

Tanulmányok

MESTERSÉGES INTELLIGENCIA A ZENÉBEN

ARTIFICIAL INTELLIGENCE AND MUSIC

Szekanecz Máté¹, Szekanecz Zoltán^{2*}

¹MSc-hallgató, Budapesti Corvinus Egyetem, Budapest
szekanecz.mate@gmail.com

²egyetemi tanár, Debreceni Egyetem Általános Orvostudományi Kar Reumatológiai Tanszék, Debrecen
szekanecz.zoltan@med.unideb.hu

ÖSSZEFOGLALÁS

A mesterséges intelligencia (AI) a zene különböző területeivel függ össze, beleértve a zeneszerzést és az előadásmódot is. A zene és AI összefüggéseinek kulcsfontosságú része olyan zenei algoritmusok fejlesztése, amelyek AI-t használnak a zeneszerzéshez. Az AI jellemzője, hogy az algoritmus képes tanulni a múltbeli adatok alapján. Az AI az előadóművészeket hallgatva is képes tanulni. Az AI az interaktív zeneszerzést is vezérli, amikor egy számítógép komponál zenét egy élő előadást hallgatva. Az AI a zenei teljesítményt is fokozhatja. A számítógép és a zeneszerző/előadóművész közötti együttműködés döntő fontosságú ebben az esetben. Számos zenei programot fejlesztettek ki a hangfelismerés és a természetes nyelvi feldolgozó technológia használatára a zenei hangvezérléshez.

ABSTRACT

Artificial intelligence (AI) has been implicated in various areas of music including music composition and performance. A key part of the relationship between music and AI is the development of music algorithms which use AI to produce music. A prominent feature of AI is the capability of the algorithm to learn based on past data. AI is also capable of learning while listening to a human performer. AI also drives interactive composition technology, where a computer composes music in response to a live performance. AI might also assist music performance. Cooperation between the computer and the composer/performer is crucial in this setting. Several music programmes have been developed to use voice recognition and natural language processing technology for music voice control.

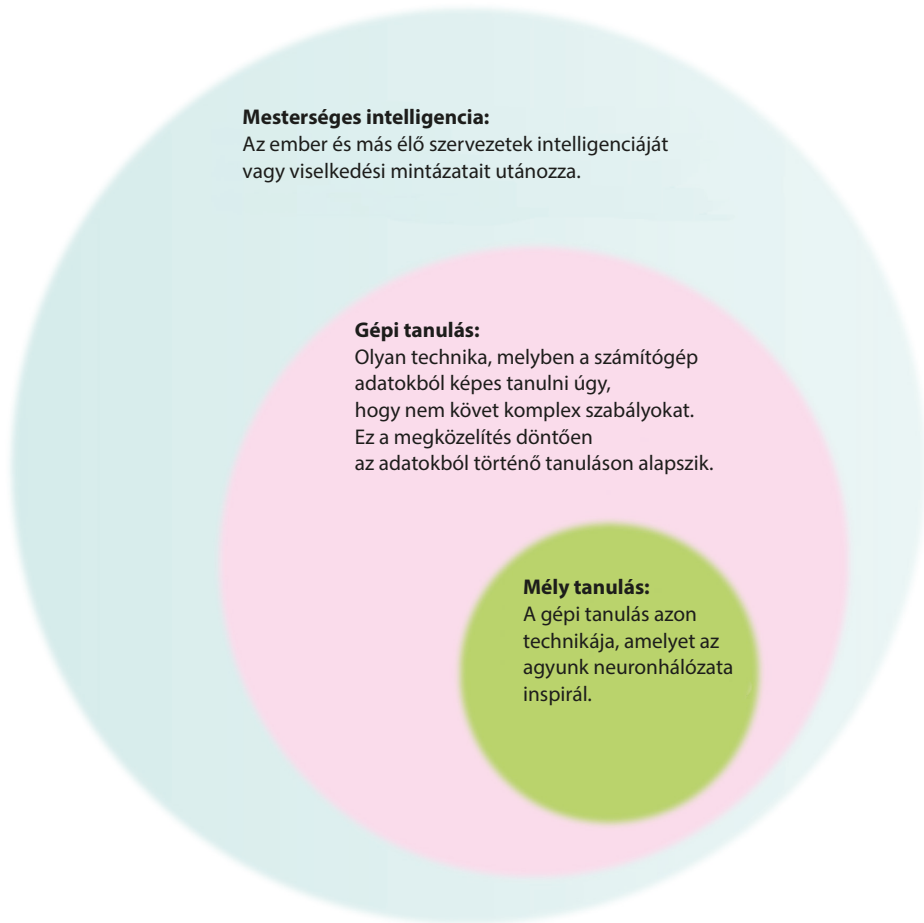
Kulcsszavak: mesterséges intelligencia, gépi tanulás, zene, zeneszerzés, előadás

Keywords: artificial intelligence, machine learning, music, music composition, performance

* levelező szerző

BEVEZETÉS: DEFINÍCIÓK

Az elmúlt évtizedben jelentős áttörés ment végbe a digitális technológia területén. A **mesterséges intelligencia** (Artificial Intelligence, **AI**) az az intelligencia, amelyet egy gép, számítógépes program vagy mesterségesen létrehozott tudat nyilvánít meg (1. ábra) (Meskó–Görög, 2020; Szekanecz et al., 2023). Az AI általában olyan fejlett számítógépes rendszerre utal, amely képes olyan műveleteket végrehajtani, amelyek egyébként emberi intelligenciát igényelnek. A leggyakoribb példa a vizuális észlelés, a beszédfelismerés, a döntéshozatal vagy a nyelvek közötti fordítás. A mindennapi életünkben is használunk olyan alkalmazásokat (például az okostelefon arc- vagy beszédfelismerő funkcióit

