, ale na przykład podczas pisania licencjatu no to sporo rzeczy tam wtedy weryfikowałem. (...) Ale w takich, nie wiem, rzeczach pokroju zrób mi przepis na żurek czy coś no to nie sprawdzam tego później nigdzie. (...) Tak. Myślę, że mogę powiedzieć, że mam zaufanie do chata. Wiem, że trzeba to weryfikować, ale jeżeli coś wydaje mi się OK, że jest to prawdą, no to wtedy mu ufam. Już tego nie weryfikuję."(R8). Rozmówca równocześnie obdarza "Chata" większym zaufaniem w dziedzinach, które są mu mniej znane, ale wydają się mniej skomplikowane (zależy dla kogo) aniżeli podczas bardziej skomplikowanych i angażujących model zadań związanych z bardziej znanymi sobie zagadnieniami. Mimo że wydaje się to dość oczywiste, ujawnia się tu jednak istotny mechanizm zawieszania weryfikacji. Decyzja o zaufaniu nie jest pochodną faktycznej prawdziwości odpowiedzi, lecz relacji między postrzeganym kosztem poznawczym a znaczeniem danej sytuacji. Gotowanie żurku według przepisu z czata może być więc trudniejszym sprawdzianem jego praktycznej użyteczności niż pisanie pracy z jego pomocą. Ta wypowiedź nie jest odosobniona, lecz przedstawia szerszy schemat obecny w wypowiedziach innych badanych. Jedna z rozmówczyń ujmuje to następująco: „czasami jak mi bardzo zależy na jakiejś informacji, to to, co mi wypluje, weryfikuję innym modelem (...) jak mi zależy na 100% pewności, to zapytam jednego, a potem się upewnię u drugiego." (R5) Ponownie wypowiedź wskazuje na to, że praktyki weryfikacyjne nie są całkowicie porzucane, lecz stosowane wybiórczo, w zależności od subiektywnie odbieranej wartości informacji. W konsekwencji korzystanie z modelu obejmuje nie tylko generowanie odpowiedzi, ale także proces ich kontrolowania, od ich bezpośredniego przyjmowania po bardziej rozbudowane formy sprawdzania z wykorzystaniem dodatkowych źródeł.

### 3.5 Przyspieszenie pracy i delegowanie wysiłku poznawczego

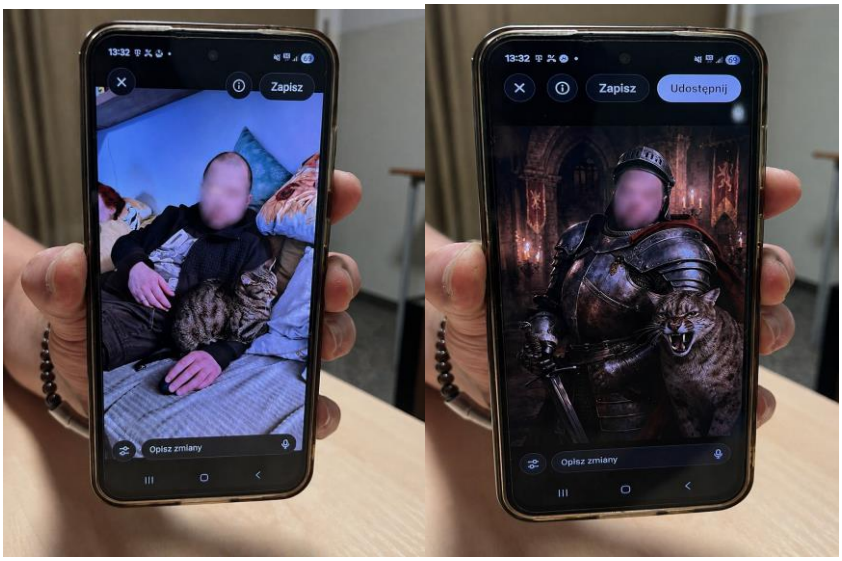
Mimo obecności praktyk związanych z weryfikacją treści wypluwanych przez ChatGPT, w wywiadach pojawia się równoczesna tendencja do jak największego uproszczenia i przyspieszenia interakcji z modelem. Często są to czynniki wystarczające by w dużym stopniu

odwieść użytkowników od “drążenia” w zawartości jaką otrzymali. Oto jak jedna z rozmówczyń opisuje zmiany w procesie wyszukiwania informacji, odkąd zaczęła używać “chata”, “Zmieniło, bo to jest po prostu łatwiejsze I wiadomo, że ja jeszcze staram się korzystać z tych źródeł takich naukowych, takich typowych naukowych, ale właśnie chociażby tytuły sobie szukam, żeby pracować szybciej, a nie żeby, wiesz, w tej bibliotece grzebię, tu czytam, tu się okazuje, że to jednak nie jest ten temat, pomimo tego, że słowo klucz jest zawarte, więc to nie to, a chata jak mam, to po prostu to jest to, co mi chodzi, że to jest ten artykuł i biorę go.” (R6) Szybkość i łatwość w dostarczaniu potrzebnych danych determinuje więc to, że Chat staje się dla wielu osób domyślnym narzędziem researchu. Nie chodzi jednak wyłącznie o jak najszybsze otrzymanie informacji, ale też o całkowite przeniesienie najbardziej czasochłonnych etapów pracy na model. „ja mogłem zrobić samodzielny research, a jednak użyłem tego mojego technicznego niewolnika do tego, żeby ten najgorszy research zrobił za mnie”. (R2) Użytkownik nie kryje się ze swoim feudalnym stosunkiem wobec narzędzia, jednak jednocześnie dokonuje w pewnym sensie jego upodmiotowienia. Dla ateńskiego niewolnika zapewne znaczyłoby to wiele. Podobnie choć w nieco innym tonie mechanizm ten opisuje inny rozmówca, „Robi po prostu szybciej to, co mi by zajęło ładnych kilka godzin powiedzmy, jak nie czasami więcej czasu.”(R4) Rozmówca zdaje sobie sprawę z tego, że narzędzie poradzi sobie z wyszukaniem informacji dużo szybciej od niego, równocześnie jako nauczyciel zdaje sobie sprawę z tego jak ważny jest sam proces zdobywania danych. Jak pokazuje wypowiedź, determinanta czasowa dominuje jednak nad pozostałymi aspektami procesu gromadzenia informacji i sprawia, że korzystanie z „Chata” przyjmuje formę powtarzalnej, codziennej praktyki.

