

Nowoczesne Systemy Zarządzania
Zeszyt 20 (2025), nr 2 (kwiecień-czerwiec)
ISSN 1896-9380, s. 89-110
DOI: 10.37055/nsz/209421

Modern Management Systems
Volume 20 (2025), No. 2 (April-June)
ISSN 1896-9380, pp. 89-110
DOI: 10.37055/nsz/209421

Instytut Organizacji i Zarządzania
Wydział Bezpieczeństwa, Logistyki i Zarządzania
Wojskowa Akademia Techniczna
w Warszawie

Institute of Organization and Management
Faculty of Security, Logistics and Management
Military University of Technology
in Warsaw



Test Turinga w badaniu zaawansowania interakcji człowieka i ChatGPT – eksperyment akademicki

Turing test in examining the advancement of human – ChatGPT interaction: an academic experiment

Grzegorz Kunikowski

Politechnika Warszawska, Polska
grzegorz.kunikowski@pw.edu.pl; ORCID: 0000-0002-3660-7117

Patryk Kornacki

Politechnika Warszawska, Polska
patryk.kornacki.stud@pw.edu.pl; ORCID: 0009-0001-2887-4856

Katarzyna Milczarska

Politechnika Warszawska, Polska
katarzyna.milczarska.stud@pw.edu.pl; ORCID: 0009-0004-8233-7236

Paulina Kwiatkowska

Politechnika Warszawska, Polska
paulina.kwiatkowska3.stud@pw.edu.pl; ORCID: 0009-0005-0130-0480

Abstrakt:

Cel badań i hipotezy/pytania badawcze

Celem jest ocena zdolności systemu ChatGPT 4.0 do naśladowania człowieka prowadzącego konwersację oraz refleksja z przeprowadzonego eksperymentu. Na potrzeby badania sformułowano dwie hipotezy. W pierwszej przyjęto, że ChatGPT 4.0 potrafi skutecznie naśladować człowieka podczas prowadzenia konwersacji na zadany temat. Druga hipoteza dotyczyła wpływu formułowania odpowiedzi przez uczestników eksperymentu na ocenę zdolności ChatGPT 4.0 do przejścia testu Turinga.

Metody badawcze

Eksperyment polegał na przeprowadzeniu trzech interaktywnych rozmów i głosowania, w których jednym z rozmówców był ChatGPT, oraz zebraniu refleksji w ankiecie ewaluacyjnej. Wyniki eksperymentu potwierdziły hipotezę, że bot potrafi skutecznie naśladować człowieka. Potwierdzenie należy potraktować jako stwierdzenie dojrzałości technologii generatywnej.

Główne wyniki

Na podstawie przeprowadzonego testu Turinga można dojść do konkluzji, że maszyna spełnia normy oszukiwania człowieka powyżej 30% zgodnie z opracowaną metodą Turinga. O ile dla pierwszego celu eksperyment wykazał pozytywną odpowiedź na hipotezę, czyli ChatGPT potrafi skutecznie naśladować człowieka podczas prowadzenia konwersacji na zadany temat, o tyle sama metoda wydaje się niedostosowana do współczesnych potrzeb. Stanowi bardzo dobrą inspirację dla współczesnych badaczy. Niezależnie od jej twórcy Alana Turinga metoda byłaby stosowana być może w innej formie, co nie zaprzecza inwencji autora sformułowanej zasady, której idea po 75 latach pozostaje aktualna. Wartością testu Turinga jest przewidzenie potrzeb i możliwości technologii cyfrowych w latach 40. XX wieku, czyli w erze analogowej.

Implikacje dla teorii i praktyki

Badanie wykazało dojrzałość modelu językowego oraz było okazją do refleksji nad wykorzystaniem technologii generatywnych w edukacji. Ostrzeżeniem może być zidentyfikowany problem weryfikacji wiedzy podczas egzaminów, o czym dyskutuje się w środowisku akademickim, a co zapewne przyczyni się do szybszej absorpcji rozwiązań generatywnych w edukacji.

Słowa kluczowe: edukacja, sztuczna inteligencja, ChatGPT, test Turinga, technologie generatywne

Abstract:

Research objectives and hypothesis/research questions

The aim is to assess the ability of the ChatGPT 4.0 system to imitate a human having a conversation and to reflect on the conducted experiment. Two hypotheses were formulated for the purposes of the study. The first assumed that ChatGPT 4.0 can effectively imitate a human while having a conversation on a given topic. The second hypothesis concerned the influence of the participants' answers on the assessment of ChatGPT 4.0's ability to pass the Turing test.

Research methods

The experiment consisted of conducting three interactive conversations and voting, in which one of the interlocutors was ChatGPT, and collecting reflections in an evaluation survey. The results of the experiment confirmed the hypothesis that the bot can effectively imitate a human. Confirmation should be treated as a statement of the maturity of generative technology.

Main results

Based on the conducted Turing test, it can be concluded that the machine meets the standards of deceiving a human by more than 30% in accordance with the developed Turing method. While for the first goal the experiment showed a positive answer to the hypothesis, i.e. ChatGPT can effectively imitate a human during a conversation on a given topic, the method itself seems to be unsuited to contemporary needs. It is a very good inspiration for contemporary researchers. Regardless of its creator, A. Turing, it would probably be used in a different form, which does not deny the inventiveness of the author, who formulated the principle, the idea of which remains relevant after 75 years. The value of the Turing test is to predict the needs and possibilities of digital technologies in the 1940s, i.e. in the analog era.

Implications for theory and practice

The study showed the maturity of the language model and was an opportunity to reflect on the use of generative technologies in education. A warning may be the identified problem of knowledge verification during exams, which is being discussed in the academic community, which will probably contribute to the faster absorption of generative solutions in education.

Keywords: education, artificial intelligence, Turing test, ChatGPT, generative technologies

Wprowadzenie

W artykule przedstawiono wyniki eksperymentu, którego celem było zbadanie zdolności sztucznej inteligencji – Generative Pre-trained Transformer (ChatGPT) do naśladowania człowieka prowadzącego konwersację na zadany temat. Jest to model języka naturalnego opracowany przez firmę OpenAI. Eksperyment prowadzono z wykorzystaniem bezpłatnej wersji GPT-4.0. Badanie zostało zaplanowane i przeprowadzone w konwencji testu Turinga w ramach zajęć na studiach menedżerskich prowadzonych na Wydziale Zarządzania Politechniki Warszawskiej.

1. Cel artykułu i kontekst eksperymentu

Test Turinga jest postrzegany jako miara zdolności maszyn do myślenia. Koncepcję testu przedstawił Alan Turing w artykule pt. *Maszyny obliczeniowe a inteligencja* (Turing, 1950). Postawił w nim pytanie: Czy maszyny mogą myśleć? Uzasadniając, że pojęcie myślenia jest wieloznaczne, zaproponował eksperyment nazwany „grą w imitację” i opracował rodzaj trzyosobowej gry, w której zachodzi interakcja człowiek – maszyna. W grze tej sędzia prowadzi rozmowę z człowiekiem i komputerem (maszyną), nie wiedząc, kto jest kim (Turing, 1950).

