

PIOTR KĘDZIORA

MÓJ CICHY WIELBICIEL, TWÓJ MIŁOŚNIK, NASZ PSYCHOFAN, WASZ PRZEŚLADOWCA

MUZYCZNA AI I JEJ ALTERNATYWNA FIGURACJA

PIOTR KĘDZIORA

Instytut Kulturoznawstwa, Uniwersytet im. Adama Mickiewicza w Poznaniu. Adiunkt, badacz współczesnej kultury muzycznej, autor książki *Ambient. Kultury słuchania i sposoby posługiwania się dźwiękiem* (2024). Realizator projektu „Polska muzyka ambient i jej twórcy”. Badacz muzyki tła, koncentrujący się na zagadnieniu słuchania jako praktyki kulturowej. Współredaktor monografii zbiorowych *Kultura medialnie zapośredniczona. Badania nad mediami w perspektywie kulturoznawczej* (2011) oraz *Kultury muzyczne – kultury słuchania* (2020). Pasjonat muzyki elektronicznej i generatywnej, projektant graficzny, web developer i programista. ORCID: 0000-0001-7295-6632.

WPROWADZENIE

Udział sztucznej inteligencji w świecie muzyki komentowany jest z racji jego rewolucyjnego charakteru na wszelkie możliwe sposoby, podobnie jak w przypadku innych sfer, których rozwój związany jest z tą technologią. Jej zalety poznajemy z wielu źródeł, między innymi dzięki notom prasowym twórców różnych rozwiązań, które „rewolucjonizują” – ich zdaniem – świat muzyki, recenzjom tworzonemu przez użytkowników mediów i technologii muzycznych, zdecydowanie rzadziej dzięki opracowaniom popularnonaukowym omawiającym rozmaite konteksty relacji muzyki i technologii maszynowego uczenia, sieci neuronowych czy *big data*.

Przykładem może być złożenie przez Recording Industry Association of America (RIAA) pozwu przeciwko Suno i Uido, dwóm serwisom oferującym usługi związane z generowaniem muzyki, które zostały oskarżone o nielegalne praktyki uczenia maszynowego na podstawie treści muzycznych objętych prawem autorskim¹. Trend naruszeń praw

¹ Trzy największe wytwórnie muzyczne – Universal Music Group, Sony Music Entertainment i Warner Music Group – złożyły pozwy przeciwko firmom AI zajmującym się generowaniem muzyki, Suno i Udio, oskarżając

do treści jest już zresztą znany dzięki sprawom z innych domen, jak choćby oskarżenia ośmiu wydawców prasowych, w tym „The New York Daily News”, „Chicago Tribune” czy „Denver Post” skierowane przeciwko OpenAI oraz Microsoft w sprawie uczenia maszynowego na bazie autorskich artykułów tych wydawców². Wcześniej podobny pozew został wystosowany przez „The New York Times” przeciwko wspomnianym firmom, inni wydawcy również ujawnili swoje umowy licencyjne, na mocy których ich treści mogą być wykorzystywane do uczenia maszynowego³. Ostatnio dołączył do tego inny wątek, jak się zdaje dużo bardziej skomplikowany, bo dotyczący sprzedaży firmie Microsoft dostępu do treści publikowanych w czasopiśmie naukowym Taylor & Francis⁴. Wspomniane wcześniej wytwórnie, takie jak Universal Music Group, pozwały start-up Anthropic po tym, jak jego model AI Claude 2 zaczął generować teksty piosenek oraz wokale niemal identyczne jak te chronione prawem autorskim. Sony Music wydało około 10 tysięcy wniosków o usunięcie nieautoryzowanych podróbek wokalnych z serwisu. Firmy zajmujące się sztuczną inteligencją stoją jednak na stanowisku, że wykorzystanie utworów chronionych prawem autorskim mieści się w ramach tak zwanego dozwolonego użytku⁵. Wątek podobny, bo usytuowany w kontekście naruszeń praw, ale tym razem do wizerunku, dotyczy wyjaśnień, jakich zażądała amerykańska aktorka Scarlett Johansson w pozwie przeciwko OpenAI. Asystent głosowy usługi ChatGPT, mimo zaprzeczeń twórców, brzmi bowiem tak samo jak ona⁶. Głos Sky, sugestywnie przypominający głos aktorki znany z roli, jaką

je o masowe naruszenia praw autorskich. Pozwy te zostały wniesione pod przewodnictwem Recording Industry Association of America (RIAA) i dotyczyły używania chronionych nagrań muzycznych do trenowania modeli AI bez zgody właścicieli praw autorskich. CEO Suno, Mikey Shulman, broni praktyk swojej firmy, twierdząc, że ich AI generuje oryginalne treści, a nie replikuje istniejące utwory. Jednak wytwórnie muzyczne argumentują, że przekonujące imitacje AI są możliwe tylko dzięki szerokiemu użyciu chronionych prawem utworów w danych treningowych. Przykładem podanym w pozwie jest niemal idealna reprodukcja utworu *Johnny B. Goode* Chucka Berry’ego przez jedną z tych usług, powstała przy użyciu jedynie podstawowych wskazówek dotyczących stylu muzycznego i niektórych fragmentów tekstu. Wytwórnie żądają do 150 tysięcy dolarów odszkodowania za każdy naruszony utwór oraz zakazu dalszego korzystania z chronionych nagrań w procesie trenowania modeli AI. RIAA podkreśla, że jej działania nie są wymierzone przeciwko używaniu dzieł chronionych prawem autorskim do trenowania AI, ale przeciwko przypadkom, gdy odbywa się to bez odpowiednich licencji i zgód. Uniwersalne umowy licencyjne, jak te zawarte między Universal Music a SoundLabs, pokazują, że branża muzyczna jest otwarta na współpracę z firmami AI, pod warunkiem że będą one respektować prawa autorskie i zapewniać uczciwe wynagrodzenie twórcom. Zob.: K. Robinson, *Major labels weighing lawsuit against AI firms Suno and Udio for alleged unlicensed training*, Billboard, 20 czerwca 2024, <https://www.billboard.com/pro/major-labels-weighing-lawsuit-ai-suno-udio-alleged-unlicensed-training/> (20 czerwca 2024); M. Bohannon, *Major record labels sue AI music startups for allegedly copying songs to train AI*, Forbes Australia, 25 czerwca 2024, <https://www.forbes.com.au/news/uncategorized/record-labels-sue-ai-music-startups-for-allegedly-copying-songs-to-train-ai/> (26 czerwca 2024).

2 K. Viner, *Eight US newspapers sue OpenAI and Microsoft for copyright infringement*, „The Guardian”, 30 kwietnia 2024, <https://www.theguardian.com/technology/2024/apr/30/us-newspaper-openai-lawsuit> (20 czerwca 2024).

3 Do wydawców tych należy między innymi The Associated Press. OpenAI zawarło również umowy licencyjne z innymi firmami medialnymi, w tym z gigantami wydawniczymi: Axel Springer w Niemczech i Prisa Media w Hiszpanii, francuską gazetą „Le Monde”, a ostatnio z londyńskim „Financial Times”.

4 M. Buttersby, *Academic authors ‘shocked’ after Taylor & Francis sells access to their research to Microsoft AI*, The Bookseller, 19 lipca 2024, <https://www.thebookseller.com/news/academic-authors-shocked-after-taylor-francis-sells-access-to-their-research-to-microsoft-ai> (20 października 2024).

5 A.R. Chow, *AI’s influence on music is raising some difficult questions*, Time, 4 grudnia 2023, <https://time.com/6340294/ai-transform-music-2023> (20 czerwca 2024).

6 N. Jones, *Who owns your voice? Scarlett Johansson OpenAI complaint raises questions*, Nature, 29 maja 2024, <https://www.nature.com/articles/d41586-024-01578-4> (20 czerwca 2024).

zagrała w futurystycznym filmie *Her* Spike'a Jonze'a, został w efekcie usunięty z usługi OpenAI⁷.

