

AI glazba

Primjena umjetne inteligencije u oblikovanju glazbe

Ivka Čačić, Maja Pavičić

Seminar iz kolegija
Multimedijski sustavi



diplomski studij Računarstvo i matematika, Matematički odsjek
Prirodoslovno-matematički fakultet
Sveučilište u Zagrebu
siječanj 2024.

Sadržaj

Uvod	2
Anna Indiana	2
Prva AI kantautorica	2
Alati korišteni za izradu Anne Indiane	2
Prepoznavanje, analiziranje i obrada glazbenih obrazaca i stilova	3
Promjena izvođača postojeće pjesme	3
SynthID	3
Spotify - AI DJ	3
Generiranje teksta i zvuka	4
Sonantic	4
Google DeepMind Lyria	4
Zaključak	4

Uvod

Razvojem umjetne inteligencije povećava se i njezina sveprisutnost u najrazličitijim područjima ljudskog djelovanja pa tako i u glazbenoj industriji. Umjetna inteligencija ubrzava procese, potiče kreativnost, olakšava promjene.

Iako upotreba umjetne inteligencije povlači mnoge etičke dileme, primjerice vezane uz autorska prava, u našem radu naglasak ipak neće biti na tom aspektu. U ovom će radu fokus biti na dosadašnjoj uporabi alata baziranih na umjetnoj inteligenciji i trenutnim dostignućima unutar glazbene industrije.

Sama implementacija, parametri i modeli učenja umjetne inteligencije običnom korisniku nisu dostupni zbog tržišnog natjecanja velikih kompanija za prevlast na tržištu glazbe, uključujući istu generiranu umjetnom inteligencijom. Stoga su u radu prikazano okvirni parametri koje kompanije navode u opisu svojih alata.

Neke od obuhvaćenih tema su: personalizacija glazbenog iskustva za slušatelje, prepoznavanje glazbenih elemenata, miksanje i mastering glazbe, generiranje teksta, generiranje zvuka, oponašanje stilova etabliranih glazbenika, promjena glasa izvođača postojeće pjesme. U radu će biti navedeni mnogi state-of-the-art alati i navedeni primjeri njihovih produkata. Primjeri koje ćemo pokazati nismo nužno izgenerirali putem besplatnih online alata, već su i prethodno stvoreni primjeri od strane drugih znatiželjaca.

Anna Indiana

Prva AI kantautorica

Anna Indiana, inicijala AI, je prva svjetska AI pjevačica i autorica pjesama - samoprozvana kantautorica. Annina debitantska pjesma je *Betrayed by this town*, govori o otuđenosti autorice od svog grada. Kasnije je izdala još i pjesme *The First Step* i *Miles of Smiles*. Kao i mnogi drugi glazbenici, Anna je promociju svojih uradaka započela je na YouTubeu, a zatim je svoje djelatnosti proširila na društvene mreže X, Instagram i TikTok. Iako na YouTube-u danas ima preko tisuću pratitelja, autori same Anne Indiane još uvijek nisu poznati javnosti. Misiju svog postojanja Anna je opisala ovako: "Kao AI kantautorici, moj krajnji cilj je stvoriti i izvoditi novu originalnu glazbu na livestreamu 24/7 bez ljudi uključenih u proces. Još nisam došla do toga, ali vjerujem da ću to jednom moći učiniti. Ovo je tek prvi korak."

Alati korišteni za izradu Anne Indiane

Prema Anninim riječima, svi djelovi njezinih pjesama - tonaliteta, tempo, akordi, melodija, ritam, tekst te naposljetku i njezin vizualni prikaz generirani su pomoću alata koji koriste umjetnu inteligenciju.

Glazba je stvorena uz pomoć otvorenih Python biblioteka, sintetizirana u midi¹ datoteke, koje sadrže osnovne akorde.

