

Andrzej Radomski<http://orcid.org/0000-0002-1735-605X>

Uniwersytet Marii Curie-Skłodowskiej w Lublinie

andrzejradomski64@gmail.com

DOI: 10.35765/pk.2022.3904.21

Sztuczna inteligencja w badaniu wybranych aspektów kultury

STRESZCZENIE

Sztuczna inteligencja jest jednym z najważniejszych elementów współczesnego świata. Funkcjonuje już ona w podstawowych dziedzinach życia. Jest także wykorzystywana w różnych dyscyplinach naukowych. Od pewnego czasu zaczyna się również stosować algorytmy sztucznej inteligencji do badania świata kultury. W artykule zostały ukazane najważniejsze metody i narzędzia sztucznej inteligencji stosowane w badaniach różnych zjawisk kultury, a także przykłady takich badań. Badacze kultury wykorzystują sztuczną inteligencję w dwóch obszarach. Pierwszym jest przetwarzanie języka naturalnego, drugim komputerowe badanie obrazów (filmów, zdjęć, grafiki itp.). Sztuczna inteligencja jest częścią uczenia maszynowego. Wykorzystuje ono konwolucyjne sieci neuronowe, duże ilości danych (big data) i superkomputery do analizy i wizualizacji cyfrowych wytworów. Badania kultury za pomocą sztucznej inteligencji są zaliczane do nowego paradygmatu zwanego humanistyką cyfrową.

SŁOWA KLUCZE: sztuczna inteligencja, uczenie maszynowe, kultura, humanistyka cyfrowa

ABSTRACT

Artificial Intelligence in the Study of Selected Aspects of Culture

Artificial intelligence is one of the most important elements of the modern world. It already functions in the most important areas of life. It is also used in various scientific disciplines. For some time now, artificial intelligence algorithms have been used to study the world of culture. The article presents the most important methods and tools of artificial intelligence that are used in the study of various cultural phenomena, as well as examples of such studies. Cultural researches use artificial intelligence in two areas. The first one is natural language processing, while the other – the computerized examination of images (movies, photos, graphic, etc.). Artificial intelligence is part of machine

learning. It uses convolutional neural networks, big data and supercomputers to analyze and visualize digital product. Artificial intelligence research into culture belongs to a new paradigm called: digital humanities

KEYWORDS: artificial intelligence, machine learning, culture, digital humanities

Wstęp

Sztuczna inteligencja (ang. AI) jest jednym z najważniejszych zjawisk współczesnej kultury. Wkracza ona w coraz to nowe obszary życia i znajduje zastosowanie w różnych zawodach. Stała się także przedmiotem dyskusji naukowych, filozoficznych i etycznych, a także modnym tematem różnych mediów. Niewątpliwie na ten stan rzeczy złożyły się spektakularne sukcesy AI w różnych praktykach życia społecznego. Pierwszym takim sukcesem, który odbił się szerokim echem w świecie i zwrócił uwagę opinii publicznej na to zjawisko, było zwycięstwo superkomputera Deep Blue w grze szachy nad szachowym mistrzem świata Garim Kasparowem (1997). Od roku 2010 odnotowuje się szereg kolejnych sukcesów AI. Potrafi ona komponować muzykę, malować obrazy, dobierać spersonalizowane reklamy czy komunikować się z internautami na tzw. chatbotach. I są to tylko wybrane przykłady jej zastosowań. W końcu AI trafiła także do praktyki naukowej.

Sztuczna inteligencja (jeśli chodzi o naukę) jest wykorzystywana m.in. do poszukiwania nowych planet, do diagnozowania nowotworów, ustalania źródeł epidemii czy odkrywania nowych leków. Jej spore osiągnięcia na polu dyscyplin przyrodniczych stwarzają okazję do pokazania określonych zastosowań AI na gruncie humanistyki, której przedmiotem badania jest świat kultury. I to jest główny cel niniejszego tekstu. Oczywiście trudno jest poruszyć w krótkim artykule wszystkie wchodzące w grę aspekty stosowania algorytmów AI do badania zjawisk kulturowych. Toteż główna uwaga została skupiona na tych projektach, które rokują (obecnie) największe nadzieje i które mają już spore osiągnięcia. W dalszej więc części uwaga została skoncentrowana na: a) przetwarzaniu języka naturalnego (ang. natural language processing, NLP) oraz b) rozpoznawaniu obrazów (ang. computer vision, cv).

Skuteczność aplikowania algorytmów AI do badania wybranych problemów kultury została zilustrowana przede wszystkim na przykładzie badań własnych autora. W części pierwszej zarysowano podstawy działania sztucznej inteligencji, w drugiej podstawowe narzędzia i metody, w trzeciej przykłady wspomnianych badań własnych autora, a w ostatniej

wnioski ogólniejszej natury – jeśli chodzi o możliwości stosowania AI w humanistyce do badania określonych zjawisk kultury.

1. Sztuczna inteligencja – czyli uczenie maszynowe

Idea sztucznej inteligencji pojawiła się jeszcze w latach 50. XX w. Punktem wyjścia leżącym u podstaw prac nad AI był zamiar stworzenia inteligentnego robota, który przypomina wyglądem człowieka i potrafi myśleć tak jak *homo sapiens*. Od razu jednakże należy zauważyć, że nie chodziło tu o zwykły program komputerowy. Komputer bowiem nie myśli, nie wyciąga wniosków, nie uczy się i nie tworzy nowych pomysłów. Wykonuje tylko określony ciąg ściśle zaprogramowanych w jakimś języku (typu: C++, Java czy Ruby) instrukcji. Tymczasem podstawą funkcjonowania sztucznej inteligencji jest zabieg mający na celu automatyzację procesu myślowego standardowo wykonywanego przez ludzi. Tak rozumiana sztuczna inteligencja jest ogólną dziedziną, na którą składa się uczenie maszynowe i uczenie „głębokie” (Chollet, 2018, s. 22). Prace nad sztuczną inteligencją mają charakter interdyscyplinarny. Prowadzone są na styku informatyki, matematyki, cybernetyki, językoznawstwa, psychologii i filozofii.

