

## Sztuczna inteligencja jako szansa dla teologii

### Artificial Intelligence as an Opportunity for Theology

MACIEJ MRÓZ

Uniwersytet im. Adama Mickiewicza w Poznaniu,  
Szkoła Doktorska Nauk Humanistycznych, Polska  
maciej.mroz@amu.edu.pl  
<https://orcid.org/0000-0003-4820-3969>

**Abstract:** This paper aims to show why and how the dynamic development of artificial intelligence can and should be of interest to theological disciplines. First, the author argues that technological artefacts that are part of the culture can be considered as new “sites” of human beings and, thus, theological sites. Reflecting on the algorithms created to mimic and replace intelligent human action leads to asking questions about the nature of man, his functioning and purpose. The author analyzes how this resonates with contemporary tasks and methods of theology. Next, the author examines the context of defining artificial intelligence and points out its implications. In the paper’s final part, the author analyzes how ethics, including ethics practised on the basis of theology, can contribute to shaping the development of artificial intelligence. The overall argument leads to the conclusion that theology, thanks to its unique perspective, can play a significant and positive role in the context of the phenomenon of artificial intelligence. This opens up new and creative opportunities for the theological disciplines to serve faith and culture.

**Keywords:** theology; artificial intelligence; ethics of artificial intelligence; philosophy

**Abstrakt:** Celem artykułu jest ukazanie, dlaczego i w jaki sposób dynamiczny rozwój sztucznej inteligencji może i powinien stanowić przedmiot zainteresowania dyscypliny nauk teologicznych. W pierwszej kolejności autor przekonuje, że artefakty technologiczne, stanowiące część kultury, mogą być traktowane jako jedno z nowych „miejsc” człowieka, a w ten sposób także miejsc teologicznych. Namysł nad algorytmami tworzonymi w celu naśladowania i zastępowania inteligentnych działań ludzi prowadzi do stawiania w nowym kontekście pytań o naturę człowieka, jego funkcjonowanie oraz przeznaczenie. Autor analizuje w jaki sposób współbrzmi to ze współczesnymi zadaniami oraz metodami

uprawiania teologii. W dalszej kolejności, autor analizuje kontekst definiowania pojęcia sztucznej inteligencji oraz wskazuje na związane z tym implikacje. W ostatniej części dokonano analizy w jaki sposób etyka, także ta uprawiana na gruncie teologii, może przyczynić się do kształtowania rozwoju sztucznej inteligencji. Całość wywodu prowadzi do wniosku, że w kontekście fenomenu sztucznej inteligencji, teologia, dzięki swojej unikalnej perspektywie, może odgrywać znaczącą i pozytywną rolę. Otwiera to dyscyplinie nauk teologicznych nowe i twórcze możliwości sprawowania służby na rzecz wiary i kultury.

**Słowa kluczowe:** teologia; sztuczna inteligencja; etyka sztucznej inteligencji; filozofia

## Wstęp

Gwałtowny rozwój technik sztucznej inteligencji (SI) przynosi zarówno wiele szans, jak i zagrożeń dla ludzkości oraz całej planety. Na gruncie nauki stanowi to wyzwanie i przedmiot badań dla rozmaitych dyscyplin naukowych. Niemniej, poza dyscypliną jaką jest informatyka, wiele obecnych badań nad SI zdaje się koncentrować przede wszystkim na ryzykach związanych z jej rozwojem. Podczas kiedy należy identyfikować i podkreślać rozmaite zagrożenia, jakie tworzą te technologie, można także dostrzec ogromną szansę, jaką postęp na polu SI przynosi dla dyscypliny nauk teologicznych. Z punktu widzenia teologii technologie stanowią ludzkie artefakty będące częścią kultury, stając się w ten sposób nowymi miejscami teologii. Dotyczy to wielu zagadnień techniki w ogóle i różnych jej zastosowań. W tym krajobrazie rozmaitych artefaktów technicznych sztuczna inteligencja jest przypadkiem szczególnym. Przede wszystkim jest technologią ogólnego przeznaczenia (jak jest nią także np. prąd elektryczny), wykorzystywaną w wielu obszarach zastosowań. Ponadto jest dla końcowego użytkownika tych technologii niewidoczna. Wielokrotnie w życiu codziennym korzystamy z rozmaitych produktów, nie zdając sobie sprawy, że narzędzia których używamy wykorzystują do swojego działania sztuczną inteligencję. Wreszcie z samej definicji SI ma naśladować i zastępować inteligentne działania człowieka. Niektórzy stawiają sobie nawet za zadanie zbudowanie takich rozwiązań, które miałyby osiągnąć poziom posiadający wszystkie atrybuty ludzkiego umysłu (tzw. ogólna lub silna sztuczna inteligencja, ang. AGI) albo nawet znacząco te możliwości przewyższający (tzw. superinteligencja lub osobliwość). Niedawno amerykańska korporacja Open AI umieściła w misji swojej firmy taki właśnie cel (Open AI, 2023). Choć istnieje ogromna pokusa ulegania w tym kontekście pewnym wizjom futurologicznym, zwłaszcza w kontekście AGI i superinteligencji, technologie te nie są jednak jeszcze dostępne, a trwa debata, czy w ogóle możliwe jest ich powstanie. Ze względu więc na fakt, że argumentacja na ten temat ma charakter głównie spekulatywny, oparty na eksperymentach myślowych, zagadnienia

te nie zostaną podjęte w ramach niniejszych rozważań. Technologie, z którymi mamy do czynienia dzisiaj, w kontraście do powyższych, noszą miano wąskiej (lub słabej), sztucznej inteligencji. Dotyczą one algorytmów wytrenowanych do poszczególnych zadań, takich jak: tłumaczenie maszynowe, rozpoznawanie obrazów, tworzenie obrazów, generowanie mowy, generowanie tekstu, różnego rodzaju predykcje, grę w gry, przewidywania struktury białek i wiele innych. SI bardzo często przewyższa ludzkie możliwości w tych obszarach, a dzięki szerokiemu zastosowaniu stawia coraz więcej pytań o samą naturę tych zjawisk oraz rolę człowieka w stechnicyzowanym świecie.