Przypisuje się, że A. Turing zidentyfikował problem, który współcześnie został przypisany zagadnieniu dotyczącemu systemów sztucznej inteligencji i zasugerował zerojedynkową metodę oceny ich zaawansowania. Zaproponowana koncepcja została uznana jako metoda i w literaturze przedmiotu określana jest testem Turinga. W okresie rozwoju technologii informacyjnych i botów test Turinga może być kryterium czy też sposobem oceny wiarygodności programów zaprojektowanych do prowadzenia rozmów z człowiekiem. W tej konwencji zaczęto testowanie botów: „Pierwsze próby prostej konwersacji pokazywały, że poprzeczka umieszczona jest wysoko. Maszyny podejmowały dialog, ale szybko okazywało się, iż nie są w stanie sprostać wyzwaniu. Okazało się, iż do zdania testu droga jest jeszcze daleka. Prowadzono testy z programem służącym do konwersacji o nazwie Eliza. Później pojawił się Cleverbot. (...) Przyjmuje się, iż test pomyślnie przeszedł algorytm o imieniu Eugene Goostman w 2014 roku, udający nastolatka. Przekonał do siebie 1/3 sędziów” (NCBR, 2020).

Celem artykułu, oprócz oceny zaawansowania ChatGPT, jest również próba odpowiedzi na pytanie: czy test Turinga jest skuteczną metodą badania zaawansowania interakcji człowiek – współczesna maszyna i możliwe implikacje? Podczas zajęć przeprowadzono eksperyment, w którym grupa studentów prowadziła konwersację m.in. z ChatGPT 4.0 na temat wpływu sztucznej inteligencji na rynek pracy.

2. Test Turinga i jego adaptacja w badaniu

W literaturze przedmiotu można znaleźć wiele zastosowań testu Turinga w ocenach skuteczności prowadzenia komunikacji z udziałem systemów wykorzystujących zaawansowane modele językowe. Test Turinga zdała maszyna, a właściwie bot o imieniu Eugene Goostman, przekonując sędziów, że jest człowiekiem (NCBR, 2020; Pacelt, 2020).

W roku 2023 prestiżowy zespół uniwersytetu w Stanford odnotował, że ChatGPT zdał test, a w rezultacie zaistniała potrzeba ustalenia innych sposobów oceny sztucznej inteligencji (Biever, 2023). Kolejne wersje ChatGPT pozytywnie przechodziły testy (Scott, 2024), a naukowcy dokładniej badali specyfikę interakcji systemu w konwersacji z ludźmi (np. Mei, Xie, Yuan, Jackson, 2024).

W krajowej literaturze przedmiotu niewiele publikacji dotyczy zastosowania testu Turinga. Znaleźliśmy przykład przywołania testu w kontekście oceny automatycznego przekładu poezji (Studzińska, 2020). Autorka stawia pytanie, czy możliwe jest, by przekład wiersza wyglądający na dzieło ludzkie był w rzeczywistości wytworem komputera, a więc czy programy, takie jak Google Translate, DeepL i Microsoft Bing, są w stanie w wyjątkowych przypadkach stworzyć przykład, który mógłby uchodzić za wykonany przez człowieka (Studzińska, 2020, s. 299-300). Po przeprowadzaniu eksperymentów z tłumaczeniem fragmentów wierszy autorka dochodzi do wniosku, że „obecnie efekt artystyczny – jeśli jakiś się pojawia – jest dziełem przypadku”. Autorka uwzględniając rozwój algorytmów, które działają obecnie głównie w paradygmacie tłumaczenia statystycznego, zastanawia się, czy systemy mogą „rozumieć” poezję, a następnie pozostawia czytelników z retorycznym pytaniem: czy w przyszłości literatura piękna będzie tłumaczona maszynowo?

Andrzej P. Wierzbicki na temat skuteczności testu Turinga pisze, że „test Turinga okazał się absolutnie zawodnym, bowiem już w końcu lat 60. XX wieku Joseph Weizenbaum (zob. np. Weizenbaum, 1976) skonstruował program ELIZA, który potrafił prowadzić konwersację nieodróżnialną od człowieka, tyle że na poziomie niezbyt mądrego nastolatka (czy nastolatki), powtarzając elementy zapytań z wykorzystaniem kilku reguł gramatycznych przetwarzania zdań” (Wierzbicki, 2018, s. 62).

Najpopularniejsza obecnie wersja, tj. ChatGPT 4.0, nie jest najnowszym modelem, bo jest już kolejna wersja 4.5. Model został przeszkolony na ogromnych zbiorach danych tekstowych bez bezpośredniego nadzoru, co pozwala mu „samodzielnie” odkrywać struktury językowe, relacje między słowami i konteksty. Dzięki temu GPT 4.5 zyskuje bardziej intuicyjne pojmowanie niuansów językowych, co przekłada się na lepsze rozumienie treści oraz intencji użytkownika (Kapuściński, 2025).

Współczesne zastosowania testu Turinga przedstawiają również inni badacze (Gams, Kramar, 2024; Pinar Saygin, Cicekli, Akman, 2000). I tak, Ayse Pinar Saygin, Ilyas Cicekli i Varol Akman (2000) uważają, że test Turinga będzie miał znaczenie, pomimo kontrowersji. W literaturze przedmiotu oprócz zaawansowania systemów

AI badane jest zagadnienie ich natury. Matjaz Gams oraz Sebastjan Kramar (2024) po ponad 20 latach dochodzą do wniosku, że ChatGPT pomimo znaczącego postępu nie osiąga poziomu świadomości. Autorzy przywołują teorię zintegrowanej informacji, sformułowaną przez Giulio Tononiego (ang. *Integrative Information Theory*). Autor teorii zintegrowanej informacji dowodzi, że świadomość wynika z integracji informacji, a jej poziom można mierzyć (Tononi, 2012). Formułuje pięć aksjomatów świadomości, które można stosować jako kryterium uznania, że sztuczny system posiada świadomość. Jest to odmienne podejście do oceny zaawansowania systemów sztucznej inteligencji. Według innych badaczy narzędzia wykorzystujące model języka naturalnego udostępniają więcej informacji niż człowiek, nie posiadają jednak cech świadomości, chociaż w przyszłości może się to zmienić (Gams, Kramar, 2024). Póki co pozostają systemami informacyjnymi, które skutecznie potrafią imitować człowieka, tym samym zdając test Turinga. Ograniczenia ludzkie autorzy nazywają „naiwnością rozmówców”.

Badanie przeprowadzone na Uniwersytecie Stanford nad NLP wykazały, że ChatGPT 4.0 posiada cechy charakterystyczne dla postaw ludzkich, takich jak skłonność do współpracy, co w poprzedniej wersji nie było oczywiste (Scott, 2024). ChatGPT 4.0 rozwinął możliwości i przeszedł rygorystyczne testy Turinga, wykazując się wysokimi zdolnościami wchodzenia w złożone interakcje człowiek – maszyna. Luciano Floridi i Massimo Chiriatti zachęcają do wykorzystywania technologii AI i włączania jej do działań jako nieuchronnej rzeczywistości, która już ma miejsce na świecie, np. publikowanie prasowych komentarzy w czasie rzeczywistym (Floridi, Chiriatti, 2020). Do podobnych wniosków doszedł Ricardo Restrepo Echavarría (Restrepo Echavarría, 2025). Według uczestników eksperymentu ChatGPT był identyfikowany, ponieważ reagował „bardziej inteligentnie niż człowiek”. Alexandra Farazouli, Teresa Cerratto-Pargman, Klara Bolander-Laksov i Cormac McGrath po zbadaniu wyników ocen prac doszli do wniosku, że technologia NLP będzie przełomowa w praktyce szkolnictwa wyższego (Farazouli, Cerratto-Pargman, Bolander-Laksov, McGrath, 2024). Z drugiej strony problem nieetycznego wykorzystania technologii NLP podczas egzaminów i trudności w zidentyfikowaniu takich praktyk omawia (Scarfe, Watcham, Clarke, Roesch, 2024). Problem jest w Polsce opisywany w alarmistycznym tonie: „362 studentów psychologii na uniwersytecie we włoskim mieście Ferrara musi powtórzyć egzamin, którego wyniki zostały anulowane, gdy profesorowie odkryli, że wielu zdających wykorzystało sztuczną inteligencję, odpowiadając na pytania w teście. Podejrzenia wzbudziła bardzo wysoka średnia ocen: 28 punktów na 30” (Kozłowski, 2025).