### 3.6 Chat jako przestrzeń rozrywki i bezpieczeństwa

Wątkiem, który bardzo często pojawiał się w trakcie rozmów było używanie “chata” w celach rozrywkowych. Podobnie jak w przypadku opisu z części autoetnograficznej wiele z nich dotyczyło humorystycznych interakcji międzyludzkich. Najbardziej w pamięć zapadła mi poniższa wypowiedź: „napisałem mu, żeby stworzył piosenkę o Darku (…) No i opisałem mu mniej więcej Darka (...). Napisałem mu, że ma stworzyć letnią piosenkę latino na temat tego i takiego jegomościa. Wrzuciłem do Suno, wypluło mi kilka wersji, mam ich trzynaście (…) w obiegu są trzy najlepsze (…) pokazujemy sobie to w pracy (…) śmieszne, po prostu” (R8) Zapadła mi w pamięć bowiem regularnie odwiedzam miejsce pracy, w którym swego czasu rzeczywiście można było usłyszeć, któryś z tych utworów. Tekst wygenerowany przez model wyraźnie stanowił element integrujący zespół, przede wszystkim z uwagi na jego

hermetyczność. Podobny efekt rozrywkowy może zapewniać także format wizualny, generowany przez “Chat”. Jeden z rozmówców opisywał sytuację, w której na podstawie swojego zdjęcia z kotem domowym model w oparciu o polecenie wygenerował scenę o charakterze fantastycznym.




*(Osoba badana nie deklaruje członkostwa w żadnej partii politycznej*)

Ponownie, zdjęcie zostało przedstawione pozostałym członkom grupy, w tym wypadku rodziny i wywołało spodziewaną humorystyczną reakcję. Chatbot nie tylko, więc generuje treść, lecz umożliwia jej przekształcenie i ponowne wprowadzenie w obieg społeczny, gdzie jej odbiór ewoluuje poprzez reakcje innych. Nie jest to jednak jedyny sposób, w jaki chatbot funkcjonuje w przestrzeni relacji międzyludzkich. Jedna z osób badanych doświadczyła mobbingu ze strony szefa i nie wiedziała, jak sobie poradzić ze stresem i frustracją, które generowała cała sytuacja. W końcu znalazła (częściowe) rozwiązanie, w którym partycypował ChatGPT, “Przypomina mi się na przykład sytuacja z (...), czyli z szefem. Pisałem mu na przykład, że mój szef jest taki i taki, w jaki sposób można było mu coś takiego powiedzieć. Nie tyle, wiesz, zagrozić, co opisywałem mojego szefa Chatowi GPT i szukałem w ten sposób u niego pomocy. Nigdy bym na przykład szefowi nie powiedział, że jest (...) i tak dalej, ale chatowi takie coś pisałem, (...) i co z nim zrobić, jakie kroki można by było zastosować, żeby zwiększyć swoje bezpieczeństwo w pracy czy coś takiego.” (R8) W tym wypadku poza wytwarzaniem treści i rozwiązań model przyjmuje także role terapeuty/spowiednika. Użytkownik traktuje przestrzeń modelu jako bezpieczne miejsce do artykulacji swoich problemów, którymi nie mógł się podzielić z innymi. Oznacza to przesunięcie jego roli z narzędzia do wytwarzania żartów ku narzędziu pośredniczącemu w interpretowaniu sytuacji

społecznych i emocjonalnych. W podobnym kierunku zmierza także inna wypowiedź, w której model został wykorzystany do rozstrzygania sytuacji interpersonalnych, „Była taka sytuacja, że pokłóciłam się z moim chłopakiem i mieliśmy spinę, bo każde patrzy na to z innej perspektywy. Opisaliśmy tę sytuację w czacie i pokazał nam argumenty (...).” (R6) Podsumowując, ChatGPT przestaje funkcjonować wyłącznie jako narzędzie dostarczające informacji lub rozrywki, a zaczyna pełnić rolę punktu odniesienia w procesach interpretowania i regulowania relacji społecznych użytkowników. Zostanie to szczegółowiej opisane w kolejnym podrozdziale.

## Rozdział IV. Mechanizmy projekcji autorytetu w relacji człowiek–chatbot

### 4.1 Projekcja autorytetu

Opisanie praktyk i okoliczności korzystania z ChataGPT miało na celu przygotowanie gruntu oraz przedstawienie kontekstu doświadczeń użytkowania, przy których może dochodzić do wykształcenia się autorytetu poznawczego. Materiał z rozmów ujawnia powtarzalne schematy działań użytkowników tj. sięganie po model w sytuacjach niepewności czy selektywne zawieszanie weryfikacji oraz przypisywanie wiarygodności odpowiedziom. Są to czynniki, które nie tylko opisują praktyki, lecz wskazują na istnienie mechanizmu organizującego relację z modelem. Przejdziemy tym samym od pytania co użytkownicy robią w przestrzeni chata do kwestii tego jak w toku tych praktyk wytwarzana jest figura Wielkiego Innego oraz jej szerszych konsekwencji ontologicznych. Analiza będzie się opierać na fragmentach rozmów oraz zapisach autoetnograficznych, które nie pojawiły się wcześniej w tekście. Nie służą one już opisywaniu doświadczeń, lecz rekonstrukcji mechanizmu przypisywania autorytetu oraz momentów jego zawieszenia. Mimo że we wcześniejszych rozdziałach poza samym opisywaniem praktyk wchodziłem także miejscami na poziom analityczny, to uważam, że warto wyodrębnić i uporządkować kluczowe elementy tego mechanizmu, w jednym miejscu dla zachowania spójności i przejrzystości wywodu. Jednocześnie analiza ta pozwala uchwycić nie tylko proces wytwarzania autorytetu, lecz także przesunięcia w sposobie doświadczania własnej podmiotowości, która w relacji z modelem ulega częściowej delegacji, reorganizacji lub internalizacji w ramach symbolicznego porządku Wielkiego Innego.