## I. Technologiczne artefakty jako „miejsca” teologii

Zachowując właściwe proporcje, i pamiętając, że dla teologii głównym jej miejscem i źródłem jest Objawienie, należy zaznaczyć, że do miejsc teologicznych (*loci theologici*) należą również inne fenomeny, w tym także różne przejawy działalności i refleksji człowieka. Wprawdzie według różnych klasyfikacji, w tym powszechnie przywoływanej – Melchiora Cano, *locus theologicus* związane z kulturą należałoby ulokować na ostatnim miejscu hierarchii ważności, to jednak należy stwierdzić, że staje się ono przedmiotem rosnącego zainteresowania dyscypliny nauk teologicznych. Dzieje się tak dlatego, że aby dobrze zrozumieć wynikające z chrześcijańskiego Objawienia orędzie zbawcze należy także zrozumieć dobrze kim jest człowiek. Podejście takie wybrzmiewa jako jedna z zasadniczych tez metody antropologicznej zaproponowanej przez Rahnera (Dzidek i Sikora 2018, 155). W tym kontekście zrozumienie kultury jako przejawu działalności człowieka, jak zauważa Kulisz, jest jednym z istotnych elementów tożsamości teologii. Tożsamość ta wymaga interpretowania i przekazywania treści Objawienia, posługując się kontekstem życia współczesnych ludzi. Prowadzi do nowych zadań teologii, do których należy służba na rzecz wiary i kultury. Albowiem to w kulturze i przez kulturę człowiek dochodzi do prawdziwego i pełnego człowieczeństwa. Jest też ona miejscem, w którym pyta on o swoje ostateczne przeznaczenie (Kulisz 2012, 254). Kulisz powołuje się przy tym zresztą na nauczanie Kościoła, zawarte w Konstytucji duszpasterskiej o Kościele w świecie współczesnym *Gaudium et spes*. Dotyczy to w ten sposób również techniki, która jest jednym z elementów kultury. Benanti mówi nawet o techno-ludzkiej kondycji (*techno-human condition*) natury człowieka (Benanti 2018). Nie chodzi mu jednak jedynie o odniesienie do jakiegoś wybranego okresu w historii ludzkiej cywilizacji. Benanti odnosi termin techno-ludzka kondycja do sposobu, w jaki ludzie od najdawniejszych czasów przeżywali i rozumieли swoje bycie w świecie: tzn. poprzez interakcję z otaczającym środowiskiem zapośredniczoną poprzez

używanie różnego rodzaju narzędzi, czyli technologicznych artefaktów (Benanti 2023). Warto przy tym zauważyć, że mamy tu do czynienia z pewnego rodzaju pętlą sprzężenia zwrotnego: nie tylko człowiek tworzy kulturę, w tym technologiczne artefakty, ale również one kształtują jego bycie w świecie. Celnie myśl tę wyraził Sir Winston Churchill w swoim przemówieniu na spotkaniu Izby Lordów w 1943 roku, mówiąc: „Najpierw to my kształtujemy nasze budynki, ale później to one kształtują nas” (“We shape our buildings; thereafter they shape us”) (Churchil 1943). W tym kontekście rozwiązania sztucznej inteligencji stanowią szczególny przypadek. Albowiem jak żadne inne artefakty SI budowana jest z tym zamiarem, żeby możliwie najbardziej naśladować, a nawet zastępować działania człowieka. Chociaż nie istnieje jedna ogólnie przyjęta definicja sztucznej inteligencji, to ich większość odnosi się do tego właśnie aspektu. Dla przykładu Encyklopedia PWN podaje następującą definicję:

„Sztuczna inteligencja (SI), ang. Artificial Intelligence (AI), dziedzina nauki zajmująca się badaniem mechanizmów ludzkiej inteligencji (psychol.) oraz modelowaniem i konstruowaniem systemów, które są w stanie wspomagać lub zastępować inteligentne działania człowieka” (Encyklopedia PWN).

Oznacza to, że sztuczna inteligencja jest pierwszym w dziejach ludzkości (odkładając na bok legendę Golema) artefaktem, który człowiek tak dosłownie próbuje skonstruować na swój obraz. Trwają również badania i budowane są systemy, które polegają na sprzęgnięciu algorytmów SI z robotyką. W rezultacie doprowadzić może to do powstania w pełni funkcjonalnych robotów humanoidalnych, które wyglądem i funkcjonowaniem bardzo przypominałyby człowieka. Niemniej, używając pewnej metafory, można powiedzieć, że bardziej niż obraz zjawiska związane z rozwijaniem sztucznej inteligencji przypominają lustro. I to w sensie bardziej dosłownym niż konstruktorzy tych systemów by sobie zapewne życzyli. Nie są one bowiem jedynie emanacją zalet człowieka. Oprócz zaprojektowanych i pożądaných skutków działania algorytmów, przynoszą one także zagrożenia. Jednym z najpoważniejszych są ich potencjalne działania dyskryminacyjne. Wynika to ze specyfiki funkcjonowania algorytmów SI. Jedną z najpowszechniej wykorzystywanych obecnie technik jest tzw. uczenie maszynowe, a w szczególności głębokie uczenie maszynowe (ang. deep learning). Co istotne to właśnie tego rodzaju technikom zawdzięczamy gwałtowny rozkwit możliwości SI w ubiegłych latach. Ten z kolei możliwy był dzięki zwiększającemu się dostępowi do ogromnych zbiorów danych. Algorytmy uczenia głębokiego funkcjonują na zasadzie przetwarzania tych wielkich zbiorów odpowiednio przygotowanych danych. W odniesieniu do tego procesu mówi się o trenowaniu i uczeniu tych algorytmów.