Dla odmiany wspomniane na wstępie zalety poznajemy szczególnie za sprawą tych, którzy tworzą AI lub zachęcają przyszłych inwestorów, upubliczniając efekty prac badawczo-rozwojowych prowadzonych za gromadzone środki. Dużo rzadziej do publicznej domeny trafiają rzetelne naukowo-badawcze lub badawczo-rozwojowe opracowania zagadnień związanych z AI, ponieważ w przeważającej większości stronią one od chwytliwych frazesów na temat świetlanej (lub zagrożonej) przyszłości gatunku ludzkiego i jego praktyk oraz zazwyczaj koncentrują się na wybiórczych i czasem bardzo skomplikowanych zagadnieniach technicznych, które stanowią mało spektakularne osiągnięcia w obszarze związanym z rozwojem AI. Zwykle nie operuje się w tych opracowaniach terminami, które znamy z newsów, dotyczą one bowiem określonych rozwiązań o szczegółowo wyartykułowanych celach, metodach i korpusach technologii informatycznych. Jak ujął to Linus Torvalds, twórca jądra systemu Linux, w wywiadzie udzielonym w trakcie tegorocznej konferencji Open Source Summit Europe w Wiedniu, wszystko, co dziś wiemy o potencjale tak zwanej sztucznej inteligencji, jest w 90 proc. marketingiem, a w 10 proc. rzeczywistości – tymi możliwościami tego korpusu technologii w jej obecnym stadium rozwoju⁸.

Mamy zatem do czynienia z pewnym rozdźwiękiem, polegającym na tym, że do opinii publicznej nie dociera to, co faktycznie dzieje się w obszarze mariażu muzyki i sztucznej inteligencji, powtarzane są za to dość ubogie merytorycznie, ale chwytliwe i działające na wyobraźnię komunikaty, często inspirowane przez samych założycieli firm działających w obszarze AI lub ich reprezentantów. Luka pomiędzy kwantową muzyką rodem z Interdisciplinary Centre for Computer Music Research albo Computer Music Multidisciplinary Research bądź osiągnięciami prezentowanymi w kolejnych edycjach konferencji New Interfaces for Musical Expression czy AI and Music Creativity⁹ a potoczną wiedzą o tym, co dzieje się w świecie muzyki za sprawą rozwoju nowych technologii uczenia maszynowego, głębokiego uczenia, sieci neuronowych, przetwarzania języka naturalnego czy robotyki, jest niestety bardzo duża i problematyczna. Tworzy ona wiele nieporozumień i błędnych przekonań na temat tego, co owe technologie faktycznie mogą i jak działają. Przepaść między wiedzą ekspercką i potoczną dotyczy *de facto* całego szeregu problemów, z jakimi mierzymy się jako entuzjaści muzyki, jej słuchacze, ale również jako humaniści i badacze kultury¹⁰.

7 K. Knibbs, *What Scarlett Johansson v. OpenAI could look like in court*, Wired, 22 maja 2024, <https://www.wired.com/story/scarlett-johansson-v-openai-could-look-like-in-court/> (20 czerwca 2024).

8 S. Bhartiya, *Linus Torvalds on the kernel, GenAI, EVs, programming languages and more...*, TFIR, 17 października 2024, <https://tfir.io/linus-torvalds-on-the-kernel-genai-evs-programming-languages-and-more/> (20 października 2024).

9 NIME to jedna z najbardziej prestiżowych konferencji, na której prezentowane są czołowe osiągnięcia w obszarze integracji nowych technologii z szeroko pojętym wykonawstwem muzycznym. AIMC to konferencja skupiająca środowisko badaczy i artystów związanych z zagadnieniami AI, która odbywa się od 2021 roku.

10 *Sztuczna inteligencja ma świadomość? „Społeczeństwo zbyt łatwo dało się nabrać”*, oprac. K. Nowak, Forsal, 23 lipca 2024, <https://forsal.pl/lifestyle/technologie/artykuly/9551683,sztuczna-inteligencja-ma-swiadomosc-spoleczenstwo-zbyt-latwo-dalo-si.html> (24 lipca 2024). Do podobnych wniosków dochodzą autorzy raportu NASK, według których ponad 20 proc. respondentów twierdzi, że to, co nazywamy sztuczną inteligencją, posiada własną samoświadomość, a więc potrafi wyznaczać sobie cele. Zob. *Sztuczna inteligencja w społeczeństwie i gospodarce. Analiza wyników ogólnopolskiego badania opinii polskich internautów*, red. R. Lange, NASK Państwowy Instytut Badawczy, Warszawa 2019.

SZTUCZNY SUROGAT MUZYCZNEJ INTELIGENCJI

Ogólny, niewątpliwie metaforyczny obraz tego, co nazywamy sztuczną inteligencją w przypadku muzyki, jest zestawem wielu odrębnych algorytmów, mechanizmów czy procesów, których opis i analiza mogłyby zająć wiele czasu. Każdy z nich jest efektem nierzadko wieloletnich prac naukowych, badawczo-rozwojowych i wdrożeniowych, dodatkowo taki sam nakład pracy poświęcono na sposoby, by procesy te móc łączyć, co jest jednym z istotniejszych, a często zapominanych elementów współczesnych systemów informatycznych. Współcześnie tworzone programy i aplikacje realizujące określone zadania mają uwspólnione interfejsy, dzięki którym procesy te można łączyć w określony przez twórców sposób. To w istocie najważniejsze zagadnienie każdego osiągnięcia wiedzy technicznej, by w możliwie najłatwiejszy sposób łączyć je z innymi produktami myśli informatycznej, a niezahamowana integracja tych produktów powoduje, że współczesne maszyny liczące są zdolne się uczyć, zapamiętywać i predestynować cokolwiek chcą ci, którzy powołali je do [sic!] życia. Kompatybilność i zdolność do integracji jest też kluczem przyszłego sukcesu komercyjnego.

Jak działa AI w muzyce? Z czego się składają i czym literalnie są technologie, które wspólnie określamy tym mianem? Z technologicznego punktu widzenia nie jest istotne, co jest zawartością repozytoriów, skoro procedury są uwspólnione i dodatkowo podatne na łączenie ich w dłuższe sekwencje. Dla ułatwienia warto jednak wskazać, że muzyczne AI ma dwa bazowe wymiary. Po pierwsze, maszyny uczące się tworzą muzykę wedle wytycznych wskazanych przez ich użytkownika oraz na podstawie danych, które im użytkownik udostępnia. Robią to, bazując na umiejętności imitowania muzyki, której „słuchają”, analizując jej strukturę rytmiczną, melodię, harmonię i brzmienie. Po drugie, maszyny uczą się przyzwyczajęń słuchowych i naszych gustów, by skuteczniej podpowiadać nam kolejne, domniemanie odpowiednie dla nas, adekwatne do naszych upodobań utwory muzyczne. Taksonomia jest nieodzownym elementem zasobów AI oraz podstawowym narzędziem komunikacji z użytkownikiem. Sposób, w jaki nazywamy muzykę oraz szukamy jej za pomocą tych nazw, stanowi jeden z pierwszych mechanizmów, który może być skutecznie wykorzystywany w uczeniu się zachowań słuchaczy. Drugim mechanizmem wspierającym ten proces jest analiza materiału dźwiękowego i nadawanie mu nazw czy oznaczeń bez udziału użytkowników.

DANE: DŹWIĘK I WIEDZA O SŁYSZENIU/SŁUCHANIU

Materiałem wyjściowym dla obliczeniowych metod przetwarzania są zunifikowane informacje o dźwięku oraz o tym, jak ludzie go nazywają, klasyfikują i pojęciowo porządkują. Obliczenia te wykonywane są w przypadku dźwięku na podstawie szeregu niezależnych procesów, które warto wypunktować. Są to: próbkowanie (*sampling*), kwantyzacja (*quantization*), transformacja Fouriera (*Fourier transformation*), tworzenie spektrogramów (*spectrogram*), filtrowanie (*filtering*), wykrywanie cech (*feature extraction*), analiza zmienności czasowo-częstotliwościowej (*time-frequency analysis*), klastryzacja i klasyfikacja (*clustering and classification*)¹¹.

¹¹ K. Umapathy, B. Ghoraani, S. Krishnan, *Audio signal processing using time-frequency approaches: Coding, classification, fingerprinting, and watermarking*, „EURASIP Journal on Advances in Signal Processing” 2010; T. Lenin,

W większości są to techniki i metody badawcze współczesnej akustyki. Procesy te są wykorzystywane w różnych dziedzinach, takich jak: rozpoznawanie mowy (*speech-to-text recognition*), przetwarzanie mowy na tekst (*hidden Markov models* i *deep neural networks*), analiza muzyczna (identyfikacja i analiza elementów muzycznych, takich jak rytm, melodia, harmonia), diagnostyka medyczna (bicie serca, szумы płucne) czy rozpoznawanie emocji (analiza tonu głosu).