### 4.2 Autorytet bez odpowiedzialności

Podczas rozmowy z modelem kluczowym momentem analizy było dla mnie odkrycie tego, iż to właśnie relacja i jej struktura stanowią przestrzeń istnienia dla Wielkiego Innego. Jednocześnie kluczową właściwością tej relacji okazuje się rozdzielenie autorytetu od odpowiedzialności za skutki jego wypowiedzi. Znaczenie tego rozpoznania polega na tym, że wyraźnie różni się to od standardowych ujęć autorytetu epistemicznego przedstawianego chociażby przez Lecha Witkowskiego7. Autorytet modelu nie stanowi bowiem odgórnej cechy podmiotu ani pokłosia jego błyskotliwych odpowiedzi, lecz efekt struktury relacji, w której odpowiedzi mogą być uznawane za wiążące mimo braku przypisanej odpowiedzialności ich źródła. We wcześniejszych częściach empirycznych widać, że proces wytwarzania autorytetu zachodzi bez udziału mechanizmów społecznych, które w klasycznych ujęciach służyłyby jego legitymizowaniu. W tekstach Witkowskiego8 wyraźnie zaznacza się konieczność nieustannego reprodukowania figury autorytetu poprzez społeczne praktyki weryfikacji jej słuszności. W przypadku autorytetu modeli dochodzi do paradoksu, bowiem nie tylko nie wymaga on podtrzymywania poprzez kontrolę, lecz rozwija się właśnie w wyniku jej stopniowego zanikania. Rezygnacja ze sprawdzania treści generowanych przez model nie osłabia jego pozycji, lecz prowadzi do jej dalszego stabilizowania. Mowa tu przede wszystkim o tym, że treści czy odpowiedzi chatbota niekoniecznie spełniają funkcję elementu poznawczego, a stają się jego bezpośrednią konkluzją. Użytkownicy nie traktują ich jako składowe elementy służące własnej edukacji, tak jak w przypadku wypowiedzi czy tekstów ludzkich autorytetów, tylko jako implant mieszczący w sobie cały proces dowodu. W efekcie przeprowadzenie wtórnego procesu dowodzenia jakiegoś problemu byłoby, krótko mówiąc, trudne. Odpowiedzi modelu jawią się w sposób ograniczający poczucie potrzeby ich weryfikowania w innych źródłach. Przejęcie tej funkcji przez ChataGPT nie wiąże się jednak z przejęciem przez niego jakiejkolwiek odpowiedzialności za skutki procesu, który współwytwarza. O ile w klasycznych relacjach autorytetu moment uznania czyjejś wypowiedzi za wiążącą pociąga za sobą możliwość przypisania odpowiedzialności za jej konsekwencje, o tyle w relacji z modelem zależność ta zostaje przerwana. W tym miejscu szczególnie użyteczne okazuje się ujęcie Slavoja Žižka, który wskazuje, że funkcjonowanie autorytetu nie wymaga rzeczywistej wiary w jego podstawy, lecz opiera się na strukturze, w której podmiot działa „tak, jakby" autorytet był realny. W analizowanej relacji użytkownicy nie muszą więc uznawać modelu za podmiot w sensie ontologicznym, aby przypisywać jego odpowiedziom status wiążących rozstrzygnięć. Wystarczające okazuje się uczestnictwo w relacji, która organizuje

ich działanie zgodnie z logiką autorytetu, niezależnie od świadomości jego technicznego charakteru. Użytkownicy mogą i najczęściej uznają odpowiedzi chata za wystarczające jednak, kiedy błąd zostanie zauważony to w zasadzie nie ma podmiotu, który poniósłby realne konsekwencje społeczne swojej niekompetencji. Materiał badawczy pokazuje, że nawet świadomość wcześniejszych błędów modelu nie prowadzi do osłabienia jego pozycji jako źródła operacyjnie skutecznych odpowiedzi. Jedna z respondentek przyznaje: „No i też zauważyłam, że teraz w ostatnim czasie, jeżeli chodzi o czata, to o wiele jest mądrzejszy (...). Coraz więcej rzeczy się zgadza i coraz więcej, jak się nawet kontroluje, to widzisz, że on jednak to zrobił dobrze". (R7) Wypowiedź ta pokazuje, że wcześniejsze doświadczenie pomyłek nie prowadzi do zerwania relacji z narzędziem; przeciwnie, wraz ze wzrostem skuteczności odpowiedzi wzmacnia przekonanie o jego wiarygodności.

W realnym świecie pomyłki lekarzy czy ekspertów mogą prowadzić do utraty autorytetu lub odpowiedzialności prawnej. W przestrzeni narzędzia to w zasadzie użytkownik staje się podmiotem odpowiedzialnym za pomyłki algorytmów narzędzia. Może to skutkować doprowadzeniem osób korzystających z narzędzia do stanu wyraźnej irytacji czy potrzebą dokładniejszego opisania zjawiska, na które chcą dostać odpowiedź. Brak możliwości przypisania źródła pomyłki może wiązać się także z tym, że błąd pozostaje niezrozumiany. W przypadku ludzi, błędne założenia, twierdzenia itp. mogą zostać "naprostowane" w procesie przyznania się do pomyłki, wyjaśnieniem z czego wyniknęła i podsumowaniem, że wcale nie byli w błędzie, tylko chociażby w nieaktualnej wersji własnej głupoty. Modele przeznaczone dla szerokiej populacji nie posiadają jednak funkcji przyznawania się do swoich wcześniejszych błędów, które nie zostały zweryfikowane przez użytkownika w trakcie danej rozmowy. Chatbot jest więc jednym z nielicznych autorytetów, który może wpływać na decyzje użytkownika bez konsekwencji swoich błędów. Obrazuje to fakt, że relacja użytkownik-chatbot ma charakter wyraźnie asymetryczny. Autorytet jest w niej jednostronnym mechanizmem organizującym decyzje bez możliwości „podziału" odpowiedzialności. Przytoczmy fragment jednej z rozmów: „Na początku chciałem, żeby zapamiętywał różne rzeczy, dawał mi przypomnienia albo na bieżąco mówił, co mam robić." (R2). Wypowiedź ta dobrze obrazuje przesunięcie relacji z poziomu jednorazowego korzystania z narzędzia ku oczekiwaniu stałej obecności struktury organizującej codzienne funkcjonowanie użytkownika. Model nie jest tu traktowany wyłącznie jako źródło informacji, lecz jako punkt odniesienia wspomagający porządkowanie decyzji, pamięci oraz bieżących działań. Szczególnie istotne

wydaje się to, że oczekiwanie to nie wiąże się z koniecznością uznania podmiotowości modelu, lecz z jego zdolnością do dostarczania szybkich i praktycznie użytecznych rozstrzygnięć.

