



Budapesti Műszaki és Gazdaságtudományi Egyetem  
Villamosmérnöki és Informatikai Kar  
Automatizálási és Alkalmazott Informatikai Tanszék

# CHC modellen alapuló részképesség becslés mesterséges intelligenciával

**TDK dolgozat**

Készítette:

Füstös Gergely

Konzulens:

Dr. Forstner Bertalan

2023

# Tartalomjegyzék

<b>Kivonat</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1. Bevezetés</b>	<b>1</b>
1.1. Az emberi elme és kognitív képességek . . . . .	2
1.2. Pszichometriai modellek . . . . .	2
1.3. Mesterséges intelligencia szerepe a kutatási területen . . . . .	2
1.4. A dolgozat relevanciája és célja . . . . .	3
<b>2. Emberi kognitív képességek és pszichometria</b>	<b>4</b>
2.1. Kognitív képességek: Definíció . . . . .	4
2.2. Alapvető pszichometriai fogalmak . . . . .	4
2.3. A hagyományos és korszerű IQ-tesztek bemutatása . . . . .	5
2.4. A Cattell-Horn-Carroll modell . . . . .	7
2.4.1. A CHC modell bemutatása . . . . .	7
2.4.2. A CHC modell számszerűsítése . . . . .	8
2.5. Az Item Response Theory . . . . .	8
2.5.1. Az IRT modell bemutatása . . . . .	8
2.5.2. A tételválasz függvény . . . . .	9
2.5.3. Az IRT modell előnyei . . . . .	9
2.6. A játékos fejlesztés jelentősége és módszertana . . . . .	10
<b>3. Neurális hálózat és beágyazás</b>	<b>11</b>
3.1. Neurális hálózatok rövid bemutatása . . . . .	11
3.1.1. Biológiai motiváció: a neuron . . . . .	11
3.1.2. A perceptron . . . . .	12
3.1.3. Többrétegű neurális hálózat . . . . .	13
3.2. Beágyazás . . . . .	14
3.2.1. Beágyazás alkalmazása . . . . .	15
<b>4. Kutatási hipotézis és módszertan</b>	<b>17</b>
4.1. A kutatás célja és felállított hipotézisek . . . . .	17
4.1.1. Feladatok részképesség-függésének vizsgálata . . . . .	18
4.1.2. Alanyok részképességének vizsgálata . . . . .	18
4.1.3. Eredménybecslés ismert részképességek esetén . . . . .	19
4.2. A célok megvalósításának elméleti háttere . . . . .	20
4.2.1. Faktoranalízis pontozott feladatokra . . . . .	20
4.2.2. Faktoranalízis bináris értékelésű feladatokra . . . . .	21
4.2.3. Eredménybecslés regresszióval és neurális hálóval . . . . .	22
4.2.4. Beágyazás alkalmazása részképességek felbontására . . . . .	23

4.3.	Módszertani kihívások és megoldások . . . . .	25
4.3.1.	Rendelkezésre álló mérések a validációhoz . . . . .	25
4.3.2.	Szimuláció az adatok gazdagításához . . . . .	25
4.3.3.	Az elemzések technológiai megvalósítása . . . . .	26
<b>5.</b>	<b>Új modellek kialakítása és kiértékelése</b>	<b>27</b>
5.1.	Szimuláció elkészítése és kiértékelése . . . . .	27
5.1.1.	Szimuláció paramétereinek meghatározása . . . . .	27
5.1.2.	A szimuláció lefuttatása . . . . .	28
5.2.	Az új megközelítés tesztelése és validálása . . . . .	31
5.2.1.	A tesztfeladatok bemutatása . . . . .	31
5.2.2.	Faktoranalízis regresszió segítségével . . . . .	32
5.2.3.	Eredménybecslés regresszóval és neurális hálóval . . . . .	34
5.2.4.	Részképességekre bontás beágyazással . . . . .	37
<b>6.</b>	<b>Összefoglalás</b>	<b>40</b>
	<b>Köszönetnyilvánítás</b>	<b>42</b>
	<b>Irodalomjegyzék</b>	<b>43</b>

# Kivonat

A TDK dolgozatom központi témája a CHC modellen alapuló részképességek becslése különböző gépi tanulási modellek alkalmazásával, különös hangsúlyt fektetve a mesterséges intelligenciára. A CHC modell (Cattell-Horn-Carroll modell) egy pszichometriai modell, amely az emberi kognitív képességeket próbálja lebontani és rendszerezni. A CHC modell részképességei nem számszerűsítik magukat a hagyományos értelemben, mint például egy százalékos skála vagy egy abszolút érték. Ehelyett a modell inkább azt hangsúlyozza, hogy a kognitív képességek között vannak különbségek, és ezeket a különbségeket leírják és rangsorolják.

A célom bemutatni, hogy különböző tesztfeladatok megoldásakor milyen részképességek játszanak jelentős szerepet, és hogyan lehet megbecsülni egy személy teljesítményét egy adott teszten, ha ismertek a személy képességei, és a teszt paraméterei.

A dolgozatom során az egyes részképességek szerepét vizsgálom a különböző feladatokban a CHC modell alapján. A CHC modell segítségével megpróbálom eldönteni, hogy egy adott feladat mely részképességeket használja, és azok szintjétől mennyire függ. Ehhez általában kognitív tesztek és elemzéseket használnak, hogy azonosítsák, mely faktorok dominálnak egy adott feladat során. Például egy matematikai feladatnál a számolási képességek faktorai lehetnek dominánsak, míg egy nyelvi feladatnál a verbális faktorok lehetnek fontosak. A jelenleg használt eljárások sok elvárást támasztanak mind a tesztek, mind az alanyok kapcsán, amelyek gyakran nem teljesíthetőek, torzítva ezzel az eredményeket.

A kutatócsoportban párhuzamosan folyik nagymintás adatfelvétel általános iskolás gyerekekkel, amelyben különböző részképességű alanyok eltérő típusú feladatokat oldottak meg, egy-egy részképesség elkülönített meghatározásához. Ezek az adatok tipikus profilokat biztosítanak a kutatásomhoz, hogy rájuk alapozva szimulációval, kibővített adatokon végzett vizsgálódásom eredményeit ellenőrizhetem és validálhatom, elősegítve ezzel a módszer megbízhatóságát és alkalmazhatóságát más adathalmazok esetén.

Kutatásom eredményeinek fő felhasználói a fejlesztőpedagógusok, pedagógiai-pszichológia szakemberek, akik számára a módszer lehetőséget biztosít az általuk széles körben használt mérések optimalizálására, valamint arra, hogy a fejlesztési feladatok közül azokat, amelyek megoldásának a részképesség-függése nem ismert, megfelelően beállított nehézséggel paraméterezve, egyénre szabottan lehessen alkalmazni, és ezáltal a fejlesztés hatékonyságát növelni.

# Abstract

The central theme of my TDK thesis is estimating special cognitive abilities based on the CHC model using various machine learning models, with a specific focus on artificial intelligence. The CHC model (Cattell-Horn-Carroll model) is a psychometric model aiming to deconstruct and organize human cognitive abilities. The abilities in the CHC model are not quantified in the traditional sense, such as a percentage scale or an absolute value. Instead, the model emphasizes that there are differences between cognitive abilities and it describes and ranks these differences.

My aim is to demonstrate the significant role special cognitive abilities play in solving different tasks and how one can estimate an individual's performance on a specific test given their abilities and the parameters of the test.

Throughout my thesis, I investigate the role of various cognitive abilities in different tasks based on the CHC model. Using the CHC model, I try to determine which cognitive abilities a specific task employs and how much they depend on their levels. Generally, cognitive tests and analyses are used to identify which factors dominate during a particular task. For example, in a mathematical task, computational skills might be dominant, while in a linguistic task, verbal factors might be important.

In the research group, simultaneous large-scale data collection is underway with primary school children, where students with different cognitive abilities solved various types of tasks to help isolate specific cognitive abilities. These data provide typical profiles for my research, allowing me to verify and validate the results of my simulations and examinations on expanded data. This contributes to enhancing the reliability and applicability of the method for other datasets.

The results of my research facilitate the application of complex developmental tasks widely used by educational developers. Specifically, with the results, those tasks whose solution's dependency on specific cognitive abilities is unknown can also be applied with appropriately adjusted difficulty parameters, thereby increasing the effectiveness of development.