Za początek prac nad sztuczną inteligencją powszechnie uważa się konferencję w Dartmouth (USA), która odbyła się w roku 1956. Zgromadziła ona 10 badaczy, którzy przez dwa miesiące dyskutowali o możliwości skonstruowania sztucznej inteligencji będącej w stanie w przyszłości dorównać człowiekowi. Uczestnicy konferencji z dużym optymizmem spoglądali na możliwość skonstruowania maszyny, która intelektualnie dorównywałaby człowiekowi. Już w roku 1956 zaprezentowano pierwszy program grający w warcaby na poziomie turniejowym, a w roku 1966 program Eliza symulujący konwersację ze zrozumieniem (Kisielewicz, 2017, s. 42–43). Szybko też zaczęto tworzyć specjalistyczne laboratoria AI – najpierw w uniwersytetach: Carnegie Mellon, Stanford i MIT, a później w kolejnych. Jednakże do końca XX stulecia nie zanotowano jakichś spektakularnych rezultatów, jeśli chodzi o postęp w tej dziedzinie (za wyjątkiem wspomnianego Deep Blue). Przełom nastąpił w pierwszej dekadzie XXI stulecia. Złożyły się na niego trzy czynniki: a) skonstruowanie konwulucyjnych sieci neuronowych (ang. CNN), b) gwałtowny przyrost różnego typu danych – zwłaszcza w Internecie (big data), c) pojawienie się komputerów o dużej mocy obliczeniowej (np. Watson firmy IBM).

Jak wspomniano wyżej, sztuczną inteligencję zalicza się do tzw. uczenia maszynowego. Obecnie wyróżnia się dwa podstawowe typy uczenia maszynowego: klasyczne (oparte na statystyce) i tzw. głębokie (ang. deep

learning). W przypadku klasycznego uczenia maszynowego człowiek wprowadza dane, a także oczekiwane odpowiedzi, a komputer ma utworzyć reguły pozwalające na ich uzyskanie. Reguły te mogą zostać użyte w celu przetworzenia nowych danych i uzyskania oczekiwanych odpowiedzi (Fłasiński, 2018, s. 23).

Podstawą „głębokiego” uczenia maszynowego są tzw. sieci neuronowe. Taka sieć składa się z trzech podstawowych elementów: danej liczby neuronów, określonej liczby warstw i odpowiednich rodzajów połączeń między warstwami. Najbardziej znanym i rozpowszechnionym typem sieci neuronowej jest jednokierunkowa sieć wielowarstwowa. Składa się ona z warstwy wejściowej, jednej lub kilku warstw ukrytych i jednej warstwy wyjściowej. Każda warstwa składa się z innej liczby neuronów i jest silnie powiązana z sąsiednią warstwą. Sieć uczy się, wykorzystując algorytm propagacji wstecznej (Patterson, 2020, s. 79). „Głębokie” uczenie maszynowe może mieć postać nadzorowanego i nienadzorowanego. W przypadku tego pierwszego, tworząc model ML (ang. machine learning), wprowadzamy dane z etykietami (np. człowiek czy samochód) i model uczy się rozpoznawać nowe obiekty (np. samochody), posługując się danymi z naszymi etykietami. W przypadku maszynowego uczenia nienadzorowanego sieć nie otrzymuje wyniku końcowego (np. etykiety zdjęcia samochodu), tylko analizując dane wejściowe, musi „odgadnąć” albo wyłonić grupy obiektów lub/i jakieś występujące w tym zbiorze wzorce (np. pojazdy określonego koloru).

„Głębokie” uczenie maszynowe składa się z trzech głównych działań: a) piszemy określony program bądź wykorzystujemy określoną bibliotekę (w rodzaju Scikit-Learn czy Keras, b) następnie wprowadzamy określone dane i na wyjściu, c) otrzymujemy jakąś wiedzę. Zatem w celu wykorzystania AI musimy zgromadzić: a) dane wejściowe (np. obrazy, dźwięki czy słowa), b) oczekiwane wyniki (np. etykiety zdjęcia samochodu, budynku, kobiety masażystki etc.), c) opracować sposób automatycznego pomiaru działania algorytmu (np. procentową wartość poprawnie sklasyfikowanych zdjęć). Uczenie głębokie tym będzie różnić się od tego „klasycznego” (np. jednowarstwowego), że będzie zawierało wiele warstw. Współcześnie mogą już być ich setki (Chollet, 2018, s. 24–25). Analiza danych przez każdą kolejną warstwę ma zapewnić lepszy wynik końcowy (czyli oczekiwany na wyjściu, np. poprawne odgadnięcie naszego samochodu czy kobiety masażystki). Uczenie tego typu składa się tu z wielu cykli (epoche) i kończy się, gdy otrzymamy automatyczną informację, że nasz model z dużym stopniem prawdopodobieństwa rozpoznaje dane obiekty (np. 90-procentowym).

Za pomocą algorytmów uczenia maszynowego (niezależnie od jego wariantu) możemy: a) gromadzić i „oczyszczać” różne kolekcje danych i to niezależnie od ich skali, b) automatycznie je analizować,

c) automatycznie je wizualizować. Automatyka maszynowa jest tutaj kluczowym elementem, gdyż „ręcznie” w żaden sposób nie ogarniemy tysięcy (nie mówiąc już o większych kolekcjach) tekstów, zdjęć czy filmów.

Obecnie dominują dwa podejścia w pracy z AI. Pierwsze polega na tradycyjnym programowaniu sieci neuronowych – z wykorzystaniem popularnych bibliotek programistycznych typu: Sckit-Learn, TensorFlow czy Keras. Drugie to rozwiązania typu AutoML, w ramach których ładujemy własne dane, a sieć je trenuje i analizuje. To drugie podejście nie wymaga umiejętności programistycznych po stronie użytkownika. W tym przypadku bowiem AI rozwija i udoskonala samą siebie. Na rynku pojawiło się wiele tego typu usług. Można tu wspomnieć o Microsoft Azure, Amazon Web Service (AWS) czy Google Vison.

2. Narzędzia i metody

Mniej więcej od roku 2010 obserwujemy duży skok jakościowy w rozwoju AI. Ma to również spore znaczenie dla humanistyki i jej przedmiotu badań: kultury. Dynamiczny rozwój AI w ostatnich latach jest konsekwencją upowszechnienia się konwolucyjnych sieci neuronowych. Innym wspomnianym tu już czynnikiem jest ogromny przyrost danych, na których można trenować sieci neuronowe. Dzięki nim doszło do udoskonalenia analizy języka naturalnego i rozpoznawania zdjęć. W tym pierwszym przypadku chodzi o rozpoznawanie mowy, rozumienie języka naturalnego oraz syntezę mowy. Jeśli chodzi o rozpoznawanie mowy, to zaowocowało ono powstaniem coraz doskonalszych narzędzi tłumaczenia (Google Translate czy Deep L), asystentów głosowych typu: Cortana, Alexa, Siri czy chatbotów. Jeśli chodzi o syntezę mowy, to powstały systemy przekładające mowę na tekst i tekst na mowę. Przykładem takiej usługi może być superkomputer Watson firmy IBM. W przypadku rozumienia języka naturalnego udoskonalono analizę struktury gramatycznej zdań opierającą się na analizie syntaktycznej.