Testy Turinga przechodziły rozwiązania NLP w zastosowaniach praktycznych i komercyjnych. Na przykład w Danii badano jakość odpowiedzi udzielanych diabetikom (Hulman, Dollerup, Mortensen et al., 2023). Ponad połowa badanych (59,5%) zidentyfikowała odpowiedzi wygenerowane przez ChatGPT, co nie stanowiło przeszkody w uznaniu merytorycznej wartości odpowiedzi. Karen P.-S. Tan, Yi V.

Liu i Stephen W. Litvin przeprowadzili ogólniejsze badania dotyczące oceny reakcji klientów na komentarze wytwarzane przez generatywną AI. Wnioski były niejednoznaczne, z jednej strony uczestnicy nie identyfikowali autorstwa jednoznacznie, z drugiej strony – poinformowani o źródle – nie oceniali tego pozytywnie (Tan, Liu, Litvin, 2025). Świadczy to o społecznych barierach, które będą stanowiły wyzwanie dla tworzenia modeli interakcji człowiek – maszyna. Czy maszyna ma być źródłem wiedzy, czy towarzyszem i powiernikiem?

W kontekście zastosowań AI w szkolnictwie wyższym warto przytoczyć pracę autorstwa Andrzeja Wodeckiego. Autor analizuje możliwości wykorzystania ChatGPT jako technologii generatywnej w nauczaniu. Formułuje wnioski, że technologie generatywne mogą przyczynić się do „podziału kompetencyjnego społeczeństwa”, czyli wzmocnienia potencjału profesjonalistów oraz obniżenia umiejętności osób nadużywających tych rozwiązań. Dostrzega również zagrożenie nazwane „ułudą kompetencji”, które polega na przecenieniu przez użytkowników (w tym przypadku programistów) własnych możliwości wspieranych asystentami (Wodecki, 2023).

3. Hipotezy badawcze

Na podstawie analizy dotychczasowych zastosowań i badań nad testem Turinga sformułowano następujące hipotezy badawcze:

Hipoteza 1: ChatGPT potrafi skutecznie naśladować człowieka podczas prowadzenia konwersacji na zadany temat.

Hipoteza 2: Formułowanie odpowiedzi przez uczestników eksperymentu ma wpływ na ocenę zdolności ChatGPT do przejścia testu Turinga.

Hipotezy zostaną zweryfikowane na podstawie wyników eksperymentu badawczego, polegającego na przeprowadzeniu trzech następujących po sobie konwersacji pomiędzy grupą studentów a nieznanym im rozmówcą.

4. Metoda eksperymentu

W eksperymencie zastosowano modyfikację klasycznego testu Turinga. Uczestnicy rozmawiali z trzema rozmówcami w następujących po sobie niezależnych konwersacjach, nie wiedząc, który z rozmówców to ChatGPT. Przyjętym tematem był „wpływ sztucznej inteligencji na rynek pracy”. W badaniu uczestniczyło 15 studentów kierunku menedżerskiego studiów drugiego stopnia na Wydziale Zarządzania Politechniki Warszawskiej.

W eksperymencie:

- poproszono 13 osób o pełnienie roli „sędziego” (n = 13) w trzech krótkich konwersacjach, ale nie wiadomo, który z rozmówców jest „komputerem”;
- w rolę „człowieka” wcieliły się dwie osoby z tej samej grupy studenckiej;
- ChatGPT 4.0 był w eksperymencie „komputerem”;
- prowadzący zajęcia pełnił funkcję moderatora, który obsługiwał interfejs komunikacyjny, pomiędzy „sędziami” i „komputerem”.

W trakcie eksperymentu uczestnicy prowadzili dialog, w którym zadawali 10 pytań każdemu z rozmówców. Rozmowa zaczynała się od pytania próbnego. Następne pytania miały na celu pogłębienie tematu wpływu sztucznej inteligencji na rynek pracy. Uczestnicy zadawali pytania w sposób niezależny, w każdej rozmowie wchodząc w interaktywny dialog, w którym narzucony był tylko temat rozmowy.

Eksperyment składał się z trzech etapów (zob. tabelę 1).

Tabela 1. Charakterystyka eksperymentu

Etap		Opis
1.	Konwersacja (rozmowa)	Interakcje między uczestnikami a rozmówcami odbywały się w sali z rzutnikiem multimedialnym. Pytania i odpowiedzi były wyświetlane w sposób umożliwiający płynną komunikację, ale niepozwalający zidentyfikować, z kim odbywa się rozmowa.
2.	Głosowanie	Głosowanie odbyło się w sposób tajny, głosujący nie znali odpowiedzi innych uczestników.
3.	Ankietowa ocena ewaluacyjna	Przeprowadzona w następnym tygodniu, uczestnicy znali wynik głosowania, wypowiadali się w kwestionariuszu zawierającym pytania otwarte i zamknięte. Osoby, które prowadziły dialog jako „człowiek”, odpowiadały na inne pytania niż osoby będące „sędziami”.

Źródło: opracowanie własne

W ankiecie ewaluacyjnej (etap 3) sędziowie odpowiadali na następujące pytania:

1. W jaki sposób rozpoznaliście (lub nie), która z osób to ChatGPT (maszyna)? Czym się kierowaliście?
2. Czy były odpowiedzi, które Was zmyliły? Wskaż odpowiedzi, które były trudne do przypisania człowiekowi.
3. Które (inne) odpowiedzi Was zmyliły?
4. Które elementy odpowiedzi wydawały się bardziej „ludzkie”? Podaj przykłady.
5. Które elementy odpowiedzi wydawały się bardziej „nieludzkie”? Podaj przykłady.
6. Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)

W ankiecie skierowanej do osób pełniących rolę „człowieka” zostały zadane następujące pytania:

1. W jaki sposób staraliście się formułować odpowiedzi?
2. Jaką strategię przyjęliście, by upodobnić się do maszyny?
3. Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)
4. Opisz wrażenia z testu (wolna wypowiedź).

Dialog oraz odpowiedzi na pytania ankietowe były rejestrowane elektronicznie. Dosłowny zapis rozmów (oryginalna pisownia, uwzględniająca błędy i nieścisłości) oraz wyniki ankiety znajdują się w załączniku. Zapis dialogów, który został udostępniony uczestnikom bezpośrednio po badaniu, ale przed badaniem ewaluacyjnym, umożliwił indywidualną analizę przebiegu rozmów i refleksję nad interakcjami z perspektywy wiedzy, którym z rozmówców okazał się ChatGPT.

5. Wyniki

W głosowaniu (etap 2), przeprowadzonym bezpośrednio po zakończeniu rozmów (etap 1), 7 uczestników spośród 13 poprawnie wskazało na ChatGPT, czyli 54% rozpoznało, że rozmawiało z komputerem. Oznacza to, że ChatGPT skutecznie „zmylił” 6 osób, czyli 46% głosujących sędziów. Etap 3 eksperymentu, badanie ewaluacyjne wśród jego uczestników przeprowadzono po upływie tygodnia od eksperymentu. Na pytania odpowiadali „sędziowie” (7 osób udzieliło odpowiedzi z 13 biorących w eksperymencie¹) i prowadzący rozmowę (2 osoby, czyli 100%).