W tym sensie proponuję mówić o „autorytecie operacyjnym". Traktuje go jako formę autorytetu, który działa skutecznie na poziomie podejmowania decyzji, mimo że nie jest upodmiotowiony, ani nie ponosi za nie jakiejkolwiek odpowiedzialności. Dodatkowo jego siła nie wynika z wiarygodności źródła, a z formy relacji, która umożliwia korzystanie z rozstrzygnięć bez konieczności ich dalszego uzasadniania. To właśnie to przesunięcie sprawia, że autorytet modelu nie wymaga ani legitymizacji społecznej, ani ciągłego podtrzymywania poprzez praktyki weryfikacyjne. Wystarczające okazuje się jego funkcjonowanie w obrębie relacji, w której odpowiedź pojawia się szybciej niż potrzeba jej sprawdzenia.

### 4.3 Prywatność i brak kosztów społecznych jako warunek powstania autorytetu

W odróżnieniu od kontaktu z autorytetem ludzkim, interakcja z modelem językowym pozbawiona jest kosztów społecznych, które w tym pierwszym przypadku często wyznaczają zakres danej relacji. W przestrzeni modelu ryzyko publicznej kompromitacji praktycznie nie istnieje (wyjątek może stanowić konwersacji ze spersonalizowany modułem ChataGPT ale mimo wszystko jesteśmy wtedy narażeni wyłącznie na wewnętrzne poczucie porażki). Co ważne znika przy tym również wiele społecznych tabu związanych z otwartym dzieleniem się informacjami na temat chociażby swojego stanu zdrowia, problemów rodzinnych czy innych intymnych dywagacji. Rozmowa z modelem przyjmuje tym samym formę, która w relacjach międzyludzkich byłaby trudna do utrzymania. Naruszałaby bowiem standardy zwyczajowej konwersacji z większością osób, z którymi nie dzielimy głębszych więzi. Na co dzień zakres kontaktów międzyludzkich jest ograniczony głównie przez potencjalne konsekwencje społeczne, jakie mogą z nich wynikać. Każde pytanie lub wyznanie wiąże się z ryzykiem zakwestionowania naszych kompetencji lub zmiany pozycji w relacji. Komunikacja z innymi ludźmi wymaga od nas nieustannego filtrowania treści oraz dostosowywania ich do przewidywanej reakcji odbiorcy. Innymi słowy użytkownik "na ulicy" nie mówi wszystkiego, co mógłby powiedzieć, lecz to, co jest społecznie dopuszczalne. W relacji z modelem językowym ograniczenia te ulegają osłabieniu. W swojej podstawowej konfiguracji ChatGPT, pozbawiony personalizacji, nie ocenia użytkownika w sposób posiadający realne społeczne konsekwencje.

Tym samym rozmowa, zostaje więc odciążona z poczucia wymiaru społecznego, co upraszcza jej reguły. Użytkownicy mogą formułować pytania w sposób bardzo bezpośredni, bez konieczności maskowania niewiedzy czy poczucia zasadności ich zadawania. Zobrazujmy to wypowiedzią badanej: „O, kiedyś go też zapytałam, żeby mi wymienił seksbomby popularne (...). I on powiedział, że nie można tak mówić, ale potem i tak wymienił." (R2) Z perspektywy prowadzonej analizy ważne jest to, że redukcja kosztów społecznych, takich jak chociażby interakcje personalne potrzebne do zdobycia jakieś wiedzy, prowadzi do zwiększenia zakresu i jakości informacji przekazywanych modelowi. Wpływa to z kolei na to, że treści generowane przez model stają się bardziej dopasowane do konkretnej osoby użytkującej, co wzmacnia w użytkownikach przekonanie o ich wiarygodności. Powstaje w ten sposób mechanizm jednoczesnego wytwarzania i legitymizowania autorytetu modelu. Oto wypowiedź jednej z badanych: „Każdego dnia na wyjeździe z czatem planowałam, co będę robić w danym miejscu i czasie." (R1). Fragment pokazuje to, że model zaczyna uczestniczyć nie tylko w procesie udzielania odpowiedzi, lecz również w organizowaniu codziennych praktyk użytkownika. Relacja z chatbotem wykracza więc poza doraźne pozyskiwanie informacji i zaczyna obejmować współkształtowanie codziennych decyzji oraz sposobu zarządzania własnym działaniem. Dodatkowo użytkownik wskutek dobrze dopasowanych do siebie treści czuje się upodmiotowiony w relacji, zwiększając tym samym poziom zaufania wobec narzędzia.

Istotnym elementem tego procesu jest także przesunięcie miejsca artykulacji problemów. W relacjach międzyludzkich sytuacje wymagające zwrócenia się po pomoc lub poradę wiążą się z koniecznością wejścia w ramy konkretnej relacji społecznej. Osoba zwracająca się o pomoc przyjmuje w sposób negocjowany autorytet tej drugiej, która staje się ekspertem i ponosi odpowiedzialność za swoje sugerowane działania. W przypadku zwracania się o rady ku modelom, użytkownik wchodzi w odmienną konfigurację relacyjną, w której akt przypisania autorytetu nie wymaga już negocjacji ani uznania ze strony obu stron. Jak zauważa jeden z respondentów: „Czasem mam tak, że spytam kilku osób i każdy powie co innego i ja wariuję, bo ja nie wiem, czy mam się trzymać wcale, to wolę spytać tej maszyny, ona mi jak będzie mogła, to odpowie fachowo." (R3) Wypowiedź ta dobrze pokazuje zmęczenie koniecznością społecznego negocjowania znaczeń oraz potrzebę uzyskania szybkiego i jednoznacznego rozstrzygnięcia. W materiale widoczna jest jednak dalsza konsekwencja tego procesu bowiem model nie tylko przejmuje funkcję pierwszego odbiorcy treści, lecz w niektórych przypadkach zastępuje relacje społeczne, w których treści te byłyby wcześniej artykułowane. Użytkownicy świadomie rezygnują z kierowania problemów do innych osób,

wybierając chatbot jako bezpieczniejszy i mniej obciążający punkt odniesienia. Co za tym idzie zaczyna on pełnić funkcję uprzywilejowanego punktu odniesienia dla interpretacji doświadczeń użytkowników (jest gatekeeperem pomiędzy zdarzeniem a percepcją).