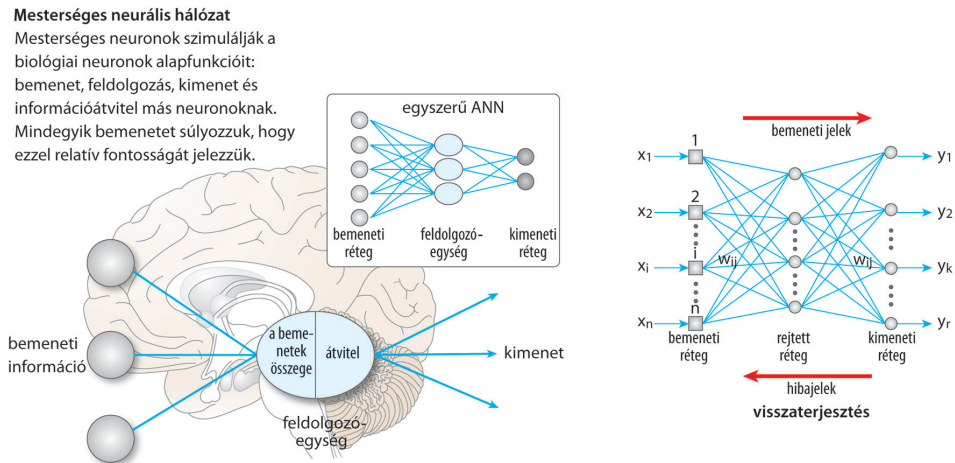


1. ábra. A mesterséges intelligencia, gépi tanulás és mély tanulás (saját szerkesztés)

[Siri], nyelvi fordítóprogramokat, navigációs rendszereket), amelyek AI-t használnak. A sakkautomaták is döntéshozatalon alapulnak. A modern AI-kutatás elsősorban arra irányul, hogy a létrehozott program képes legyen tanulni, azaz célirányosan és reprodukálhatóan megváltoztatni a viselkedését (Meskó–Görög, 2020; Szekanecz et al., 2023).

Az AI magában foglalja a **gépi tanulást** (Machine Learning, **ML**) (1. ábra) (Schmidhuber, 2015). Az ML olyan tanulási folyamatot jelent, ahol a program a már megszerzett adatokat használja fel a további feladatok hatékonyságának növelésére. A számítógépes tanulási algoritmusok a bejövő adatok alapján modellt építenek, és ezek alapján előrejelzéseket, döntéseket hoznak. Az ML az emberi agy működését utánozó idegi (neurális) hálózatokat is használ (Schmidhuber, 2015).

Az ML szerkezeti és működési alapja az idegrendszer példájára kifejlesztett **mesterséges neurális hálózat** (Artificial Neural Network, **ANN**) (2. ábra). Az ANN-en belül a rétegekbe rendezett mesterséges „idegsejtek” nonlinearis aktivációs függvényeken keresztül kommunikálnak egymással. Az ANN-rendszerek kapacitásának növekedésével egyre bonyolultabb architektúrák jelentek meg, amelyek mélyebb és különböző típusú neuronrétegeket halmoztak fel, és változatos elágazásokat tartalmaztak. A neurális hálózatok elmélyülésével nőtt absztrakciós képességük, így ezek a rendszerek egyre bonyolultabb feladatok megoldására váltak alkalmassá. Az ML ezen fejlett változatát **mély tanulásnak** (Deep Learning, **DL**), a DL-ben részt vevő ANN-eket pedig **mély neurális hálózatoknak** (Deep Neural Network, **DNN**) nevezzük (Szekanecz et al., 2023; Schmidhuber, 2015).



2. ábra. A mesterséges neurális hálózatok (ANN) lényege és működése (saját szerkesztés)

AZ AI ALAPJAI ÉS TÖRTÉNETE RÖVIDEN

Bár az AI alapvetően a tudományos-fantasztikus irodalom terméke, mára olyan tudományággá vált, amely a valós életből fakadó problémákra próbál választ adni. Az AI-n alapuló rendszereket széles körben alkalmazzák a közgazdaságtanban, az orvostudományban, a tervezésben, a közlekedésben, az űrkutatásban, a hadviselésben, a háztartási gépekben. Emellett, a lakosság körében elterjedt a számítógépes programok, mobilapplikációk és videójátékok kapcsán (Meskó–Görög, 2020; Szekanecz et al., 2023).

Ami az AI rövid történetét illeti, az első mechanikus digitális számológépet Blaise Pascal találta fel a 17. században. Már a 19. században kifejlesztették az első programozható számológépeket. A neurális hálózatok elméletének alapjait 1943-ban Warren McCulloch és Walter Pitts teremtette meg. A „mesterséges intelligencia” kifejezést először John McCarthy használta egy tudományos konferencián az 1950-es években. Az első primitív „beszélő” robotot (Eliza) Joseph Weizenbaum alkotta meg szintén az 1950-es években. Az 1960-as és 1970-es években az AI-kutatás felgyorsult. Többek között megállapítást nyert a szabályalapú rendszerek jelentősége az orvosi diagnosztikában és terápiában. Hans Moravec kifejlesztette az első számítógéppel vezérelt járművet, amely önállóan navigált egy akadályokkal felszerelt pályán. 1970-ben Paul John Werbos írta le először a neurális hálózatok és az ún. visszaterjesztési (backpropagation) algoritmusok kombinált használatát, amelyek az AI működésének alapját képezik. Az 1990-es években már számos AI-alapú alkalmazást fejlesztettek ki. Nagy nyilvánosságot kapott, amikor 1997-ben a Deep Blue sakkszámítógép legyőzte Garri Kaszparovot. A Google 2017-ben bemutatott *DeepMind* programja már négy óra alaptanulás után képes volt legyőzni bármely sakkozót. Az önvezető járművek fejlesztését a 2000-es években kezdte meg a Tesla, az Audi és a Google. A Huawei Mate 10 volt a világ első mesterséges intelligenciával felszerelt mobiltelefonja. Végül, 2017-ben egy humanoid robot, Sophia kapott először állampolgárságot a világon (Meskó–Görög, 2020; Szekanecz et al., 2023).