Mówiąc o uczeniu się maszyn trzeba przy okazji zaznaczyć, że począwszy już od samej nazwy SI aż do rozmaitych fenomenów z nią związanych, mamy do czynienia ze swego rodzaju antropomorfizacją maszyn, co rodzi z kolei całą masę różnego rodzaju wyzwań. Wracając jednak do ogromnych zasobów danych, które leżą u podstaw sukcesu współczesnych algorytmów, trzeba zaznaczyć, że zdecydowana większość tych zbiorów danych pochodzi z zapisu aktywności ludzi. Prowadzi to w rezultacie do przejawiania przez algorytmy uprzedzeń. Naturalnie uprzedzone są nie same algorytmy jako jakiegoś tajemnicze byty. Dyskryminacje są jedynie wynikiem ich działalności jako agentów w świecie. U podstaw natomiast leżą właśnie „uprzedzone” dane. Aby to lepiej zrozumieć, prościej będzie prześledzić to na konkretnym przykładzie. W 2018 roku „Reuters” opublikował doniesienia o tym, że algorytm przygotowany przez firmę Amazon w celu wyszukiwania najlepszych kandydatów na pracowników okazał się uprzedzony względem kobiet (Dastin 2018). Polegało to na tym, że opracowano zautomatyzowany system do oceny zgłoszeń kandydatów do pracy. Algorytm wytrenowano z użyciem zgłoszeń do pracy z ubiegłych 10 lat. Obserwacje wyniku działania systemu dowiodły, że preferuje on kandydatów będących mężczyznami. Biorąc pod uwagę ówczesną strukturę zatrudnienia, zdominowaną przez mężczyzn oraz to, że mamy do czynienia z gigantem z obszaru technologii, nietrudno się domyślić skąd pochodzą preferencje algorytmu względem kandydatów mężczyzn. Istniejące dane, użyte w procesie trenowania algorytmu, wskazywały na to, że to mężczyźni są częściej wybieranymi kandydatami do pracy. Pokazuje to, w jaki sposób algorytmy uczenia maszynowego mogą stać się narzędziami automatyzacji uprzedzeń na ogromną skalę. Przykłady można by zresztą mnożyć. Nie chcę jednak wliczać jedynie problemów i wyzwań, ponieważ istnieje już wiele opracowań tego typu. W przedstawionym przypadku istotny jest fakt, że analiza wyniku działania algorytmu uwypukliła problem, leżący *de facto* w funkcjonowaniu ludzi. Analogicznie patrząc na inne algorytmy, możemy dowiadywać się nowych rzeczy i stawiać nowe pytania o naturę człowieka, jego funkcjonowanie oraz przeznaczenie. Sztuczna inteligencja może stanowić zatem swego rodzaju lustro, w które spoglądając, możemy dostrzec rzeczy, które trudno byłoby sobie inaczej uzmysłowić. Artefakty technologiczne sztucznej inteligencji mogą więc stać się nie tylko narzędziami, których używamy, ale właśnie nowymi „miejscami” człowieka, i miejscami teologii (Benanti 2019). Na gruncie teologii badań odnoszących się do sztucznej inteligencji ciągle jeszcze niewiele się dzieje. Zwiększenie wysiłków na tym polu postuluje m.in. Puzio, stawiając jednocześnie 10 tez na temat teologii technologizacji (Puzio 2023, 30-33):

1. Technologizacja jest istotna dla teologii.
2. Teologia jest istotna dla technologizacji.

3. Dokładne badania naukowe i rzetelne zbadanie krajobrazu dyskursu stanowią punkt wyjścia do teologicznego dialogu z technologią, a polaryzacji należy unikać.
4. Teologia technologizacji jest interdyscyplinarna, międzywyznaniowa i międzynarodowa. Eksperymentuje i odważnie posuwa się naprzód, tworząc nowe obszary.
5. Religia i teologia są przekształcane przez technologizację.
6. Technologizacja stawia wyzwania i kwestionuje teologiczne koncepcje i teorie, co skutkuje koniecznością opracowania nowych podejść teologicznych.
7. Technologizacja musi być zawsze postrzegana zarówno z perspektywy jej wyzwań, jak i możliwości; oferuje wiele szans dla teologii i religii.
8. Teologia musi odważyć się zmierzyć z pomijanymi i uciszonymi tematami oraz badać zupełnie nowe sposoby myślenia.
9. Teologia musi być zaangażowana w projektowanie i rozwijanie technologii.
10. Teologia powinna stać się wpływową dziedziną (dosł. influencerem).