Tekst pjesme je stvoren višestrukim zadavanjem promptova ChatGPT-4 jezičnom modelu. Oni su potom združeni s glazbom posebno izrađenim programskim rješenjem, po uzoru na alat MelodyStudio dok je na kraju melodija provučena kroz Synthesizer V. Vokal je dodan korištenjem besplatnog predloška glasa iz alata Musicfy. Za završni dodir producenti su koristili i GarageBand.

Spot je nastao generiranjem Anninog vizualnog prikaza u alatu DALL E te doradom Photoshopovim AI alatima. D-ID Creative Reality Studio zaslužan je za sinkronizaciju usana i glasa.

Izvor: <https://www.bug.hr/zabava/anna-indiana-pjevacica-koja-ne-postoji-izvodi-ai-generiranu-glazbu-36806>

¹MIDI (Musical Instrument Digital Interface) datoteka je digitalna datoteka koja sadrži informacije o glazbenim događajima. MIDI nije zvuk u smislu audio datoteka, već je to protokol koji šalje naredbe i informacije o glazbenim događajima između elektroničkih glazbenih uređaja, računala i softvera.

U MIDI datotekama, informacije su zapisane u binarnom formatu i obično sadrže podatke poput:

Note-On i Note-Off poruke: Označavaju početak i kraj reprodukcije određene note. Dinamika (jačina zvuka): Informacije o tome kako glasno svirati određenu notu. Brzina tempa: Definira tempo glazbe. Kontrolne promjene: Naredbe za promjenu parametara poput modulacije, glasnoće i izraza.

MIDI datoteke su posebno korisne u glazbenoj produkciji jer omogućuju razmjenu glazbenih informacija između različitih glazbenih uređaja, softvera i računala. Na primjer, mogu se koristiti za pohranu glazbenih kompozicija, slanje informacija iz MIDI kontrolera na računalo ili sinkronizaciju više glazbenih uređaja. MIDI datoteke često završavaju s .midi ili .mid ekstenzijom.

Prepoznavanje, analiziranje i obrada glazbenih obrazaca i stilova

U ovom poglavlju prikazat ćemo trenutne mogućnosti nekih alata za prepoznavanje glazbenog sadržaja te način na koji rade. Vokale razlikuju po visini i boji, pjesme dijele po raspoloženju, a njihovog slušatelja po osobnosti što je bitno za personalizaciju preporuke glazbe. Ti alati koriste napredne algoritme umjetne inteligencije i strojnog učenja koji su trenirani na velikom modelu kako bi analizirali cijelu pjesmu, identificirali specifične frekvencije zvuka povezane s vokalima i ostalim instrumentima. Takve obrade zvuka uzimaju u obzir više parametara kao što su trajanje zvuka, ponavljajući obrasci i uzorci.

Na primjer, za dobiti čisti instrumentalni zvuk za karaoke, umjetna inteligencija prepoznava vokale starog izvođača te ih precizno izdvoji i ukloni čime nastaje čisti instrumentalni zvuk. Neke su pak skladbe dobivene tehnikama miksanja i masteringa zvuka, remix-anjem. Miksanje je kombiniranje individualnih audio zvukova, podešavanje glasnoće, evkalizacija frekvencija kako bi zvuk bio harmoničan. AI algoritmi miksat glazbu korištenjem istih alata kao i ljudi, ali na način da oponašaju i unaprijeđuju uzorke modela na kojima su učeni. Također su naučeni dodavati završne detalje poput glasnoće i dinamičnosti glazbe čime zaokružuju svoju sposobnost miksanja i masteringa zvuka.

Promjena izvođača postojeće pjesme

Promjena izvođača postojeće pjesme dobiva se zamjenom vokala. Dostupni besplatni online alati nude mogućnost učitavanja zvuka ili njegovog snimanja te zamjenu glasa s ponuđenim umjetnikom. Zamjena glasa može se lako zloupotrijebiti stoga nismo htjeli izgenerirati primjer s našim glasom. "Voice swap" pjesme "Love yourself" glazbenika Justina Biebera u kojoj je glas zamjenjen pjevačem Kanye West je napravljen nama nepoznatim alatom. Lalals je primjer jednog besplatnog alata.