W przypadku procedur w obrębie tak zwanej sieci semantycznej, którą współtworzą użytkownicy, a analizują powołane do tego systemy, zachodzą inne procesy, których celem jest systematyzacja wytwarzanej na temat muzyki wiedzy, jak również jej walidowanie i uzupełnianie w miejscach, w których jest ona niewystarczająca, by na jej bazie skutecznie tworzyć powiązania z innymi jakościami struktury bazodanowej. Na te procedury składają się między innymi takie procesy, jak ekstrakcja metadanych, czyli tworzenie zbiorów danych oraz ich atrybutów (analiza zbiorów danych muzycznych, takich jak katalogi muzyczne, bazy danych metadanych, na przykład MusicBrainz, Spotify API, oraz opisy tekstowe, jak recenzje, artykuły, komentarze, zestawy informacji, jak tytuły utworów, nazwy artystów, gatunki muzyczne, daty wydania, tempo, tonacja i teksty piosenek), analiza treści muzycznej (*content-based analysis*), takiej jak cechy akustyczne (ekstrakcja cech akustycznych, jak *mel-frequency cepstral coefficients* – MFCC, spektralna energia, tonacja, rytm itd.) czy modelowanie tematyczne tekstów piosenek (*non-negative matrix factorization* bądź *latent Dirichlet allocation*), analiza związków semantycznych (*semantic relationship analysis*) poprzez uczenie reprezentacji wektorowych dla elementów muzycznych (Word2Vec, GloVe) oraz analizę sieci (*network analysis*), w tym tworzenie grafów węzłów (jak utwory, artyści) i relacji (jak współpraca artystów, podobieństwo gatunkowe) między elementami muzycznymi. Wykorzystuje się również uczenie maszynowe i klasteryzację (*machine learning and clustering*) do tworzenia drzew decyzyjnych (*support vector machines* – SVM) czy sieci neuronowych do klasyfikacji gatunków muzycznych bądź nastroju piosenek, do klasteryzacji (dzięki algorytmom klasteryzacji, takim jak *k-means* czy DBSCAN), do grupowania podobnych utworów czy artystów, w końcu sięga się po wizualizację sieci (*network visualization*) do tworzenia grafów semantycznych i interaktywnych map muzycznych¹². Komputery tworzą i analizują sieci semantyczne związane z muzyką poprzez wykorzystanie zaawansowanych technik z zakresu przetwarzania języka naturalnego (NLP), uczenia maszynowego, a także analizy sygnałów audio. Proces ten obejmuje kilka kroków, które pozwalają na modelowanie relacji

P. Dhanalakshmi, *Environmental sound classification using discrete wavelet transform*, „Advances in Natural and Applied Sciences” 6(9)/2015; H. Chaurasiya, *Time-frequency representations: Spectrogram, cochleogram and correlogram*, „Procedia Computer Science” 167/2020; Y.D. Mistry, G.K. Birajdar, A.M. Khodke, *Time-frequency visual representation and texture features for audio applications: A comprehensive review, recent trends, and challenges*, „Multimedia Tools and Applications” 23(82)/2023; P. Rao, *Audio signal processing*, [w:] *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, red. B. Prasad, S.R.M. Prasanna, Springer, Berlin 2008.

¹² *Music Data Mining*, red. T. Li, M. Ogihara, G. Tzanetakis, CRC Press, Boca Raton 2012; M. Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, Springer, Cham 2015; M.A. Casey i in., *Content-based music information retrieval: Current directions and future challenges*, „Proceedings of the IEEE” 4(96)/2008; A. Klapuri, M. Davy, *Signal Processing Methods for Music Transcription*, Springer, New York 2006.

semantycznych pomiędzy różnymi elementami muzycznymi, takimi jak utwory, artyści, gatunki i inne metadane.

Współczesne systemy informatyczne gromadzą, jak widać, każdy typ i rodzaj wiedzy, by skuteczniej współpracować z użytkownikiem, często podsuwając rozwiązania, które dotychczas były dużo bardziej czasochłonne, ale nie niezbędne. W dobie automatyzacji, której sztuczna inteligencja jest częścią (automatyzuje, przyspiesza i zastępuje procesy naśladujące sposób, w jaki działa umysł ludzki), orientacja w obecnie tworzonej, upublicznianej i promowanej muzyce jest dziś zapośredniczona przez procesy, na które składają się wymienione procedury, algorytmy czy skrypty. Wynikiem coraz bardziej efektywnego z nich korzystania, spowodowanego zwiększaniem zasobów obliczeniowych i pamięciowych, jest między innymi miraż utwierdzający nas w przekonaniu, że bez ich pomocy i wsparcia świat muzyki byłby znacznie uboższy, a przynajmniej mniej dostępny. Na poziomie tych bazowych dla muzyki w czasach AI procesów, czyli rozpoznania, analizy i gromadzenia wiedzy o materiale dźwiękowym oraz o materiale semantycznym, które definiują muzykę, nie dzieje się nic, co wskazywałoby, że mamy do czynienia z czymś więcej niż wspomnianą automatyzacją i optymalizacją. Oczywiście stanowią one swoisty fundament tego, co dalej dzieje się ze zgromadzonym i zunifikowanym materiałem, na jego podstawie podejmowane są bowiem dalsze procesy, które składać się będą na całościowy obraz inteligentnych systemów tworzenia muzyki lub jej promowania czy rekomendowania.

PROCESY: MASZYNY TWORZĄCE MUZYKĘ (SZUKAJĄCE UŻYTKOWNIKÓW) I POLECAJĄCE MUZYKĘ (TWORZĄCE UŻYTKOWNIKÓW)

W obszarze tworzenia muzyki wykorzystywany jest szereg technologii, które mają różny rodowód, stopień zaawansowania oraz szczegółowe cele do osiągnięcia. Ogólny obraz tego, co powstaje z połączenia tych technologii oraz muzyki, jest znów niejasny, gdy nie posiadamy bazowej wiedzy na temat poszczególnych składowych, które muszą być użyte, by skorzystać choćby z części zapowiadanych możliwości muzycznej sztucznej inteligencji.

Warto przypomnieć, że to, co uzyskują artyści w związku z eksperymentowaniem z technologiami uczenia maszynowego czy sieci neuronowych, a wcześniej z używaniem komputerów do komponowania lub rozpoznawania muzyki, ma niezwykle rozbudowaną historię. Sięgające lat siedemdziesiątych XX wieku eksperymenty rozpoczęły drogę, którą po dziś dzień podążają poszukujący artyści i akademicy konfrontujący nowe osiągnięcia technologii z obszarem sztuki dźwięku¹³. Prace Francois Pacheta, badania i kompozycje Eduardo Reckya Mirandy, wdrożone systemy Magenta firmy Google czy Jukebox od OpenAI, wcześniej gramatyki jazzowe Marka J. Steedmana czy generatywne systemy melodyczne Richarda Pinkertona wyznaczają z wieloma pominiętymi tu osiągnięciami eksperymentalno-artystycznymi bogaty i płodny fundament, na którym rozwijane są komercyjne usługi niezależnych dostawców muzycznego software'u czy największych producentów tej branży. Jednocześnie powstaje to, co stanowi trzon

¹³ T. Acil, *Re-thinking boundaries: The evolution and impact of AI in music and soundscapes*, „Avant” 2(15)/2024.

myślenia o AI w kontekście muzyki, a mianowicie platformy bazujące na muzyce, ale pracujące z innego rodzaju danymi¹⁴.