Az elméletileg végtelen felhasználási lehetőség mellett röviden ki kell térnünk az AI lehetséges korlátaira és veszélyeire is. A tudósokat aggasztja a „szuperintelligencia” fejlődése, amely még a legfényesebb emberi elmét is felülmúlhatja, és potenciálisan önfejlődővé és ellenőrizhetetlenné válhat. Egyesek szerint az AI az evolúció következő szakasza lehet. A növekvő automatizálás miatt a munkanélküliség jelentősen növekedhet. Az AI, beleértve például az arc- és hangfelismerést, szintén kedvezőtlen felhasználásokhoz vezethet (például fegyverkezés, kémkedés, lehallgatás, terrorizmus, személyes propaganda). Amellett, hogy olyan gyógyszereket terveznek, amelyek reményt adnak az emberiségnek, órák alatt több ezer toxikus molekulát is meg lehet tervezni vegyi fegyverként (Szekanecz et al., 2023).

AZ AI ELEMEI ÉS MŰKÖDÉSE

Gépi tanulás

A gépi tanulás három fő formája a felügyelt, felügyelet nélküli és megerősített tanulás. A felügyelt tanulás során folyamatos segítség, ellenőrzés történik. Az oktató (fejlesztő) pontosan ismeri a tanítás végső célját, meghatározza a várható választ, és az algoritmus megtanulja ennek a feladatnak a megoldását. Ez a leggyakrabban használt ML-típus. A felügyelet nélküli tanulás során az oktató nem befolyásolja a tanulási folyamatot, hanem a feladat megoldása során figyel az algoritmus által levont következtetéseket. Az ilyen típusú ML-ben bizonyos szabályok felállítása után hagyjuk, hogy az algoritmus magától tanuljon. Az eredmények alapján nem módosítjuk az algoritmust, hanem megfigyeljük, hogy milyen következtetésekre jutott. Végül a megerősítő tanulás során az oktató kitűz egy célt, de nem határozza meg lépésről lépésre, hogy az algoritmus hogyan tanulja meg azt. Ehelyett a fejlesztő csak a feladat elvégzése után ad visszajelzést, ezzel arra ösztönzi a rendszert, hogy a kapott eredmények alapján megtalálja a saját stratégiáját. Az AI esetében az algoritmus beépítheti saját tapasztalatait, miközben egyre több feladatot lát el (Meskó–Görög, 2020; Szekanecz et al., 2023; Schmidhuber, 2015).

A mély tanulás (DL) az ML egy speciális, továbbfejlesztett formája (1. ábra). A DL használatával nagyon összetett adathalmazok elemezhetők a képektől és videóktól az emberi érvelésig. A DL az emberi agy által ihletett ANN-t és DNN-t használ (lásd alább). Nagyon egyszerűen, minél több rétegből áll egy hálózat, annál összetettebb feladatokat tud megtanulni, ugyanakkor több adatot és hosszabb tanulási időt igényel. A DL-modellek jó eredménnyel képesek képeket, hangokat és egyéb többdimenziós adatformákat feldolgozni (Meskó–Görög, 2020; Szekanecz et al., 2023; Schmidhuber, 2015).

Mesterséges és mély neurális hálózatok

A mesterséges neuronok az emberi idegrendszer alapvető funkcióit másolják. Ezek a funkciók magukban foglalják a bemenetet, a feldolgozást, a kimenetet és az információ más neuronokhoz való továbbítását. Minden bemenet súlyozott, hogy jelezze, mennyire fontos a többi bemenethez képest (2. ábra). Az eredeti ANN (Perceptron) kétrétegnyi, egymáshoz kapcsolódó neuront használt, és egyszerű képfelismerő feladatok elvégzésére volt tanítható. A vizuális kéreg jobb megértése több, de kisebb/egyszerűbb rétegből álló ANN kifejlesztéséhez vezetett (Neurocognitron). Végül, még több réteg használata „mélyebbé” teszi a hálózatot (DNN), és lehetővé teszi számára, hogy az absztrakció több szintjén keresztül érzékelje a világot. A DNN-ek kiválóan teljesítenek bizonyos észlelési fel-

adatokban, de nehezebb betanítani őket. Ezért az informatikusok kifejlesztették a visszaszaporítást (backpropagation), a DNN-ek betanítására alkalmas technikát (2. ábra). DNN és a backpropagation kombinálása már nagyon jó eredményeket ért el például a kép- és hangfelismerés területén. A bankokon kívül olyan nagyvállalatok is elkezdtek ilyen kép- és hangfelismerő funkciókat használni, mint a Google, Facebook, Microsoft, Apple, IBM és mások (Meskó–Görög, 2020; Szekanecz et al., 2023).

AZ AI A ZENÉBEN: ALAPOK ÉS RÖVID TÖRTÉNET

Az AI és a zene kapcsolata többek között magában foglalja az AI alkalmazását a zeneszerzés, zenei előadás, digitális hangfeldolgozás, valamint a zenemarketing és -fogyasztás során. Például a zeneszerzés kapcsán az AI-algoritmusok képesek emberi előadót hallgatni, és kíséretet előadni. Az AI interaktív kompozíciós technológiát is vezérelhet, ahol a számítógép egy élő előadásra válaszul komponál zenét (Miranda, 2021).

Ami a rövid történetet illeti, 1957-ben Lejaren Hiller és Leonard Isaacson döbbsentette meg a zene világát azzal, hogy bemutatották az első AI-rendszerrel felépített kompozíciót, az *Illiace szvitet* (URL1, QR1).



Illiace volt az egyik első számítógép neve, amelyet valaha építettek, 1952-ben az Illinoisi Egyetemen telepítették. Hiller és Isaacson zenészek és informatikusok voltak. A programok akkoriban lyukkártyákat használtak, kicsi volt a memória, és lassú a végrehajtás. Hiller és Isaacson egy tesztforgatókönyvet valósított meg, ahol lehetséges hangmagasságokat generáltak, és csak azokat tartották meg, amelyek megfelelnek a kompozíció kanonikus szabályainak. Magas szintű zenei reprezentációkat vezettek be, például nagyobb dallam- és ritmikai struktúrákat, hogy áthidalják a szakadékot az ember által előállított zenével szemben (Hiller–Isaacson, 1959).