W propozycji Puzio na uwagę zasługuje w szczególności ujęcie pozytywnej roli, jaką wobec coraz powszechniejszego użycia technologii może odgrywać teologia. Nie powinna ona postrzegać technologizacji jedynie w kontekście zagrożenia, ale wręcz przeciwnie, starać się wnieść do tego procesu swoje unikalne wartości i perspektywę. Puzio argumentuje, że: „pomimo znacznego spadku znaczenia kościołów chrześcijańskich, religie nie powinny być niedoceniane jako aktorzy kultury. Dzięki swoim instytucjom, organizacjom i narracjom wpływają na nasze wartości, zasady przewodnie, światopoglądy, relacje społeczne i społeczność. Jednak adekwatność nie jest czymś oczywistym: zawsze musi być rozumiana jako zadanie, jako stawianie się adekwatnym” (Puzio 2023, 31). Taki pozytywny i dialogiczny charakter badań na styku teologii i sztucznej inteligencji zgodny jest ze sposobem uprawiania teologii, o jakim mówi papież Franciszek w konstytucji *Veritatis gaudium*. W kontekście reformy studiów kościelnych postuluje on dialog, otwierający na wymianę darów pomiędzy osobami różnych przekonań religijnych i humanistycznych, a także naukowcami różnych dyscyplin, zarówno wierzącymi, jak i niewierzącymi. Jak podkreśla papież dialog ten powinien być nie jedynie taktyczny, ale autentyczny jako nieodłączny wymóg, by zyskać wspólnotowe doświadczenie radości prawdy i pogłębiać jej praktyczne implikacje (Franciszek 2017,4b). Z tego wynika dla Franciszka kolejny aspekt odnowienia studiów kościelnych, a mianowicie transdyscyplinarność (coś więcej niż interdyscyplinarność), która wszelką wiedzę umiejscawia w obrębie mądrości płynącej z Objawienia. Co eschatologicznie dane jest w Jezusie Chrystusie w całym bogactwie mądrości

i wiedzy, w historii odsłania się w wielopostaciowym bogactwie rzeczywistości widzianej w świetle „zamyśłu Chrystusa”. Jak pisze Zatwardnicki wg Franciszka: „W ten sposób studia kościelne miałyby służyć nie tyle integracji pokawałkowanej wiedzy, ile nadawać pewne ukierunkowanie przy poszanowaniu istniejących napięć” (Zatwardnicki 2018, 243). Badania na styku teologii i sztucznej inteligencji stanowią ogromną szansę dla tak rozumianej inter- i trans- dyscyplinarności. Na gruncie samej teologii można zadawać pytania o sztuczną inteligencję w obrębie różnych jej dziedzin: antropologii, teologii fundamentalnej, teologii moralnej, filozofii, etyki oraz katolickiej nauki społecznej. Dyskurs na tym polu już trwa, niemniej chciałbym przedstawić w dalszej części propozycję podjęcia tematów, które wydają się obecnie najbardziej interesujące. Opracowań w tych obszarach jest ciągle mało. Proponowana lista zagadnień nie jest bynajmniej wyczerpująca, ale ma za zadanie zilustrowanie przykładami możliwych obszarów badawczych na gruncie dyscypliny nauk teologicznych.

## 2. Ani sztuczna, ani inteligentna

Dyskurs wokół różnorodnych zagadnień techniki, a sztucznej inteligencji w szczególności, naznaczony jest w dużym stopniu różnego rodzaju antropomorfizacją i animizacją. Jeśli weźmiemy pod uwagę przedstawiony już kontekst projektowania systemów SI jako naśladowania inteligentnych działań człowieka, jest to najzupełniej zrozumiałe. Na gruncie różnych dyscyplin naukowych powstają nieustannie neologizmy, starające się wydobywać nowe wymiary prezentowanych zagadnień. Nadal jednak w większości używamy wobec nowych fenomenów, związanych z technologicznymi artefaktami, pojęć, które oryginalnie powstały w odniesieniu do ludzi i zwierząt. Używamy więc do opisu zjawisk związanych z algorytmami, takich pojęć jak: uczenie się, autonomia, tworzenie, a nawet odpowiedzialność, wnioskowanie, czy wreszcie inteligencja. Ta antropomorfizacja posuwa się jeszcze dalej w języku potocznym i mówimy o inteligentnych zegarkach czy lodówkach lub zastanawiamy się komu sztuczna inteligencja może odebrać pracę. Nadawanie cech ludzkich algorytmom odbywa się więc na zasadzie *sui generis* redukcji pojęć. Wprawdzie istnieją systemy filozoficzne na gruncie których redukcja ta jest nie istotowa czy jakościowa, a jedynie ilościowa. Niemniej do czasu powstania AGI, które jak podkreślałem wcześniej, ma charakter mocno spekulatywny, redukcja taka zawsze będzie występować. Z tego powodu definiowanie różnych pojęć związanych z SI, zwłaszcza w dyskursie naukowym, wydaje się szczególnie istotne. Zjawisko antropomorfizacji sztucznej inteligencji choć naturalne i często użyteczne z jednej strony, bywa naznaczone wieloma pu-

łapkami z drugiej. Istnieje bowiem ryzyko, że sztuczną inteligencję będziemy traktować jako wyabstrahowane z człowieka, zaimplementowane technicznie, subiektywnie najlepsze jego własności. Crawford w swojej książce *Atlas of AI* w tym kontekście argumentuje przeciwko zbyt zawężonemu rozumieniu tych własności. Ponadto przekonuje, że systemy SI do swojego funkcjonowania potrzebują całego szeregu realnych i naturalnych zasobów, co rodzi wiele poważnych wyzwań. Doprowadza w ten sposób do sformułowania tezy, że sztuczna inteligencja *de facto* nie jest ani inteligentna, ani sztuczna (Crawford 2021, 7). Floridi natomiast zwraca uwagę na fakt, że obserwacja rozwoju algorytmów SI prowadzi do wniosku, że sztuczna inteligencja nie jest małżeństwem, ale rozwodem między zdolnością do rozwiązania problemu lub poradzenia sobie z zadaniem, a potrzebą bycia inteligentnym podczas wykonywania tego zadania (Floridi 2023, XIV). Dlatego wg Floridiego, w przypadku sztucznej inteligencji, lepiej jest mówić o agencji niż inteligencji. Agencja jest tu rozumiana jako zdolność do oddziaływania w świecie, w sposób, w którym agent potrafi:

1. odbierać i wykorzystywać dane ze środowiska za pomocą czujników lub innych form wprowadzania danych;
2. podejmować działania w oparciu o dane wejściowe, autonomicznie, aby osiągnąć cele, za pomocą urządzeń lub innych form wyjścia;
3. poprawiać swoją wydajność, ucząc się na podstawie swoich interakcji (Floridi 2023, 10).