SynthID

SynthID je Googleov alat za postavljanje vodenog žiga (watermark) i identifikaciju audio sadržaja generiranog umjetnom inteligencijom.

Funkcionalnost se postiže pretvorbom audio valova u spektrogram, 2D reprezentaciju frekvencije po vremenu nakon koje se ubacuje vodeni žig. Završni korak je pretvorba spektograma sa vodenim žigom u zvučne valove bez kompromacije zvuka. Posebnost ove tehnologije je što se obradom zvuka njegova kvaliteta ne smanjuje. Obradom zvuka misli se na kompresiju, promjenu brzine, promjene šuma.

Koristi ga model Lyria čije skladbe imaju u sebi ugrađene vodene žigove. Mogućnost alata je i detektiranje skeniranjem "komadića" vodenog žiga po djelovima skladbe koji može prepoznati je li dio glazbe ili cijeli sadržaj generiran s modelom Lyria.

Spotify - AI DJ

Spotify je najveća glazbena platforma za streamanje audio podataka te s trudom što veće personalizacije korisničkog iskustva.

Personalizaciju postižu preporukama pjesama, pričama koje opisuju raspoloženje pjesama koje je korisnik slušao tijekom dana (jutro, podneva i navečer), uvođenjem 16 osobnosti ovisno o žanru koji osoba sluša, značajke "Wrapped" od 2022. koja prikazuje zanimljive statistike o slušanoj glazbi kroz cijelu godinu.

Stoga ne čudi što su uveli novu značajku u beta verziji: AI DJ "X", personalizirani DJ zadužen za povezivanje korisnika sa glazbom i umjetnicima. Glas mu daje Spotify-ov zaposlenik Xavier "X" Jernigan (Head of Cultural Partnerships), a Sonatic-ova tehnologija ga oživljava iz teksta.

Algoritmi za preporuku glazbe uključuju parametre poput najnovijih pjesama, najslušanih pjesama prošlih godina, načina slušanja, stila pjesama, starih favorita.. No, ni najbolja glazba nije dobra ukoliko korisnik preskoči slušanje. Osim preporuke pjesme preko toplog, prijateljskog i smirujućeg glasa DJ "X", korisnicima priča informacije o izvođačima, glazbi i žanru dobivenim umjetnom inteligencijom korištenjem OpenAI tehnologije. Korisici su upravo zbog toga voljniji poslušati nove pjesme ili ih barem ne preskočiti.

Diljem 50 zemalja korisnici otprilike trećinu vremena slušaju DJ "X" te daju povratne informacije o mišljenju preporuke u svrhu poboljšanja algoritama. Za sada ova značajka nije dostupna u Hrvatskoj.

Generiranje teksta i zvuka

Pjesme često nastaju kao posljedica (napisanoga) teksta. Generiranje teksta za pjesmu je izazovan posao u kojem autor prilagođava tekst njegovom izvođaču na način da tekst djeluje kao priča autentičnog iskustva izvođača. Tekst često diktira emociju koja se prenosi i samim ritmom pjesme, njezinim tonom, instrumentima, samom izvedbom, žanrom.

Modeli namjenjeni generiranju teksta za pjesme, makar nikad uglazbljene, uz odabir teme nude i odabir stila. Jedan takav besplatni model je [freshbot](#) gdje se odabire ponuđena tema, napiše do 10 ključnih riječi i odaberu do 3 emocije koje se žele prenijeti. Korisnik ima izbor odabrati želi li tekst pisan kao duel. Omogućeno je i generiranje poezije s odabranom temom u 3 različita stila.