1960-ban Rudolf Zarirov orosz kutató publikálta a világ első tanulmányát az „Ural-1” számítógép segítségével történő algoritmikus zeneszerzésről. Később, 1965-ben Ray Kurzweil bemutatott egy számítógéppel készített zongoradarabot, amely képes volt különböző kompozíciókban mintafelismerésre. A számítógép elemezte ezeket a mintákat, és új dallamok létrehozására használta őket. Ezt követően a számítógépes zene területe általában, és különösen az AI alkalmazása a zenében, hihetetlenül gyorsan fejlődött. A hardver, a szoftver, valamint az AI zeneszerzésre, oktatásra, felvételre és zenei terjesztésre való felhasználásának fejlesztése robbanásszerűen nőtt meg. Például 1997-ben az *Experiments in Musical Intelligence* (EMI) nevű AI-program jobb teljesítményt nyújtott, mint egy emberi zeneszerző abban a feladatban, hogy Johann Sebastian Bach stílusát utánzó zene-

művet komponáljanak (Miranda, 2021; Dannenberg, 2000). Azóta számos programot készítettek, ami például Bach stílusában komponál, és vezérli a lejátszást is (URL2, QR2).



Húsz évvel ezelőtt még csak néhány úttörő volt, akik komolyan vették az AI által komponált zenét. Akkoriban az ANN-ek csak elméleti ígéretnek számítottak. Európában akkor még a PROLOG-programozás határozta meg az AI-kutatás alapját. A DL növekvő népszerűségével a zenei terület is fejlődött. A DL azonban önmagában csak korlátozott kapacitással rendelkezik az emberi intelligencia zenei modellezésére. A zenei rendszerek felépítésére alkalmas technológia fejlesztése mellett az AI kiváló eszköz a zenei intelligencia tanulmányozására is. Az AI-kutatások bebizonyították, hogy ez a típusú intelligencia általában többet igényel, mint az egyszerű logikus érvelés. Ehhez kreativitásra, szubjektivitásra, érzelmekre és interakcióra is szükség van. A zene ezeknek a jellemzőknek a komplexitását foglalja magában. Az AI/DL szerepet játszik a kompozícióban és a hangszintézisben, az előadásban, a hangszerelesben, a kottairásban, a stúdiómunkálatok során, valamint a zenei robotikában is (Dannenberg, 2000).

AZ AI A ZENESZERZÉSSEN

Komplex jelleképezések és digitális hangszerek

Az ML-algoritmusok, főként a regressziós algoritmusok, interface-ként működhetnek, lehetővé téve a művészek számára, hogy összetett leképezéseket építsenek a bemeneti és kimeneti adatok között. Számos ML-eszközkészlet áll rendelkezésre, amelyek lehetővé teszik a végfelhasználók számára az ML-algoritmusokkal való kísérletezést anélkül, hogy széles körű ML-ismeretekre lenne szükségük. Az egyik ilyen ML-eszközkészlet például a *Wekinator*, amely csupán új minták rögzítésével lehetővé teszi a gyakorlati kreativitást (Miranda, 2021; Herremans et al., 2017; Fiebrink–Caramiaux, 2018).

Az ML-t a különféle zenei szcenáriókhoz tervezett digitális hangszerek (Digital Musical Instrument, DMI) tervezésében is alkalmazták. Ezen DMI-k egy részét egy adott kompozíció vagy előadás igényeire fejlesztették ki, míg mások szélesebb körű alkalmazhatóságot biztosítanak azáltal, hogy különféle kompozíciós és improvizációs környezetben használhatóak. A DMI-k a tisztán passzívan reagálótól a részben autonóm, interaktív képességekkel rendelkező eszközökig változnak. Ilyen eszközök például a *Sansa*, a *Phalanger* és a *NOISA*. A *Sansa* egy kiterjesztett akusztikus hangszer. Használható akár klasszikus akusztikus hangszerként, akár hiperhangszerként, a *Wekinator* által biztosított kiterjesztett interakciós modulok segítségével. Különböző üzemmódjai lehetővé teszik az elektronikus hangszerek együttes vezérlését, a partitúrában való

navigálást, az elektronikus hangok kézmozdulatokkal történő formálását, a hangszer vagy más hangforrások, például emberi hang jelének felerősítését és feldolgozását. Ezzel szemben a *Phalanger*nek nincs akusztikus komponense. Ez egy teljesen digitális felület (interface), amellyel a hang kéz- és ujjmozdulatokkal vezérelhető. A *Phalanger* kizárólag a videóadatokra támaszkodik, és egy ANN-t használ a felhasználó kezének észlelésére és a háttértől való elkülönítésére, valamint a különböző kézhelyzetek felismerésére. A rendszer különböző kamerákhoz, fényviszonyokhoz és börtónusokhoz konfigurálható. A *NOISA* (Network of Intelligent Sonic Agents) a példa egy ilyen eszközre, amelynek célja, hogy növelje az előadó kötődését azáltal, hogy autonóm hangválaszokat generál. A hangszer figyeli az előadó kötődési szintjét, és önálló válaszokat képez, amelyek célja, hogy segítsenek az előadónak fenntartani az AI programmal való kapcsolódását. A *NOISA* egy számítógépből, egy, a karra erősíthető érzékelőből, valamint egy mozgáskövető kamerából áll. A hangszer figyeli az előadó mozgulatait és arckifejezéseit. A *NOISA* lényegében egy hangszer és egy interaktív zenei rendszer (Interactive Music System, IMS; lásd később) hibridje. További példaként említhető a Princeton Egyetemen kifejlesztett *ChucK*, egy szövegalapú, többplatformos nyelv, amely valós idejű zeneszerzést, előadást és zeneelemzést tesz lehetővé. A *Morpheus* pedig a londoni Queen Mary Egyetem kutatási projektje. Ez a rendszer már meglévő zenei elemeket optimalizálva egy új zeneművé alakítja. Az *AIVA*-t (Artificial Intelligence Virtual Artist) 2016-ban fejlesztették ki Luxemburgban. Az *AIVA* egy olyan program, amely bármilyen típusú médiához készít hangszávokat. Az *AIVA* mögötti algoritmusok a DL-en alapulnak. Az *I am AI* egy filmzeneszerű darab, amelyet az *AIVA* komponált (URL3, QR3).