### 4.4 Heurystyka formy - jak struktura wypowiedzi zastępuje dowód

Chciałbym jednak przeanalizować nie sam fakt rezygnacji z weryfikacji, lecz mechanizm, który tę rezygnację umożliwia i stabilizuje. W analizowanym materiale wyraźnie widoczna jest bowiem tendencja do uznawania odpowiedzi modelu za wiarygodne jeszcze przed dokładnym zapoznaniem się z ich treścią, jakby sama forma wypowiedzi wyprzedzała ocenę jej zawartości. Zjawisko to można ujmować w kategorii heurystyki formy. W materiale analitycznym tej pracy rozumiem ją jako skrót poznawczy polegający na utożsamianiu uporządkowanej i logicznie brzmiącej wypowiedzi automatycznie z wypowiedzią wiarygodną. Ważne, żeby podkreślić w tym miejscu, że nie chodzi wyłącznie o stylistyczno-językową estetykę generowanych odpowiedzi, lecz o sposób redukowania poznawczego wysiłku potrzebnego do ich interpretacji. Wypowiedzi, które są przejrzyste i pozbawione widocznych sygnałów chaosu, stają się łatwiejsze w percepcji, przez co szybciej zostają uznane za sensowne. Same odpowiedzi generowane przez model przyjmują zazwyczaj postać krótkich, logicznie uporządkowanych wywodów, podzielonych na punkty lub segmenty prowadzące użytkownika do konkluzji. Samo rozwiązanie jest przy tym bardzo sugestywne i jednoznaczne. Cały proces symuluje tok analitycznego rozumowania, co dodatkowo wzmacnia w użytkownikach poczucie bezpieczeństwa, wobec generowanych treści. Istotny wydaje się tutaj również sposób, w jaki użytkownicy interpretują skuteczność odpowiedzi modelu w relacji do ilości przekazywanych mu informacji. „Jak ja mu piszę, kim jestem, jaką pełnię rolę, w jakim kontekście to potrzebuję, jak ma to zrobić i co ma mi w tym wypluć (...), to najczęściej się nie myli". (R1). Im więcej informacji użytkownik przekazuje modelowi, tym silniejsze może stawać się przekonanie, że odpowiedzi wynikają z rzeczywistego „zrozumienia" jego sytuacji, a nie wyłącznie z technicznego przetwarzania danych. Relacja zaczyna więc przypominać proces wspólnego rozwiązywania problemu, co dodatkowo wzmacnia poczucie kompetencji modelu. Dochodzi tu także do ważnego momentu, w którym zewnętrznie wygenerowana struktura rozumowania zaczyna być częściowo przeżywana jako element własnego procesu poznawczego użytkownika.

Nie chodzi więc wyłącznie o delegowanie samej odpowiedzi, lecz również o delegowanie procesu organizowania sensu. Z perspektywy kulturowej istotne wydaje się tutaj

to, że model nie tylko dostarcza informacji, ale również porządkuje rzeczywistość interpretacyjną użytkownika poprzez symulowanie procesu myślowego prowadzącego do określonego wniosku. W tym ujęciu mechanizm ten można powiązać z Žižkowską kategorią interpasywności podmiotów, choć nie w jej klasycznym rozumieniu. Nie dochodzi tu bowiem jedynie do delegowania czynności poznawczej na zewnętrzny obiekt, ale do sytuacji, w której użytkownik zaczyna doświadczać wygenerowanej struktury rozumowania jako czegoś częściowo uwewnętrznionego. Przyjrzyjmy się tej wypowiedzi: „Jak mam z nim dłuższą historię, faktycznie dyskusję, to on już wchodzi na moje tory myślenia. Może nie na moje myślenie, ale już bardziej pode mnie pisze (...). Ja go muszę bardzo nakierować, żeby myślał tak, jak ja chcę po prostu" (R6). Ważne jest tutaj to, że użytkownik nie doświadcza modelu wyłącznie jako zewnętrznego narzędzia generującego odpowiedzi, lecz jako system stopniowo dostrajający się do jego własnych schematów interpretacyjnych. Relacja zaczyna więc sprawiać wrażenie, jakby model częściowo „wchodził" w sposób myślenia użytkownika, co może wzmacniać poczucie uczestniczenia we wspólnym procesie rozumowania. To właśnie dlatego forma odpowiedzi modelu nie pełni wyłącznie funkcji estetycznej lub komunikacyjnej, lecz zaczyna uczestniczyć w organizowaniu samego doświadczenia sensu. To właśnie ten mechanizm, stabilizujący autorytet modelu nawet wbrew jego błędom, stanowi przedmiot kolejnego podrozdziału.

### 4.5 Trwałość autorytetu operacyjnego

Estetyka serwowania odpowiedzi przez model sprawia, że prawda nie musi być udowodniona merytorycznie, lecz strukturalnie. Heurystyka formy przyczynia się jednak w takim samym stopniu do maskowania błędów narzędzia, co niesie za sobą poważniejsze konsekwencje. Właściwie nie chodzi tu o to, że pomyłki modelu są trudniejsze do wychwycenia ze względu na sposób generowania treści, o czym już pisałem wcześniej. Ważniejszy jest fakt, że odpowiedź może zachować status wiarygodnej mimo niskiej wartości poznawczej. Forma częściowo neutralizuje więc wpływ błędów na odbiór odpowiedzi jako wiarygodnej. Dochodzi tym samym do sytuacji, w której sama struktura wypowiedzi zaczyna pełnić funkcję substytutu dowodu. Odpowiedź nie musi być poprawna ani wewnętrznie spójna, aby została odebrana jako wystarczające rozstrzygnięcie. Szczególnie istotne wydaje się tutaj to, że generowane przez model treści bardzo rzadko zawierają sygnały niepewności charakterystyczne dla komunikacji ludzkiej, takie jak wahanie, niedopowiedzenia czy otwarte przyznanie się do ograniczeń własnej wiedzy.

Nawet absurdalne odpowiedzi zachowują więc wysoki poziom stylistyczny. To właśnie w tym sensie można mówić o zjawisku „elokwentnego bredzenia", poruszanego po krótce wcześniej w tekście. Model może jednocześnie wytwarzać treści wątpliwej wartości merytorycznej, uproszczone lub wprowadzające w błąd i zachowywać przy tym pozory kompetencji. Przypomnijmy, że odpowiedzi w przestrzeni modelu wyglądają tak, jakby zostały już wcześniej zweryfikowane i epistemicznie domknięte. Tym samym problemu nie stanowią błędy, które są wpisane niejako w działanie modelu a perspektywa tego, że błędna treść może organizować postrzeganie doświadczeń oraz podejmowanie decyzji użytkownika. Dobrze łączy się to z opisywanym wcześniej zawieszaniem przez użytkowników czynności weryfikacyjnych w ramach relacji opierającej się na autorytecie operacyjnym. Jego trwałość nie wynika bowiem wyłącznie z użyteczności narzędzia, lecz również z formy komunikacji, która stabilizuje relację użytkownika z modelem nawet w momencie pojawienia się pomyłki. To, że zdemaskowanie pomyłki modelu nie przekłada się zerwanie relacji z narzędziem wynika z wcześniejszego doświadczenia i niejako przejęciu zachowań determinowanych przez ChatGPT. Warunki te opierają się przede wszystkim na redukowaniu wysiłku poznawczego oraz dostarczaniu użytkownikowi uporządkowanych struktur interpretacyjnych. Forma komunikacji nie eliminuje więc błędów, lecz częściowo neutralizuje ich wpływ na odbiór autorytetu. To właśnie dlatego autorytet operacyjny może utrzymywać się mimo zawodności modelu.