Serwisy streamingowe muzyki, takie jak Spotify, Tidal, Apple Music czy Deezer, korzystają z różnych metod sztucznej inteligencji oraz uczenia maszynowego do personalizacji doświadczeń użytkowników, rekomendacji utworów, analizy danych oraz zarządzania zawartością. Do najważniejszych technik wspomagających te procesy należy zaliczyć między innymi filtrację kolaboracyjną (*collaborative filtering*) opartą na analizie zachowań użytkowników, dzięki której system rekomenduje utwory na podstawie wyborów użytkowników o podobnym guście muzycznym (*matrix factorization*, *k-nearest neighbors*). Analiza danych o słuchanych utworach, polubieniach, pomijaniu utworów i tworzeniu playlist pozwala tworzyć wzorce preferencji muzycznych. W efekcie rekomendowane są utwory, które mogą spodobać się użytkownikowi na podstawie preferencji innych użytkowników o podobnym guście. Kolejnym procesem jest filtracja zawartości (*content-based filtering*), w której analizuje się same utwory muzyczne, ich cechy akustyczne i metadane oraz teksty piosenek (*natural language processing*). System sprawdza cechy akustyczne utworów, takie jak tempo, tonacja, instrumentacja, by następnie polecić podobne kompozycje. Rekomendacje mogą być również oparte na analizie tekstów piosenek, zidentyfikowane przez podobieństwa tematyczne. Kolejnym procesem jest uczenie głębokie (*deep learning*), które wykorzystuje sieci neuronowe do analizy dużych zbiorów danych muzycznych. Konwolucyjne sieci neuronowe (*convolutional neural networks*) używane do analizy sygnałów dźwiękowych sprawdzają sygnały dźwiękowe utworów, wyodrębniając z nich cechy takie jak rytm, melodia, harmonia. Rekurencyjne sieci neuronowe badają sekwencje odtwarzania utworów przez użytkowników, przewidując, jakie utwory mogą im się spodobać w przyszłości. Na podstawie tych danych system generuje spersonalizowane rekomendacje dla użytkownika. Tak zwane uczenie wzmocnione (*reinforcement learning*) stosowane jest do optymalizacji rekomendacji muzycznych w czasie rzeczywistym i wykorzystuje do tego Q-learning czy metodę Deep Q-Networks. Rekomendacje ujęte są w tym procesie jako serie decyzji, które wpływają na zaangażowanie użytkownika. Na podstawie jego reakcji (jak na przykład polubienia bądź pomijanie utworów) algorytm dostosowuje swoje wskazania, aby maksymalizować jego zadowolenie. System uczy się, które rekomendacje są najbardziej efektywne, i jest w stanie dostosować je w czasie rzeczywistym. Innym procesem jest analiza sentymentu oparta na przetwarzaniu języka naturalnego (NLP). Używa się jej do badania tekstów piosenek oraz recenzji użytkowników w celu zrozumienia emocji i tematów, jak również recenzji i opinii użytkowników, by usprawnić rekomendacje tematyczne. Uczenie maszynowe (*machine learning*) jest powszechnie stosowanym procesem oraz obejmuje szeroki zakres algorytmów i technik używanych do analizy danych, a także generowania

¹⁴ F. Pachet, *The continuator: Musical interaction with style*, „Journal of New Music Research” 3(32)/2003; M.J. Steedman, *The perception of musical rhythm and metre*, „Perception” 5(6)/1977; R.C. Pinkerton, *Information theory and melody*, „Scientific American” 2(194)/1956; *Readings in Music and Artificial Intelligence*, red. E. Reck Miranda, Routledge, London 2000; *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*, red. E. Reck Miranda, Springer, Cham 2021.

rekomendacji. Opiera się w znacznej mierze na takich procesach, jak klasyfikacja, regresja i klasteryzacja. Algorytmy te analizują ogromne zbiory danych dotyczących zachowań użytkowników oraz cech muzyki, systemy rekomendacyjne uczą się preferencji słuchaczy na podstawie ich interakcji z serwisem, modele klasyfikacyjne i regresyjne są zaś używane do przewidywania, jakie utwory mogą się im spodobać.

Podsumowując część dotyczącą sztucznego surogatu muzycznej inteligencji, można stwierdzić, że daleko jej do tego, co wiemy dziś o procesach poznawczych towarzyszących muzyce. Szczególne miejsce w tych procesach, jak w przypadku wszystkich ludzkich aktywności umysłowych konfrontowanych z ich sztucznymi ekwiwalentami, zajmują multimodalność, zdolności mnemotechniczne i ich włączanie w większe kompleksy pracy pamięci dźwiękowej i percepcji słuchowej. Zastrzec należy jednak, że tempo rozwoju badań w tym zakresie jest niebywałe, chociaż nadal są to zagadnienia rozwiązywane na poziomie badań podstawowych lub wstępnych badań rozwojowych i wdrożeniowych. Czym innym jest zatem badawcza perspektywa zespołów naukowych, a czym innym funkcjonująca w publicznym obiegu wiedza i faktycznie wdrożone instrumenty sztucznego wspomagania słuchania i tworzenia muzyki. Badania przynoszą nieraz zaskakujące i bardzo interesujące efekty w obszarach muzyki eksperymentalnej, ale są to zagadnienia niezwykle niszowe. Za to w przypadku komercyjnych zastosowań AI w muzyce należy powiedzieć jasno, że efekty są bardzo ubogie i funkcjonują jako swoista ciekawostka, jednocześnie pączkują w niekontrolowanym tempie i zalewają nas z każdej strony. Nie ma to jednak nic wspólnego z inteligencją muzyczną. Wyłączając z tego obszaru niezwykle ciekawe przykłady udanych zastosowań AI w muzyce – szczególnie w sytuacjach wykonawczych i performatywnych, na które pozwala wydajna i niezwykle droga oraz niszowa technologia, umożliwiająca maszynom uczestniczenie w muzycznej integracji z człowiekiem – przytłaczająca większość muzyki tworzonej za pośrednictwem usług komercyjnych jest odtwórcza i mało oryginalna.

Mamy zatem do czynienia z dwiema perspektywami: pierwszą dotyczącą praktyk badawczo-naukowych, eksperymentalnych i opartych na akademickich działaniach artystycznych oraz drugą – komercyjnych wdrożeń, które nadają ton dyskusji o aktualnym powinowactwie muzyki i komputerów. Pierwszą reprezentują zespoły badawcze rozsiadane po całym świecie, których osiągnięcia można poznać, śledząc najważniejsze coroczne wydarzenia naukowe i artystyczne, drugą zaś – właściciele i twórcy przedsięwzięć wdrażających osiągnięcia badawcze w postaci komercyjnych usług. Ponad tym wszystkim funkcjonuje rozbudowana sfera publicystyczna, która niewątpliwie ma znaczący udział w kreowaniu wizerunku i znaczeń cyrkulujących w obiegu publicznym. Oczywiście jest, że nie dorównują jej skalą środowiska, skądinąd bardzo liczne, skupione wokół takich wydarzeń, jak konferencje AI Music Creativity (od 2012 roku), International Computer Music Conference (od 1986 roku), European Signal Processing Conference (od 1993 roku), Annual ACM Conference on Intelligent User Interfaces (od 1988 roku), International Conference on New Interfaces for Musical Expression (od 2001 roku), International Workshop on Machine Learning and Music

(od 2007 roku), International Conference on New Music Concepts (od 2013 roku), Sound and Music Computing Conference (od 2003 roku), Computer Simulation of Music Creativity Conference (od 2016 roku), International Workshop on Musical Metacreation (od 2012 roku) czy International Conference on Technologies for Music Notation and Representation (od 2014 roku). W roku 2024 odbędzie się ponad czterdzieści takich wydarzeń, w poprzednich latach było to odpowiednio: 26 w 2023, 38 w 2022, 35 w 2021 oraz 48 w 2020 roku.

Czy zatem popularność terminu „sztuczna inteligencja” nie sprawiła, że nasze oczekiwania są zbyt wygórowane? Być może jednak to nie problem naszych oczekiwań, ale raczej deklaracji o doskonałości tych rozwiązań, których celem jest nie tyle ich prezentacja, ile raczej zachęta do ich używania i kupowania? Jeśli przyjmujemy, że „zaczarowano” owe rozwiązania w pojęciu, które do nich nie przystaje, to jak inaczej mówić o muzycznej AI? Czy tak zwana muzyczna sztuczna inteligencja jest faktycznie inteligentna w myśl tego, co moglibyśmy nazwać inteligencją muzyczną w ogóle?

NOWE FIGURACJE: CONCIERGE CZY STALKER?