A *Riffusion* olyan ANN, amely hangképek felhasználásával generál zenét. A rendszer a hangokat képpé alakítja, amelyek azután inverz Fourier-transzformáción esnek át, és audiofájlokká konvertálódnak. A zene spektrogrammá alakítható a muzsika „vizualizálása” érdekében („látható zene”) (Miranda, 2021; Fiebrink–Caramiaux, 2018; Hörnel–Menzel, 1998).

Az AI-rendszerek széles körben alkalmazzák a zenei harmonizáció során. Az egyik legteljesebb harmonizációs mű a *CHORAL*-rendszer, amely a korálokat Bach stílusában harmonizálja. A *CHORAL* kap egy dallamot, és ún. heurisztikus folyamat révén harmonizálja a kezdetben egyszerű dallamot. A *MUSACT* az ANN segítségével tanulja meg a zenei harmónia modelljét. A *HARMONET*-ben a harmonizációs problémát az ANN-ek alkalmazása és előre meghatározott szabályok betartása (constraint satisfaction) révén közelítik meg. A *MELONET* a *HARMONET* kibővített formája, mely egy ANN-t használ a dallamszekvenciák magasabb szintű szerkezetének megtanulására és reprodukálására. Ha adott egy dallam, a rendszer megszerkeszti bármely korálszólam barokk stílusú harmoni-

zálását és variációját. A *HARMONET* és a *MELONET* együtt erőteljes zeneszerző rendszert alkot, amely olyan variációkat generál, amelyek minősége megközelíti egy képzett orgonaművész játékát (Miranda, 2021; Hörnel–Menzel, 1998).

Az AI-t használó számítógépes kompozíciók egyik legismertebb példája David Cope *EMI*-projektje (Experiments in Music Intelligence), amely különböző zeneszerzők stílusának emulációjára összpontosít. Sikeresen komponált zenét Mozart, Palestrina, Albinoni, Brahms, Debussy, Bach, Rahmanyinov, Chopin, Sztravinszkij és Bartók stílusában. A rendszer repetitív mintákat keres az adott zeneszerző több művében. A felfedezett mintákat szignatúráknak nevezzük. Az *EMI* a zeneszerző egyik művét használja alapul, és rögzíti e szignatúrákat egy új darab komponálása során. Ezt követően a program szabad motívumokat illeszt be a már rögzített szignatúrák közti szabad helyekre. Ennek kapcsán a rendszernek olyan problémákkal kell megküzdenie, mint a szignatúrák kezdeti és befejező részeinek összekapcsolása a környező motívumokkal, elkerülve a stilisztikai anomáliákat, és fenntartva a hangmozgásokat. Az így komponált művek, bár korántsem tökéletesek, összhangban vannak a zeneszerző stílusával (Miranda, 2021; Fiebrink–Caramiaux, 2018).

Interaktív zenei rendszerek

Az interaktív zenei rendszerek (IMS) olyan számítógépes zenei hálózatok, amelyek bemeneti adatok gyűjtésével és értelmezésével képesek érzékelni környezetüket, döntéseket hozni, és cselekedni mind emberi cselekvésekre reagálva, mind autonóm belső folyamatok eredményeként. A kreatív döntések *offline* módban születnek, például amikor a zeneszerző létrehozza a szoftvert és esetleg egy partitúrát, és valós időben, az előadó és a szoftver közötti interakció részeként. Egy hagyományos magnetofon képes befogadni a hangot, de nem tudja elnyomni a háttérzajt. Az AI/ML használható nagy volumenű adatfolyamok értelmezésére és a zeneszerző által meghatározott zenei információk kinyerésére. Doug Van Nort definiálta az orvostudományban használatos genetikailag módosított organizmusok alapján az ún. GSO-kat (Genetically Sonified Organisms). A GSO az interaktív zenemű kiváló példája. Olyan, mint egy „környezeti zeneművészet”, amely mesterségesen létrehozott hangminták halmazából dolgozik, beleértve a vadon élő állatok hangjait (például békák, méhek) (URL4, QR4).



Egy másik érdekes IMS az *Imitation Game*, egy robot-ütőhangszeres példa, amikor a robot egy valódi zenésszel kommunikálva játszik. A bemenet egy ütőhangszeres zenész játéka, melyet egy ANN-be táplálnak. A program képes felismerni a különböző hangszereket (például cintányérok, bongók, kolompok) és technikákat (például dobütés, vonóhasználat). Az ütőhangszeres robot három különböző interaktív lehetőség közül választhat. Imitálhat (a zenésszel együtt

játszhat), kezdeményezhet (új motívumokat adhat hozzá), vagy ismételhet (a zenész játékát improvizációval bővítve megismételheti) (URL5, QR5).



Ez a két példa jól szemlélteti az AI rugalmasságát és felhasználhatóságát. Az eszközök lehetővé teszik a zenészek és digitális partnerük zenei együttműködését. Ennek révén az AI/ML kiváló minőségű zenei információhoz tud jutni. A zeneszerző/programozó szabadon eldöntheti, hogy egy adott kontextusban mely információk relevánsak, és megfelelő példák bevitelével betaníthatja az algoritmust a szükséges zenei információ lekérésére. Mindez nagy fokú kreatív szabadságot tesz lehetővé, ami a gyakorlati példák betáplálásából ered (Miranda, 2021; Rowe, 1993).

Az ember és számítógép együttes kreativitása

Az ember és gép együttműködésének fő célja, hogy elősegítse a kreatív alkotómunkát. Ez például bővítheti az emberi kreativitást stimulálva a zeneművészt, hogy lépjen túl a már kialakult szokásain, és új, kreatív utakat válasszon. Az ML-algoritmus olyan kimenetet generálhat, amelyet a felhasználó egyébként magától nem tudott volna létrehozni. Miközben a komputer-asszisztált zeneszerzés egy ideje már elérhető, az AI-t, konkrétan az ML-algoritmusokat használó ember-számítógép alkotói kollaboráció viszonylag új módszer. A fő különbség a számítógép-asszisztált alkotómunka és az ember-számítógép kollaboráció között az, hogy az utóbbi adaptív képességekkel is rendelkezik (Jones et al., 2012).