Mamy więc do czynienia z behawioralnym lub funkcjonalnym ujęciem definicji sztucznej inteligencji w odróżnieniu od ujęcia ontologicznego. Rozważania nad ontologiczną naturą inteligencji stanowią osobny temat. Jak zauważa Benanti pole pytań o różnice pomiędzy czymś sztucznym, a naturalnym stanowią w świecie artefaktów przestrzeń coraz bardziej skomplikowaną. Jako pewną linię demarkacyjną, a nawet początek nowej epoki podaje datę 16 lipca 1945 roku, kiedy to wyprodukowano pierwszy sztuczny diament. Różnił się on od naturalnego tylko wygrawerowanym laserowo numerem seryjnym oraz tym, że pozbawiony był wszystkich niedoskonałości (Benanti 2021, 184). Dla filozofii, nauk humanistycznych i teologii rodzi się nieuchronne pytanie: czy zmierzamy w kierunku rzeczywistości, w której rozróżnienie między tym, co naturalne, a tym, co sztuczne, jest skazane na zanik? Jeśli tak, to jakie będą konsekwencje tego nowego rozumienia rzeczywistości? A jakie perspektywy otworzy? – pyta Benanti. Na polu badań sztucznej inteligencji, w przypadku podejmowania prób takiego rozróżnienia pomiędzy naturalnym a sztucznym, istnieją zresztą rozmaite podejścia. Do najpopularniejszych należą różnego rodzaju testy. Najsłynniejszym z nich jest „Test Turinga”. Turing postawił pytanie: czy maszyny potrafią myśleć? I jako drogę rozwiązania tego dylematu



zapropował stworzenie gry nazwanej w oryginale *imitation game*. W dużym uproszczeniu: jeśli maszyna konwersująca z człowiekiem na odległość potrafiłaby wywołać w nim przekonanie, że rozmawia z innym człowiekiem test byłby zdany, a w rezultacie maszyna powinna być uznana za posiadającą zdolność myślenia (Turing 1950, 433-460). Istnieją również inne podejścia. Dla przykładu „Lovelace Test” skupia się na zdolności algorytmów do przejawiania kreatywności (Bringsjord et al. 2001), a Schneider proponuje różne sposoby badania istnienia świadomości u maszyn (Schneider 2019, 51-65). Pojawiają się także argumentacje odwołujące się do intuicyjnego poznania. Jednym ze znaczących tego typu przypadków jest historia Blake’a Lemoine’a, inżyniera i byłego pracownika firmy Google. Lemoine, na bazie własnych konwersacji z jednym z Wielkich Modeli Językowych (ang. Large Language Models) – LaMDA, doszedł do przekonania, że algorytm ten uzyskał świadomość (ang. sentience). Co ciekawe powoływał się przy tym na swego rodzaju poznanie religijne. Lemoine, który odnosi się do siebie jako chrześcijański duchowny (ang. christian priest) napisał: „Nie ma żadnych naukowych dowodów na to, czy LaMDA jest świadoma, ponieważ nie istnieje żadna zaakceptowana, naukowa definicja „czucia (sentience)”. Wszyscy zaangażowani, łącznie ze mną, opierają swoją opinię na temat tego, czy LaMDA jest świadoma, na ich osobistych, duchowych i/lub religijnych przekonaniach” (Lemoine 2022). Te różnorodne podejścia do prób określania, czy mamy do czynienia z naturalną inteligencją czy jedynie z, jak je nazywa Schneider – *AI Zombie*, spotykają się oczywiście także z krytyką. Tutaj również można przytoczyć przykłady rozmaitych podejść. Dreyfus w odniesieniu do pytania o możliwość istnienia świadomości u sztucznej inteligencji wyklucza ją, wywodząc swoje stanowisko od myśli Heideggera. Argumentuje, że algorytmy w istocie są jedynie cyfrowymi modelami świata, podczas gdy najlepszą reprezentacją świata jest świat sam w sobie, a komputerowe algorytmy nie posiadają zdolności doświadczania świata, którą Heidegger nazywa *dasein* (Dreyfus 1992, xxxi). Z kolei Marks argumentuje, że powstanie świadomości u maszyn jest niemożliwe odwołując się do faktu, że istnieją na gruncie algorytmiki tzw. problemy nierozstrzygalne, do których zalicza on m.in. właśnie świadomość oraz kreatywność (Marks 2022, 22-29). Przytaczam te zagadnienia jedynie w dużym skrócie po to, by unaocznic ich złożoność, różnorodność oraz interdyscyplinarny charakter. Sednem tych dyskursów wydaje się być natomiast przywoływane już rozróżnienie Floridiego pomiędzy inteligencją, a agencją lub w innym ujęciu: między istnieniem, a funkcjonowaniem. Tutaj jednak ujawnia się kolejny poziom złożoności zagadnienia. Ponieważ pomiędzy tymi pojęciami istnieje napięcie, które Benasayag tak opisuje: „Problemem dzisiaj jest fakt, że w ramach tej zintegrowanej całości chciałoby się w sposób całkowicie sztuczny oddzielić procesy funkcjonowania od procesów