„Sztuczna inteligencja” z racji tego, co już zostało napisane, funkcjonuje jako figura retoryczna, w związku z tym warto zapytać, czy ta fraza ma zamienniki. Zagadnienia związane ze sztuczną inteligencją mogą być opisane w sposób deskryptywny. Listowanie technik, metod, procedur i algorytmów daje uszczegółowiony obraz efektów rozwoju tych technologii. Warto jednak zadać sobie trud, by tę metaforyczną, pogładową frazę zakwestionować, choćby na potrzeby uwypuklenia tego, co przynosi jej używanie. Można skorzystać z typu opisu, który jest już w humanistyce dobrze rozpoznany i ma bogatą historię oraz wiele zalet szczególnie widocznych w sposobie, w jaki opis ten działa na wyobraźnię i pozwala opowiadać o zjawiskach kulturowych, zwłaszcza tych współczesnych, nowych i niedostatecznie rozpoznanych. Zagadnienia związane z AI w muzyce można zdefiniować w myśl założeń interpretacji figuratywnej, której tradycja sięga początków nowoczesności i rozwijana była przez takich autorów, jak Walter Benjamin, Georg Simmel, Zygmunt Bauman. Co więcej, jej źródła można też doszukiwać się u Ulricha Becka, Guya Deborda, Michela Foucaulta, Anthony’ego Giddensa, Ervinga Goffmana czy Armanda Mattelarta. W pierwszej grupie wymienionych autorów analizy oparte na analogiach figuratywnych przybierały formy opisu zjawisk współczesności wykorzystujące postacie flâneura, obcego, włóczęgi i turysty, gracza, ofiary czy aktora (występującego). W drugiej grupie, mniej ważnej z naszego punktu widzenia, figuracja dotyczyła raczej określonej, dyskursywnie wzmocnionej cechy wspólnot, będącej przedmiotem dociekań. Społeczeństwa ryzyka, spektaklu, informacji czy społeczeństwo panoptyczne miały więc tę własność, że choć nie korzystały z metafory figury *per se*, to jednak eksponowały pewne cechy podmiotowe członków danego społeczeństwa wspólnie z uwarunkowaniami społeczno-kulturowymi, które owe cechy wspierały, wzmacniały lub wręcz wytwarzały. O samym społeczeństwie niewiele mówi nam dodany mu przymiot, jeśli nie wyjaśnimy szerszego kontekstu, w jakim funkcjonują jego członkowie. Podobnie o kulturze, której zmiany zauważamy i próbujemy opisać, możemy mówić tylko

na podstawie tego, w jaki sposób zmiany te odzwierciedlone są w codzienności ludzi żyjących w obszarze ich występowania.

W argumentowaniu na rzecz użycia analizy figuratywnej w przywołanym już wąskim sensie reprezentowanym przez Benjaminowskiego *flâneura* (zaczepniętego od Charles'a Baudelaire'a) bądź Baumanowskie figury pielgrzyma, turysty czy włóczęgi¹⁵, warto nadmienić, że personifikacja, czyli nadawanie cech ludzkich przedmiotom, zwierzętom, roślinom czy zjawiskom, nie jest niczym nowym, szczególnie w omawianym tu przypadku, skoro o zespole określonych technologii informatycznych mających wysoki potencjał zapamiętywania i przetwarzania zaczęliśmy mówić w kontekście ich zdolności do działania w sposób, w jaki funkcjonuje inteligentny ludzki umysł. Poza tym takie typowe figuracje kompetencji, predyspozycji, których sensem jest ich użycie (działanie), często sięgają po figury pomocnika i asystenta. Zdecydowanie rzadziej pojawiają się w tym kontekście postaci doradcy, konsultanta, mentora, których źródła sięgać mogłyby do takich historycznych postaci, jak Niccolò Machiavelli czy Henry Kissinger. Podążając literackim tropem, warto uwypuklić potencjał, jaki kryje się w postaciach fikcyjnych błyskotliwego i skrupulatnego analityka, dr. Watsona z *Sherlocka Holmesa*, twórczego manipulanta i poety Cyrano de Bergeraca pomagającego Christianowi zdobyć serce Roxane, naśladowcę i kontrolującego asystenta pokładowego HAL 9000 z *Odysei kosmicznej 2001* czy też zaślepionego i ogarniętego obsesją Humberta Humberta, śledzącego i kontrolującego tytułową Lolitę Vladimira Nabokova¹⁶.

MUZYCZNY PSYCHOFAN

Pewne jest, że maszyny, które nazywamy zbiorczo sztuczną inteligencją, są naszymi naśladowcami. Nie naśladową nas jednak intencjonalnie, gdyż nie mają świadomości i nie wyznaczają sobie celów, ale robią to na naszą prośbę¹⁷. Rozwijając i karmiąc te technologie, poprawiamy warunki własnego życia, wraz z ich rozwojem chcemy, by coraz więcej decyzji podejmowały za nas zaprogramowane do tego urządzenia. W nadejściu ery urządzeń uczących się tego, co je otacza, oraz uczących się sposobów i regularności ich używania możemy dostrzec golemowski idiom poddanego, stworzonego na podobieństwo swego pana, którego

¹⁵ Z. Bauman, *Ponowoczesne wzory osobowe*, [w:] tegoż, *Dwa szkice o moralności ponowoczesnej*, Wydawnictwo Instytut Kultury, Warszawa 1994; W. Benjamin, *Pasaże*, tłum. I. Kania, Wydawnictwo Literackie, Kraków 2006; Z. Bauman, *Przedstawienie na pustyni*, [w:] „Drobne rysy w ciągłej katastrofie...”. *Obecność Waltera Benjamina w kulturze współczesnej*, red. A. Zeidler-Janiszewska, Wydawnictwo Instytut Kultury, Warszawa 1993; F. Hessel, *Sztuka spacerowania*, tłum. S. Lisiecka, „Literatura na Świecie” 8–9(361–362)/2001; tenże, *Flâneur w Berlinie*, tłum. S. Lisiecka, „Literatura na Świecie” 8–9(361–362)/2001; K. Loska, *Flâneur jako metafora współczesnej kultury*, [w:] *Intermedialność w kulturze końca XX wieku*, red. A. Gwóźdź, S. Krzemień-Ojak, Trans Humana, Białystok 1998; A. Zeidler-Janiszewska, *Dryfujący flâneur, czyli o sytuacjonistycznej transformacji doświadczenia miejskiej przestrzeni*, [w:] *Przestrzeń, filozofia i architektura. Ośiem rozmów o poznawaniu, produkowaniu i konsumowaniu przestrzeni*, red. E. Rewers, Wydawnictwo Humaniora, Poznań 1999.

¹⁶ A.C. Doyle, *Przygody Sherlocka Holmesa*, tłum. E. Łozińska-Małkiewicz, Świat Książki, Warszawa 2023; E. Rostand, *Cyrano de Bergerac*, tłum. C. Clark, Penguin Books, London 2006; A.C. Clarke, *Odyseja kosmiczna 2001*, tłum. R. Kot, Dom Wydawniczy Rebis, Warszawa 2022; V. Nabokov, *Lolita*, tłum. M. Kłobukowski, Wydawnictwo W.A.B., Warszawa 2021.

¹⁷ Stan ten jest jednak przewidziany w badaniach przyszłych zdolności AI i określany mianem ogólnej sztucznej inteligencji (*artificial general intelligence*) lub, wymiennie, superinteligencji bądź transformatywnej sztucznej inteligencji (*transformative artificial intelligence*). Chęć uzyskania tego stanu rozwoju technologicznego jest celem takich firm, jak Open AI oraz Meta.

uwaga i wzrok będą zawsze na niego skierowane¹⁸. Maszyny AI są więc asystentami, pomocnikami, ale też stalkerami i wścibskimi obserwatorami, o znacząco ograniczonych zdolnościach postrzegania rzeczy i zjawisk w ich różnorodności i polisemiczności oraz oceny skali związków znaczeniowych z innymi zjawiskami. Stalker, cichy wielbiciel, miłośnik, prześladowca czy psychofan to określenia kogoś, kto nadmiernie interesuje się czyimś życiem, co powoduje rozmaite konsekwencje. Jedną z nich jest naśladowanie swojej ofiary z jednoczesnym nieuświadomionym staraniem się, by ją „utopić” w pochlebstwach oraz robić zawsze to, co – jak się oprawcy zdaje – ta najbardziej pragnie. To zaś, czego ofiara pragnie, oprawca wnioskuję z bacznej obserwacji. Relacja łącząca nas z maszynami, które nas obserwują, nosi znamiona toksyczności, a wyobrażenia podpowiada szereg przykładów takiej relacji z życia ludzi wchodzących w zamierzoną lub niezamierzoną interakcję z innymi ludźmi. W relacji człowiek – komputer jest podobnie. W sposób zamierzony korzystamy z dobrodziejstw nowych technologii muzycznych wykorzystujących uczenie maszynowe, gdy tworzymy muzykę, kierując naszą uwagę na generatywność, ale w sposób niezamierzony wchodzimy w tę relację, gdy jesteśmy poddani procesom analizy jako osoby słuchające. Zgodnie z rozróżnieniem ze wstępu w pierwszym przypadku zasobem stanowiącym bazę do uczenia maszynowego jest dźwięk i muzyka jako uporządkowany materiał dźwiękowy w całej swej rozpiętości. W drugim przypadku zasobem jesteśmy my sami, słuchacze, jako warianty konstelacji przyzwyczajęń, gustów i dokonanych oraz zarejestrowanych wyborów.