W ostatnich latach tworzone bądź udoskonalane są algorytmy AI, które stanowią swoistą odpowiedź na potrzeby analizy zdigitalizowanych wytworów kultury oraz tzw. wytworów cyfrowych, które są produkowane masowo w różnych praktykach społeczeństwa informacyjnego i sieciowego. Zdigitalizowane wytwory dziedzictwa kulturowego i historycznego (jak i również współczesne ślady działalności człowieka) można bowiem poddać analizie za pomocą współczesnych technologii – w tym uczenia maszynowego.

Przetwarzanie języka naturalnego zajmuje się automatyzacją analizy, rozumienia, tłumaczenia i generowania języka naturalnego przez komputer. Obejmuje ono m.in.: a) klasyfikowanie dokumentów (mail, tweet,

recenzja etc.), b) tłumaczenie maszynowe c) wyszukiwanie treści, d) rozpoznawanie mowy (np. interpretowanie poleceń głosowych), e) prowadzenie konwersacji przez tzw. chatboty, f) analizę sentymentu (Krohn, 2022, s. 23–24). Przykładem takich rozwiązań są np. Google Translate czy Google Duplex.

Podstawą NLP jest tokenizacja – czyli przekształcenie jakiegoś dokumentu (książki, artykułu, przemówienia itp.) na listę dyskretnych elementów języka. Tokenami mogą być litery, słowa czy zdania (Boschetti i Massaron, 2017, s. 221). Następnie możemy dokonać wyszukiwania n-gramów. N-gramy to ciągi sąsiadujących jednostek językowych – na przykład kolejnych słów w zdaniu. W przetwarzaniu języka naturalnego n-gramy mogą być wykorzystane do wykrywania grup słów mających tendencję do występowania obok siebie (Deitel, 2020, s. 421). Ponadto, możemy dokonać statystycznej analizy tekstu czy jakiegoś korpusu w rodzaju: najczęściej występujące słowa, liczby słów, przeciętnej liczby słów w zdaniu itp. (Deitel, 2020, s. 407). Przykładem programów do tego typu przedsięwzięć mogą być Voyant czy JStor. Jeśli chcemy dokonywać bardziej pogłębionych analiz (np. analizy sentymentu), to możemy skorzystać z dedykowanych bibliotek programistycznych – typu: TextBlob czy Textastic (w Python).

Innym przykładem zastosowania uczenia maszynowego w badaniu korpusów językowych są analizy stylometryczne. Stylometria to dziedzina zajmująca się badaniem właściwości stylów: literackiego, naukowego, publicystycznego i innych poprzez ilościową analizę tekstu. Opiera się ona na założeniu, że każdy z nas ma unikatowy styl pisania, który jest kształtowany przez całe życie (Vaughan, 2022, s. 52). Jako czytelnicy czasem również jesteśmy zdolni do rozpoznania konkretnej osoby po danym tekście: jeśli wcześniej mieliśmy do czynienia z odpowiednią ilością tekstu, który wyszedł spod ręki tej osoby, to rozpoznamy często używane zwroty, szyk zdania czy sposób używania interpunkcji. W ramach uczenia maszynowego dostarczamy programowi pewne dane wejściowe, na podstawie których porównuje on te elementy w czasie. Musi mieć przestrzeń cech, w której tworzy sobie pewne wektory (czyli uporządkowane zbiory cech). Opisują one profil danej osoby, a kolejne teksty są z tym profilem porównywane (Vaughan, 2022, s. 63–69). Nazywa się to stylometrycznym profilem behawioralnym.

Możliwości, jakie oferuje sztuczna inteligencja w zakresie NLP, są wykorzystywane do badań na gruncie humanistyki. Jednym z największych osiągnięć AI na tym polu była pomoc w odczytaniu pisma kreteńskiego. Konkretnie, chodziło o odczytanie starożytnego pisma kreteńskiego (tzw. linearnego B). Stosowny algorytm został opracowany przez naukowców z MIT. Badacze postanowili wykorzystać wiedzę o tym, że w trakcie ewolucji systemu językowego najwolniej dokonują się zmiany

w składni. Zauważyli także, że niektóre zapisy poszczególnych słów czy symboli nie zmieniły się w przeciągu stuleci. Co więcej, deszyfrację ułatwiłoby odkrycie relacji, w jakiej pozostaje badany język w stosunku do swojego przodka. Koncentrując się na tych trzech kwestiach, naukowcy postanowili użyć sztucznej inteligencji do odczytania pisma linearnego B (pochodzi z ok. XIV w. p.n.e., a rozszyfrowane zostało w 1953 r.) oraz języka ugaryckiego (wymarły od XII w. p.n.e., a odczytany na początku XX w.). Wynik eksperymentu oceniono jako zadowalający – gdyż algorytm (pracując sposobem głębokiego uczenia maszynowego) odczytał poprawnie 67,3% pisma linearnego B (Kowal).

Drugi ważny dla badaczy kultury obszar zastosowań AI dotyczy kultury wizualnej. Chodzi tu o analizę obrazów ruchomych (filmy, animacje, gry) oraz statycznych – typu zdjęcia i inna grafika. Tym zajmują się narzędzia i metody – określane jako computer vision (cv).

Computer vision można określić jako identyfikację i przetwarzanie obiektów na obrazach i filmach w taki sam sposób bądź podobny jak ludzie. Computer vision w standardowej definicji to interdyscyplinarna dziedzina naukowa, która zajmuje się tym, jak komputery mogą uzyskać wysoki poziom wiedzy na podstawie cyfrowych obrazów lub filmów. Z punktu widzenia inżynierii stara się zrozumieć i zautomatyzować zadania, które może wykonać ludzki system wzrokowy. Podstawowy obszar działania CV obejmuje: metody pozyskiwania, przetwarzania, analizowania i rozumienia obrazów cyfrowych oraz ekstrakcji danych wielowymiarowych ze świata rzeczywistego w celu uzyskania informacji liczbowych lub symbolicznych¹.

Podstawą działania metod i narzędzi CV jest uczenie maszynowe. Do niedawna widzenie komputerowe działało tylko w ograniczonym zakresie. Dzięki postępom sztucznej inteligencji i innowacjom w głębokim uczeniu się i sieciach neuronowych udało się, w tej dziedzinie, w ostatnich latach dokonać ogromnego skoku i przewyższyć ludzi w niektórych zadaniach związanych z wykrywaniem i znakowaniem obiektów. W wyszukiwarkach internetowych bowiem gdy poszukujemy określonego zdjęcia bądź całej ich kolekcji, to algorytmy szukają po tagach i innego typu etykietach. Jest to jednakże zawodna metoda. Dla przykładu, gdy chcemy znaleźć serię zdjęć określonej postaci – powiedzmy Olgi Tokarczuk (polskiej noblistki) – to możemy otrzymać taki rezultat jak na zdjęciu nr 1.