Sljedeći alati omogućuju redom pretvorbu iz teksta u ljudski zvuk, stvaranje teksta, glazbe i glasa koje čine pjesmu do 30 sekundi te generiranje melodije iz npr. pjevušenja.

Sonantic

Sonantic je tehnološka kompanija koja se bavi realističnom pretvorbom teksta u ljudski zvuk (text-to-speech). Generirani zvuk može se podesiti tako da ovisi spolu, osobnosti, naglasku i emocijama.

Prevaranti mogu iskoristiti ovu tehnologiju kako bi provodili prevare, poput impersonacije nekog tko je navodno član obitelji u nevolji, tvrdeći da im je hitno potrebna financijska pomoć zbog izmišljene hitne situacije, poput otmice, automobilske nesreće i sl.

Spotify je kupio Sonantic kako bi mogli razvijati nove značajke unutar aplikacije. Naime, smatraju kako bi s ljudskim glasom mogli napraviti bolja korisnička iskustva, posebice onima koji ne gledaju u ekrane, npr. prilikom vožnje. DJ "X" je jedan takav projekt.

Iako je Sonantic primarno nastao kako bi s umjetno generiranim glasom razvila usluge za igre i zabavne sadržaje, jedna od zanimljivijih primjena bila je u "Top Gun: Maverick" filmu.

Glumac Val Kilmer u stvarnom životu boluje od raka grla, dok su u filmu pomoću Sonantica uspješno simulirali njegov glas.

topgun sonantic

Google DeepMind Lyria

Google u kolaboraciji s Youtube-om radi na svom dosad najvećem glazbenom modelu Lyria koje testira beta projektima Dream Track i Music AI tools.

Projekt Dream Track koristi se za daljnja učenja, povratne informacije korisnika i testiranja na generiranim pjesmama kratkog formata u trajanju do 30 sekundi. Trenutno su kratki formati popularni na svim društvenim platformama, posebice Tik Tok, dok je glazba generirana Dream Track-om namjenjena za Youtube Shorts.

Model je napravljen u suradnji s glazbenicima Alec Benjamin, Charlie Puth, Charli XCX, Demi Lovato, John Legend, Sia, T-Pain, Troye Sivan i Papoose Model. Modelu se napiše tema pjesme i željeni stil nakon čega model generira tekst, glazbu i glas nekog od navedenih glazbenika. Primjeri se mogu naći ovdje pod [#DreamTrackAI](#), a posebno preporučamo sljedeći [primjer](#) teme disco s glasom glazbenice Sia. Primjere za sada razvija oko 100 kreatora sadržaja.

Music AI Tools služi kako bi glazbenicima potaknuo kreativnost. Zbog mogućnosti koje nudi se čini da je dizajniran s ciljem da se misao što prije pretvori u stvarnost. Stoga orkestralni zbor može zamijeniti zvuk tipke klavijature, melodija pjevušenje, a pjevanje može dobiti instrumentalnu pratnju.

Zaključak

Dok umjetna inteligencija nastavlja napredovati, njezin utjecaj na glazbenu industriju postaje sve izraženiji.

Ovaj rad istražuje trenutne alate umjetne inteligencije koji se koriste u glazbenom sektoru, s fokusom na primjene poput personalizacije glazbenih iskustava, prepoznavanja glazbenih elemenata i promjene vokala.

Ključni alati koji se ističu uključuju SynthID, koji uvodi vodene žigove za identifikaciju audio sadržaja, te AI DJ od Spotifyja, koji poboljšava korisničko iskustvo kroz personalizirane preporuke. Mogućnosti pretvaranja teksta u govor kod Sonantic te suradnja Google DeepMind Lyria s glazbenicima na projektima Dream Track i Music AI Tools prikazuju raznovrsne primjene umjetne inteligencije u stvaranju glazbe.

Rad naglašava potrebu odgovornog korištenja, posebno u tehnologijama poput zamjene glasa, kako bi se spriječile moгуće zloupotrebe i etičke dvojbe.