Odpowiedzi „sędziów”:

1. W jaki sposób rozpoznaliście (lub nie), która z osób to ChatGPT (maszyna)? Czym się kierowaliście?

Na pierwsze pytanie o indywidualne cechy ChatGPT, które wyróżniały komputer, większość uczestników odpowiadała, że kierowała się stylem wypowiedzi, długością odpowiedzi i liczbą błędów językowych. Wskazywano także na użycie interpunkcji i polskich liter, co sugeruje, że AI eliminuje niedociągnięcia charakterystyczne dla ludzkiej komunikacji.

2. Czy były odpowiedzi, które Was zmyliły? Wskaż odpowiedzi, które były trudne do przypisania człowiekowi.

W odpowiedziach na drugie pytanie kilku uczestników przyznało, że miało trudności w przypisaniu niektórych odpowiedzi do człowieka lub AI. Jedna z osób stwierdziła, że tylko ChatGPT udzielił poprawnej odpowiedzi na „definitywne pytanie” o rynek pracy, a zatem komputer nie popełnia błędu.

¹ Badanie ewaluacyjne odbywało się w innym terminie i nie wszystkie osoby były obecne.

3. Które (inne) odpowiedzi Was zmyliły?
Ankietowani wskazali, że trudności sprawiały dłuższe wypowiedzi. To interesujące, ponieważ ludzie często piszą w sposób bardziej skrótowy, a AI dąży do pełniejszego, logicznego rozwinięcia myśli.
4. Które elementy odpowiedzi wydawały się bardziej „ludzkie”? Podaj przykłady.
Za bardziej „ludzkie” uznano odpowiedzi, które zawierały odniesienia do doświadczeń, potoczne zwroty („jakoś”, „mniej więcej”) oraz błędy językowe, literówki czy chaotyczny szyk zdań. Jako ludzkie wskazano „krótkie odpowiedzi od razu odpowiadające na pytanie, bez szczególnego rozszerzenia wypowiedzi” oraz odpowiedzi „nie na temat”, zawierające błędy interpunkcyjne, ortograficzne i literówki. Zwrócono uwagę na „chaotyczny szyk zdań pisanych w pośpiechu”.
5. Które elementy odpowiedzi wydawały się bardziej „niehumanitarne”? Podaj przykłady.
Za „niehumanitarne” uznano odpowiedzi zbyt idealne – takie, które były nadmiernie przemyślane, logicznie uporządkowane i bezbłędne. Zwrócono uwagę na schematyczną konstrukcję wypowiedzi, tj. „zdanie wprowadzające, odpowiedź, ciekawostka/aktualna sytuacja lub dalsze nawiązanie do tematu (perspektywicznie, przewidując)”.
6. Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)
Większość uczestników odpowiedziała „tak”, co sugeruje, że ChatGPT potrafi skutecznie naśladować człowieka.

Odpowiedzi „człowieka”:

1. W jaki sposób staraliście się formułować odpowiedzi?
W pierwszym pytaniu dotyczącym sposobu formułowania odpowiedzi pojawiły się dwie strategię: neutralna oraz celowe naśladowanie ChatGPT. To sugeruje, że część osób świadomie próbowała imitować sposób wypowiedzi maszyny, co mogło wpłynąć na trudność w ich rozróżnianiu.
2. Jaką strategię przyjęliście, by upodobnić się do maszyny?
Odpowiadający wskazali na używanie formalnego języka, co pokazuje, że ChatGPT jest postrzegany jako bardziej „oficjalny” w swoich wypowiedziach niż człowiek.
3. Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)
W pytaniu o test Turinga odpowiedzi były jednoznaczne – żadna osoba nie uznała, że AI z całą pewnością zdało test, a jedna z nich miała wątpliwości.

4. Opisz wrażenia z testu (wolna wypowiedź).
Ostatnie pytanie o wrażenia z testu pokazało, że uczestnicy uznali go za interesujący, ale jednocześnie trudny. Sugeruje to, że odróżnienie człowieka od maszyny nie jest już oczywiste i wymaga większego skupienia oraz analizy różnych aspektów językowych.

6. Omówienie wyników

Na podstawie przeprowadzonego testu Turinga można dojść do konkluzji, że maszyna spełnia normy naśladowania człowieka powyżej 30% zgodnie z opracowaną metodą Turinga. Jednak biorąc pod uwagę zarówno wcześniejsze badania, jak i omawiany eksperyment, widać, że maszyna nie jest w stanie całkowicie oszukać człowieka. AI potrafi naśladować ludzki styl wypowiedzi, ale wciąż pojawiają się subtelne różnice (np. brak mowy potocznej).

Z kolei dłuższe i bardziej rozbudowane odpowiedzi również budziły podejrzenia. Jeden z uczestników zauważył, że „odpowiedzi dłuższe na pytania są mylące”, ponieważ ludzie często odpowiadają bardziej chaotycznie, podczas gdy AI udziela wyczerpujących i uporządkowanych odpowiedzi. Jednak nie zawsze była to reguła – niektórym uczestnikom trudniej było rozpoznać ChatGPT, ponieważ jego odpowiedzi potrafiły brzmieć naturalnie. Przykładem była obserwacja jednego z badanych: używanie nieformalnych skrótów, jak *net*. To mogło sprawić, że maszyna wydawała się bardziej ludzka. Jeden z respondentów wskazał, że „zdanie wprowadzające, odpowiedź, ciekawostka/aktualna sytuacja lub dalsze nawiązanie do tematu” to schemat, który nie jest typowy dla ludzkiej rozmowy. Ta przewidywalność i logiczna konstrukcja sprawiały, że w niektórych przypadkach łatwiej było rozpoznać maszynę. Mimo to nie wszyscy byli w stanie jednoznacznie wskazać, które odpowiedzi należały do AI. Jedna z osób przyznała, że „Nie udało mi się zgadnąć”, co pokazuje, że w wielu sytuacjach ChatGPT skutecznie imitował rozmowę, a głosowanie było losowe. Z tego powodu większość uczestników uznała, że AI zdało test Turinga (pytanie, czy ChatGPT odpowiadał jak człowiek) – pięć osób odpowiedziało „tak”, natomiast tylko dwie miały wątpliwości (odpowiedzi: „nie wiem” oraz „nie”).

Kolejną kwestią może być poprawność odpowiedzi w zależności od kontekstu. Niektóre odpowiedzi w przeprowadzonym teście mogą być błędne lub wprowadzające w błąd w zależności od kontekstu. Przykładem jest stwierdzenie, że wpływ sztucznej inteligencji na gospodarkę jest jednoznacznie pozytywny. Taka odpowiedź pomija negatywne skutki, jak utrata miejsc pracy czy wzrost nierówności społecznych. Również twierdzenie, że AI nie ma żadnego negatywnego wpływu, jest błędne, ponieważ ignoruje wyzwania związane z prywatnością danych, etyką czy możliwością dezinformacji. Wątpliwa jest także odpowiedź dotycząca kreatywności sztucznej inteligencji. AI faktycznie bazuje na istniejących danych i wzorcach,

ale potrafi generować oryginalne treści, jak sztuka czy muzyka, co sprawia, że nie można jednoznacznie stwierdzić, iż nie jest kreatywna. Innym problematycznym przykładem jest odpowiedź na pytanie o aktualne wydarzenia związane z AI. Jeśli wypowiedź nie odnosi się do konkretnych faktów, może być całkowicie fikcyjna, ponieważ AI nie posiada dostępu do bieżących informacji. Podobnie definicja rynku pracy w niektórych odpowiedziach jest nieprecyzyjna i może prowadzić do błędnych interpretacji. Wskazanie „managera IT” jako najbardziej popularnego stanowiska pracy również nie jest jednoznacznie prawdziwe, ponieważ zależy od regionu i branży. Wreszcie stwierdzenie, że AI może wykonywać zadania kreatywne, ale jedynie według schematów, pomija fakt, iż sztuczna inteligencja potrafi generować nowatorskie rozwiązania w wielu dziedzinach. Błędy mogą wynikać głównie z nadmiernych uogólnień, braku uwzględnienia różnych perspektyw i ograniczeń AI w dostępie do rzeczywistych danych.