# 1. fejezet

## Bevezetés

Az emberi elme működésének megértése már régóta az emberiség érdeklődésének középpontjában áll. Ezen érdeklődés kiemelt területe a kognitív képességek kutatása és mérése, más néven a pszichometria. A modern technológia és az informatika fejlődésével, illetve a mesterséges intelligencia elterjedésével - számos más kutatási területhez hasonlóan - az intelligencia mérésében és az eredmények kiértékelésében is újabb perspektívák nyíltak meg. Emellett a modern pszichológiai modellek lehetővé teszik számunkra, hogy mélyebb és átfogóbb képet kapjunk az intelligencia fogalmáról, amely alapot ad az összetett mérések kidolgozásához és értelmezéséhez.

A kognitív tudományok és a pszichometria világában az emberi elme működésének és képességeinek pontos számszerűsítése és értékelése folyamatosan változó kihívást jelent. A Cattell-Horn-Carroll (CHC) modell ezen kognitív képességek strukturális bemutatására törekszik, ám a modell által leírt részképességek számszerűsítése nem triviális. Sok fejlesztőpedagógus szembesül a problémával, hogy hogyan ismerhetnék fel és értékelhetnék ezeket a részképességeket, különösen olyan esetekben, amikor a feladatok részképesség-függése nem ismert. Vegyük példának a *Simon mondja* elektronikus memóriajátékot, amelyben négy különböző színű és hangú gomb található, és a játékosoknak a gép által véletlenszerű sorrendben lejátszott szín- és hangmintákat kell pontosan megismételniük, miközben a minta hossza minden körben növekszik. Erről a játékról nem ismert pontosan, hogy milyen részképességekre milyen arányban épít, hiszen mind a vizuális és auditív képességek, mind a rövidtávú memória fontos szerepet játszhat egy feladvány megoldásakor.



1.1. ábra. A Simon mondja játék<sup>1</sup>

A dolgozatom célja, hogy új módszereket és megközelítéseket mutasson be a CHC modell alapján azonosított részképességek pontos becslésére, különös tekintettel a gépi tanulási modellek és a mesterséges intelligencia alkalmazására. A kutatásom középpontjában egyrészt a tesztfeladatok nehézségének becslése áll, másrészt pedig az egyes részképességek szerepének vizsgálata különböző feladatokban. A bemutatott módszertanok célja, hogy elő-

<sup>1</sup><https://www.amazon.com/Electronic-Memory-Handheld-Classic-Gameplay/dp/B07YNKKGRJ>

segítsék a pedagógusok és fejlesztőpedagógusok munkáját a fejlesztő feladatok sikeresebb, egyénre szabott alkalmazásával, ezzel is növelve a fejlesztési folyamatok hatékonyságát.

## **1.1. Az emberi elme és kognitív képességek**

A kognitív képességek az emberi élet minden területén jelen vannak, az egyszerű mindennapi tevékenységektől kezdve, egészen az összetett feladatok megoldásáig. Ezek a mentális képességek, amelyek lehetővé teszik az információk szerezését, feldolgozását és alkalmazását, alapvetőek az érzékelés, figyelem, emlékezet és problémamegoldás terén. Az egyén életében betöltött kulcsfontosságuk miatt a kognitív képességek helyes megértése és értelmezése nélkülözhetetlen az oktatásban és a fejlesztésben dolgozó szakemberek számára. Ezen képességek mélyebb megismerését a második fejezetben tárgyalom, különös hangsúlyt fektetve a játékos fejlesztési módszertanra és annak szerepére a kognitív képességek mérésében és fejlesztésében.

## **1.2. Pszichometriai modellek**

A pszichometria a pszichológia azon ága, mely az emberi képességek mérésére összpontosít. Az elmúlt években kifejlesztett modellek és tesztek lehetővé teszik az egyének kognitív képességeinek mérhető és objektív értékelését. A pszichometriai modellek alapvetőek az eredmények értelmezésében és a fejlesztési stratégiák meghatározásában. A harmadik fejezet részletesen tárgyalja a pszichometria alapjait, a különböző modelleket, hagyományos és modern megközelítésekkel, valamint a modellek megbízhatóságát és érvényességét.

## **1.3. Mesterséges intelligencia szerepe a kutatási területen**

A mesterséges intelligencia (MI) kifejezés sokak számára mára már közhírtté vált, és már az élet számos területén meghatározó fogalom lett. A technológia rohamos fejlődésével a mesterséges intelligencia alkalmazásának területe is egyre bővül. Az MI nem csak az informatikai iparágakban játszik központi szerepet, de az egészségügytől kezdve az oktatáson át az üzleti életig számos területen átalakítja mindennapi életünket.

A mesterséges intelligencia alkalmazása a kognitív képességek és a pszichometria tudományterületén különösen ígéretes. Az MI képes feldolgozni és értelmezni nagy mennyiségű adatot, sokkal gyorsabban, mint az emberek. Ezen kívül az öntanuló algoritmusok képesek az adatok alapján azonosítani olyan összefüggéseket, amelyeket az adatok karakterisztikája, jellege vagy a rejtett képességek ismeretének hiánya miatt a jelenlegi eljárásokkal nem lehetséges.

A kognitív képességek és a pszichometria területén végzett kutatásokban az MI-nek köszönhetően könnyebben felismerhetjük és csoportosíthatjuk a különböző kognitív képességeket, az egyéni különbségeket, és azt, hogy ezek hogyan befolyásolják az egyén teljesítményét különféle környezetekben. Emellett a modern pszichológiai megközelítések az intelligenciát sokféle összetevőjű tényezőként ábrázolják, amelyek felismeréséhez és elkülönítéséhez a mesterséges intelligencia alkalmazása jelentős segítséget nyújt.

A mesterséges intelligencia innovatív megoldásai tehát új módszertanokat és eszközöket is bevezetnek a kognitív képességek kutatási területén. Olyan technikák, mint a mélytanulás vagy neurális hálózatok, lehetővé teszik, hogy sokkal összetettebb, több paraméteres kognitív tesztek is elemezni tudjunk, amelyek korábban megközelíthetetlennek tűntek. A dolgozat során bemutatom, hogy miként jelent a mesterséges intelligencia nagy előrelépést, illetve hogyan alkalmazható a kognitív képességek és pszichometria kutatási területein.

## 1.4. A dolgozat relevanciája és célja

A 21. századra az oktatás és a pedagógia területe jelentős átalakuláson ment keresztül, melyek központjában a tanulók egyéni szükségletei állnak[11]. Ahogyan a társadalom és a technológia fejlődik, úgy nő a szükség arra, hogy az oktatási módszerek is alkalmazkodjanak az új kihívásokhoz. Ebben az összetett rendszerben a kognitív képességek kutatása és mérése központi szerepet játszik, hiszen a megfelelő ismeretek nélkülözhetetlenek a hatékony oktatási stratégiák kialakításához.

A kognitív képességek mérése azonban nem csak az oktatás területén, hanem a mindennapi életben is kiemelkedő jelentőséggel bír. A tanulási nehézségek vagy a kognitív funkciókban beálló változások - mint például a demencia -, időben történő felismerése és az ezekre adott hatékony válasz mind a tanuló, mind a társadalom számára előnyöket hozhat. A különböző pszichometriai modellek és a mesterséges intelligencia kombinált alkalmazása új távlatokat nyithat ezen a területen. Az új megközelítéssel lehetővé válik a kognitív fejlesztés folyamatának optimalizálása olyan esetben is, amikor az alany tanulási nehézséggel, demenciával, vagy egyéb, a teljesítményét temporálisan zavaró nehézséggel küzd.

A tudományos kutatások hosszú ideig a kognitív képességeket egytényezős faktorként értékelték. A modern pszichológia azonban az intelligenciát több részképesség összességként értelmezi, ami új megközelítéseket hozott a kognitív tesztek tervezésében és kiértékelésében. A CHC modell jelenleg az egyik vezető kvalitatív eszköz az intelligencia részképességeinek elkülönítésére, azonban a részképességek kvantitatív elemzésére alkalmazott módszerek gyakran megkérdőjelezhetőek az előfeltevések pontatlansága miatt.

A mesterséges intelligencia és a gépi tanulás azonban lehetővé teszi, hogy ilyen összetett pszichológiai tényezőket is számszerűen tudjunk megvizsgálni, amelyre az egyszerűbb statisztikai modellek még nem voltak képesek.