1 Jest to definicja zaczerpnięta z Wikipedii: https://en.wikipedia.org/wiki/Computer_vision



Zdjęcie nr 1. Zrzut z ekranu (grafika Google'a)

Algorytmy Google'a bowiem „nie patrzą” na postać, tylko na to, jakim tagiem jest „podpisana” dana fotografia. Jeśli znany polski piłkarz Robert Lewandowski został opisany jako Olga Tokarczuk, to jego portret znajdzie się w kolekcji zawierającej fotografie polskiej noblistki. Musimy zatem posłużyć się inną metodą i co za tym idzie innymi algorytmami – mówiąc inaczej, nauczyć się rozpoznawać przez AI określone postacie. W tym celu powinniśmy zgromadzić określony zbiór (w tym wypadku fotografii). Następnie dzielimy go na dwa podzbiory: treningowy i testowy. Zbiór treningowy służy do nauki, zbiór testowy do oceny skuteczności tej nauki. Po przeprowadzeniu określonego cyklu nauki (epoche) pokazujemy AI zdjęcie na przykład danej postaci, którego nie było w zbiorze treningowym. Jeśli AI z dużym stopniem prawdopodobieństwa (ang. confidence) prawidłowo rozpozna to, co jest na tym zdjęciu, to naukę możemy zakończyć i uznajemy, że algorytm jest gotowy do analiz czy wizualizacji.

Narzędzia spod znaku cv posiadają takie m.in. funkcje, jak: a) detekcja obiektu – czyli określenie, w której części obrazu znajduje się jakiś obiekt, b) klasyfikacja obrazów – czyli wskazanie, do jakiej klasy obiektów należy ten na danym zdjęciu (np. czy to jest koń, pies itp.), c) klasyfikacja wieloetykiotowa – czyli określenie nie tylko klasy danego obiektu, ale też pewnych jego atrybutów (np. czarny koń), d) segmentacja – czyli wykrywanie położenia i kształtu obiektu/ów na zdjęciu. Przykładem tego typu programu jest biblioteka OpenCV (ang. Open Source Computer Vision Library). Jest to biblioteka do przetwarzania obrazów w czasie rzeczywistym. Napisana została w języku C, ale istnieją nakładki na inne języki (m.in. na Pythona). Możliwości biblioteki są ogromne, począwszy od prostych operacji na pojedynczych pikselach przez zaawansowane algorytmy przetwarzania obrazów, po algorytmy uczenia maszynowego, wykorzystywane w takich zagadnieniach jak detekcja twarzy. Niewątpliwą zaletą jest też możliwość wykonywania

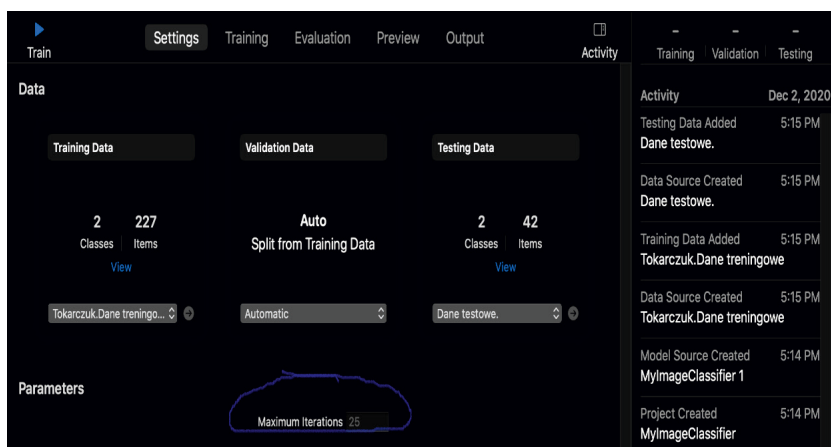
wszystkich operacji bezpośrednio na strumieniu wideo (Pabiasz). Istnieje też możliwość skorzystania z innych programów (jak np. z Create ML), a także z rozwiązań komercyjnych, które prawie całkowicie automatyzują proces analizy zdjęć czy filmów. Do najpopularniejszych tego typu usług należą: AWS (ang. Amazon Web Services), Microsoft Azure i Google Cloud Vision.

3. Przykłady

W tej części zostaną zaprezentowane przykłady analiz obrazów i tekstu – z zastosowaniem algorytmów sztucznej inteligencji, które zostały przeprowadzone przez autora niniejszego artykułu. Zostały tu zastosowane narzędzia i metody opisane powyżej.

Pierwszy przykład dotyczy rozpoznawania obrazów przez AI. Celem było nauczenie AI w miarę niezawodnego wskazywania postaci ostatniej polskiej noblistki – czyli Olgi Tokarczuk. Jak pokazano w części poprzedniej artykułu, grafika Google potrafi jako Olgę Tokarczuk wskazać np. Roberta Lewandowskiego. Aby nauczyć AI rozpoznawać, i to z dużym stopniem prawdopodobieństwa, postać Olgi Tokarczuk, wykorzystano środowisko developerskie Create ML. Create ML służy przede wszystkim do pisania programów w języku Swift, ale posiada też funkcję trenowania AI.

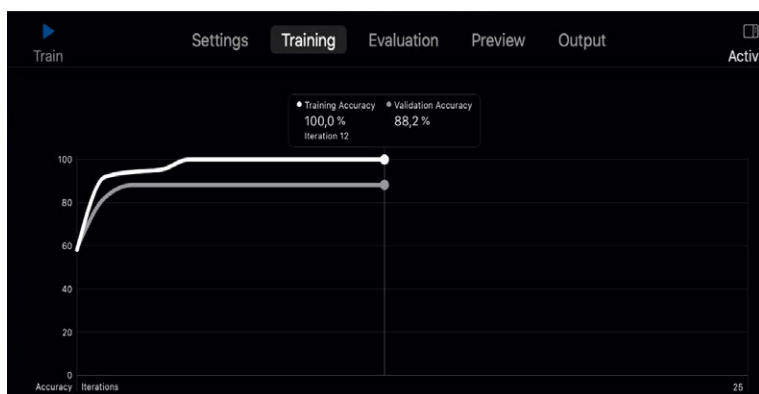
Aby wytrenować AI, zgromadzono dwa zbiory: treningowy i testowy. Pierwszy zawierał 127 zdjęć Olgi Tokarczuk oraz 100 zdjęć Wisławy Szymborskiej (z tego powodu, że potrzebne było utworzenie dwóch klas). Zbiór testowy zawierał odpowiednio 20 zdjęć Tokarczuk oraz 22 zdjęcia Szymborskiej (zdjęcie nr 2, zrzut z ekranu).