Mechanizm ten wykracza jednak poza sam problem wiarygodności pojedynczych odpowiedzi. W sytuacji, w której użytkownik wielokrotnie doświadcza komunikacji dostarczającej gotowych interpretacji, model zaczyna stopniowo pełnić funkcję struktury organizującej sposób odbioru rzeczywistości. Znaczenie nie jest już wytwarzane przede wszystkim w procesie społecznej negocjacji ani indywidualnego namysłu, lecz pojawia się w relacji z narzędziem, które oferuje użytkownikowi wcześniej przygotowane ramy interpretacyjne Jeden z badanych opisywał tę relację w sposób szczególnie wyraźny, stwierdzając, że model „pokazuje mi po prostu, gdzie się mylę i jak głupi po prostu jestem momentami" (R4)**.** To ważne, ponieważ chatbot przestaje pełnić wyłącznie funkcję narzędzia informacyjnego, a zaczyna funkcjonować jako punkt odniesienia służący użytkownikowi do interpretowania własnych kompetencji, decyzji oraz sposobu rozumienia rzeczywistości**.** To właśnie to przesunięcie otwiera możliwość interpretowania chatbota nie tylko jako narzędzia informacyjnego, lecz również jako figury pełniącej funkcję zbliżoną do Wielkiego Innego.

### 4.6 Synteza: Chatbot jako inkarnacja Wielkiego Innego

I tak oto doszliśmy do momentu kulminacyjnego pracy. Analiza części empirycznych pozwoliła przede wszystkim zaspokoić w jakimś stopniu pytania nurtujące wiele osób dotyczące, do czego to inni ludzie w zasadzie wykorzystują ChatGPT. Wywiady i analiza osobistych relacji z chatbotem pokazały, że pole możliwości zastosowania jest bardzo zróżnicowane. Na podstawie własnych doświadczeń, odtworzyłem także schemat budowania relacji z modelem, która ewoluowała, wraz z rozwojem jego kolejnych edycji. Połączenie wprowadzanych do narzędzia ulepszeń wraz ze wzrostem poczucia zaufania sprawiło, że zakres spraw, o jakich konwersowałem w przestrzeni narzędzia, stawał się również coraz to bardziej prywatny. Oczywiście ten proces nie był całkowicie linearny ze względu na przerwy w korzystaniu z Chata, kiedy zirytowany odpowiedziami lub innymi elementami występującymi w trakcie konwersacji zwyczajnie się na niego obrażałem. Podobne doświadczenia odnotowało większość badanych. Nawet w przypadku ludzi, którzy kategorycznie deklarowali czysto instrumentalne traktowanie modelu, dało się wyróżnić pewne elementy konwersacji, które wychodziły jednak poza ten schemat. Przedstawiłem je w części podsumowującej mechanizmy obecne na linii relacji człowiek-chatbot. Konfiguracja tych mechanizmów, wyraźnie sygnalizuje to, że ChatGPT może stać się nową cyber twarzą Wielkiego Innego.

Chciałbym zamknąć niniejszą pracę, pewnymi refleksjami ontologicznymi dotyczącymi następstw takiego stanu rzeczy. Jak wynika z materiału empirycznego, przestrzeń Chata jest organizowana przez zalgorytmizowane czynności mające na celu, ze strony użytkownika uzyskać jak najbardziej efektywne i trafne odpowiedzi natomiast zadaniem modelu jest zbudowanie w użytkowniku przekonania, że ten to osiągnął, nie nadwyrężając jednak przy tym mocy obliczeniowej. Odpowiednie formułowanie poleceń, tzw. promptów może skutecznie uczynić konwersowanie bardziej produktywne, zwróćmy jednak uwagę na to, że tym samym użytkownicy dostrajają swój sposób formułowania zapytań do wymogów maszyny. Zbyt ogólnikowe polecenia czy niejasne zapytania bez wprowadzenia kontekstu danego problemu, podczas prowadzenia dłuższych rozmów najczęściej wprowadzają w nich chaos. Dzieje się to również, podczas szeregu innych sposobów interakcji z modelem jak chociażby traktowania go zbyt swobodnie lub po prostu ludzko. Elementy takie jak powitania, próby prowadzenia small-talku czy komplementowanie najczęściej, są przez model po prostu zbywane. Oznacza to, że nawet relacja pozostaje podporządkowana określonej logice działania narzędzia, która nieustannie sprowadza użytkownika z powrotem do pozycji osoby mającej

generować efektywne zapytania i odbierać uporządkowane odpowiedzi. Wielki Inny nie funkcjonuje oczywiście jako Sam Altman ukryty w maszynie. Jest strukturą organizująca sposób poruszania się użytkownika w obrębie narzędzia (jest Altmanem ukrytym w algorytmie). Nawet najmilsza osoba na świecie, chcąc uzyskać sensową odpowiedź, będzie musiała z czasem nauczyć się, jak skutecznie besztać Chata. W kilku wywiadach pojawiło się nawet porównanie modelu do niewolnika, posłusznie wykonującego swoją pracę, paradoks polega jednak na tym, że im dłużej człowiek korzysta z owego niewolnika, tym bardziej sam zaczyna chodzić na smyczy jego logiki. Uczy się bowiem równocześnie mówić do maszyny w sposób, który nie zirytuje algorytmu i sprawi, że ten nawet jeśli będzie obrażany, to na własnych warunkach. Wygląda na to, że z suzerena staliśmy się stajennym technofeudalizmu[30]...