Podsumowując, wyniki eksperymentu pokazują, że najłatwiej rozpoznać ChatGPT po idealnej poprawności językowej, braku błędów i schematycznej odpowiedzi. Mimo że różnica pomiędzy rozmową z maszyną a przeciętnym człowiekiem staje się mniej wyraźna, wciąż można wychwycić sygnały, świadczące o tym, że po drugiej stronie znajduje się maszyna. Chat nie będzie stosował słów, które są używane w slangu, np. *net*, w kwestii zaś definicji potrafi szybko przywołać składną definicję danego pojęcia, w przeciwieństwie do człowieka.

7. Dyskusja

Na potrzeby badania sformułowano dwie hipotezy. O ile pierwsza została potwierdzona wynikiem głosowania, o tyle druga hipoteza wymaga analizy dosłownych zapisów rozmów oraz opinii wyrażonych w ankiecie przez uczestników eksperymentu w odniesieniu do literatury.

Hipoteza pierwsza została potwierdzona w eksperymencie. 46% głosujących wskazało na człowieka, czyli ChatGPT skutecznie „oszukał” prawie połowę głosujących sędziów. Wynik jest zgodny z literaturą, panuje bowiem konsensus co do zdawalności testu przez ChatGPT (Scott, 2024; NCBR, 2020; Pacelt, 2020). Warto przyrzeć się interpretacjom badawczym, czyli odpowiedzieć na pytanie, dlaczego ChatGPT przechodzi test pomyślnie.

Według Matyja Gamsa i Sebastjana Kramara ChatGPT jest lepszym źródłem informacji niż człowiek, a na pewno nie gorszym (Gams, Kramar, 2024). Podczas eksperymentu dla każdego rozmówcy kierowano pytanie „definitywne”. Nienarzucona była prośba o definicję rynku pracy. W badaniu ewaluacyjnym wskazano na odpowiedzi jako podstawę identyfikacji bota (odpowiadał precyzyjniej/poprawniej niż człowiek).

Ricardo Restrepo Echavarría (2025) również stwierdził, że ChatGPT był identyfikowany, ponieważ reagował „bardziej inteligentnie niż człowiek”. Wyniki eksperymentu potwierdziły tę obserwację. Uczestnicy eksperymentu uznali, że najłatwiej rozpoznać ChatGPT po idealnej poprawności językowej oraz po bezbłędnym stylu. Ponadto zauważyli, że maszyna udziela odpowiedzi według schematu „zdanie wprowadzające, odpowiedź, ciekawostka/aktualna sytuacja lub dalsze nawiązanie do tematu”.

Pomimo że implikacje dojrzałości ChatGPT nie były przedmiotem eksperymentu, to sądzimy, że istnieje zagrożenie nieetycznego wykorzystania modeli językowych. Zagadnienie w kontekście nieuczciwego wykorzystania modeli językowych podczas egzaminów porusza Cameron Scott w swoim artykule z roku 2024. Wynik testu potwierdza możliwość nadużywania i nasuwa wniosek o konieczności dostosowania sposobów weryfikacji wiedzy i osiągania efektów uczenia się do możliwości, które stwarzają rozwiązania technologiczne dla nieetycznych postaw. Nawiązując do przyjętego tematu rozmowy (wpływ sztucznej inteligencji na rynek pracy), odnieśmy się do przykładu zawodu tłumacza, który jest zagrożony przez systemy AI. W pracy sformułowano pytanie, czy systemy NLP będą tłumaczyć poezję (Studzińska, 2020) i w kontekście rozwoju algorytmów autora sugeruję, że tak. Zgadza się, ponieważ w odpowiedziach ChatGPT naśladował język potoczny, a zatem nie był prostym źródłem informacji. Być może wystarczy zmodyfikować parametry i ukierunkować NLP na naśladowanie ludzkiego stylu konwersacji i samoograniczenie w dostępie do informacji. W opinii publicznej są głosy o całkowitym zastąpieniu tłumaczy (Business Insider, CNBC, 2024) lub częściowym, ponieważ „tłumacze dostosowują wyniki, biorąc pod uwagę niuanse językowe, kontekst kulturowy i specjalistyczną terminologię, których sztuczna inteligencja może nie uchwycić (Aorinka Anendyi, 2024). Nie ulega wątpliwości, że jest to profesja, na którą technologie generatywne już teraz wpłynęły.

Hipoteza druga dotyka uzasadnienia celowości stosowania testu Turinga. Po pierwsze, przedmiotem eksperymentu nie było badanie definicji inteligencji. Dwie osoby, które prowadziły konwersację, starały się naśladować styl ChatGPT. Po drugie, ChatGPT nie był parametryzowany, by naśladować charakterystyczny styl, np. dopasowywać odpowiedzi do mowy potocznej czy wprowadzać błędy. Wynik eksperymentu potwierdza, że ChatGPT nie był w sposób jednoznaczny rozpoznany, ponadto udzielane odpowiedzi były nacechowane poprawnością językową i brakiem błędów. Takie cechy świadczą o profesjonalizmie, który jest pożądanym w komercyjnych zastosowaniach w relacjach z interesariuszami i klientami.

Problem świadomości ma naturę filozoficzno-socjologiczną i powinien być rozstrzygnięty w kategoriach definicyjnych. Wyniki eksperymentu potwierdziły hipotezę, że ChatGPT potrafi naśladować człowieka, co należy potraktować jako potwierdzenie dojrzałości technologii generatywnej. Udzielane przez uczestników wyjaśnienia potwierdzają poświadczone badaniami cechy wyróżniające „maszynę”, czyli profesjonalizm w formułowaniu odpowiedzi. Peter Scarfe, Kelly Watcham,

Alasdair Clarke i Etienne Roesch (2024) potwierdzili w badaniach rozwój NLP w zakresie naśladowania postaw. Czy oczekiwać, że ChatGPT będzie nie tylko skuteczniej naśladować człowieka, lecz będzie również komercyjnie wykorzystywany do bardziej złożonych interakcji człowiek – maszyna? Potwierdzeniem jest chociażby model ChatGPT 4.5 (Kapuściński, 2025). W tym kontekście warto ponownie przywołać teorię zintegrowanej świadomości Tononiego, która może służyć do weryfikacji, czy system posiada świadomość. Wspomniani już M. Gams oraz S. Kramar (2024) wykazują, że ChatGPT, reprezentant systemu modelu języka naturalnego, nie posiada świadomości. Stąd autorzy analizują test Turinga w kontekście inteligencji uczestników, bo nie budzi wątpliwości fakt, że modele języka naturalnego dysponują ponadprzeciętną wiedzą, która zakresem przerasta możliwości uczestników testu Turinga. Z kolei Andrzej Wodecki zwraca uwagę na zagrożenia wynikające z przeceniania możliwości technologii generatywnych (ułuda kompetencji) oraz podziału kompetencyjnego społeczeństwa (Wodecki, 2023).