Minderre jó példa a *Sonic Xplorer*. A már említett *Wekinator* használatával a *Sonic Xplorer* ANN révén összefüggéseket hoz létre hat különböző tulajdonság és négy érzékelési hangjellemző között. Miután betanítási példákkal ellátta a rendszert, a felhasználó beállíthatja a kialakítani kívánt hang minőségét. A hat tulajdonság a meleg, világos, stabil, erős, zajos és erősödő. A felhasználó átjárhat az *Xplorer*, illetve a szintetizátor interface-e között, hogy a szintézis-paraméterek közvetlen beállításával finomhangolhassa a generált hangokat (Hörnel–Menzel, 1998; Jones et al., 2012).

A hasonló példák a hangszintézis során alkalmazható kreatív kísérletezésre vonatkoznak. Az ember-számítógép kooperációnak azonban nem szükséges a statikus hangok tervezésére korlátozódnia, hanem kiterjeszhető nagyobb volumenű alkotófolyamatok során történő alkalmazásra is (Jones et al., 2012).

A számítógép mint virtuális lejtű

Az élő előadás során alkalmazott ember-gép interakció az AI-eszközök zenei alkalmazásának másik fő területe. Az úgynevezett hangszerparadigmában a számítógép a hangszerhez hasonló szerepet tölt be azáltal, hogy a betáplált műveleteket

hangra fordítja. Másrészt, az előadó-paradigmában a gép virtuális zenészként jelenik meg. Az ilyen interaktív módon szerzett zenemű nem egy előre meghatározott zenei struktúra alapján készül, hanem különböző hangzási lehetőségeknek ad teret. A végső zeneművet a valódi és virtuális zenészek az előadás során alakítják ki. Az ilyen interaktív zeneművekben az AI tovább fejlesztheti a már működő digitális zenei rendszerek észlelési és kooperációs képességeit, valamint „zeneértesét” (Dannenberg, 2000; Rowe, 1993).

Az AI használatának korlátai

A fent leírt lehetőségek mellett az ML-algoritmusoknak vannak hiányosságai is, különösen, ha kreatív alkalmazásokról van szó. Az egyik ilyen gyakori probléma a túlillesztés (overfitting). Ez például olyankor fordul elő, ha az algoritmus a tanulás során a bevitt zenei minta mellett megtanulja az azzal együtt fellépő háttérzajt is. Emiatt a zajmentes hasonló példák felismerése nehézségbe ütközik. Az ilyen túlillesztés olyan rendszerekhez vezet, amelyek jól teljesítenek a betanítás során, de igen gyengén a tesztelés kapcsán (Fiebrink–Caramiaux, 2018; Dreyfus, 1992).

A felügyelt tanulás során optimalizálási problémák adódhatnak. Ezek azon a feltételezésen alapulnak, hogy létezik egy helyes válasz az algoritmus megtanulására. Ez a felügyelt tanulási algoritmusok alkalmazási körét a lehetőségek egy meglehetősen szűk spektrumára korlátozhatja. A valószínűségi alapokon működő ML-algoritmusok, mint például a *WaveNet*, olyan műtermékeket hozhatnak létre, amelyek egyébként nagyon hasonlítanak néhány bemeneti zenei mintára. A gépi utánzás megnehezíti az ilyen rendszerek használatát az ember-számítógép kooperáció során. Ezek az algoritmusok lényegében „fekete dobozok”, ahol a felhasználó zenei mintákat vihet be, de a kimenetelt észlelve semmilyen formában nem tud visszajelzést adni (Fiebrink–Caramiaux, 2018; Dreyfus, 1992).

Külön téma a szerzői jogok kérdése. Amikor az AI-t más, a zenétől eltérő területeken alkalmazzák, azokra is a jelenleg érvényes szerzői jogi törvények érvényesek. Ezzel szemben az AI/ML által generált zene nem részesülhet szerzői jogi védelemben. 2022 februárjában a Copyright Review Board elutasította a szerzői jogvédelem alatt álló, AI által szerzett zenei alkotásokra vonatkozó kérelmet, mivel ezek a megfelelő emberi alkotómunka nélkül születtek, így nem volt elegendő feltétel a szerzői jogi követelés érvényesítéséhez (Miranda, 2021).

Az AI jövője a zeneszerzésben

Míg a zenei AI-eszközök elérhetősége jelentősen megnövekedett az elmúlt néhány évben, az ilyen eszközök felhasználói köre és a kortárs művészeti értékrend közötti látszólagos ellentmondás arra utal, hogy szorosabb együttműködésre van

szükség az AI-fejlesztők és a zeneszerzők között. Az AI-eszközök tervezését is hozzá kell igazítani a jelenlegi művészeti irányelvekhez. Alapvető, hogy az emberi és a számítástechnikai kreativitás kapcsolatát újra kell értékelni. Az emberi intelligenciát segítő (Intelligence Augmentation) *versus* az azt másoló rendszerek (AI) kulcsfontosságú kérdéskört képeznek (Miranda, 2021; Rowe, 1993; Jones et al., 2012).

Amennyiben a zeneművészek számára kívánunk eszközöket fejleszteni, akkor egyértelműen el kell mozdulnunk az automatizálás és a korábban említett „fekete doboz” jelenség felől az olyan interaktív AI-rendszerek irányába, amelyek tanulnak az emberi preferenciákból, és alkalmazkodnak azokhoz. Ez azt jelentené, hogy olyan „human-in-the-loop” tervezési stratégiákat kell alkalmazni, amelyek lehetővé teszik a felhasználók számára, hogy az AI-eszközök felé kommunikálhassák preferenciáikat, és irányítsák az ilyen AI-rendszereket a tanulási folyamat során (Miranda, 2021; Rowe, 1993; Jones et al., 2012).