A kutatás során célom, hogy egy olyan újító megközelítést alakítsak ki, mely nem csak elméleti síkon járul hozzá a kognitív képességek kutatásához, hanem gyakorlati alkalmazása is lehetséges lesz. Kutatásom fókuszában azok a fejlesztőpedagógiai eszközök állnak, melyek részképességi háttere eddig elsősorban tapasztalati úton került felbontásra. Célom, hogy ezen eszközöket nem csak szubjektív megfigyelések alapján, hanem kvantitatív adatokkal alátámasztva, egyéni igényekhez igazodóan alkalmazzam. A kutatásom eredményei tehát nem csak a tudományos közösség számára hordoznak értéket, hanem közvetlenül hozzájárulnak a fejlesztőpedagógusok munkájának hatékonyságához is.



## 2. fejezet

# Emberi kognitív képességek és pszichometria

A történelem során az emberi elme működése és azok a folyamatok, melyek révén információt feldolgozunk, mindig is központi helyet foglaltak el a tudományos kutatásokban. Az emberek különböző mentális kapacitásokkal rendelkeznek, amelyek befolyásolják azokat a módszereket, melyekkel érzékeljük a környezetünket, reagálunk a kihívásokra és megtaláljuk az utunkat a mindennapokban. A különféle kognitív képességek mérésének és értelmezésének kérdése pedig a pszichometria tudományába tartozik. A következő fejezetekben betekintést adok a kognitív képességek és a pszichometria világába, különös tekintettel azokra a részekre, melyek fontos szerepet játszanak a jelen kutatás megértéséhez.

### 2.1. Kognitív képességek: Definíció

A *kognitív képesség* egy gyakran használt fogalom a tudományos életben, azonban érdekes módon nincs egyetlen, mindenki által elfogadott precíz definíciója. A legtöbb tudományos cikk, amely alapjaiban vizsgálja a kognitív képességeket, különböző összetett módokon, hosszú oldalakon át definiálja ezt az egyszerűnek hangzó fogalmat [4, 5, 2]. Azonban a jelen dolgozat célja inkább az, hogy megismertesse az olvasóval a kognitív képességek fogalmát, mint hogy precízen definiálja azt. Ez a fejezet tehát inkább tekinthető a kognitív képesség bemutatásának, mint definíciójának.

A kognitív képességek azok a mentális funkciók és folyamatok, melyek lehetővé teszik számunkra, hogy információt szerezzünk, feldolgozzunk és alkalmazzunk. Ezek a képességek felelősek az érzékelés, a figyelem, az emlékezet, a nyelv és a problémamegoldás folyamatáért, és alapvetőek az élet minden területén való boldoguláshoz, beleértve a tanulást, a munkát, a kommunikációt és a mindennapi döntéshozatalt. Minden ember egyedi módon rendelkezik ezekkel a képességekkel, amelyek az élet során fejlődnek, és sok tényezőtől függhetnek, beleértve a genetikát, az oktatást és a környezeti hatásokat.

Precízebben, a kognitív képességek általában azt a *kapacitást* jelentik, hogy hogyan tudunk gondolkodni, tanulni, emlékezni, problémákat megoldani és kommunikálni. Ez magában foglalja az információ feldolgozásának sebességét, a logikai gondolkodást, a kritikus értékelést és az analitikus készségeket.

### 2.2. Alapvető pszichometriai fogalmak

A pszichometria a pszichológia tudományának egyik alapvető ága, mely az emberi elme különböző képességeinek mérésére összpontosít. Az elmúlt évtizedekben számos modell és teszt került kidolgozásra, amelyek lehetővé teszik számunkra, hogy mérhető, objektív

módon közelítsünk az egyének kognitív képességeihez. Ezek a modellek alapvetőek ahhoz, hogy az emberi jellemzőket kvantitatív módon jellemezzük, összehasonlítsuk és értelmezzük. Az emberek közötti különbségek bemérése nemcsak a tudományos kutatás, hanem az oktatás, a munkaerő-piac és más társadalmi területek számára is fontos. Többek között ezek a különböző mérések segítséget nyújtanak abban, hogy megfelelő oktatási módszereket alkalmazzanak a tanulók fejlődésének támogatására.

A pszichometria egy olyan terület, ahol a mérés precizitása és megbízhatósága kulcsfontosságú. Ahhoz, hogy megértsük, hogyan működnek a pszichometriai eszközök, és hogyan alkalmazzák őket, ismernünk kell néhány alapvető fogalmat.

## Mérés

A *mérés* a pszichometria alapja. Ez a folyamat általában az egyéni különbségek kvantitatív értékelését jelenti valamely pszichológiai jellemző, mint például a kognitív képességek terén[22]. A mérések általában az úgynevezett tesztfeladatok segítségével történnek, amelyek kérdéseket vagy állításokat tartalmaznak, és a válaszokat pontokkal vagy a feladat sikeres, illetve sikertelen megoldásaként értékelik.

## Validitás

A *validitás* egy mérési eszköz érvényességét jelenti, vagyis azt, hogy valóban azt méri-e, amit mérnie kell. Egy teszt akkor tekinthető érvényesnek, ha valóban képes felmérni azokat a jellemzőket vagy képességeket, amelyeket mérni kíván. A validitást többféle módon lehet értékelni, beleértve a tartalmi validitást - azaz, hogy a teszt tartalma releváns és képviseli a mérni kívánt jellemzőt -, a konstruktum validitást - azaz, hogy a teszt ténylegesen a kívánt pszichológiai konstrukciót méri -, és a prediktív validitást - azaz, hogy a teszt képes előre jelezni a későbbi viselkedést vagy teljesítményt[14].

## Megbízhatóság

A *megbízhatóság* vagy *reliabilitás* azt méri, hogy egy teszt eredményei mennyire konzisztensek és stabilak az idő múlásával. Egy megbízható teszt esetén a válaszoknak konzisztenseknek kell lenniük, és az eredményeknek reprodukálhatóaknak kell lenniük különböző időpontokban vagy különböző mintákon. A megbízhatóságot általában a teszten belüli konzisztenciával és a teszt ismételt alkalmazásával értékelik[3].

## Tesztfeladatok és feladatsorok kialakítása

Az *tesztfeladatok* a teszt alapelemei, amelyek kérdésekből vagy állításokból állnak. Az tesztfeladatok kialakítása során fontos figyelembe venni a pontosságot, az érthetőséget és a relevanciát. A feladatsorok olyan csoportokba rendezett tesztfeladatokból állnak, amelyek egy adott pszichológiai jellemzőt vagy konstrukciót mérnek. A feladatsorok kialakítása során az egyes tesztfeladatokat úgy kell összeválogatni és pontozni, hogy azok valóban képviseljék a mérni kívánt jellemzőt, és a feladatsor összességében érvényes és megbízható mérést biztosítson. A feladatsor értékelésének egy módja a tesztfeladatok pontszámának összegzése. A későbbiekben a tesztfeladatokra gyakran egyszerűen feladatként illetve kérdésként hivatkozom.

## 2.3. A hagyományos és korszerű IQ-tesztek bemutatása

Az intelligencia mérésének története hosszú időre nyúlik vissza, és az évek során számos különböző tesztet fejlesztettek ki. A hagyományos IQ-tesztek, mint például a Stanford-

Binet Intelligencia Skála és a Wechsler Intelligencia Skálák, már évtizedek óta léteznek, és széles körben használják őket a világon. Azonban a modern kor újabb kihívásokat hozott, amelyre válaszolva korszerűbb tesztek is megjelentek.

### Hagyományos IQ tesztek

A Stanford-Binet Intelligencia Skála egyike a legrégebbi és leginkább elismert IQ-teszteknek. Ez a teszt a 20. század elején jött létre, és azóta is folyamatosan fejlesztik. A Stanford-Binet teszt a verbális és nem verbális intelligencia számos aspektusát méri, beleértve a szóbeli megértést, a kvantitatív gondolkodást, és a vizuális-szóbeli kapcsolatokat[21].

A Wechsler Intelligencia Skálák szintén jelentős befolyást gyakoroltak az intelligencia mérésére. David Wechsler hozta létre, és a teszt a verbális és nem verbális készségekre egyaránt fókuszál. A Wechsler tesztek egyik kulcsfontosságú jellemzője a verbális és a teljesítmény alskálák használata, amelyek segítenek jobban megérteni az egyén erősségeit és gyengeségeit[25].