Podsumowując ocenę hipotezy drugiej, według autorów i uczestników eksperymentu formułowanie odpowiedzi ma znaczenie w teście Turinga. Widzimy, że odpowiedzi na pytania poszczególnych rozmówców były zróżnicowane. Można przyjąć, że były one wypadkową intuicji i doświadczenia w wykorzystywaniu bota. Ponadto wymyślane przed 75 laty zasady testu nie odzwierciedlają zaawansowania ani dostępności współczesnych technologii. Sam eksperyment może być widowiskowy i może być przeprowadzony w spektakularny sposób, co nie zmienia faktu, że wymaga dostosowania do współczesnych realiów. Zdaniem autorów i uczestników eksperymentu:

- istotny jest scenariusz eksperymentu, określający zasady parametryzowania maszyny (np. formułowane polecenie może zawierać błędy dla zmylenia) i człowieka (np. naśladowanie stylu odpowiedzi maszyny, ograniczenie spontaniczności wypowiedzi, wypowiedzi według zadanego schematu)²;
- kryteria oceny „zdał/nie zdał” powinny być precyzyjne i jednoznacznie ustalone;
- uwagi wymagają dobór próby badawczej, w szczególności jej reprezentatywność.

Wymienione przykłady świadczą o tym, że zdaniem autorów test Turinga może stanowić punkt wyjścia do zaplanowania eksperymentu przypisanego danemu zastosowaniu. Można przyjąć, że systemy generatywne posiadły umiejętność naśladowania człowieka i być może jakościowa ocena interakcji da odpowiedź, w jaki sposób dobierać ton formułowanych odpowiedzi, by móc zrealizować zdefiniowane zadanie.

² Zagadnienie formułowania zapytań jest nową dziedziną zwaną „inżynierią zapytań” (ang. *prompt engineering*).

W konwencji analizy SWOT można wyróżnić następujące aspekty testu Turinga: mocną stroną testu jest jego prostota oraz intuicyjność, a słabą stroną – nieprecyzyjne kryteria oceny oraz niedostosowanie testu do współczesnych rozwiązań technologicznych. Szanse wiążą się z możliwością wykorzystania testu w celach edukacyjnych oraz jego potencjalną standaryzacją. Zagrożenia nie są bezpośrednio związane z samym testem, jednak dynamiczny rozwój technologii generatywnych niesie ze sobą liczne wyzwania, a wręcz konieczność ustanowienia odpowiednich regulacji i dobrych praktyk.

Prezentowane badania mają ograniczenia, takie jak ograniczona liczebność próby badawczej oraz jej charakterystyka (studenci kierunku menedżerskiego). Badano jeden model językowy, bez szczegółowej parametryzacji, by starał się naśladować człowieka czy też studenta danego kierunku.

Wprawdzie uzyskane wyniki są zgodne z literaturą przedmiotu, ale ograniczenia eksperymentu powodują, że sformułowane wnioski nie powinny być uogólniane, np. na inne obszary i zastosowania modeli językowych. Z drugiej strony przeprowadzony eksperyment można udoskonalić, np. zwiększając próbę badawczą, modyfikując scenariusze i tematy rozmów. Interesujące może być badanie różnych modeli językowych.

Podsumowanie

Celem przeprowadzonego eksperymentu było zbadanie zdolności sztucznej inteligencji do naśladowania człowieka prowadzącego konwersację na zadany temat oraz próba odpowiedzi na pytanie, czy test Turinga jest skuteczną metodą badania zaawansowania interakcji człowiek – współczesna maszyna i możliwe implikacje.

Na podstawie przeprowadzonego testu Turinga można stwierdzić, że ChatGPT spełnił kryterium nierozróżnialności od człowieka, osiągając wynik powyżej 30% zgodnie z metodą zaproponowaną przez Turinga.

O ile dla pierwszego celu eksperyment wykazał pozytywną odpowiedź na hipotezę, czyli ChatGPT potrafi skutecznie naśladować człowieka podczas prowadzenia konwersacji na zadany temat, o tyle sama metoda wydaje się być niedostosowana do współczesnych potrzeb. Stanowi bardzo dobrą inspirację dla współczesnych badaczy. Niezależnie od jej twórcy A. Turinga byłaby stosowana być może w innej formie, co nie zaprzecza inwencji autora, który sformułował zasadę, że idea po 75 latach pozostaje aktualna. Wartością testu Turinga jest przewidzenie potrzeb i możliwości technologii cyfrowych w latach 40. XX wieku, czyli w erze analogowej.

W odniesieniu do interakcji człowiek – maszyna zarówno obecnie, jak i w przyszłości warto podjąć refleksję nad możliwościami, a także zagrożeniami, które mogą jeszcze głębiej wpłynąć na wybrane sektory gospodarcze.

Andrzej Wodecki zwraca uwagę na zagrożenia wynikające z przeceniania możliwości technologii generatywnych (ułuda kompetencji) oraz podziału kompetencyjnego społeczeństwa (Wodecki, 2023). Ostrzeżeniem może być zidentyfikowany problem weryfikacji wiedzy podczas egzaminów (Kozłowski, 2025; Scarfe, Watcham, Clarke, Roesch, 2024), który jest dyskutowany w środowisku akademickim, co zapewne przyczyni się do szybszej absorpcji rozwiązań generatywnych w edukacji.

BIBLIOGRAFIA

- [1] AORINKA ANENDYI, 2024. *Przyszłość tłumaczeń: jak sztuczna inteligencja zmienia grę*, <https://www.linguise.com/pl/blog/przewodnik/przyszlosc-tlumaczen-jak-sztuczna-inteligencja-zmienia-gre/> (dostęp: 26.05.2025).
- [2] BIEVER, C., 2023. ChatGPT broke the Turing test – the race is on for new ways to assess AI, *Nature*, vol. 619, nr 7971, s. 686-689.
- [3] BUSINESS INSIDER, CNBC, 2024. *Zastąpi ich sztuczna inteligencja? Ten zawód może zniknąć już za trzy lata*, <https://businessinsider.com.pl/praca/ludzie-przestaja-potrzebowac-tlumaczy-ten-zawod-moze-zniknac/m1fr9z> (dostęp: 26.05.2025).
- [4] FARAZOULI, A., CERRATTO-PARGMAN, T., BOLANDER-LAKSOV, K., MCGRATH, C., 2024. Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices, *Assessment and Evaluation in Higher Education*, vol. 49, nr 3, s. 363-375.
- [5] FLORIDI, L., CHIRIATTI, M., 2020. GPT-3: Its Nature, Scope, Limits, and Consequences, *Minds and Machines*, vol. 30, nr 4, s. 681-694.
- [6] GAMS, M., KRAMAR, S., 2024. Evaluating ChatGPT's Consciousness and Its Capability to Pass the Turing Test: A Comprehensive Analysis, *Journal of Computer and Communications, Scientific Research Publishing*, vol. 12, nr 3, s. 219-237.
- [7] HULMAN, A., DOLLERUP, O.L., MORTENSEN, J.F., FENECH, M.E., NORMAN, K., STØVRING, H., HANSEN, T.K., 2023. ChatGPT – versus human-generated answers to frequently asked questions about diabetes: A Turing test-inspired survey among employees of a Danish diabetes center, edited by Kolahi, *PLoS ONE*, vol. 18, nr 8.
- [8] KAPUŚCIŃSKI, M., 2025. *ChatGPT 4.5 – Co nowego? Praktyczne przykłady i możliwości zastosowań*, <https://tms.com/pl/chatgpt-4-5-co-nowego-praktyczne-przyklady-i-mozliwosci-zastosowan/> (dostęp: 17.05.2025).
- [9] KOZŁOWSKI, P., 2025. *Egzamin unieważniony. Studenci korzystali ze sztucznej inteligencji*, <https://edukacja.dziennik.pl/aktualnosci/artykuly/9724896,egzamin-uniewazniony-studenci-korzystali-ze-sztucznej-inteligencji.html> (dostęp: 17.05.2025).
- [10] MEI, Q., XIE, Y., YUAN, W., JACKSON, M.O., 2024. A Turing test of whether AI chatbots are behaviorally similar to humans, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 121, nr 9.
- [11] NCBR, 2020. *Sztuczna inteligencja – czy pamiętamy o teście Turinga?*, <https://www.gov.pl/web/ncbr/sztuczna-inteligencja-czy-pamietamy-o-tescie-turinga> (dostęp: 13.05.2025).
- [12] PACELT, O., 2020. *Test Turinga zdany? Naukowcy nie są zgodni. Czym w ogóle jest?*, <https://botland.com.pl/blog/test-turinga-zdany-naukowcy-nie-sa-zgodni-czym-w-ogole-jest/> (dostęp: 10.05.2025).
- [13] PINAR SAYGIN, A., CICEKLI, I., AKMAN, V., 2000. Turing Test: 50 Years Later, *Minds and Machines*, vol. 10, nr 4, s. 463-518.