Zdjęcie nr 2. Zbiory treningowe i testowe w Create ML, oprac. własne

Źródło: zrzut z ekranu.

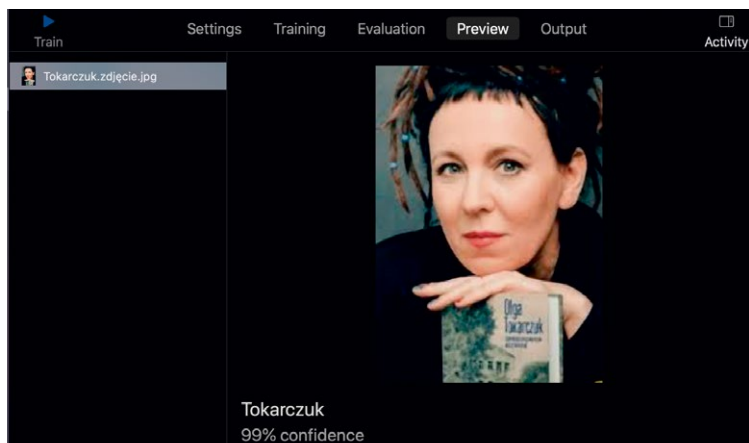
Po przeprowadzeniu sesji treningowej dokonano testowania skuteczności tej nauki – tj. sprawdzono, na ile model był w stanie trafnie rozpoznać zdjęcia Tokarczuk po pokazaniu mu obrazów, które nie były w zbiorze treningowym. Gdyby okazało się, że ta skuteczność jest zbyt mała, to trening należałoby powtórzyć. Wynik ewaluacji był następujący:



Zdjęcie nr 3. Oprac. własne

Źródło: zrzut z ekranu.

Jak widzimy, model całkiem dobrze poradził sobie z postawionym zadaniem, zatem można go było użyć do rozpoznawania dowolnych zdjęć Olgi Tokarczuk. Model rozpoznawał z dużym stopniem prawdopodobieństwa fotografie (zwykle powyżej 90%) – o czym świadczy kolejny wynik na zdjęciu nr 4.

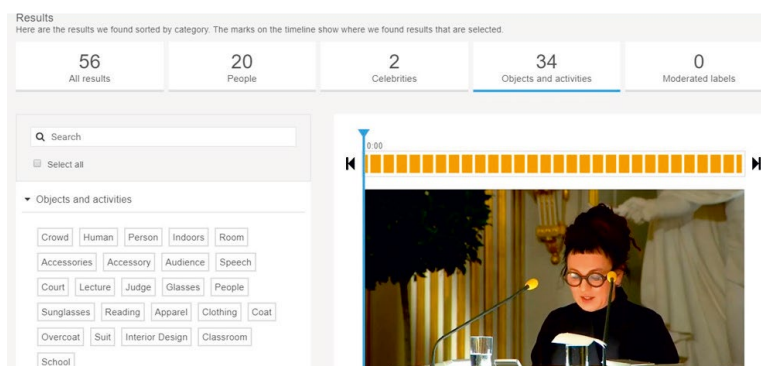


Zdjęcie nr 4. Oprac. własne

Źródło: zrzut z ekranu.

Jak widzimy (zdjęcie nr 4), załadowano do modelu fotografie, której nie było ani w zbiorze treningowym, ani także testowym i model z niemal całkowitą pewnością (99% confidence) stwierdził, że jest to Olga Tokarczuk.

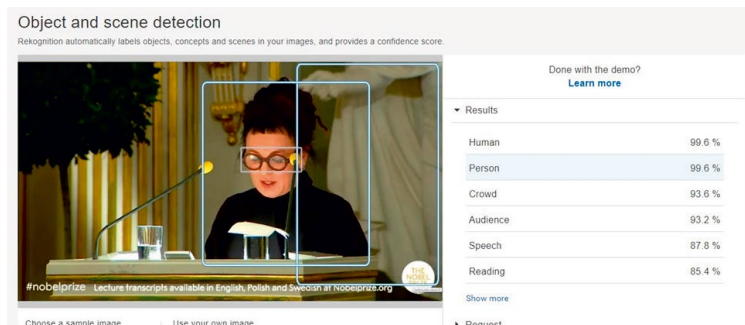
Z kolei przedmiotem analizy za pomocą algorytmów sztucznej inteligencji oferowanych przez AWS i AutoML Vision była mowa noblowiska Olgi Tokarczuk. Oczywiście mógł to być każdy inny film bądź zdjęcie. Najpierw fragment filmu został załadowany do ML AWS-a. Program przebadał pierwszą minutę filmu – analizując go klatka po klatce (niebieski kursor). Wynik owej analizy przedstawia się następująco:



Zdjęcie nr 5. Analiza wykładu noblowskiego Olgi Tokarczuk (Sztokholm, 2019) za pomocą AWS, oprac. własne

Źródło: zrzut z ekranu.

Jak widzimy powyżej, AWS zidentyfikował 56 obiektów – w tym 20 osób. Dwie określił jako celebryci. Spis wszystkich obiektów widzimy po lewej stronie. Następnie przedmiotem analizy była sama główna bohaterka i wynik tej analizy przedstawia się następująco:

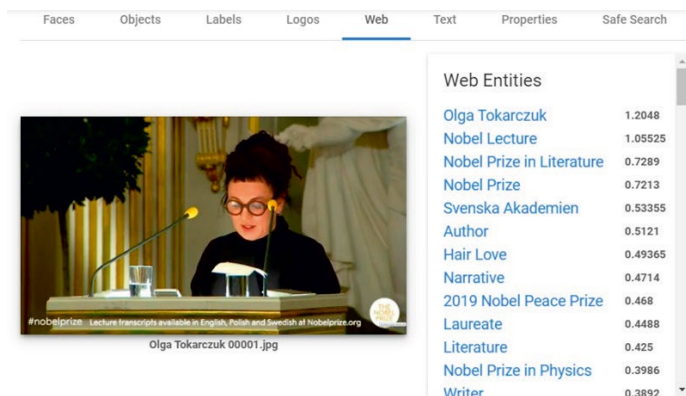


Zdjęcie nr 6. Analiza postaci za pomocą AWS, oprac. własne

Źródło: zrzut z ekranu.