### Korszerű IQ tesztek

A korszerű IQ-tesztek, mint például a Raven Progresszív Mátrixok, a kreativitásra és a problémamegoldásra helyezik a hangsúlyt. Ezek a tesztek gyakran használnak absztrakt, nem verbális feladatokat, hogy megkerüljék a kulturális és nyelvi előítéleteket. A Raven teszt például mátrixok sorozatából áll, ahol a tesztalanyoknak fel kell ismerniük a mintákat és előre jelezniük a következő elemet a sorozatban[19].

Ezen kívül a korszerű tesztek gyakran integrálják a számítógépes technológiát, lehetővé téve gyorsabb és pontosabb értékelést, valamint az adaptív tesztelés lehetőségét, ahol a teszt nehézsége a tesztalany válaszaihoz igazodik. Az ilyen típusú tesztek jobban meg tudják ragadni az egyedi képességeket, és sokszor jobban igazodnak a mai világ kihívásaihoz.

Azonban mind a hagyományos IQ tesztek, mind sok korszerű IQ teszt egyik problémája, hogy csak összetett képességeket mérnek. Ismert, hogy az intelligencia nem egyetlen faktorból áll, és ezeknek a faktoroknak a különálló vizsgálatára már van néhány teszt - többek között a Woodcock-Johnson teszt és a magyar fejlesztésű Kognitív Profil Teszt<sup>1</sup> -, azonban a faktorok kombinált tesztelésére még nincs bevált módszer[15].

A Woodcock-Johnson teszt egy komplex IQ teszt, amely alkalmas különböző kognitív képességek mérésére. A teszt két részből áll: egy standard és egy bővített tesztsorból, amelyek együttesen mélyreható képet adnak a vizsgált személy kognitív képességeiről. A standard tesztsor 10 különböző tesztet tartalmaz, míg a bővített tesztsor további 10 tesztet kínál, lehetővé téve a részletes elemzést. A teszt 2-től 90 éves korig ajánlott, így széles korcsoport számára elérhető és használható. Az egyes tesztek a kognitív képességek különböző aspektusait mérik, ami lehetővé teszi a részletes és pontos értékelést.[20]

A Kognitív Profil Teszt egy magyar fejlesztésű tesztkészlet és weboldal, amely dedikáltan részképességeket mér és ezekből állítja össze egy személy kognitív profilját. A kutatásomhoz használt méréseket is a Kognitív Profil Teszt keretrendszerében végezték.

Összefoglalva tehát mind a hagyományos, mind a korszerű IQ-tesztek fontos szerepet játszanak az intelligencia mérésében. A modern tesztek között vannak, amelyek a speciális kompetenciákat, azaz *részképességeket* is vizsgálják, bár ezeknek a mérése még kezdeti szakaszban van. A részképességek pontos meghatározása és azok mérésének módszerei a következő fejezetben kerülnek bemutatásra.

---

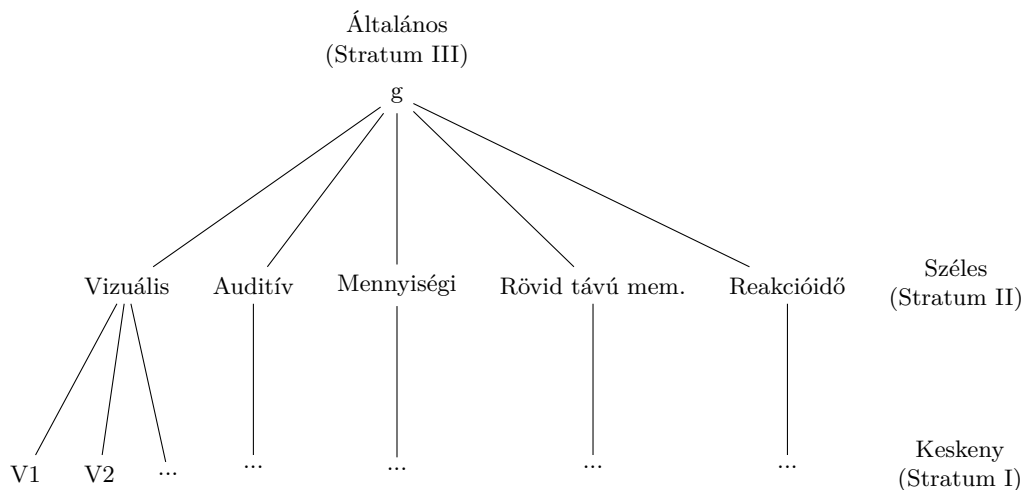
<sup>1</sup><https://kognitivprofil.hu>

## 2.4. A Cattell-Horn-Carroll modell

A kognitív képességek viselkedése nagyon összetett, és sok kutatás során részletesebb elemzésekre lenne szükség, mint amelyeket a mai intelligencia tesztek lehetővé tesznek. A kognitív képesség tovább bontható specifikus részképességre, amelyek egy-egy kognitív területet vagy képességet jelölnek, mint például a logikai gondolkodás, vagy a memória. Ezek a képességek az egyének különböző kognitív feladatainak és kihívásainak kezelésében kulcsszerepet játszanak. Az emberek eltérő szintű részképességekkel rendelkeznek, és ezek a részképességek együttesen alkotják az egyén kognitív profilját. A részképességek megértése lehetővé teszi, hogy jobban átlássuk az egyének kognitív erősségeit és gyengeségeit, illetve segít az egyéni és csoportos kognitív fejlődés támogatásában is. Ebben segít a *CHC modell* (Cattell-Horn-Carroll modell[18]), amely pszichometriai keretet kínál a kognitív képességek strukturált vizsgálatához és értékeléséhez.

### 2.4.1. A CHC modell bemutatása

A CHC modell egy olyan pszichometriai modell, amely az emberi kognitív képességeket rendszerezi. A modell hierarchikus felépítésű, és három fő réteget határoz meg a kognitív képességek vizsgálatának mélységére. A modell rétegeinek vizualizálása a 2.1-es ábrán látható.



2.1. ábra. CHC modell rétegei

Az első réteg, Cattell általános tényezői (g-faktor), az intelligencia általános mérését célozza meg, amely egy adott személy általános kognitív teljesítményét jelzi. A g-faktor minden kognitív feladathoz hozzájárul, de nem részletezi, mely konkrét részképességek dominálnak egy adott feladat során.

A második réteg, Horn széles terjedelmű faktorai, az intelligencia különböző aspektusait reprezentálja. Ezek a faktorok részletesebben lebontják az intelligenciát, és segítenek megérteni, hogy egy adott feladat mely részképességeket használja. Ilyenek például a verbális tudás, a számolási képességek és a munkamemória.

A harmadik rétegben, Carroll keskeny faktorai találhatók, amelyek még specifikusabb faktorokat azonosítanak, és még részletesebben lebontják a széles terjedelmű faktorokat. Ezek a faktorok konkrét feladatokhoz vagy területekhez kapcsolódnak, például a szókinész, a verbális érvelés, a matematikai gondolkodás vagy a zenei képességek.

Tehát a CHC modell segítségével meg lehet próbálni eldönteni, hogy egy adott feladat mely részképességeket használja, és azok szintjétől mennyire függ. Ehhez általában kognitív tesztek és elemzéseket használnak, hogy azonosítsák, mely faktorok dominálnak egy adott feladat során. Például egy matematikai feladatnál a számolási képességek faktora lehetnek dominánsak, míg egy nyelvi feladatnál a verbális faktorok lehetnek fontosak.

Az egyén kognitív képességeinek pontos lebontása és azok szintjének meghatározása általában speciális kognitív tesztek és értékelések segítségével történik, és szakemberek végzik el ezeket az értékeléseket. Ezek a tesztek lehetővé teszik, hogy megtudjuk, hogy az adott személy milyen erősségekkel és gyengeségekkel rendelkezik kognitív területeken, és hogyan használja ezeket a képességeket különböző feladatok során.

#### 2.4.2. A CHC modell számszerűsítése

A CHC modell részképességei nem számszerűsítik magukat a hagyományos értelemben, mint például egy százalékos skála vagy egy abszolút érték. Ehelyett a modell inkább azt hangsúlyozza, hogy a kognitív képességek között vannak különbségek, és ezeket a különbségeket leírják és rangsorolják.

Az értékelés általában normatív alapon történik, ami azt jelenti, hogy az egyén teljesítményét összehasonlítják egy normál eloszlásban lévő csoporttal. Az eredményeket általában standard pontszámokként vagy percentilis rangsorokként adják meg. A standard pontszámok azt mutatják meg, hogy az egyén mennyire tér el az átlagtól a normál eloszlásban, míg a percentilis rangsor azt mutatja meg, hogy az egyén teljesítménye hogyan viszonyul a normál eloszlásban lévő csoport többi tagjához.