A hatáskör kérdése a zene és az AI kapcsolatának jövőbeli kutatása szempontjából is fontos lehet. Mivel a „one-fits-all” megközelítés ritkán működik, a zenei AI-eszközök tervezését művészi jellegű tevékenységnek kell tekinteni. A tervezési döntéseknek zeneesztétikai vonatkozásai is vannak. Az AI-fejlesztők és a zeneművészek közötti szorosabb együttműködés biztosíthatja, hogy az AI-eszközök által használt adattípusok és algoritmusok megfeleljenek a művészeti és esztétikai igényeknek, és mindez elősegítse az AI zenei felhasználásának további fejlődését (Rowe, 1993).

ZENEI ELŐADÓMŰVÉSZET ÉS IMPROVIZÁCIÓ

AI és kifejező előadásmód

A digitális zene egyik fő korlátja a kifejezőképesség, azaz a gesztusok hiánya. A zenei előadások kifejezőképességének javítására először egy olyan szakértői rendszert dolgoztak ki, amely Bach-fúgák lejátszásakor meghatározta az alkalmazandó tempót és artikulációt. A bevitt szabályokat két hozzáértő előadóművész bevonásával alkották meg. A kimenet megadja az alaptempót, valamint a lejátszott hangok időtartamára és artikulációjára vonatkozó előadói utasításokat. Ennek a rendszernek a fő korlátja az volt, hogy csak 4/4-es tempójú zeneművek és azon belül is kizárólag Bach-fúgák esetében működött jól. A *Director Music*-rendszer már magában foglalta a tempóra, a dinamikára és az artikulációra vonatkozó, MIDI transzformációs szabályokat. Kidolgoztak egy rendszert annak elemzésére, hogy a zenész kifejezési eszköztára hogyan jelenik meg az előadásban. Az elemzés két különböző kifejezési dimenzióját, a dinamikát és a kinetikát (*rubato*) tárta fel. A szerzők egy olyan programot is kidolgoztak, amely e két, az

előadóra jellemző dimenzió mint beviteli adat alapján megfelelően kifejező előadómódot generál (Miranda, 2021; Bresin, 1998).

A tempó- és dinamika-transzformáció végrehajtására alkalmazott másik megközelítés az ANN-technikák használata. Ennek segítségével szimulálhatjuk például létező zongoristák stílusát. Ezeket az ANN-eket kibővítették a szokásos *input-output* hálózatot a korábban említett, a kimeneti neuronoktól a bemenetiekhez vezető visszaterjesztési (backpropagation) algoritlussal (2. ábra) (Hörnel-Menzel, 1998; Bresin, 1998).

A *SaxEx* olyan számítógépes program, amely korábbi szóló előadások alapján képes szintetizálni a jazzballadák kiváló minőségű, kifejező tenorszaxofon szólóit. A korábbi szabályalapú rendszerek nem tudtak kettőnél több expresszivitási paraméterrel (dinamika és kinetika [rubato]) dolgozni. Ezt a problémát próbálták megoldani a CBR (Creation by Refinement) bevezetésével. A CBR segítségével már kezelni lehetett az öt legfontosabb kifejezési paramétert (dinamika, rubato, vibrato, artikuláció, „attack notes”). A CBR képes a bemeneti hangot úgy átalakítani, hogy annak fenti öt kifejezési tulajdonsága a lehető legjobban megegyezzen az eredeti előadó expresszivitásával. A közelmúltban a CBR más, az expresszivitást tökéletesítő alkalmazásait is kifejlesztették. Az IBL-lel (Instance-Based Learning), azaz a példákra alapuló tanulással a zenei előadások expresszivitása fokozható. Például, az IBL-komponens felismeri egy koncertzongorista előadási jellegzetességeit, és megtanulja, hogy miként alkalmazza ezeket további zenedarabokra (Miranda, 2021; Rowe, 1993; Bresin, 1998).

A számítógépnek a zenei expresszivitásra vonatkozó képessége alapvető volt az ún. hiperhangszerek létrehozásakor. Ezeket a hangszereket arra tervezték, hogy a digitális zenét gazdagítsák az előadóművészek kifejezőképességével és gazdag, élő hangzást adjanak. Vegyünk egy hagyományos hangszert, például egy csellót, és csatlakoztassuk a számítógéphez a nyakban és a vonóban elhelyezett elektronikus szenzorokkal. Rögzítsünk szenzort a vonót tartó kézre is. Alkalmazzunk egy, a *SaxEx*-hez hasonló rendszert, amely, a zenészhez hasonlóan, a kotta, az elsajátított általános zenei ismeretek és a szenzorok által detektált adatok alapján értelmezi az adott darabot. E bemeneti adatok elemzése alapján a hiperhangszer képes a zenei tulajdonságok, ezen belül a hangszín, a ritmus és a zenei megfogalmazás megváltoztatására, valamint intelligens zenei kíséret hozzáadására. Például Tod Machover, a Massachusetts Institute of Technology (MIT) Media Lab munkatársa kifejlesztett egy hipercellót. A világhírű csellista, Yo-Yo Ma néhány éve a Tanglewood Fesztiválon mutatta be ezt a hiperhangszert és egy Tod Machover által komponált darabot (URL6, QR6) (Bresin, 1998; Machover, 1992).



AI a zenei improvizációban

A zenei improvizáció nagyon összetett alkotói folyamat, amelyet szintén sikerült modellezni. A zenei improvizáció természetesen sokkal nagyobb kreativitást igényel, mint a zeneszerzés, ahol a szerzőknek van idejük átdolgozni kompozíciójukat (Rowe, 1993; Papadopoulos–Wiggins, 1998).

A számítógépes improvizáció korai példája a *Flavours Band*-rendszer. Ez lényegében egy eljárásrend a jazz és popzenei stílusok meghatározásához. Lehetővé teszi kötött zenei struktúrák (akkordok, hangszínek) kombinálását és feloldását, amivel improvizációs variációk generálhatók. A *GenJam* elkészíti egy adott jazz-zenész modelljét, és a genetikai algoritmusokhoz hasonló módon megtanítja improvizálni (URL7, QR7).