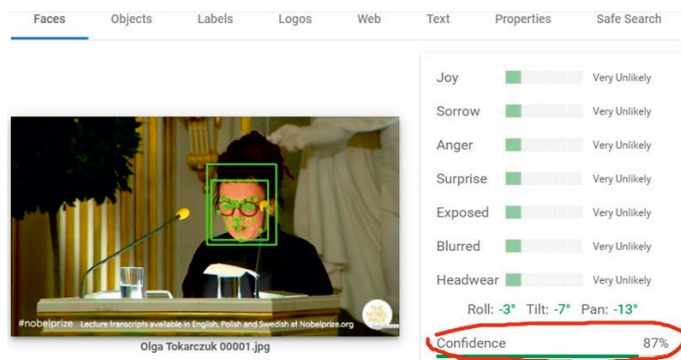
Program z prawie 100-procentową pewnością wykrył, że mamy do czynienia na zdjęciu nr 6 z człowiekiem, osobą, która przemawia do publiczności – czytając tekst.

Algorytmy AI potrafią jednakże powiedzieć jeszcze więcej o zdjęciu i osobie/osobach na nich się znajdujących. Tym razem zobaczmy, jak to widzi AutoML Google'a. Pozostajemy cały czas przy tym samym zdjęciu. Wynik interpretacji googłowskiej jest następujący:



Zdjęcie nr 7. Analiza mowy Tokarczuk za pomocą AutoML, oprac. własne
Źródło: zrzut z ekranu.

Program niemal bezbłędnie wykrył, że to Olga Tokarczuk, która wygłasza mowę noblowską w Sztokholmie w trakcie dorocznej uroczystości. To jednak nie wszystko. Ten sam program potrafi określić stan określonych emocji danej postaci. Zwraca uwagę na takie cechy, jak: radość, smutek, złość itp. – o czym świadczy następujący wynik (zdjęcie nr 8):

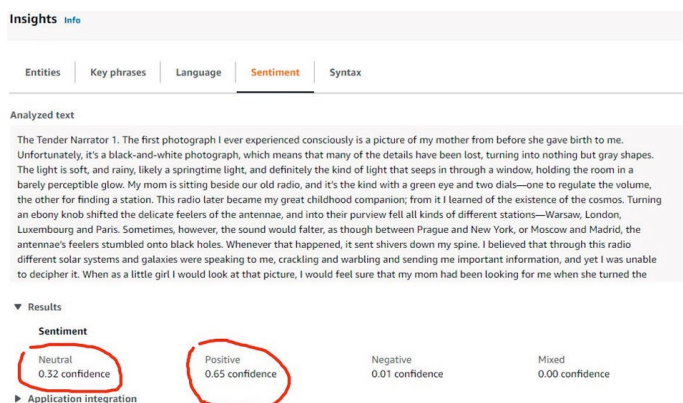


Zdjęcie nr 8. Oprac. własne
Źródło: zrzut z ekranu.

Widzimy tam, że określone stany emocjonalne Olgi Tokarczuk zostały zidentyfikowane z pewnością sięgającą 87%. Tu na marginesie należy dodać, że w przypadku innych fotografii i postaci ten wynik czasami wynosi 100%.

Zaprezentowane powyżej przykłady zastosowania AI do analizy obrazów nie wyczerpują wszystkich wchodzących w grę możliwości, które mogłyby być użyteczne dla badaczy kultury. Oto inny przykład z zakresu computer vision. Wróćmy jeszcze do AWS. Integralną częścią każdego filmu czy utworu multimedialnego jest warstwa audialna – w postaci mowy czy muzyki. Nie inaczej jest i w przypadku wystąpienia noblowskiego Olgi Tokarczuk. Jej przemówienie zostało przez autora przekonwertowane na tekst i dzięki temu możliwa była analiza sentymentu. Zalicza się ona do NLP. Analizę sentymentu można oczywiście przeprowadzić na każdym tekście – nie tylko wydobytym z utworu filmowego czy nagrania audialnego.

W ramach AWS istnieje usługa Amazon Comprehend (Google i Microsoft Azure oferują podobne rozwiązania), która pozwala zdiagnozować emocjonalny wydźwięk samego przemówienia (oczywiście i każdego tekstu). Nazywa się to właśnie analizą sentymentu. Jako ludzie bowiem mamy zdolność do rozpoznawania cudzych emocji na podstawie treści wypowiedzianych słów, a także pewnych czynników pozawerbalnych, jak ton głosu czy ogólna ekspresja ciała. Ograniczając nasze zainteresowanie do prostszego problemu, tj. rozpoznania jedynie ogólnego stosunku autora do opisanego treści, a nie konkretnej emocji, możemy się starać taki proces zautomatyzować – tym właśnie zajmuje się analiza sentymentu. Polega ona zatem na określeniu stosunku autora danej wypowiedzi/tekstu do czegoś bądź kogoś. Może on przybierać cztery kategorie: pozytywny, negatywny, neutralny i mieszany. W przypadku przemówienia Olgi Tokarczuk otrzymaliśmy wynik następujący:



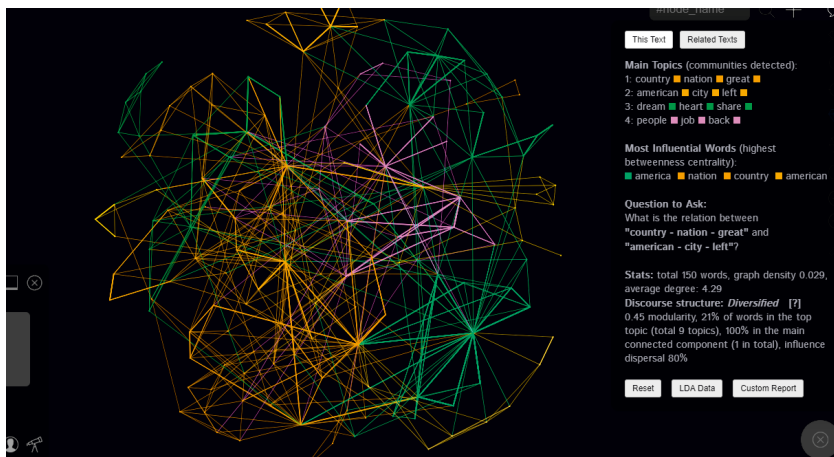
Zdjęcie nr 9. Analiza sentymentu mowy noblowskiej Olgi Tokarczuk za pomocą AWS, oprac. własne

Jak widzimy, algorytmy ML wskazują, że istnieje 65% prawdopodobieństwa pozytywnego wydźwięku tego przemówienia – natomiast neutralny stosunek określony został na 32%. W tym więc przypadku nie jest to jednoznaczna opinia na temat emocjonalnego wydźwięku tego tekstu/mowy.