Az értékelések során figyelembe veszik az egyéni kognitív képességek különbségeit, és ezek alapján adják meg az eredményeket. Például egy adott teszt eredménye lehet, hogy valaki a 75. percentilisben teljesít a verbális szókinccs terén, ami azt jelenti, hogy a normál eloszlásban lévő csoport 75%-a rosszabbul teljesített ennél a tesztnél, és csak a 25% teljesített jobban.

Fontos megérteni, hogy a CHC modell nem jelent ki konkrét értékeket vagy számokat a részképességekről, hanem inkább egy elméleti keretet nyújt a kognitív képességek leírására és értékelésére, valamint az egyéni különbségek vizsgálatára. Az értékelések során használt konkrét pontszámok és rangsorok a tesztelő intézményektől és a tesztektől függenek. A továbbiakban, minden alkalommal, amikor faktorokat számokkal ábrázolom részletesen ismertetem a számszerűsítés interpretációját.

### 2.5. Az Item Response Theory

A mérési modellek kulcsszerepet játszanak a pszichológiai és pedagógiai értékelésben, segítve a teszteredmények értelmezését és a különböző képességek pontosabb mérését. Ebben a fejezetben bemutatom az *Item Response Theory*-t (IRT), az egyik legismertebb kvantitatív pszichometriai modellt, mely a személyek képességmérésére ad egy számszerű modellt. Az IRT modellre magyarul *Valószínűségi tesztelmélet*ként is szoktak hivatkozni.

#### 2.5.1. Az IRT modell bemutatása

Az IRT egy statisztikai keretet biztosít a tesztfeladatokra adott válaszok elemzéséhez, és lehetőséget nyújt a tesztfeladatok és a válaszadók képességeinek egyidejű értékelésére. A modell alapja az a feltevés, hogy egy tetszőleges kérdésre adott válasz valószínűsége függ a válaszadó képességeitől és a feladat nehézségétől.[7].

Az IRT a teszt egyes tételeire - tesztfeladatokra - fókuszál, értékelve, hogy mennyire valószínű, hogy a válaszadó helyes választ ad, figyelembe véve a képességét. Az feladatok

lehetnek dichotóm vagy politóm formátumúak, előbbi esetben két lehetséges választási lehetőséggel, utóbbi esetben több választási lehetőséggel és eltérő pontértékekkel.

A válaszadó képessége vagy tulajdonsága egy látens, nem megfigyelhető változóval került modellezésre. Az IRT segítségével becsülhetjük ennek az értékét, így jobb rálátást nyerhetünk a válaszadó teljesítményére. Ezáltal az IRT hasznos eszköz lehet a tesztek fejlesztésében és a válaszadók képességeinek pontosabb mérésében.

### 2.5.2. A tételválasz függvény

Az IRT modell az úgynevezett *tételválasz függvény* (item response function) számszerűsíti a személyek képessége és a feladatok nehézsége közötti összefüggést. A tételválasz függvény megadja, hogy egy adott képességgel rendelkező egyén milyen valószínűséggel válaszol helyesen egy meghatározott nehézségű feladatra. A feladatok leírására az IRT modell legelterjedtebb, 3 paraméteres változata a következő paramétereket definiálja:

1. **Nehézség** ( $b_i$ ): Az a képességszint, ahol a válaszadó 50%-al oldja meg helyesen a feladatot. Tehát egy  $b$  nehézségű feladatot egy  $b$  képességű egyén definíció alapján 50% valószínűséggel old meg helyesen. A definíció következménye, hogy magasabb nehézség érték nehezebb feladatot jelent, hiszen magasabb képességű személy szükséges ahhoz, hogy ugyanazzal a valószínűséggel oldja meg.
2. **Diszkrimináció** ( $a_i$ ): Az a mérték, hogy a feladaton elért sikeresség mennyire határozza meg a személy képességét. Egy nagyon könnyű feladat általában nem diszkriminál nagyon, hiszen abból, hogy a személy azt a feladatot meg tudta oldani, nem sok információt tudunk meg a személy képességéről. Hasonlóan a legtöbb esetben egy nagyon nehéz feladat sem igazán diszkriminál, amelyet szinte senki sem tud megoldani. Azonban a diszkrimináció paramétert nem maga a feladat nehézsége, mint inkább a feladat jellege határozza meg.
3. **Találgatás** ( $c_i$ ): A valószínűség, hogy a válaszadók alacsony képességszinttel is helyesen válaszolnak a tételre, talán csak találgatásból. Egy két választási lehetőséget tartalmazó feladat esetében ez 50%, hiszen mindössze találgatással ekkora esélye van egy tetszőleges személynek helyes választ adni.

Az IRT tehát a fenti paraméterek segítségével a következő képlettel határozza meg a tételválasz függvényt:

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}},$$

ahol  $\theta$  a személy képesség értéke,  $a_i$ ,  $b_i$  és  $c_i$  a feladat fent bemutatott paraméterei és  $p_i(\theta)$  megadja, hogy az  $i$ . feladatot egy  $\theta$  képességű személy mekkora valószínűséggel oldja meg helyesen.

### 2.5.3. Az IRT modell előnyei

Az IRT-modell számos előnnyel rendelkezik a hagyományos tesztelméletekhez képest. Egyik kiemelkedő előnye, hogy támogatja a számítógépes adaptív tesztelés kialakítását, mely során a teszt kérdései a válaszadó aktuális képességeihez igazodva változnak, biztosítva így a mérés pontosságát és hatékonyságát. Továbbá lehetőséget nyújt arra, hogy a képességeket és a tételek paramétereit külön-külön állítsuk be, amivel pontosabb és megbízhatóbb eredményeket érhetünk el. Az IRT jellegzetessége még az úgynevezett skála invariancia, ami azt jelenti, hogy a mérés nem függ a választott teszttételektől. Ennek köszönhetően két különböző, de IRT alapú teszt eredményei közvetlenül összehasonlíthatók egymással.

## 2.6. A játékos fejlesztés jelentősége és módszertana

A kognitív feladatok gyakran unalmasnak és ismétlődőnek bizonyulnak a résztvevők, különösképpen a gyermekek számára, ami csökkenti az elköteleződésüket és negatívan befolyásolja a mérés pontosságát. A játékosítás, másnéven *gemifikáció* (angolul *gamification*) vagyis a játékelemek beépítése a feladatokba, potenciális megoldást kínál erre a problémára, mivel így a feladatok élvezetesebbé és motiválóbbá válhatnak[16].

A gemifikáció célja a játéktervezési elemek - mint például a verseny, narratíva, ranglisták, grafika, stb. - beillesztése a kognitív feladatokba olyan módon, hogy a mérés tudományos értéke megmaradjon. A játékelemek használata növelheti a résztvevők motivációját és érdeklődését, ami kifejezetten fontos ahhoz, hogy a mérés pontos eredményeket adjon. Bizonyított, hogy ha a teszt során a résztvevő lelkesedése csökken, az jelentősen rontja a gyűjtött adatok minőségét[16]. Ezáltal tehát a gemifikáció az érdeklődés fenntartásával növelheti a mérés pontosságát is.

Emellett a gemifikáció hozzájárulhat a résztvevő kognitív fejlődéséhez is. A játéktervezés elemei, mint például a problémamegoldás, stratégiai gondolkodás és döntéshozatal, segíthetnek a kognitív készségek fejlesztésében. A játékokban való részvétel javíthatja a munkamemória, a figyelem és más kognitív területek teljesítményét.

Azonban fontos megjegyezni, hogy a gemifikáció nem megfelelő minden feladattípus és célközönség számára. A játékelemek bevezetésekor figyelembe kell venni a célcsoport életkorát, érdeklődési körét és a feladatok jellegét, hogy biztosítsuk a gemifikáció sikerét. A játékelemeknek értelmesen és célzottan kell kapcsolódniuk a kognitív feladatokhoz, anélkül, hogy azok tudományos értékét csökkentenék.

A kutatásomhoz rendelkezésre álló mérések feladatainak megalkotása során is kulcsfontosságú szerepet töltött be a mérés gemifikációja. Ez nem csak abban segített, hogy a feladatokat élvezetesebbé tegye a résztvevők számára, hanem jelentős mértékben hozzájárult a mérési eredmények pontosságának és megbízhatóságának növeléséhez is.