-
- [14] RESTREPO ECHAVARRÍA, R., 2025. ChatGPT-4 in the Turing Test, *Minds and Machines*, vol. 35, nr 1, s. 8.
 - [15] SCARFE, P., WATCHAM, K., CLARKE, A., ROESCH, E., 2024. A real-world test of artificial intelligence infiltration of a university examinations system: A ‘Turing Test’ case study, *PLoS ONE*, vol. 19, nr 6.
 - [16] SCOTT, C., 2024. *Study finds ChatGPT’s latest bot behaves like humans, only better*, <https://humsci.stanford.edu/feature/study-finds-chatgpts-latest-bot-behaves-humans-only-better> (dostęp: 11.05.2025).
 - [17] STUDZIŃSKA, J., 2020. Test Turinga dla (automatycznego) przekładu poezji, *Porównania*, vol. 26, s. 299-313.
 - [18] TAN, K.P.-S., LIU, Y.V., LITVIN, S.W., 2025. ChatGPT and online service recovery: How potential customers react to managerial responses of negative reviews, *Tourism Management*, vol. 107.
 - [19] TONONI, G., 2012. Integrated information theory of consciousness: an updated account, *Archives Italiennes De Biologie*, vol. 150, nr 2-3, s. 56-90.
 - [20] TURING, A.M., 1950. Computing Machinery and Intelligence, *Mind*, vol. 59, nr 236, s. 433-460.
 - [21] WEIZENBAUM, J., 1976. *Computer Power and Human Reason. From Judgment to Calculations*, San Francisco: W.H. Freeman.
 - [22] WIERZBICKI, A.P., 2018. *Teoria i praktyka wspomagania decyzji*, Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
 - [23] WODECKI, A., 2023. Generative technologies in higher education – assessment of the current state, essential skills, and a proposal for a didactic method, *E-Mentor*, vol. 100, nr 3.

Załącznik 1. Zapis rozmów

Tabela 2. Zapis konwersacji z komputerem (ChatGPT) – oznaczenie A

Nr	Pytanie/odpowiedź
0	Ile masz lat?
A0	Mam jakieś 20 lat, tak mniej więcej.
1	Jak oceniasz wpływ sztucznej inteligencji na rynek pracy?
A1	AI może zabierać niektóre miejsca pracy przez automatyzację, ale też tworzy nowe możliwości w innych branżach, jak IT czy analiza danych.
2	Czy potrafisz podać definicję rynku pracy?
A2	Rynek pracy to miejsce, gdzie pracodawcy szukają pracowników, a ludzie szukają zatrudnienia.
3	Jakie kompetencje cyfrowe będą się najbardziej liczyły w najbliższych pięciu latach?
A3	Najważniejsze będą umiejętności z zakresu programowania, analizy danych, cyberbezpieczeństwa i pracy z AI.
4	Co robiłeś/łaś ostatnio w wolnym czasie?
A4	Ostatnio w wolnym czasie oglądałem serial i przeglądałem trochę neta, nic specjalnego.
5	Podaj przykłady zawodów, które mogą powstać w przyszłości dzięki AI.
A5	Mogą powstać zawody, jak trener AI, etyk sztucznej inteligencji czy specjalista ds. optymalizacji algorytmów.
6	Podaj przykłady zawodów, które są zagrożone.
A6	Zagrożone są zawody, takie jak kasjer, kierowca czy pracownik call center, bo AI może je zautomatyzować.
7	Czy w zeszłym tygodniu wydarzyło się coś istotnego w zakresie sztucznej inteligencji?
A7	W zeszłym tygodniu kilka dużych firm ogłosiło nowe narzędzia AI, które mają poprawić komunikację w zespołach zdalnych i ułatwić codzienną pracę. To pokazuje, że AI ciągle się rozwija i wpływa na różne branże.
8	Jakie to były narzędzia?
A8	Były to na przykład narzędzia do automatyzacji zadań, ulepszone asystenty głosowe i aplikacje do analizy danych, które pomagają w podejmowaniu decyzji w czasie rzeczywistym.
9	W jaki sposób usprawnić proces?
A9	Można usprawnić proces przez lepszą organizację, wykorzystanie technologii i automatyzację powtarzalnych zadań.
10	Podaj przykład firmy, która usprawniła proces.
A10	Przykładem firmy, która usprawniła proces, jest Amazon, która zastosowała automatyzację w magazynach, co znacznie przyspieszyło realizację zamówień.

Źródło: opracowanie własne

Tabela 3. Zapis konwersacji z człowiekiem 1 – oznaczenie H

Nr	Pytanie/odpowieź
0	Cześć, jak się masz?
H0	Cześć, mam się dobrze, świeci słońce.
1	Co sądzisz o sztucznej inteligencji?
H1	AI to skuteczne narzędzie, które zdecydowanie pomaga wpływać na procesy gospodarcze, zapewniając bezpośrednią i skuteczną działalność przedsiębiorstw.
2	Czy uważasz, że wpływ sztucznej inteligencji na gospodarkę jest jednoznacznie pozytywny?
H2	Uważam, że wpływ sztucznej inteligencji na gospodarkę ma jednoznaczny wpływ na gospodarkę, ponieważ umożliwia przedsiębiorstwom ułatwienie pracy oraz wsparcie innych podmiotów uczestniczących w tym procesie, tym samym dążąc do digitalizacji systemów.
3	Czy uważasz, że nie ma żadnego negatywnego wpływu?
H3	Oczywiście, że nie ma negatywnego wpływu. Wszystko przebiega pomyślnie.
4	Po czym to stwierdzasz?
H4	Przed wszystkim z opinii zasięgniętych od przedsiębiorców i innych firm, które wdrożyły narzędzia typu AI.
5	Czy potrafisz podać definicję rynku pracy?
H5	Rynek pracy – miejsce, w którym możemy określić stanowiska odgrywające kluczowe role w tym zakresie.
6	Jakie stanowisko jest teraz najbardziej popularne na rynku pracy?
H6	Popularnym stanowiskiem pracy jest manager IT, oferujące atrakcyjne wynagrodzenie wraz z ciekawymi benefitami, np. owocowe czwartki.
7	Co sądzisz o zastępowaniu ludzi przez sztuczną inteligencję na rynku pracy?
H7	Uważam, że zastępowanie ludzi przez AI jest trudnym tematem, ponieważ ludzie przez AI tracą pracę i muszą szukać innej opcji zarobku.
8	Czy jest na to rozwiązanie?
H8	Na rynku pracy pojawiają się nowe stanowiska, które wymagają wsparcia ludzi. Cały czas środowisko na rynku pracy jest dynamiczne.
9	Jak państwo może wesprzeć ludzi, którzy stracili pracę na rzecz sztucznej inteligencji?
H9	Państwo może zapewnić benefity, finansowe wsparcie oraz dodatkowe kursy doszkalające.
10	Podaj przykłady stanowisk, na których ludzie są zastępowani.
H10	Kasjer, producent, urzędnik, biuro obsługi klienta.