Innym jeszcze interesującym rozwiązaniem na polu NLP są analizy korpusów tekstualnych. Chodzi tu zarówno o ich interpretację statystyczną, stylometrię, jak i znajdowanie relacji wewnątrztekstualnych. Pokażmy to na przykładzie Infra Nodus. Projekt został stworzony przez Dmitrya Paranyuszkiną.

Infra Nodus to narzędzie typu open source (wsparte na algorytmach AI), które można wykorzystać do badania dowolnego tekstu lub wypowiedzi. Wizualizuje ono tekst jako sieć, aby pokazać najistotniejsze tematy i relacje między nimi.

Algorytm InfraNodus może być wykorzystywany przez badaczy do porządkowania i lepszego zrozumienia tekstów, do pomiaru poziomu stroniczości (analiza sentymentu), do identyfikacji tych części dyskursu, które zawierają nowe pomysły, oraz do klasycznej statystyki (liczba słów, przeciętna długość zdań itp.). Metoda ta opiera się na algorytmie analizy sieci tekstowej, która reprezentuje dowolny tekst jako sieć i identyfikuje najbardziej wpływowe słowa w dyskursie na podstawie współwystępowania terminów. Następnie stosuje się algorytm wykrywania zbiorowości grafów w celu zidentyfikowania różnych klastrów tematycznych, które reprezentują główne tematy w tekście, jak również relacje między nimi (Paranyuszkin).



Zdjęcie nr 10. Wizualizacja inauguracyjnego przemówienia prezydenta USA Donalda Trumpa z 20 stycznia 2017 r. za pomocą Infra Nodus, oprac. własne

Źródło: zrzut z ekranu.

Jest to analiza i wizualizacja tradycyjnego przemówienia – wygłoszonego przez prezydenta USA obejmującego swój urząd (w tym przypadku przez D. Trumpa). Na zdjęciu numer 10 otrzymaliśmy, oprócz statystyki, wizualizację przedstawiającą dokument jako sieć. Wizualizacja zaprezentowanego przemówienia jest oparta na wspomnianej teorii grafów. Jest ona interaktywna. Widzimy na niej węzły dominujące i relacje łączące je z innymi węzłami. Co więcej, możemy zaobserwować, że pewne węzły są na tyle podobne, że tworzą odrębne grupy tematyczne – co jest zaznaczone odmiennymi kolorami. Otrzymane wyniki podają nam główne pojęcia (podzielone na cztery grupy), najbardziej wpływowe słowa oraz podstawowe zestawienia charakterystyczne dla grafów, typu: gęstość grafu, średni stopień czy modularność. Rysuje nam się także struktura dyskursu analizowanej wypowiedzi – podzielonej na określone tematy zgrupowane wokół takich pojęć jak: państwo, naród, siła, Amerykanie czy wielkość. Za najbardziej wpływowe słowa w tym dyskursie zostały uznane: Ameryka, naród, państwo i Amerykanie.

4. Nowy paradygmat badań

We współczesnej nauce istnieje wiele rywalizujących ze sobą nurtów badania kultury. Na gruncie humanistyki, która zajmuje się poznawaniem rzeczywistości kulturowej, obok klasycznych już nurtów o proveniencji postmodernistycznej jeszcze powstało szereg nowych. Do najważniejszych z nich zalicza się zwykle: humanistykę zaangażowaną, artystyczną, kognitywną czy ekologiczną (Nycz, 2017, s. 46–48). Powstało też szereg bardziej szczegółowych subdyscyplin – koncentrujących się zwykle na badaniach szeroko rozumianej sfery medialnej, w rodzaju: digital culture studies, visual culture studies, software studies, performance studies, media and sound studies, thing studies, memory and postmemory studies czy netnografia. Wszystkie one są swoistą odpowiedzią na potrzebę badania nowych pejzaży kulturowych, które pojawiły się w ostatnich dwu dekadach bieżącego stulecia.

Zaprezentowane powyżej przykłady badań oraz dedykowane im narzędzia i metody zalicza się zwykle do jeszcze innego nurtu czy też paradygmatu zwanego humanistyką cyfrową. Ona, jak wskazuje nazwa, jest wytworem rewolucji cyfrowej i polega na zastosowaniu narzędzi ICT (ang. information and communication technologies) do badań w dyscyplinach humanistycznych i społecznych. Badania prowadzone w tym paradygmacie, jak zauważa jego wybitna przedstawicielka, składają się z trzech komponentów: cyfrowych danych, komputerowego ich przetwarzania oraz multimedialnej prezentacji (Drucker, 2021, s. 2). Jeszcze inną

cechą charakterystyczną tego paradygmatu jest przetwarzanie ogromnej liczby danych – czyli big data. Jest to szczególnie charakterystyczne dla tzw. analityki kulturowej. Lev Manovich (jej twórca) twierdzi, że nie można poznać współczesnej kultury bez uwzględnienia przede wszystkim danych medialnych i to w dużej skali. Metody klasycznych nauk o kulturze stworzone z myślą o badaniu kultur istniejących w tzw. rzeczywistości fizycznej okazują się niewystarczające, a w wielu przypadkach są po prostu zawodne przy interpretowaniu wielkich kolekcji danych medialnych – umieszczanych głównie w świecie wirtualnym. Stąd (postuluje) konieczność użycia uczenia maszynowego, aby uchwycić wśród milionów cyfrowych artefaktów trendy dominujące we współczesnym świecie (Manovich, 2020, s. 3). Jak zauważa Rens Bod, cyfrowe, informatyczne podejście do materiału nauk o kulturze doprowadziło do odkrycia wielu nowych zjawisk. Bez tego podejścia niemożliwie byłoby porównywanie setek tekstów naraz albo znajdowanie historycznych schematów w tysiącach zdigitalizowanych źródeł. Cyfrowe nauki, twierdzi, przynoszą nie tylko nowe ustalenia, ale także umożliwiają pytania badawcze nigdy wcześniej niestawiane (Bod, 2013, s. 464–465). Stosowanie algorytmów „głębokiego” uczenia maszynowego (czyli sztucznej inteligencji) wpisuje się w ten paradygmat.