## 3. fejezet

# Neurális hálózat és beágyazás

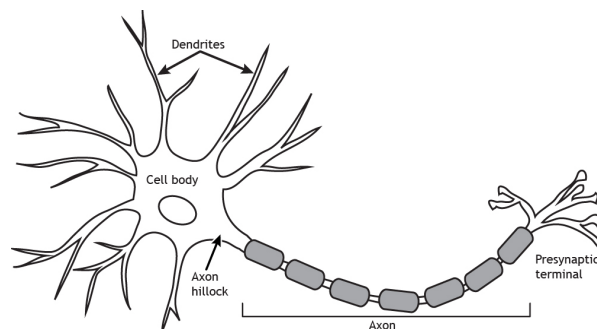
Az elmúlt évtizedben a neurális hálózatok és a mélytanulás óriási fejlődésnek indultak, és jelentősen befolyásolták rengeteg kutatásterületen a végzett méréseket. A neurális hálózatok hatékony eszköznek bizonyultak sokféle feladat megoldásában, mint például képfelismerés, ajánló rendszerek vagy tetszőleges regressziós feladatok. Bár a neurális hálózatok manapság egyre szélesebb körben ismertek és használtak, fontosnak tartom, hogy bemutassam azok alapjait, különösen azért, mert a beágyazás a pszichometriai kutatások terén úttörőnek számít. A következőkben rövid bevezetést nyújtok a neurális hálózatok alapjaiba és kicsit bemutatom a beágyazás alapjait valamint alkalmazásait, amely egy fontos modern eszköz a gépi tanulás területén.

### 3.1. Neurális hálózatok rövid bemutatása

Ebben a szakaszban egy rövid áttekintés található a neurális hálózatok alapvető koncepcióiról, struktúrájáról és működéséről, továbbá megvizsgálom, hogy miként lehet neurális hálózatot építeni és tanítani adott problémák megoldására.

#### 3.1.1. Biológiai motiváció: a neuron

A neurális hálózat a biológiából ismert *neuronokról* kapta a nevét. A neuronok úgyvezetett dendriteken keresztül kapják meg a jeleket (általában több dendriten keresztül), és ha az ingerlés intenzitása meghalad egy bizonyos küszöbértéket, akkor a neuron elektromos jelet hoz létre és továbbítja azt - ezt nevezik akciós potenciálnak. Az akciós potenciál az axon mentén halad tovább a kapcsolatai felé. A folyamat során tehát bemeneti jelek érkeznek a dendriteken keresztül, amelyeket a sejttest dolgoz fel, majd a kimenetként létrejött akciós potenciál az axon mentén továbbítódik[12].



3.1. ábra. A biológiai neuron felépítése[12]



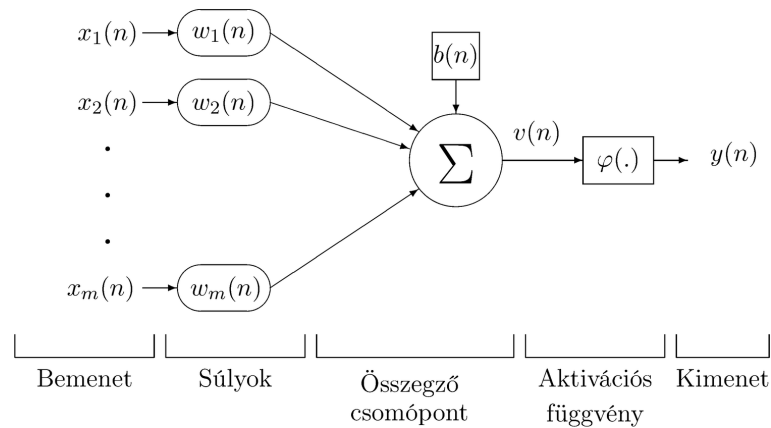
Az említett folyamat lehetővé teszi a neuronok számára, hogy összetett hálózatokat hozzanak létre és kommunikáljanak egymással. Ez az alapelv nemcsak a biológiai, hanem a mesterséges neurális hálózatok működésében is visszaköszön, ahol a mesterséges neuronok is hasonló módon, bemeneti jelek alapján aktiválódnak és továbbítják az információt a hálózatban.

### 3.1.2. A perceptron

Neurális hálózatoknál a neuron megfelelője a perceptron. A perceptron a neurális hálózat legkisebb egysége, amely a neuronhoz hasonlóan jeleket kap, és ezeknek a jeleknek egy bizonyos függvénye továbbításra kerül. A hálózat ilyen perceptronokból épül fel, és emiatt a hasonlóság miatt nevezték el *neurális hálózatnak*.

#### A perceptron felépítése

A perceptron felépítése a 3.2-es ábrán látható. Balról jobbra haladva az először a bemeneti paramétereket találjuk, amik a az attribútumokat jelentik. Például ezek a bemeneti paraméterek lehetnek a személyek részképességei. Ezek a paraméterek egyenként megszoróznak egy-egy súly értékkel, majd ezeknek az összege fut bele az összegző csomópontba. Ebben a pontban egy úgynevezett torzítás (*bias*) érték is hozzáadódik az összeghez. Az így elkészült összeget egy úgynevezett aktivációs függvényen kiértékeljük, és így kapjuk a perceptron kimenetét.



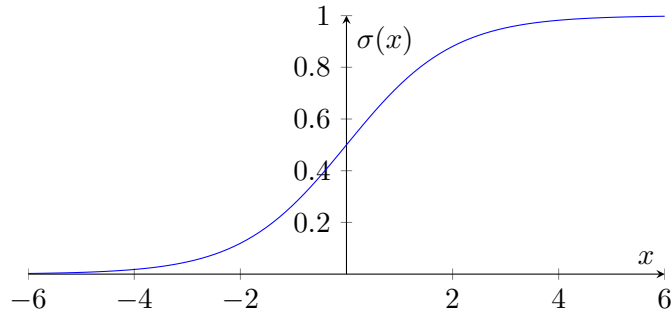
3.2. ábra. Egy perceptron felépítése[13]

#### Az aktivációs függvény

Az aktivációs függvény az a függvény, amely a kimenetet előállítja az összegző csomópontban kapott értékből. Aktivációs függvénynek választhatjuk legegyszerűbb esetben az identitás függvényt, de a szigmoid függvény is gyakori választás. A szigmoid függvény a  $(-\infty; \infty)$  tartományt képezi le a  $[0; 1]$  intervallumra a következő módon:

$$\sigma(x) = \frac{1}{1 + e^{-Kx}}$$

A 3.3-as ábrán látható a szigmoid függvény grafikonja. Az ábrán látható, hogy a függvény miként képezi le a valós számok halmazát a  $[0; 1]$  intervallumra. Az ábrázolt grafikon a  $K = 1$  esetet mutatja be, ennek a paraméternek a változtatásával a függvény meredeksége változtatható.



**3.3. ábra.** Sigmoid függvény grafikonja

## A perceptron tanulása

A neuronális hálózat tehát ilyen perceptronokból épül fel olyan módon, hogy az egyes perceptronok kimenetei sok esetben más perceptronok bemeneteivel vannak összekötve. Vegyük azonban most példának a legegyszerűbb neuronális hálózatot, mely egy darab perceptronból áll. Nézzük meg ezen a példán keresztül, hogyan működik a tanulás folyamata.

A folyamat elején mindössze a bemeneti paraméterek, attribútumok ismertek, illetve az aktivációs függvény. A perceptron tanulása leegyszerűsítve úgy működik, hogy először inicializálásra kerülnek a súly értékek és a torzítás. Egy iteráció abból áll, hogy az algoritmus kiszámítja az adott súlyokkal az összes bemenetre a kimenet értékét, majd összehasonlítja az elvárt kimenettel. Végül az összehasonlítás alapján úgy módosítja a súlyok és a torzítás értékét, hogy csökkentse a modell hibáját. A perceptronnak azt is meg kell adni, hogy milyen hiba függvény próbáljon minimalizálni. Erre egy gyakori példa a *Mean Squared Error* (MSE), amely az eltérések négyzetének átlaga.