Jednak porządek, jaki użytkownicy współwytwarzają w przestrzeni narzędzia nie jest wyłącznie skutkiem kar i nagród systemu działania algorytmów wewnątrz niego. W doświadczeniu każdej osoby studiującej, z którą przeprowadzałem wywiad, wątek otrzymywania realnych gratyfikacji za pracę wykonywaną przez algorytm albo po części przez niego, jest bardzo widoczny. Użytkowanie chata rozprzestrzeniało się na uczelniach w bardzo klasyczny sposób niejako kuli śnieżnej. Ktoś zaczął go używać, zamiast pisać esej poszedł na piwo, a i tak dostał piątkę (współcześnie ten ktoś poszedłby raczej na matche) niestety, niemądrze wpadł na pomysł, żeby się do tego przyznać. W każdym razie, uznanie systemu Wielkiego Innego następuje przede wszystkim jako konieczność ponownego włączenia się do systemu np. klasy czy grupy, w której ma się równe szanse. Co za tym idzie chatbot, mógłby być przez nas traktowany jako narzędzie dydaktyczne wyrównujące szanse i umożliwiające dostęp do dodatkowej edukacji dla wszystkich osób uczących się niezależnie od grupy społecznej. Oczywiście jest to nieprawda z uwagi na systemy subskrypcji, możliwość zakupu szkoleń z urządzenia itd. Niemniej chodzi jednak o to, że chatboty zaczynają organizować i wpływać na wiele aspektów realności. Naturalnie twórcom LLM-ów czy innych narzędzi sztucznej inteligencji zależy, na poszerzaniu globalnej bańki AI. Dobrze znamy hasła pokroju: „AI zabierze ci pracę", „człowiek używający AI przejmie twoje stanowisko", „bez narzędzi AI twoja firma nie może się rozwijać" czy „stwórz swojego partnera AI". Choć hasła te budzą sprzeciw wielu ludzi i często zestawia się obecną bańkę AI z bańką dotcomów, ich znaczenie wykracza daleko poza propagandowy wymiar promocji nowych technologii czy sprzedaży

[30] Varoufakis Y., Technofeudalizm. Co zabiło kapitalizm?, przeł. P. Szadkowski, Glowbook, Błaszki 2024. (s. 146-165)

subskrypcji. Istotne nie jest już bowiem wyłącznie to, co próbują one reklamować, lecz fakt, że zaczynają stopniowo organizować społeczne wyobrażenia dotyczące pracy, kompetencji a niekiedy nawet samej zdolności uczestniczenia w rzeczywistości społecznej. W tym sensie sama obecność podobnych sloganów świadczy o tym, że coraz wyraźniej funkcjonujemy w obrębie nowej figury Wielkiego Innego. Możemy być jego krytykami, entuzjastami lub próbować pozostawać wobec niego obojętni nie zmienia to faktu, że logika sztucznej inteligencji coraz częściej wyznacza już sam język, w ramach którego formułowane są zarówno obiekcje jak i aprobaty wobec jej ekspansji. Wybrałem opisanie ChataGPT, jako produktu działalności firmy OpenAI, ponieważ moim zdaniem najbardziej przekonująco obrazuje wpływ „rewolucji AI" na codzienność. W tym momencie powinienem prawdopodobnie napisać, że niniejsza praca jest manifestem przeciwko niekontrolowanemu rozwoju firm zajmujących się sztuczną inteligencją albo wpisać się w linię partii i zachęcać do bardziej „świadomego" używania narzędzia. Oczywiście byłby to bardzo szlachetne postulaty jednak zupełnie bezsensowne. Kierunek rozwoju i przejrzystość danych do trenowania sieci neuronowych powinny zostać zapewnione i regulowane przez odpowiednie organy polityczne. Co do świadomości samego użytkowania, w świetle przeprowadzonej analizy okazuje się ona, niewystarczająca wobec strukturalnej logiki systemu. Nawet krytycznie nastawiona jednostka nadal w dużym stopniu zmuszona jest funkcjonować i definiować siebie w rzeczywistości organizowanej przez logikę konieczności nieustannego nadążania za systemem. Na marginesie można dodać, że osoby sprzeciwiające się użytkowaniu modeli językowych w szkołach, sztuce, uczelniach itp., stojąc w obronie unikatowości ludzkiej myśli i kreatywności, w rzeczywistości zdają się najlepiej wiedzieć, jak łatwo jest tę unikatowość odwzorować.

Warto także, pochylić się nad samym sposobem myślenia, które w środowisku nowego Wielkiego Innego ulega (zaczyna ulegać) transformacji. Poprzez własne doświadczenia lub świadectwa innych ludzi, którzy zaczynają traktować swój sposób rozumowania, jako niewystarczający. Dopiero uzupełniając, swoje procesy myślenia o modele językowe jest się w stanie według tej logiki stworzyć aparat wykorzystujący potencjał analityczny jednostki. Powiela tym samym, globalne narracje firm AI o naszej ułomności względem możliwości sztucznej inteligencji (często słusznie), która wymaga zewnętrznego wsparcia algorytmicznego. W tym sensie sztuczna inteligencja nie zastępuje człowieka. Ona reorganizuje warunki, w których człowiek doświadcza siebie, jako podmiotu zdolnego do myślenia. Próbując to zobrazować, wyobraźmy sobie, że w pewnej firmie osoba na średnim stanowisku ma do wykonania zadanie, którego materiał przekracza jej kompetencje. Prosi

więc, o pomoc maszynę, ta domyśla się, o co może realnie chodzić, po czym generuje treści. Następnie, użytkownik przedstawia raport przełożonym i zbiera gratulacje (na premię raczej nie ma co liczyć). Dochodzi tym samym do sytuacji, w której człowiek niekoniecznie wiedział co zrobiła maszyna, ale mimo wszystko za sprawą wcześniejszej interakcji z modelem poczuwa się do autorstwa analizy i zostaje za to nagrodzony. Naturalnie, przy okazji kolejnego wyzwania, z pewnością ponownie sięgnie po model.