istnienia, posuwając się nawet do skutecznego zanegowania tych ostatnich. Tekst ten nie broni idei, że należy wybrać jeden lub drugi z tych wymiarów, ale raczej konieczności powrotu do tej złożonej jedności. Między tymi dwoma wyimaginowanymi biegunami istnieje wszystko” (Benasayag 2019, 14). Z jednej strony to człowiek tworzy technologiczne artefakty, ale ponieważ w stechnicyzowanym świecie doświadcza rzeczywistości w sposób coraz bardziej przez nie zapośredniczony – to pod wpływem tego procesu – gruntownej przebudowie ulega także sam świat. Floridi nazywa to zjawisko reontologizacją. Jest to bardzo radykalna forma przekształcenia, która nie tylko projektuje, konstruuje lub buduje system (np. firmę, maszynę lub artefakt) na nowo, ale także fundamentalnie przekształca jego wewnętrzną naturę. W tym sensie, na przykład, nanotechnologie i biotechnologie nie tylko przeprojektowują, ale także reontologizują nasz świat (Floridi 2007, 4). Użycie artefaktu przekształca aktywność, dla której został zaprojektowany. Transformacja ta dotyczy zarówno reorganizacji percepcyjno-motorycznych sposobów interakcji z otoczeniem, jak i sposobów planowania działań i relacji społecznych. Dlatego też spojrzenie na artefakty jest sposobem spojrzenia na człowieka i jego istotę, sposobem badania ludzkiej natury (Benanti 2016, 17). Odnosząc to do działania systemów sztucznej inteligencji, można zauważyć, że mamy do czynienia z pewnym procesem algorytmizacji życia człowieka, polegającym na tym, że zarówno jego działanie, jak i istnienie stają się funkcją optymalizacji celu. Oprócz zjawisk pozytywnych i neutralnych niesie to ze sobą wiele wyzwań. Można to dla przykładu zaobserwować w niektórych środowiskach pracy, gdzie miast zastępować uciążliwą lub niebezpieczną pracę ludzi pracą robotów, za pomocą systemów sztucznej inteligencji nadzoruje się i optymalizuje pracę ludzi, prowadząc niejako do ich „robotyzacji”. To przykład dość skrajny. Istnieją również bardziej subtelne, znane z codziennego życia, kiedy to np. dostosowuje się do działania maszyn całe środowisko życia ludzi, jak wtedy, kiedy ktoś mebluje i urządza mieszkanie w sposób ułatwiający działanie tzw. robotom sprzątającym. Technologie sztucznej inteligencji mogą także w sposób głębszy, a przez to mniej jeszcze dostrzegalny kształtować nasze myślenie o świecie i ludziach. Można to zaobserwować we współczesnym, głównie zachodnim, stosunku do starości i ludzi starszych. W zalgorytmizowanym świecie to optymalizacja wydaje się głównym celem, a zwiększanie efektywności główną wartością. W takim świecie rola oraz wartość osób starszych ulegają coraz większej degradacji. Zbliżający się do schyłku życia ludzie tracą jednak wiele ze swej wydajności. Nie ma tu miejsca na słabość, a nawet na kruchość. Tymczasem życie w hiperpołączonym i stechnicyzowanym świecie coraz częściej uświadamia nam potrzebę i znaczenie relacji. W tej sytuacji być dla kogoś może znaczyć daleko więcej niż tylko funkcjonować wydajnie dla kogoś. Teologia, która docenia życie człowieka w całej jego złożoności, a więc także

w kruchości, może odegrać tutaj swoją istotną rolę. Także jako platforma do przekraczania utartych schematów myślenia, otwierając drogę, na której procesów starzenia się i pęknięć nie można oceniać w kategoriach zysku lub straty (Benasayag 2019, 14). Może to stanowić zadanie dla takich dyscyplin teologicznych, jak antropologia, teologia fundamentalna i filozofia. Teologia, oparta na chrześcijańskim przesłaniu o zbawieniu, posiada unikalną wizję człowieka i jego przeznaczenia. Wychodzi daleko poza traktowanie ludzi jako sprawnych maszyn, na których głównym ich nośniku informacji – mózgu – znajdują się zapisane dane, mające stanowić największy walor człowieka. Natomiast myślenie teologiczne może otworzyć człowieka na jego szerszy wymiar i dostrzec w Jezusie z Nazaretu wzorcową istotę ludzką, jedyną, która może objawić człowieka człowiekowi. W rzeczywistości jest On ostatnią istotą ludzką, eschatycznym Adamem, ponieważ wprowadza ludzi w ich przyszłość (Benanti 2019). Teologia wątki tej refleksji może więc wprowadzać w zupełnie nowe wymiary i otwierać nowe perspektywy. Może dyskusję wokół sztucznej inteligencji prowadzić na tory, gdzie zamiast technologizacji człowieka, następują procesy autentycznej humanizacji technologii.

### 3. Etyczne maszyny

Bardzo istotnym wątkiem pytania o sztuczną inteligencję jest wymiar etyczny. Choć SI może działać jako agencja w świecie, systemy te są tworzone przez ludzi w określonym celu i w określony sposób. Rodzi się na tym gruncie pytanie, jakie wartości chcemy brać pod uwagę przy ich projektowaniu oraz jaki paradygmat etyczny w nich implementować (Dignum 2019, 94). Naprzeciw tym pytaniom wychodzą wytyczne i kodeksy etyczne proponowane przez różnego rodzaju instytucje. Serwis AlgorithmWatch, katalogujący takie wytyczne według stanu na kwiecień 2020 roku, zawierał 167 tego rodzaju dokumentów (AlgorithmWatch 2020). Jedną z propozycji na tym polu jest apel o etyczny rozwój sztucznej inteligencji, przygotowany przez Papieską Akademię Życia *Rome Call for AI Ethics*. Zawiera on wezwanie o taki rozwój tych technologii, który służy każdemu człowiekowi i całej ludzkości; który szanuje godność osoby ludzkiej, tak aby każda osoba mogła czerpać korzyści z postępu technologicznego. Proponuje on w tym zakresie sześć zasad, którymi powinno kierować się budowanie tego typu rozwiązań:

1. przejrzystość (ang. „transparency”): systemy SI muszą być zrozumiałe dla wszystkich;
2. włączenie (ang. „inclusion”): systemy te nie mogą nikogo dyskryminować, ponieważ każdy człowiek ma równą godność;

3. odpowiedzialność (ang. „accountability”): zawsze musi być ktoś, kto bierze odpowiedzialność za to, co robi maszyna;
4. bezstronność (ang. „impartiality”): systemy SI nie mogą podążać za uprzedzeniami ani ich tworzyć;
5. niezawodność (ang. „reliability”): sztuczna inteligencja musi być niezawodna;
6. bezpieczeństwo i prywatność (ang. „security and privacy”): systemy te muszą być bezpieczne i szanować prywatność użytkowników (Rome Call 2020).

Pierwszymi sygnatariuszami apelu w Rzymie w lutym 2020 roku były takie podmioty, jak: Papieska Akademia Życia, Microsoft, IBM, FAO – Organizacja Narodów Zjednoczonych do spraw Wyżywienia i Rolnictwa oraz ze strony włoskiego rządu: minister ds. innowacji technologicznych. Wydarzenie w kolejnym roku było umieszczone w dorocznym raporcie Uniwersytetu Stanforda *AI Index* jako jeden z najczęściej komentowanych tematów w kontekście etycznego wykorzystania technologii SI w 2020 roku (Zhang et al. 2021, 131). Dokumenty takie, jak rzymski apel, bywają krytykowane jako formułujące zalecenia odnośnie etycznego podejścia do rozwoju technologii na zbyt ogólnym poziomie. Niemniej należy zaznaczyć, że są one tworzone przez różne instytucje nie jako szczegółowe instrukcje dziedzinowe, a jedynie określenie obszarów wyzwań oraz wyznaczenie kierunków rozwiązań. Co więcej wytyczne takie mogą stanowić platformę do efektywnego dialogu społecznego, politycznego, a nawet religijnego. Dobrze ilustruje to także rzymski apel, który stał się podstawą inicjatywy zorganizowanej 10 stycznia 2023 roku w Rzymie zatytułowanej: *AI Ethics: an Abrahamic commitment to the Rome Call*, czyli dołączenia do apelu o etyczny rozwój SI także przedstawicieli innych religii abrahamowych, tj. judaizmu i islamu.

Na gruncie etyki sztucznej inteligencji, oprócz dyskursu wokół podejścia do projektowania i zarządzania tymi systemami, pojawiają się także pomysły, aby w same algorytmy wbudowywać swego rodzaju „intuicje” moralne. Mowa tu o tzw. *Artificial Moral Agents* (AMA), czyli pewnego rodzaju sztucznej agencji moralnej. To podejście jest szeroko krytykowane ze względu na wiele czynników. Główny jednak argument zasadza się na wątpliwości, czy istotnie możliwe w ogóle jest zaimplementowanie czegoś na kształt ludzkiej moralności w algorytmach oraz na zagrożeniach wiążących się z nieudanymi próbami w tym zakresie (van Wynsberghe 2018). Niezależnie od tej krytyki, podejmowane są wysiłki zmierzające do stworzenia AMA. Jednym z przykładów jest próba wytrenowania „uprzejmego robota”. Co ciekawe, autorzy tego pomysłu powołują się na etykę cnót oraz koncepcje zawarte w pracach chrześcijańskiego filozofa Dallasa Willarda (Crook and Cornelli 2021). Budzi

jednak wątpliwości czy przyjęta przez zespół naukowców, jak piszą: „metoda mapowania ludzkiej jaźni wg Willarda na koncepcje neuronauki, a stamtąd na możliwe implementacje” doprowadziła do powstania w pełni funkcjonalnego modelu. To znaczy, czy mamy tu do czynienia istotnie z mapowaniem, czy jedynie ze znaczącą redukcją. Ponownie prowadzi to nas, jak lustro, w stronę pytań o człowieka i jego naturę. Zagadnienia te wydają się szczególnie interesujące do podjęcia przez dyscyplinę nauk teologicznych. Dla przykładu, chrześcijański personalizm zakłada unikalne własności ludzkiej moralności. Jej istotnym elementem jest samostanowienie, będące w pewien sposób aktem twórczym (Stachewicz 2020, 158). Powstaje uzasadnione pytanie, czy algorytmy SI kiedykolwiek będą zdolne osiągnąć tego typu własności. W tym miejscu jedynie skrótowo sygnalizuję te wątki, albowiem o wiele bardziej szczegółowo podejmuję je w ramach przygotowywanej przeze mnie rozprawy doktorskiej, dotyczącej pytania o możliwość podmiotowości moralnej sztucznej inteligencji.

## Zakończenie

Fenomeny, związane ze sztuczną inteligencją, mogą stanowić nowe „miejsca” człowieka i „miejsca” teologii. Oprócz przedstawionych w niniejszym artykule istnieje jeszcze cały szereg tematów, które stanowią szanse dla teologicznej refleksji. Należą do nich m.in. także pytania o wpływ funkcjonowania tych technologii na środowisko naturalne, zagadnienia związane z pracą oraz demokratyzacją dostępu do dóbr, które wytwarzają. Zagadnienia związane z SI mogą być podejmowane przez różne dyscypliny teologiczne, a ich analiza prowadzi nas ponownie do pytań o naturę oraz przeznaczenie człowieka. Istotne jest, aby teologia podejmowała te nowe zadania, do których należy służba na rzecz wiary i kultury, w sposób twórczy. Nie powinna ona jedynie przyjmować postawy konfrontacyjnej czy apologetycznej, ale dostrzegać płynący z nich potencjał.