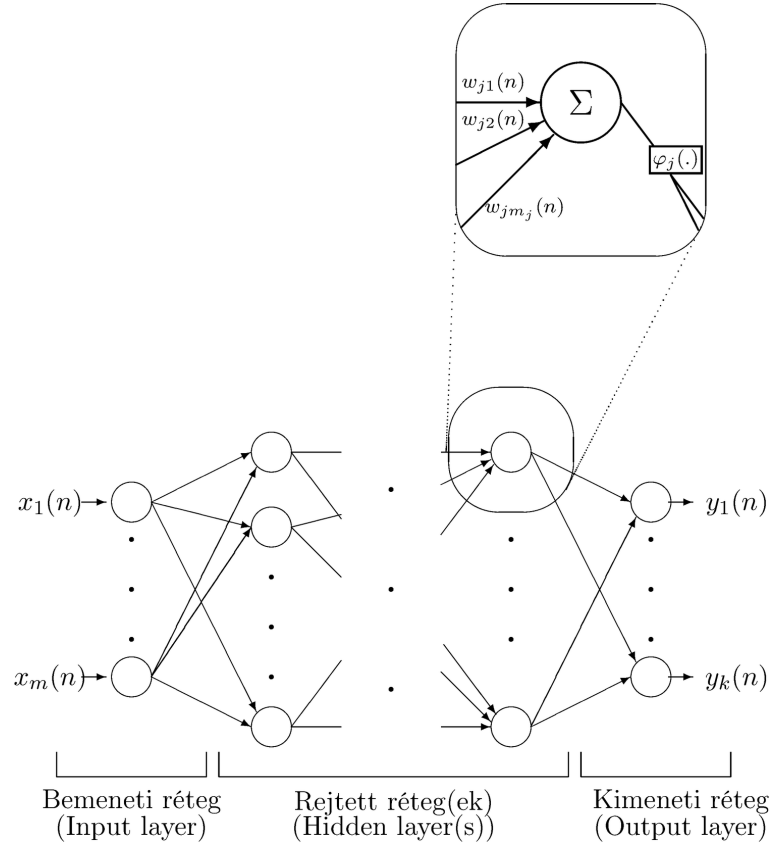
A perceptron tehát ilyen iterációk sorozatával határozza meg a súlyok és a torzítás értékét, majd amikor a hiba már kellően alacsony, vagy befejeződött az előre definiált mennyiségű iteráció, akkor a tanulás leáll, és a kapott modell a leállás pillanatában lévő súly értékeket fogja alkalmazni.

### 3.1.3. Többrétegű neuronális hálózat

A perceptronokból álló egyszerű hálózatok azonban korlátozottak abban, hogy milyen típusú problémákat képesek megoldani. Ezt a problémát lehet orvosolni azzal, ha több perceptront rétegekbe rendezünk, többrétegű perceptronokat, azaz több rétegű neuronális hálózatokat létrehozva. Egy többrétegű hálózatban a perceptronok kapcsolatba lépnek egymással, és az információ áramlik a hálózat bemenetétől a kimenetéig. A többrétegű neuronális hálózat felépítése a 3.4-es ábrán látható.

#### Többrétegű hálózatok felépítése

Egy többrétegű hálózatban három fő rétegtípust különböztetünk meg: van egy bemeneti réteg, lehetnek rejtett rétegek és van egy kimeneti réteg. A **bemeneti réteg** az, ami közvetlenül kapcsolatban áll a hálózat bemenetével. Ennek a rétegnek a csomópontjai a bemeneti paraméterek. A **rejtett rétegek** a bemeneti és a kimeneti réteg között találhatóak. Egy hálózat tartalmazhat egy vagy több rejtett réteget is. A rejtett rétegek száma és mérete nagymértékben befolyásolja a hálózat képességét. A **kimeneti rétegben** találhatóak a hálózat kimenetei. A kimeneti réteg neuronjainak száma a megoldandó feladattól függ.



**3.4. ábra.** Egy többrétegű neurális hálózatfelépítése[13]

### Tanulási folyamat

A tanulási folyamat a többrétegű hálózatok esetében is hasonló a perceptron esetében leírtakhoz. Azonban itt a hálózat komplexitása miatt a hibát a kimeneti rétegtől kezdve a bemeneti réteg irányába terjedve számoljuk vissza, egy úgynevezett hátraterjesztési (*backpropagation*) algoritmus segítségével. A súlyokat és a torzításokat a hálózatban a hiba minimalizálása érdekében módosítjuk.

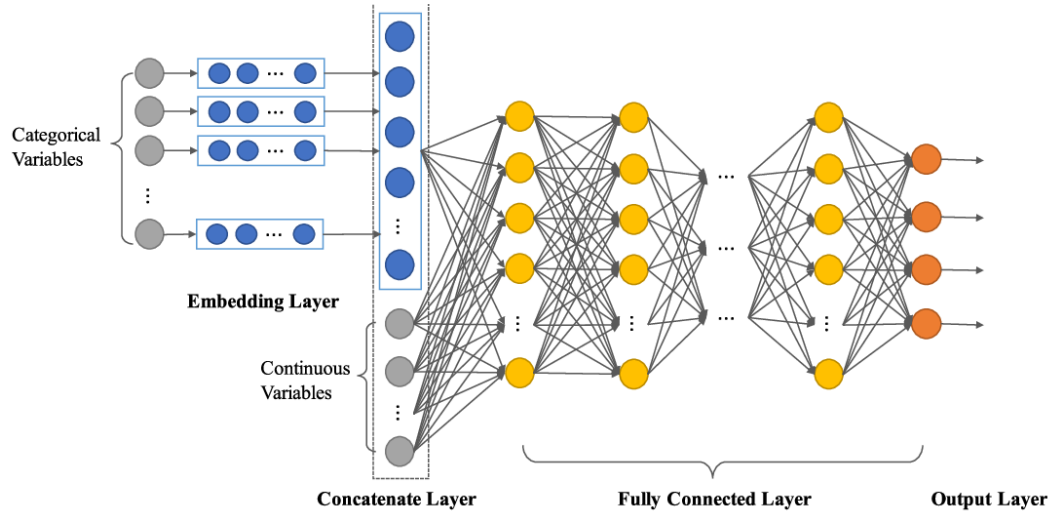
### 3.2. Beágyazás

A *beágyazás*, angolul *embedding* egy olyan leképezés, amely egy diszkrét (kategórikus) változót folytonos számokból álló vektorra alakít. Neurális hálózatok kontextusában a beágyazások alacsony dimenziós, tanult folytonos reprezentációi a diszkrét változóknak. A neurális hálózati beágyazások azért hasznosak, mert képesek csökkenteni a kategórikus változók dimenzióját illetve reprezentálni a kategóriákat több független paraméter segítségével.

A neurális hálózati beágyazásnak három fő célja van:

1. A legközelebbi szomszédok megtalálása a beágyazással készült vektorok terében. Ezeket fel lehet használni a kategóriák csoportosítására vagy ajánlások készítésére a felhasználói érdeklődés alapján.
2. Bemenetként egy felügyelt feladathoz tartozó gépi tanulási modellhez.
3. A fogalmak és a kategóriák közötti kapcsolatok vizualizálására.

A beágyazás működését legegyszerűbben a 3.5-ös[17] ábra segítségével lehet szemléltetni. A neurális hálózat elején a kategórikus bementi változókat felbontjuk folytonos értéket tartalmazó vektorokra, és a vektorok elemei lesznek a neurális hálózat többi részének bemenetei. Tehát egy adott kategórikus változót megfigyelve, minden lehetséges értékhez rendelünk egy vektort, és a továbbiakban a vektor értékei szimbolizálják az adott kategórikus változót.



3.5. ábra. Beágyazás felépítése[17]

A vektorok kialakítása hasonló módon zajlik, mint a neurális hálózat tanulása. A vektorok értékeit először inicializáljuk egy tetszőleges értékkel, majd lefuttatunk egy iterációt a tanulási folyamatból. A tanulási folyamat végén a súlyokon és torzításokon kívül a beágyazás vektorokat is aszerint módosítjuk, hogy a neurális hálózat hibáját minimalizáljuk. Ezáltal a tanulási folyamat végén a kategórikus változók összes lehetséges értékéhez kapunk egy vektort, amely több egymástól független valós értékkel reprezentálja a kategórikus változót.

### 3.2.1. Beágyazás alkalmazása

A beágyazásokat számos különböző területen és alkalmazásban használják fel, kihasználva a magas dimenziós kategórikus adatok folytonos, alacsony dimenziós reprezentációjának előnyeit.

### Szövegelemzés és Természetes Nyelvfeldolgozás (NLP)

A beágyazások talán legismertebb alkalmazása a szövegelemzés és a természetes nyelvfeldolgozás területén található. A szavak, mondatok vagy akár egész dokumentumok beágyazásával képesek vagyunk megőrizni a szövegelemek közötti szemantikai és kontextuális kapcsolatokat. A Word2Vec, GloVe és BERT olyan ismert technikák, amelyek ezt a megközelítést használják.