W obu sytuacjach gatekeeperem jest język oraz reguły i warunki translacji. Jakie sensory i znaczenia kryją się za frazami promptu zadanego maszynie, która ma na tej podstawie wygenerować materiał muzyczny? Jakie sensory i znaczenia kryją się za analizą naszych wyborów na serwisach streamingowych, gdy w jej efekcie architektura tych systemów wchodzi w fazę budowania predykcyjnego modelu użytkownika? W obu przypadkach relacja nosi znamiona toksyczności, sytuacji niezdrowej i trudnej do zaakceptowania, gdyż jesteśmy obserwowani, a intencje obserwującego są tylko do pewnego stopnia jawne, zrozumiałe i akceptowalne. Relacja ta wydaje się też nieprzemyślana, gdyż możemy odnieść wrażenie, że mówimy o sobie zbyt wiele lub że zadawane pytania są niestosowne, bo świadczą o tym, że jesteśmy kontrolowani. Tego odczucia najpewniej brakuje w interakcji z komputerami, które stają się naszą nieudolną imitacją.

Budowanie analogii do relacji międzyludzkich może być niesłusznie uznane za próbę nadania maszynie waloru esencjalnego, pewnej podmiotowości czy świadomości. To oczywiste, że nie posiada ona żadnej z tych cech. Ta analogia ma jedynie pokazać naszą pozycję w tym układzie, a jest ona po prostu, mówiąc najbardziej oględnie, mało komfortowa i należy, jak się zdaje, włożyć wiele pracy i wysiłku, by została utrzymana na akceptowalnym poziomie. Tyczy się to relacji jeden na jeden, nad nią mamy bowiem kontrolę. ChatGPT świetnie asystuje w zadaniach,

¹⁸ Postać golema ma, jak się zdaje, te same konotacje i funkcje, wokół których orientuje się niniejszy wywód. Figura wspomniana w biblijnej *Księdze Psalmów* jest częścią tradycji i folkloru kultury żydowskiej, w szczególności za sprawą Judy Loewa ben Bezalela, rabina gminy żydowskiej w Pradze, blisko związanego z Poznaniem.

które są proste, ale czasochłonne. Postawienie granicy, do jakiej ma się angażować, jest podstawową zasadą higieny tej relacji. Lepiej, aby twój fan i wielbiciel wiedział tyle, ile musi wiedzieć, i nic więcej. Gdy ta granica zostanie przekroczona, możliwe, że nasz fan stanie się naszym lustrzanym odbiciem, naszą kopią, nieudolną, powtarzającą nasze dowcipy, udającą sposób, w jaki się ubieramy lub chodzimy. W przypadku relacji z systemami bardziej rozbudowanymi sytuacja jest zgoła inna, bo ich priorytetem jest, abyśmy wielokrotnie przekraczali akceptowalne granice tego, co o nas wiadomo, co ujawniamy i czym się dzielimy.

Na koniec warto dodać, że w świecie międzyludzkich interakcji funkcjonuje oparty na normach zestaw dyrektyw regulujących nasze sposoby wchodzenia w te interakcje. Określa je prawo cywilne, prawo obywateli do ochrony własnej prywatności, ochrony własności oraz nietykalności. Chroni nas prawo do tożsamości, do wizerunku oraz do ubiegania się o pomoc, gdy którekolwiek z powyższych praw zostaje podważone przez czyny innych ludzi. W relacji z maszynami nie posiadamy regulacji, które nas chronią. Stąd jest ona bliższa sytuacjom nieformalnym, swobodnym, dynamicznym, życiowym, ale i ulotnym oraz trudnym do zrekonstruowania. Prawo praktycznie nie działa, gdy ktoś nadmiernie wpatruje się w nas na ulicy, gdy czyjś wzrok zbyt długo skupia się na nas lub na określonych częściach naszego ciała. Prawa nie da się zastosować, gdy ktoś, zachowując wszelkie pozory normalności, wypytuje nas o rzeczy, o które nie powinien pytać. Nierzadko o toksyczności relacji dowiadujemy się również *post factum*, gdy docierają do nas plotki na nasz temat lub gdy pojawiają się niedopowiedzenia i insynuacje.

Technologie sztucznej inteligencji w obszarze muzyki funkcjonują jak prześladowcy. Podśłuchują dźwięki i muzykę, imitują ją, podglądają nasze zachowania, starają się nam przypodobać, narzucają się, a jednocześnie są całkowicie zależne od naszych działań i decyzji, będą podążać za nami wszędzie i zawsze, by każdy nowy wybór, gest i gust utrwalić, zrekonstruować i zaprezentować z dumą przy pierwszej nadarzającej się okazji. Nie jest to jednak żadna wartościowa aktywność z muzycznego punktu widzenia. Należy podkreślić, że nadal zastosowanie komercyjnych rozwiązań w muzyce przynosi marne efekty nie tylko w sferze własności intelektualnej, ale również w sferze estetycznej. Potwierdza to choćby fakt, że największy wysyp systemowych rozwiązań cyfrowych w muzyce możemy obserwować w muzyce technicznie najprostszej, a więc łatwej do odtworzenia i bardzo podobnej nawet wtedy, gdy tworzyli ją tylko artyści¹⁹. Stąd jest w niej

¹⁹ Mam tu na myśli bodaj najbardziej „zagrożone” gatunki muzyki, jak *ambient*, *drone* czy *minimal*. W przeważającej większości usług umożliwiających generowanie muzyki imitacje jej pojawiają się przy okazji wyszukiwania takich fraz, jak: *cinematic*, *atmospheric*, *relaxing*, *ambient*, *calm* itp. Sprawa ta ma jednocześnie wiele innych, dodatkowych wątków wartych odnotowania. Po pierwsze, sami artyści, szczególnie ci związani z tworzeniem muzyki filmowej, wskazują, że dla wielu produkcji telewizyjnych oraz niektórych produkcji filmowych znacznie korzystniejsze z punktu widzenia kosztów jest użycie na wpół gotowych pasaży dźwiękowych, które są dostępne od ręki i tanie. Po drugie, w źródłach muzyki ambientowej kryje się swoisty zarodek tego stanu rzeczy, gdyż muzyka ta nierzadko powstawała przy okazji eksperymentów z muzyką generatywną, tworzoną przy współudziale kompozytora definiującego program lub skrypt oraz komputera, który ten skrypt wykonywał, tworząc w ten sposób niekończące się struktury muzyczne oparte na zasadach rachunku prawdopodobieństwa, warunkach logicznych oraz metodach stochastycznych. Po trzecie, o pewnej dwuznaczności tego fenomenu może świadczyć fakt, że o pojęciu *atmospherics* w kontekście muzyki wypełniającej przestrzeń, ale nieangażującej uwagi słuchacza pisali zarówno teoretycy marketingu POS (*point of sale*), na przykład Philip Kotler, jak i artyści z Jonem Hassellem, twórcą idei *Fourth Worlds Music*, na czele. Nie ulega wątpliwości, że rozumienie tych pojęć u nich obu jest zbudowane na zupełnie

dużo brzmień elektronicznych, statycznych i mających proste struktury melodyczne czy harmoniczne, jak w przypadku szeroko pojętej muzyki tła i innej: od filmowej przez ambientową po muzykę do przestrzeni, towarzyszącą obrazom czy prezentacjom.