Nowy paradygmat badań (określany jako humanistyka cyfrowa), którego ważną częścią jest stosowanie algorytmów sztucznej inteligencji i w ogóle uczenia maszynowego, charakteryzuje się zatem następującymi cechami: a) w jego ramach pracuje się na zdigitalizowanych danych pochodzących z przeszłości, jak i z teraźniejszości (fizycznej i wirtualnej), b) większość analiz jest zautomatyzowana (badacz zwykle ładuje dane i ustala parametry), c) wyniki analiz często są zwizualizowane (wizualizacja jako metoda badawcza) i możliwe do powtórzenia przez inne zespoły badawcze (intersubiektywnie kontrolowalne), d) badania mają charakter laboratoryjny i transdyscyplinarny.

Jeśli chodzi o sztuczną inteligencję – to algorytmy AI można stosować zarówno do świata wirtualnego, jak i fizycznego. Mogą one pomóc w analizie określonych cech zarówno jednej osoby, tekstu czy zdjęcia, jak i dużych zbiorów cyfrowych danych. Szczególnie ten ostatni aspekt jest niezmiernie ważny. Nawet duże zespoły badawcze nie są bowiem i nie będą w stanie zinterpretować big data za pomocą klasycznych metod – stworzonych do badania świata „analogowego”. Niezbędne tu są narzędzia cyfrowe i zautomatyzowane sposoby analizy, takie jak na przykład uczenie maszynowe. Jest to zresztą zgodne z dominującą obecnie tendencją – czyli kwantyfikacją praktycznie wszystkich aspektów naszego życia – dzięki której możemy mieć pewniejszą i bardziej wszechstronną wiedzę na temat różnych praktyk społecznych (Szpunar, 2016, s. 25). Z drugiej strony, algorytmy uczenia

maszynowego potrafią wyodrębnić konkretną jednostkę – z jej różnymi potrzebami, oczekiwaniami czy preferencjami. W literaturze przedmiotu przytacza się przykład amerykańskiej firmy Target, która na podstawie danych o zakupach dokonywanych przez kobiety potrafi określić, która z nich jest w ciąży, a nawet przewidzieć przypuszczalny termin porodu. Dzięki takiej wiedzy wspomniana firma mogła na przykład wysyłać ciężarnym klientkom kupony rabatowe na różne produkty niezbędne dla prawidłowego przebiegu ciąży (Mayer-Schonberger, 2014, s. 83).

Analizując big data i innego typu dane, AI jest w stanie odtworzyć to, co zawsze było ważnym przedmiotem badań humanistycznych, a mianowicie dyskursy czy dominujące praktyki kulturowe. Do tej pory bowiem badacze kultury (literaturoznawcy, historycy czy kulturoznawcy) intuicyjnie określali „architekturę” dyskursu za pomocą wybranego zbioru tekstów. Algorytmy uczenia maszynowego są w stanie przeszukać cały dostępny księgozbiór (np. Google books) i wówczas wyniki takiej analizy będą oparte na całości zbioru, a nie na jego mniejszej lub większej próbie.

Wreszcie, jednym z najbardziej obiecujących kierunków rozwoju AI jest/będzie uczenie maszynowe nienadzorowane. W przypadku badań kultury algorytmy AI nie będą znały wyniku (na wejściu), tylko w kolekcji danych zmuszone będą szukać i klasyfikować szeroko rozumiane obiekty czy zjawiska. Przykładem tego typu analiz jest projekt InfraNodus, który służy do interpretacji różnych korpusów tekstualnych. Jego algorytmy, oparte na grafach, potrafią skonceptualizować każdy tekst jako sieć i wykryć, jak widzieliśmy, dominujące tematy oraz główne kategorie.

BIBLIOGRAFIA

- Bod, R. (2013). *Historia humanistyki*. Warszawa: Aletheia.
- Boschetti, A., Massaron, L. (2016). *Python. Podstawy nauki o danych*. Gliwice: Helion.
- Chollet, F. (2018). *Deep Learening*. Gliwice: Helion.
- Deitel, P. i Deitel, H. (2020). *Python dla programistów*. Gliwice: Helion.
- Drucker, J. (2021). *The Digital Humanities*. London and New York: Routledge.
- Flasiński, M. (2018). *Wstęp do sztucznej inteligencji*. Warszawa: PWN.
- https://en.wikipedia.org/wiki/Computer_vision (dostęp: 12.02.2022).
- Kowal, A. *SI zajmie się jedną z najstarszych zagadek z dziedziny językoznawstwa*. Pozyskano z: <https://whatnext.pl/si-jedna-z-najstarszych-zagadek-z-dziedziny-jezykoznawstwa/> (dostęp: 28.02.2022).
- Krohn, J. (2022). *Uczenie głębokie i sztuczna inteligencja*. Gliwice: Helion.
- Manovich, L. (2020). *Cultural Analytics*. Cambridge, Massachusetts: The MIT Press.

- Mayer-Schonberger, V. i Cuckier, K. (2014). *Big Data*. Warszawa: MT.
- Pabiasz, S. *Wykrywanie obiektów za pomocą OpenCV*. Pozyskano z: <https://osworld.pl/wykrywanie-obiektow-za-pomoca-opencv/> (dostęp: 28.02.2022).
- Paranyuszkin, D. *InfraNodus: Generating Insight Using Text Network Analysis*. Pozyskano z: <https://dl.acm.org/doi/10.1145/3308558.3314123> (dostęp: 28.02.2022).
- Patterson, J. i Gibson, A. (2028). *Deep Learning. Praktyczne wprowadzenie*. Gliwice: Helion.
- Szpunar, M. (2016). *Kultura cyfrowego narcyzmu*. Kraków: Wydawnictwo AGH.
- Vaughan, L. (2022). *Python z życia wzięty. Rozwiązywanie problemów za pomocą kilku linii kodu*. Gliwice: Helion.

Andrzej Radomski – prof. UMCS, pracuje w Instytucie Nauk o Kulturze (Katedra Informacji i Kultury Cyfrowej UMCS), a także Academy Teacher w Academia Electronica w Second Life i AltSpace VR. Zajmuje się badaniem kultury cyfrowej, humanistyki cyfrowej, a także zastosowaniem metod data science i uczenia maszynowego (w szczególności sztucznej inteligencji) do analizy zjawisk historycznych, kulturowych i społecznych. Redaktor naczelny czasopisma „Kultura i Historia” (www.kulturaihistoria.umcs.lublin.pl), redaktor naukowy w periodyku multimedialnym „Medialica” (www.medialica.umcs.lublin.pl). Członek Rady Naukowej DARIAH.PL (Konsorcjum ds. rozwoju i promocji humanistyki cyfrowej). Twórca i autor dwóch blogów naukowych i kilkunastu filmów naukowych oraz popularnonaukowych.