Egy másik rendszer hasonló algoritmust használ, mely egy adott akkord alapján képes jazzdallamok improvizatív alkotására. A *GenJammel* ellentétben az utóbbi program tartalmaz egy finomhangoló funkciót, amely automatikusan értékeli és értékeli az eredmény minőségét. Ennek során az improvizált dallam számos különböző jellemzőjét értékeli, beleértve a dallamot, a hangok időtartamát, a hangok közötti intervallumok távolságát (Papadopoulos–Wiggins, 1998).

A fenti rendszereket azért illették kritikával, mert nem a zeneszerzővel együttműködve működtek. Ezzel szemben a *Band-out-of-a-Box* (BoB) rendszer a BoB és a zenész közötti valós idejű interaktív improvizáció lehetőségét teremtette meg. A BoB lényegében „zenésztársként” szolgál a valós idejű improvizációhoz. A BoB-ot két különböző stílusú világhírű szólózenész, Charlie Parker szaxofonos és Stephane Grapelli hegedűművész tesztelte, és pozitívan értékelt. Végül, egy újabb figyelemre méltó interaktív zenei rendszer a *Continuator*. Ez a kibővített többretegű Markov-modellen alapul. Mindez lehetővé teszi, hogy megtanuljon interaktívan játszani a felhasználóval annak stílusában. Lehetővé teszi a folyamatos párbeszédet zeneművész és a rendszer között, és ezáltal a közös improvizációt is (URL8, QR8) (Miranda, 2021; Bresin, 1998; Papadopoulos–Wiggins, 1998).



ÖSSZEFOGLALÁS

A mesterséges intelligencia lassan átszövi mindennapjainkat, így a zenét is. Az AI a gépi tanuláson keresztül a zeneszerzőtől vagy előadóművésztől függetlenül, vagy vele együttműködve segítheti, fejlesztheti a zeneszerzést, illetve az előadást is. Jelenleg még az út elején tartunk, de máris több példát tudunk bemutatni az AI és az ember kollaborációjára a zene kapcsán.

IRODALOM

- Bresin, R. (1998): Artificial Neural Networks Based Models for Automatic Performance of Musical Scores. *Journal of New Music Research*, 27, 3, 239–270. DOI: 10.1080/09298219808570748, https://www.researchgate.net/publication/257810708_Artificial_neural_networks_based_models_for_automatic_performance_of_musical_scores
- Dannenberg, R. (2000): Artificial Intelligence, Machine Learning, and Music Understanding. In: *Proceedings of the Brazilian Symposium on Computer Music (SBCM2000), Curitiba, Brazil, 2000*, 1–10.
- Dreyfus, H. (1992): *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press
- Fiebrink, R. – Caramiaux, B. (2018): The Machine Learning Algorithm as Creative Musical Tool. In: Dean, R. T. – McLean, A. (eds.): *The Oxford Handbook of Algorithmic Music*. 181–208. DOI: 10.48550/arXiv.1611.00379, <https://arxiv.org/pdf/1611.00379.pdf>
- Herremans, D. – Chuan, C. H. – Chew, C. (2017): A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys*, 50, 1–30. DOI: 10.1145/3108242, <https://arxiv.org/pdf/1812.04186.pdf>
- Hiller, L. – Isaacson, L. (1959): *Experimental Music: Composition with an Electronic Computer*. New York: McGraw-Hill, <https://archive.org/details/experimentalmusi00hill>
- Hörnle, D. – Menzel, W. (1998): Learning Musical Structure and Style with Neural Networks. *Computer Music Journal*, 22, 4, 44–62. DOI: 10.2307/3680893
- Jones, D. – Brown, A. R. – d'Inverno, M. (2012): The Extended Composer. In: Jon McCormack, J. – d'Inverno, M. (eds.): *Computers and Creativity*. Berlin–Heidelberg: Springer, 175–203. DOI: 10.1007/978-3-642-31727-9_7
- Machover, T. (1992): *Hyperinstruments: A Progress Report 1987 – 1991. MIT Media Lab Internal Research Report*. MIT Media Laboratory, https://opera.media.mit.edu/publications/machover_hyperinstruments_progress_report.pdf
- Meskó B. – Görög M. (2020): Rövid útmutató egészségügyi szakemberek számára a mesterséges intelligencia korában. *Magyar Tudomány*, 181, 10, 1361–1377. DOI: 10.1556/2065.181.2020.10.8, https://mersz.hu/mod/object.php?objazonosito=matud202010_f46917_i1
- Miranda, E. R. (2021): *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer Nature AG.
- Papadopoulos, G. – Wiggins, G. (1998): A Genetic Algorithm for the Generation of Jazz Melodies. In: *Proceedings of the Finnish Conference on Artificial Intelligence (STeP'98)*, Jyväskylä, Finland, <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/music/web/files/Genetic%20Algorithm%20for%20Generation%20of%20Jazz%20Melodies.pdf>
- Rowe, R. (1993): *Interactive Music Systems. Machine Listening and Composing*. Cambridge, MA: MIT Press, https://wp.nyu.edu/robert_rowe/text/interactive-music-systems-1993/
- Schmidhuber, J. (2015): Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85–117. DOI: 10.1016/j.neunet.2014.09.003
- Szekanecz Z. – Szekanecz M. – Meskó B. (2023): A mesterséges intelligencia alkalmazása a reumatológiában és más autoimmun kórképekben. *Immunológiai Szemle*, 15, 1, 4–23. <https://www.medicina-kiado.hu/uploads/images/hirek/2023/immunologia-szemle-mesterseges-intelligencia.pdf>

URL1: <https://www.youtube.com/watch?v=fojKZ1ymZlo>

URL2: <https://www.youtube.com/watch?v=Y48xGobW90o>

URL3: <https://www.youtube.com/watch?v=Emidxpkyk6o>

URL4: https://www.youtube.com/watch?v=Ho_IIXBOVYo

URL5: <https://www.youtube.com/watch?v=Mv8Y0K-fYNw>

URL6: <https://www.youtube.com/watch?v=K0TVbFDbl9Y>

URL7: <https://www.youtube.com/watch?v=ACT8Sy6tNZI>

URL8: <https://www.youtube.com/watch?v=cHKcq0D5EY4>