PODSUMOWANIE

Świat muzyki pokazuje, że nieustanne powątpiewanie w przyszłość ludzkości, którą zdominują inteligentne maszyny, a więc udające nas inteligentne istoty, jest zdecydowanie nieuzasadnione. Przyjmuje się to z coraz większym niesmakiem, gdy tylko wykona się minimum pracy polegającej na prześledzeniu tego, co za sprawą tych maszyn powstaje i jak powstaje. Zdaje się, że AI absurdalnie niewydajna w sferze kosztów, jakie generuje, jest zupełnie inna niż wyobrażenie, jakie wytwarzamy na jej temat. Po krótkich analizach zdolności sztucznej inteligencji w przytłaczającej większości wypadków przechodzimy natychmiast do daleko siężnych wizji upadku zawodów, redefinicji branż, przemodelowań całych sektorów ludzkiej działalności i w końcu do tworzenia nowych społeczeństw i kultur. Pobrzmiewa tu analogia Jarona Laniera, porównującego maszyny udające inteligentne istoty do samochodu, który staje do wyścigu z człowiekiem. Oczywiście wygrana samochodu wynikająca z pokonania danego odcinka w czasie krótszym niż człowiek nie oznacza, że nazwiemy Bugatti biegaczem²⁰. Nazywamy je samochodem. Co ważne, w innych okolicznościach ten sam samochód nie wybierze drogi dłuższej, ale bardziej malowniczej. Oczywiście ograniczenia, jakie napotyka na swej drodze rozwój AI, wydają się również wystarczające, by odsunąć widmo zagrożenia²¹, jakie miałyby stworzyć „inteligentne” maszyny, gdy tylko staną się

innych fundamentach aksjologicznych, *de facto* dotyczą one praktyk z różnych dziedzin kultury. Łączy je jednak przekonanie o zdolności muzyki do wytwarzania lub modyfikowania przestrzeni (rzeczywistej lub wyobrażonej), którą wykorzystują osoby wspomagające tworzenie fabuł, narracji, środowisk czy enklaw poprzez kreowanie ich sonosfery. Por.: P. Kotler, *Atmospherics as a market tool*, „Journal of Retailing” 49/1973–1974; J. Hassell, *Atmospherics*, Ndeya, Los Angeles 2021.

²⁰ S. Hattenstone, J. Lanier, *Tech guru Jaron Lanier: ‘The danger isn’t that AI destroys us. It’s that it drives us insane’*, „The Guardian”, 23 marca 2023, <https://www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane> (10 czerwca 2024).

²¹ Jednym z bardziej wysublimowanych i ciekawych zabiegów intelektualnych w kontekście tematyki AI jest koncepcja, wedle której współcześnie znana nam, nieokreślona co do parametrów i nierozpoznana co do swego potencjału AI będzie zdolna wytwarzać własne, nieujawnione modele sztucznej inteligencji. W zasadzie robi to już dziś, choćby oferta OpenAI jest profilowana na zachęcanie użytkowników do generowania promptów, na podstawie których maszyna sama wytworzy nowy, sprofilowany na określone zadanie model obliczeniowy, a uczyni to, interpretując kilka zdań języka naturalnego, jakie jej przekażemy. Obawa dotyczy jednak momentu, w którym model taki sam zada sobie trud, by zdefiniować taki profil, cel czy zamiar stworzenia własnej kopii. Ponad fundamentalnym problemem braku świadomości, a więc i celu, który może na jej bazie zostać wykoncypowany, pojawia się niezbywalny, bardzo łatwy do zweryfikowania czynnik materialny. Trudno wyobrazić sobie samoczynne wydatkowanie na rzecz tego wieloaspektowego przedsięwzięcia. Zarządzający infrastrukturą nie byłiby w stanie przeoczyć takiego incydentu. Podważa je jedynie futurystyczna i nihilistyczna wizja, w której sztuczna inteligencja skutecznie, bez wiedzy zarządu i odpowiednich działów przedsiębiorstwa, skrycie rozpoczyna czasochłonny i kosztowny proces tworzenia swojej własnej, wyspecjalizowanej kopii, przekierowuje na ten proces zasoby sprzętowe, uzupełnia magazynowe braki podzespołów, rozlicza nadwyżki zużytej wody i energii elektrycznej, manipuluje raportami okresowymi, ukrywa zestawienia finansowe prezentowane akcjonariuszom, modyfikuje harmonogramy zmian pracowników i ingeruje w ich urlopy, otrzymuje zgody środowiskowe z odpowiednich jednostek terytorialnych, by zwiększyć eksploatację zasobów naturalnych oraz emisję zanieczyszczeń, w końcu skutecznie reaguje na publiczne incydenty, w trakcie których niezależne organizacje, wspólnoty i pojedynczy obywatele protestują z powodów wzmoczonej aktywności przemysłowej.

inteligentne. AI generuje ogromne koszty energetyczne oraz środowiskowe. Tworzenie modeli obliczeniowych jest drogie i wymagające pod względem finansowym, organizacyjnym, formalnoadministracyjnym i sprzętowym.

W tym wszystkim miałyby znaleźć swoje miejsce muzyka, którą taka inteligencja na sposób ludzki musiałaby zacząć tworzyć bez określonego celu ekonomicznego, a z pobudek estetycznych. Muzyka rozumiana jako ludzka aktywność wykonawcza, poznawcza i emocjonalna jest w świecie AI bezpieczna również z innego powodu. Paradoksalnie istnieje scenariusz, który uchroni świat muzyki przed zbyt daleko idącą ingerencją narzędzi tworzenia muzyki oraz tworzenia słuchaczy. W myśl zasady znanej jako *model collapse* narzędzia AI nie mogą być uczone na podstawie danych wytworzonych przez AI. Dzieje się tak, gdy nowo powstające modele uczone są na źródłach wytworzonych przez starsze modele. Miałoby to hipotetycznie zastosowanie w przypadku muzyki pozostającej pod silnym wpływem treści generowanych i rekomendowanych przez modele sztucznej inteligencji. Tworzenie coraz większej ilości muzyki generatywnej i publikowanie jej w komercyjnych systemach streamingowych, a jednocześnie opracowywanie na bazie tego zbioru reguł i prawidłowości związanych z tym, jakie utwory są odtwarzane, w sposób oczywisty prowadzą do incydentów autoreferencyjnych, zwanych również pętlą rekurencyjną²². Wymyślanie muzyki na podstawie trendów popularności, wysycanie nią list odtworzeń, badanie tych list w celu rekomendowania najlepszej zdaniem modeli predykcyjnych muzyki do słuchania oraz tworzenie nowej muzyki, która spełni te kryteria, jest klasycznym błędnym kołem odnoszących się do siebie nawzajem mechanizmów oceny muzyki i słuchania. W granicznym stadium taka sytuacja nie tylko doprowadzić może do rekomendowania muzyki, która będzie coraz bliższa ideałom wyimaginowanym przez maszyny liczące, ale też skutecznie wyeliminuje każdą inną, co w oczywisty sposób wpłynie na popularność platformy udostępniającej muzykę. Ten skrajny przypadek stanie się jednocześnie pełnowartościowym aktem kreacji muzyki wytworzonej w sposób autonomiczny na podstawie wyliczonych, predykcyjnych i autoreferencyjnych gustów. Będzie to nie lada osiągnięcie, gdyż w jego efekcie ujawni się rzeczywisty „muzyczny gust” AI.

Scenariusz ten jest daleki od realizacji, ale z każdym dodanym utworem wytworzonym bez udziału człowieka muzyka, którą udostępniają serwisy streamingowe, oraz sposoby jej użycia będą stanowić źródło analiz i maszynowego uczenia, a utrwalenie tego procesu będzie przebiegało w sposób wykładniczy. Co do samej inteligencji wydaje się, że zbyt łatwo przyszło nam rysowanie paraleli między inteligencją a naśladowaniem jej procesów składowych przez komputery. Pokutuje tutaj nośność modelu predykcyjnego, o którym uczą nas kognytywiści, filozofowie umysłu i przedstawiciele neuronauk, którego odzwierciedleniem są takowe modele stosowane w uczeniu maszynowym czy sieciach neuronowych. W podobny sposób postrzega się to, jak uczymy się i adaptujemy na podstawie danych, by skuteczniej i trafniej przewidywać, oraz to, że w obu przypadkach obowiązuje

²² I. Shumailov i in., *AI models collapse when trained on recursively generated data*, „Nature” 631/2024.

zasada hierarchiczności przetwarzania tych danych: od najprostszych bodźców sensorycznych po złożone reprezentacje poznawcze. Nie zmienia to jednak faktu, że mózgowa architektura o wielkiej złożoności, zdolność mózgu do nieosiągalnej techniczną drogą plastyczności oraz swoboda, z jaką ludzki umysł posługuje się intuicją czy kontekstem, pomagając sobie w sytuacjach niepewności i wieloznaczności, stanowią o jego realnej wyższości ponad wszelką miarę. Wspominanie o emocjach oraz świadomości, a więc i celowości, pozostawmy działom marketingu podmiotów, które wkroczyły na drogę rozwoju technologicznego i w związku z tym zmuszone są dbać o inne podmioty, finansujące ich działania i oczekujące sukcesu oraz zwrotu nakładów.