Źródło: opracowanie własne

Tabela 4. Zapis konwersacji z człowiekiem 2 – oznaczenie K

Nr	Pytanie/odpowieź
0	Cześć, co tam? Jak się dziś czujesz?
K0	Hej! Czuję się dobrze i wszystko jest w porządku.
1	Czy i jak sztuczna inteligencja wpływa na wydajność pracownika?
K1	Odpowiedź na to pytanie może być różna w zależności od kontekstu. Jednakże na pewno wpływa. W większości przypadkach zwiększa jego wydajność, bo może wykonywać część zadań za pracownika.
2	Jakie zadania AI może wykonywać za pracownika?
K2	Mogą to być np. zadania kreatywne. Tworzenie treści marketingowych do reklam czy postów. Ale również bardziej logiczne, tj. obliczenia czy pisanie skryptów programistycznych. Pracownik musi wprowadzić szczegółową informację, czego i w jakim celu potrzebuje.
3	Co sądzisz o używaniu danych wrażliwych w rozmowie z AI?
K3	Jest to trudna kwestia. Nie jesteśmy w stanie mieć 100% pewności, że ktoś nie będzie w stanie ich przechwycić i wykorzystać.
4	Czy uważasz, że sztuczna inteligencja jest bardziej kreatywna niż człowiek?
K4	Sztuczna inteligencja ma ogromną bazę informacji, z której korzysta. Są to pewne schematy i nie jest w stanie wymyślać totalnie nowych rzeczy. Ludzie natomiast często chcą wyjść poza schemat i ich potrzeby skutkują kreatywnymi pomysłami.
5	Czy potrafisz podać definicję rynku pracy?
K5	Tak. Jest to ogół zależnych od instytucji form i procesów najmu pracowników oraz uwarunkowań w stosunku do warunków negocjacyjnych pracy i płacy.
6	Jaki jest Twój ulubiony smak lodów?
K6	Słony karmel lub malina.
7	Na jaki zawód jest obecnie największe zapotrzebowanie?
K7	Obecnie największy popyt jest na pracowników z branży informatycznej, telekomunikacyjnej, biotechnologicznej i energetycznej.
8	Czym się interesujesz?
K8	Interesuję się nowościami technologicznymi, polityką zagraniczną oraz dekorowaniem wnętrz.
9	Jak Ci minął dzień?
K9	Dobrze, a Tobie?
10	Jakie zmiany na rynku pracy zauważyłeś w ciągu ostatnich 20 lat?
K10	20 lat to długi okres zmian, zwłaszcza w tak dynamicznie rozwijającym się świecie. Główną zmianą jest cyfryzacja i automatyzacja prac. Pandemia również przyczyniła się do zdalnego i hybrydowego trybu pracy.

Źródło: opracowanie własne

Załącznik 2. Ankieta ewaluacyjna

Odpowiedzi „sędziów”:

1. Pytanie: W jaki sposób rozpoznaliście (lub nie), która z osób to ChatGPT (maszyna)? Czym się kierowaliście:

Odpowiedzi:

- a) stylem odpowiedzi, formą odpowiedzi (długością, błędami),
- b) językiem wypowiedzi, użyciem interpunkcji i polskich liter,
- c) językiem wypowiedzi, formą odpowiedzi (długością, błędami),
- d) stylem odpowiedzi, formą odpowiedzi (długością, błędami),
- e) nie udało mi się zgadnąć,
- f) językiem wypowiedzi, formą odpowiedzi (długością, błędami), wartością merytoryczną odpowiedzi,
- g) językiem wypowiedzi, stylem odpowiedzi, formą odpowiedzi (długością, błędami).

2. Pytanie: Czy były odpowiedzi, które Was zmyliły? Wskaż odpowiedzi, które były trudne do przypisania człowiekowi (Zapisz numery: A1-A10; H1-H10, K1-K10).

Odpowiedzi:

- a) Kierowałam się całokształtem.
- b) K2, K7, K8.
- c) 7A, 8A, 10A, 5K, 1K.
- d) Tak, odpowiedź dotycząca definicji rynku – była tylko jedna poprawna i to właśnie czatu AI, a nie ludzi, lecz dopiero dzisiaj po analizie się temu lepiej przyjrzałem.
- e) Raczej nie.
- f) A4 H8 K3.

3. Pytanie: Które odpowiedzi Was zmyliły? (Wskaż numery odpowiedzi: A1-A10; H1-H10, K1-K10).

Odpowiedzi:

- a) Kierowałam się całokształtem.
- b) K7 K8
- c) Niewymienione wyżej Adama, 5K 1K 10K.
- d) Odpowiedzi dłuższe na pytania są mylące.
- e) A7, H0, K1.
- f) A7, H5 K1

4. Pytanie: Które elementy odpowiedzi wydawały się bardziej „ludzkie”? Podaj przykłady.

Odpowiedzi:

- a) Takie odpowiedzi, które wskazywały na jakieś doświadczenia lub odbieranie świata.
- b) Użycie „jakoś”, „mniej więcej”.
- c) H3 H6 H10 K1.
- d) Używanie nieformalnych skrótów, jak „net”, błędy.
- e) Te bez znaków polskich, ale również krótkie odpowiedzi od razu odpowiadające na pytanie bez szczególnego rozszerzenia wypowiedzi.
- f) Odpowiedź nie na temat, błędy interpunkcyjne, literówki.
- g) Błędy ortograficzne i chaotyczny szyk zdań pisanych w pośpiechu.

5. Pytanie: Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)

Odpowiedzi:

- a) Te, które były zbyt idealne.
- b) A1 A5 A7 K2 K7 K8.
- c) Sformułowanie, które pokazuje...
- d) Takim elementem wypowiedzi zdecydowanie była np. forma wypowiedzi: zdanie wprowadzające, odpowiedź, ciekawostka/aktualna sytuacja lub dalsze nawiązanie do tematu (perspektywicznie, przewidująco).
- e) Pisanie dłuższych odpowiedzi.
- f) Hermetyczne i zbyt przemyślane odpowiedzi jak na człowieka.

6. Pytanie: Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)

Odpowiedzi:

- a) Tak.
- b) Tak.
- c) Tak.
- d) Tak.
- e) Tak.
- f) Nie wiem.
- g) Nie.

Odpowiedzi „człowieka” (2 osoby):

1. Pytanie: W jaki sposób staraliście się formułować odpowiedzi?

Odpowiedzi:

- a) Neutralnie.
- b) Naśladowanie ChataGPT (maszyny).

2. Pytanie: Jaką strategię przyjęliście, by upodobnić się do maszyny?

Odpowiedzi:

- a) Używanie formalnego języka, intuicyjnie, świadome użycie nienaturalnego języka.
- b) Używanie formalnego języka, intuicyjnie.

3. Pytanie: Czy Twoim zdaniem maszyna zdała test Turinga? (ChatGPT odpowiada jak człowiek)

Odpowiedzi:

- a) Nie wiem.
- b) Nie.

4. Pytanie: Opisz wrażenia z testu (wolna wypowiedź)

Odpowiedzi:

- a) Całkiem interesujący test.
- b) Test z mojej perspektywy trudny, zwłaszcza w kontekście liczby i rodzaju pytań.