## BIBLIOGRAFIA

- AlgorithmWatch. 2020. Dostęp: 05.09.2023. <https://inventory.algorithmwatch.org>.
- Benanti, Paolo. 2016. *Homo Faber. The Techno-Human Condition*. Bologna: EDB.
- Benanti, Paolo. 2019. “Artificial Intelligences, Robots, Bio-engineering and Cyborgs: New Challenges for Theology?” *Concilium* (00105236), Issue 3: 34-47.
- Benanti, Paolo. 2021. *La grande invenzione. Il linguaggio come tecnologia, dalle pitture rupestri al GPT-3*. San Paolo Edizioni.
- Benanti, Paolo. 2023. “The urgency of an algorethics.” *Discover Artificial Intelligence*, 3: 11. Springer. <https://doi.org/10.1007/s44163-023-00056-6>.

- Benasayag, Miguel. 2019. *Funzionare o esistere?* Vita e Pensiero.
- Bringsjord, Selmer, Paul Bello and David Ferrucci. 2001. "Creativity, the Turing Test, and the (Better) Lovelace Test." In: *Minds and Machines*, 11: 3-27. <https://doi.org/10.1023/A:1011206622741>.
- Churchil, Winston. 1943. Cytowany w: Volchenkov, Dimitri. 2018. „Grammar of Complexity: From Mathematics to a Sustainable World, World Scientific Publishing Company”. [https://www.worldscientific.com/doi/10.1142/9789813232501\\_0007](https://www.worldscientific.com/doi/10.1142/9789813232501_0007).
- Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. New Heaven and London.
- Crook, Nigel and Corneli. 2021. "The Anatomy of moral agency: A theological and neuroscience inspired model of virtue ethics." *Cognitive Computation and Systems*, Volume 3, Issue 2: 109-122. <https://doi.org/10.1049/ccs2.12024>.
- Dastin, Jeffrey. 2018. *Amazon scraps secret AI recruiting tool that showed bias against women*. Dostęp: 05.09.2023. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
- Dignum, Virginia. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-030-30371-6>.
- Dzidek, Tadeusz i Piotr Sikora. 2018. „Metody”. W: *Poznanie teologiczne*, red. T. Dzidek, Ł. Kamykowski i A. Napiórkowski, 151–174 (*Teologia Fundamentalna*, 5). Kraków: Uniwersytet Papieski Jana Pawła II w Krakowie. Wydawnictwo Naukowe.
- Dreyfus, Hubert. 1992. *What Computers Can't Do: The Limits of Artificial Intelligence*. The MIT Press.
- Encyklopedia PWN. *Sztuczna inteligencja*. Dostęp: 05.09.2023. <https://encyklopedia.pwn.pl/haslo/sztuczna-inteligencja;3983490.html>.
- Floridi, Luciano. 2007. "A Look into the Future Impact of ICT on Our Lives." In: *The Information Society*, Volume 23: 59-64. <https://doi.org/10.1080/01972240601059094>.
- Floridi, Luciano. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. OUP Oxford.
- Franciszek. 2017. *Konstytucja Apostolska „Veritatis gaudium” o uniwersytetach i wydziałach kościelnych*.
- Lemoine, Blake 2022. *Scientific Data and Religious Opinions*. Dostęp: 05.09.2023. <https://cajundiscordian.medium.com/scientific-data-and-religious-opinions-ff9b0938fc10>.
- Marks, Robert J. 2022. *Non-Computable You: What You Do That Artificial Intelligence Never Will*. Discovery Institute Press.
- OpenAI. 2023. „About”. Dostęp: 05.09.2023. <https://openai.com/about>.
- Puzio, Anna. 2023. "Theology Meets AI: Examining Perspectives, Tasks, and Theses on the Intersection of Technology and Religion." In: *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz Theologie und Künstliche Intelligenz*, Vol. 1. Darmstadt: wbg Academic.
- Rome Call for AI Ethics*. 2020. Dostęp: 05.09.2023. <https://www.romecall.org/the-call/>.
- Schneider, Susan. 2019. *Artificial you: AI and the future of your mind*. Princeton University Press. New Jersey & Oxfordshire.
- Stachewicz, Krzysztof. 2020. „Karol Wojtyła's philosophy of freedom". *Teologia i Moralność*, 15(2020), nr 1(27). DOI: 10.14746/tim.2020.27.1.10.
- Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind*, vol. LIX, no. 236.
- Wynsberghe van, Aimee and Robbins Scott. 2018. "Critiquing the Reasons for Making Artificial Moral Agents." *Sci Eng Ethics*, 25(2019): 719-735. <https://doi.org/10.1007/s11948-018-0030-8>.
- Zatwardnicki, Sławomir. 2018. „Wizja teologii w wypowiedziach papieża Franciszka". *Rocznik Teologii Katolickiej*, t. XVII, 3: 237-257. DOI: 10.15290/rtk.2018.17.3.17.
- Zhang, Daniel et al. 2021. *The AI Index 2021 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford.

**MACIEJ MRÓZ** – doktorant w Szkole Doktorskiej Nauk Humanistycznych Uniwersytetu im. Adama Mickiewicza w Poznaniu, dyscyplina: nauki teologiczne. Magister teologii oraz magister informatyki. Zawodowo zajmuje się rozwijaniem systemów przetwarzania dużych zbiorów danych (Big Data) oraz systemów uczenia maszynowego. Przygotowuje rozprawę doktorską z obszaru etyki sztucznej inteligencji pytającą o możliwość moralnej podmiotowości sztucznych agentów.