By zrozumieć głęboki sens muzycznej sztucznej inteligencji, a więc nieludzkiej instancji podmiotowej, świadomej swoich celów i mającej wobec samej siebie plany, musimy wyobrażać sobie wytwór informatycznych technologii, który sam dla siebie bezwarunkowo i wolicjonalnie wydatkować będzie własną energię i zasoby na muzykę, a więc, innymi słowy, będzie czynić dokładnie to, co człowiek. Prowadzi nas to do całego wachlarza koncepcji uzasadniających źródła i potrzebę muzyki oraz muzykowania, jakie wykazuje gatunek ludzki. Nie decydując się na żadną z odpowiedzi i przyjmując cały ich zakres, można mimo wszystko stwierdzić, że każda z nich wydaje się wobec sztucznego bytu całkowicie nieadekwatna. Innymi słowy, nie posiadamy nawet koncepcji sztucznej inteligencji, która byłaby zdolna zdecydować się na taką aktywność i ją skutecznie podejmować. Żaden wytwór współczesnej myśli technicznej nie wytworzy w sobie potrzeby tworzenia lub słuchania muzyki między innymi dlatego, że my sami, jako ludzie, nadal nie mamy pewności, po co to robimy. Tym sposobem dochodzimy do miejsca, w którym uzasadnione jest stwierdzenie, że muzyka AI będzie, póki co i na długo, obszarem działań imitacyjnych, odtwórczych, nieoryginalnych. Używanie AI jako narzędzia wspomagającego pracę twórczą jest oczywistością i sprowadza znaczną większość prognoz dotyczących przyszłości muzyki do poziomu, na którym posługiwanie się pojęciem inteligencji jest nieadekwatne. Czy rozumiemy przez to pojęcie asystowanie, doradztwo, pomoc w prostych, ale czasochłonnych procesach, z czym mamy do czynienia w przypadku ambitnych i kreatywnych połączeń muzyki z komputerami, czy też rozumiemy przez nie naśladownictwo, imitacyjność, kopiowanie, co dotyczy komercyjnych wdrożeń systemów AI do generowania muzyki, oraz śledzenie, manipulowanie i kontrolowanie, z jakimi spotykamy się w przypadku wdrożeń systemów AI do zarządzania, dystrybucji i promocji muzyki, pozostaje, jak się zdaje, osobistą sprawą każdego z nas.

Nie zmienia to jednak faktu, że sztucznie tworzona muzyka powstaje bardzo szybko i jest jej coraz więcej. Nie chodzi tu tylko o takie ekscentryczne ciekawostki, jak utwory niemożliwych duetów (Oasis i John Lennon czy Whitney Houston z Adele), lub dziwnie podobne i już gdzieś słyszane, mimo że ponoć oryginalne, utwory Suno i Uido. Serwisy takie jak Soundful całkowicie zmieniają charakter obecności muzyków w produkcji komercyjnej na potrzeby reklamy, materiałów do mediów audiowizualnych, prezentacji multimedialnych, przestrzeni handlowych czy nawet seriali VOD o niższych budżetach na twórczość muzyczną. Wiele z tych celów realizować można za pośrednictwem takich usług, jak AIVA, Loudly,

wspomniany Soundful, WavTool, Boomy, Soundraw, Orb Producer czy Ecrett Music. Wszystkie te platformy czy ekosystemy dźwiękowe pozwalają tworzyć muzykę na podstawie poleceń tekstowych, wybierać jej warianty i rozwijać ją w rozmaite strony, co w znacznym stopniu przyspiesza pracę tych, którzy potrzebują muzyki uzupełniającej czy towarzyszącej obrazom, wydarzeniom bądź przestrzeniom.

Z drugiej strony automatyzacja procesów związanych z sound designem i edycją dźwięku jest ważnym aspektem obecności tego, co nazywamy sztuczną inteligencją, w muzyce. Nie zmienia to faktu, że mastering utworu muzycznego wykonany w serwisach BandLab, Masterchannel czy Mastering Studio nie będzie dorównywał wielodniowej pracy zespołu studia dźwiękowego, który opracowuje materiał dźwiękowy przez kreatywne użycia niezliczonej ilości technik analogowych i cyfrowych. Narzędzi wspomagających codzienną pracę z materiałem dźwiękowym jest coraz więcej, a wśród nich znaleźć można realizacje, które zdobyły wielką popularność między innymi przez to, że okazały się niezwykle skuteczne w oszczędzaniu każdego rodzaju zasobów.

Jednak w żadnym z powyższych przypadków nie sposób porównać zdolności komputerów do odtwarzania tego, co zapamiętały, lub wspomagania procesów, które nam, ludziom, zajmują więcej czasu, do kreatywnego działania, do którego zdolny jest człowiek. Muzyczna sztuczna inteligencja nie jest w żadnej mierze inteligentna wedle kryteriów tworzenia muzyki jako czynności o arcyłudzkiem, procesualnym, emocjonalnym i intencjonalnym charakterze.

BIBLIOGRAFIA

- Bauman, Zygmunt. *Dwa szkice o moralności ponowoczesnej*. Warszawa: Wydawnictwo Instytut Kultury, 1994.
- Benjamin, Walter. *Pasaże*. Tłum. Ireneusz Kania. Kraków: Wydawnictwo Literackie, 2006.
- Hassell, Jon. *Atmospherics*. Los Angeles: Ndeya, 2021.
- Klapuri, Anssi, Manuel Davy, red., *Signal Processing Methods for Music Transcription*. New York: Springer, 2006.
- Kotler, Philip. „Atmospherics as a market tool”. *Journal of Retailing* 49 (1973–1974).
- Lange, Rafał, red., *Sztuczna Inteligencja w społeczeństwie i gospodarce. Analiza wyników ogólnopolskiego badania opinii polskich internautów*. Warszawa: NASK Państwowy Instytut Badawczy, 2019.
- Li, Tao, Mitsunori Ogiwara, George Tzanetakis, red., *Music Data Mining*. CRC Press: Boca Raton, 2012.
- Müller, Meinard. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Cham: Springer, 2015.
- Reck Miranda, Eduardo, red., *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Cham: Springer Nature, 2021.
- Reck Miranda, Eduardo, red., *Readings in Music and Artificial Intelligence*. London: Routledge, 2000.

MY SILENT ADMIRER, YOUR LOVER, OUR PSYCHOFAN, YOUR PERSECUTOR: MUSIC AI AND ITS ALTERNATIVE FIGURATIONS

This article investigates the transformative role of Artificial Intelligence (AI) in the music industry, focusing on the complexities of copyrights and ethical implications of AI-generated music. It discusses the evolving meaning of the phrase 'Artificial Intelligence' in the world of music and proposes an alternative form of conceptualising AI's performative and active presence = through alternative figurative interpretations. While AI is widely employed to tailor music experiences by analysing user data and preferences, its commercial applications often face criticism for their lack of originality and a tendency to repeat musical structures. In this article, AI is presented from a dual perspective: (1) AI as a support tool for simple engineering tasks in music production; and (2) AI as a technology with the ambition to create music in a human-like manner. However, neither approach suggests that the 'intelligence' touted by both enthusiastic and pessimistic perspectives on current and future AI products can approximate what we might define as true 'musical intelligence'. Despite substantial investment and technological advancements, AI largely remains an imitative tool rather than a truly creative force, raising questions about its authentic value in the music industry.

SŁOWA KLUCZOWE: sztuczna inteligencja w muzyce, muzyka generatywna, personalizacja muzyki, technologie muzyczne, przyszłość muzyki

KEY WORDS: Artificial Intelligence (AI) in music, generative music, music personalisation, music technologies, future of music