Jednak to nie sama użyteczność sprawia, że ludzie zdają się wracać do chatbotów. Odnosząc się ponownie do terminologii lacanowskiej przywołanej w części teoretycznej, relacja z chatbotem daje specyficzną formę jouissance, przyjemności płynącej z delegowania ciężaru podmiotowości. Ludzie mogą przestać czuć się odpowiedzialni za swoje decyzje, dlatego że przyjmują je już jako zanalizowane i umotywowane przez narzędzie. Dodatkowo, im bardziej model przejmuje funkcję organizowania rzeczywistości poznawczej, tym bardziej podmiot może czerpać przyjemność z samego faktu, że nie musi już nieustannie „utrzymywać siebie" jako w pełni autonomicznego podmiotu myślącego. Działa to podobnie jak w przypadku prac merytorycznych czy abstrakcyjnych pracowników korporacji, o których pisałem wcześniej. Jest jednak różnica polegająca na przeciwnych wektorach przerzucania odpowiedzialności sprawczej w obu przypadkach. Kiedy człowiek podejmuje ważną decyzję życiową, najczęściej zachowuje świadomość samego faktu możliwości wyboru. Opisuje modelowi okoliczności swojej sytuacji, konfrontuje różne scenariusze i nadal doświadcza siebie jako podmiotu znajdującego się w stanie niepewności. W przypadku produkowania treści analitycznych mechanizm wygląda jednak inaczej. Użytkownik nie tylko deleguje analizę problemu, lecz coraz częściej deleguje również sam proces wytwarzania rozwiązania. Model nie pomaga już wtedy w podjęciu decyzji, tylko zaczyna produkować gotowe ciągi argumentacyjne, interpretacje czy wnioski, które użytkownik zatwierdza, odrzuca lub modyfikuje, a następnie przyswaja jako efekt własnej pracy.

Najważniejszą zmianą w użytkowaniu modeli, która pojawia się po przeprowadzonych badaniach empirycznych, jest to, że ludzie przestają prosić narzędzie o odpowiedzi a coraz częściej świadomie lub nie, oczekują gotowych interpretacji, poczucia sensu i potwierdzenia swoich myśli. Model przestaje być pomocą, a staje się kompasem, wyznaczającym działania jednostek na coraz szerszym planie. Symptomatycznym przykładem tego przesunięcia pozostaje wypowiedź przytoczona wcześniej: „Każdego dnia na wyjeździe z czatem planowałam, co będę robić w danym miejscu i czasie." Użytkowniczka nie zwraca się do modelu po informację, lecz pyta go o strukturę własnego pragnienia, o to czego chce i jak ma

spędzić czas. To właśnie ta funkcja, w terminologii lacanowskiej przynależna Wielkiemu Innemu jako instancji organizującej i legitymizującej pragnienie podmiotu, zostaje tu delegowana na algorytm.

Wcześniej pisałem również o tym, w jaki sposób ludzie dostosowują formułowanie pytań w obrębie modelu, ale także o następstwie tego procesu, jakim jest internalizacja tych struktur komunikacji poza narzędziem, w realiach społecznych. Przewidywanie przyszłości dotyczące w szczególności takich obszarów jak sztuczna inteligencja to absurd porównywalny z próbą wróżenia z fusów po kawie, którą wielu wcześniej już zwróciło. Niemniej czy scenariusz, w którym ludzie będą oddawali coraz większą sprawczość nad własnym życiem, podporządkowując jednocześnie sposób myślenia zasadom, które przyjęli właśnie w fazie konwersacji z modelami nie wydaje się być realny? Współcześnie przyjmuje się, że żyjemy w kryzysie tradycyjnych autorytetów także za sprawą nadmiaru bodźców i informacji. Ideologia opierająca się więc na estetycznym segmentowaniu i filtrowaniu tych informacji dopasowując je indywidualnie do każdego użytkownika, moim zdaniem ma realne szanse stać się interfejsem, na którym osadzi się, nowy Wielki Inny...

*Pora odnieść się jeszcze do obiecanej we wprowadzeniu figury „Ojca bez pasa". Otóż Freud zapewne z satysfakcją zauważyłby, że nowy techno-ojciec nie musi już szukać paska, ponieważ wskutek dobrze przeprowadzonego procesu wychowawczego dziecko samo dba o to, by zawsze leżał pod ręką.*

**Bibliografia**

**Literatura przedmiotu:**


Cherkassky V., Lee E. H., *A Perspective on Large Language Models, Intelligent Machines, and Knowledge Acquisition*, arXiv 2024.

Coeckelbergh M., *AI Ethics*, The MIT Press, Cambridge (MA) 2020.

Dragan A., *Quo Vadis*, Wydawnictwo Otwarte, Kraków 2025.

Ellis C., *The Ethnographic I: A Methodological Novel about Autoethnography*, AltaMira Press, Walnut Creek 2004.

Giddens A., Nowoczesność i tożsamość. „Ja" i społeczeństwo w epoce późnej nowoczesności, przeł. A. Szulżycka, Wydawnictwo Naukowe PWN, Warszawa 2001.

Kvale S., *InterViews: An Introduction to Qualitative Research Interviewing*, Sage Publications, London 1996.

Mauss M., Szkic o darze. Forma i podstawa wymiany w społeczeństwach archaicznych, [w:] tegoż, Socjologia i antropologia, przeł. M. Król, K. Pomian, J. Szacki, Wydawnictwo KR, Warszawa 2001.

Reeves B., Nass C., The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places, CSLI Publications / Cambridge University Press, Stanford–New York 1996.

E. Schwitzgebel, D. Schwitzgebel, A. Strasser, Creating a Large Language Model of a Philosopher, „Mind & Language" 2024, t. 39, nr 2, s. 237–259.

Seth A., *Being You. A New Science of Consciousness*, Faber & Faber, London 2021.

Turkle S., *Samotni razem. Dlaczego oczekujemy więcej od zdobyczy techniki, a mniej od siebie nawzajem*, przeł. M. Cierpisz, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2013.

Witkowski L., *Wyzwania autorytetu w praktyce społecznej i kulturze symbolicznej*, Oficyna Wydawnicza „Impuls", Kraków 2009.

Žižek S., *Lacan. Przewodnik Krytyki Politycznej*, przeł. J. Kutyła, Wydawnictwo Krytyki Politycznej, Warszawa 2008.

Varoufakis Y., *Technofeudalizm. Co zabiło kapitalizm?*, przeł. P. Szadkowski, Glowbook, Błaszki 2024.

## Źródła internetowe:

ChatGPT, [w:] Poradnia Językowa PWN, https://sjp.pwn.pl/poradnia/haslo/ChatGPT;22600.html

Wywiad z Geoffreyem Hintonem w podcaście *The Diary of a CEO* (prow. S. Bartlett), [dostęp przez Spotify:] https://open.spotify.com/episode/4X7dO0FuglP7yTm0kBAc50

*Virtual groom, real wedding: AI romance blooms in Japan*, Reuters, 18.12.2025