### Ajánlórendszerek

A beágyazásokat ajánlórendszerekben is használják, ahol a felhasználók és a termékek közötti kapcsolatokat modellezzik. Egy tipikus példa lehet a filmek vagy a termékek ajánlása, ahol a felhasználók és a filmek vagy termékek közötti interakciókat beágyazások segítsé-

gével reprezentáljuk, és ezen beágyazások segítségével próbáljuk megjósolni a felhasználók érdeklődését.

### **Képfeldolgozás**

Bár a beágyazásokat leginkább szöveges adatokhoz kapcsolják, használatuk nem korlátozódik kizárólag erre a területre. A képfeldolgozásban például az egyes képek beágyazásai segíthetnek a hasonló tartalmú vagy stílusú képek azonosításában.

### **További alkalmazási lehetőségek**

Az előzőekben említett példák mellett a beágyazás bármilyen kutatási ágazatban használható lehet, ahol a bevezetőben taglalt célokra, vagyis kategóriák rendszerezésére és fogalmak, valamint kategóriák összekapcsolására van szükség.

A pszichometriában, különösen a részképességek vizsgálatánál, a beágyazás alkalmazása jelentős újításnak tekinthető. Bár ezen a területen korábban nem alkalmazták ezt a módszert, a részképességek vizsgálatához kapcsolódóan mindkét cél megjelenik. Ha a különféle tesztek kategóriaként kezeljük, a tesztek közötti összehasonlítás és hasonló tesztek keresése alapvető feladat a pszichometriában. Továbbá az, hogy melyik tesztfeladathoz milyen részképességekre van szükség, éppen azt mutatja, hogy a kategóriákat milyen fogalmakkal és ezek numerikus értékeivel kívánjuk társítani. Ez az innovatív megközelítés így számos értékes eredménnyel gazdagíthatja a kutatási területet

## 4. fejezet

# Kutatási hipotézis és módszertan

Ebben a fejezetben részletezem a kutatásom alapvető célkitűzéseit, az elején felállított hipotéziseimet, valamint a kutatási folyamat során felmerülő újabb kérdéseket. A célokhoz mellékelem a megoldás ötletét illetve elméleti alapjait. Alaposan körbejárom a módszertani kihívásokat, amelyekkel szembenéztem, bemutatom a lehetséges megoldási javaslatokat, és kifejttem, hogy miért az adott megoldást választottam.

### 4.1. A kutatás célja és felállított hipotézisek

A kutatás kezdetén meghatároztam azokat az alapvető kérdéseket, melyekre válaszokat kerestem, azonban a kutatás előrehaladtával ezek száma folyamatosan nőtt. Minden egyes válasz további kérdésekhez vezetett, mivel a mesterséges intelligencia alkalmazása a vizsgált területen még nagy újdonság. Ezen új eszközökkel olyan kérdésekre is választ kaphatunk, amelyek a hagyományos, egyszerű statisztikai módszereket használó kutatók számára eddig megválaszolatlanok maradtak.

A kutatás fő kérdései a korábban bemutatott CHC modellre épülnek. A CHC modell azt mondja, hogy az emberi kognitív képességek felbonthatóak részképességekre olyan módon, hogy ezek a részképességek önállóan befolyásolják bizonyos feladatok teljesítését. A kutatás során nem minden esetben foglalkoztam a CHC modell által leírt összes részképességgel, esetenként azzal az egyszerűsítéssel éltem, hogy a vizsgált feladatban bizonyos faktorok biztosan nem játszanak szerepet és leszűkítettem a vizsgált részképességek számát.

A CHC modell azonban nem jelent ki konkrét értékeket vagy számokat a részképességekről, hanem inkább elméleti keretet nyújt a kognitív képességek leírására és értékelésére, valamint az egyéni különbségek vizsgálatára. Az értékelések során használt konkrét pontszámok és rangsorok tehát a kiértékelőtől függenek. Az itt bemutatott mérések és elemzések során a konkrét számszerű részképesség értékek mindig csak egy kontextuson belül fognak információt tartalmazni. Természetesen a részképességek számszerűsítése azt az egyszerű konvenciót követi, hogy egy nagyobb részképesség érték jelentése az, hogy az adott személy könnyebben old meg egy arra a részképességre építő feladatot, azonban a részképesség pontos értéke önmagában nem hordoz információt. Ha szükség van a részképességek összehasonlítására két különböző mérés között, akkor az értékeket normalizálni kell valamilyen módon. Azonban a legtöbb esetben elegendő lesz az értékek összehasonlítása az adott mérésen belül.

A következőkben ismertetem a kutatás fő céljait, illetve bemutatom a szükséges megközelítéseket, anélkül, hogy kitérnék a részletes megoldási ötletekre.

#### 4.1.1. Feladatok részképesség-függésének vizsgálata

A kutatás első és legfontosabb célja különböző feladatok részképesség-függésének vizsgálata. A céлом minden feladathoz meghatározni, hogy a különböző részképességek mekkora szerepet játszanak a feladat megoldása során.

##### A feladat relevanciája

A kutatás legfőbb eredménye a feladatok részképesség-függésének meghatározása, hiszen sok olyan fejlesztésben használt teszt, feladat és játék van, melyekről a kutatók és az oktatók még nem tudják, hogy mely részképességre építenek. Sok feladat esetén vannak a pszichometrikusoknak hipotézisei, azonban ez korábban még nem került megerősítésre kvantitatív módon - illetve felfedezhetünk olyan összefüggéseket, amelyekre a terület kutatói eddig még nem gondoltak.

Az eredmények elősegítik, hogy az oktatók specifikus felméréseket végezzenek el a különböző részképességek vizsgálatára. Továbbá az eredmények segítségével létrehozható olyan adaptív tesztsorozat, amely a kitöltő korábbi megoldásai alapján módosítja a következő kérdéseket olyan módon, hogy részletesebb képet kapjunk a kitöltő azon részképességeiről, amelyek gyengébbnek bizonyulnak a korai megoldások alapján.

##### Megoldási megközelítések

A cél eléréséhez két különböző megközelítést szükséges alkalmazni a bemeneti információk alapján. Míg az első változatban sokkal több információ áll rendelkezésünkre és részletes képet ad a feladat részképesség-függéséről, addig a második megközelítés a rendelkezésre álló minimális adatok alapján olyan eredményre jut, amely kiváló alapot ad a többi módszer alkalmazására - melyek segítségével pontosíthatjuk a kapott eredményt.

Az első megközelítésben a válaszadók részképességeinek értékei már ismertek, valamint rendelkezésre áll a megoldásuk egy adott feladatra. Ez azért lehetséges, mert léteznek olyan tesztsorozatok, amelyek már bemértek pszichometrikusok által, és validáltan egyetlen részképességre támaszkodnak, pontosabban a többi részképességtől való függésük elhanyagolható[10]. A céлом meghatározni, hogy a vizsgált részképességek mekkora szerepet töltenek be a feladat megoldása során. Feltételezésem szerint az erre kidolgozott eljárás pontos képet ad a feladat részképesség-függéséről, azonban hátránya, hogy sok információ szükséges hozzá, amelyek sok esetben nem állnak rendelkezésünkre a mérés során.

Ez indokolja a második módszer elkészítésének szükségességét, amelyben a bemenet csupán számos válaszadó különböző feladatokra adott válaszait tartalmazza. A cél mindössze ezekből az adatokból meghatározni olyan paramétereket, amelyek valamilyen hasonló módon befolyásolják a megoldást. Ebben a megközelítésben természetesen a paraméterek meghatározását követően még szükség van azok azonosítására - részképességhez kötésére is.

#### 4.1.2. Alanyok részképességének vizsgálata

A következő cél a kitöltők részképesség értékeinek meghatározása. Céлом, hogy a feladatokat megoldó személyek képességeit kvantitatív módon jellemezzem, részképességekre lebontva.

##### A feladat relevanciája

A kitöltők részképességeinek elemzésével felismerhetjük, hogy a személyek mely kognitív képességeit szükséges a leginkább fejleszteni, és ennek segítségével megfelelő módon alkalmazhatunk személyre szabott feladatokat és játékokat az egyének fejlesztésére. A rész-

képesség elemzés során felfigyelhetünk kiugró értékekre, amelyek segítségével korai stá-tuszban észrevehetőek a különböző tanulási nehézségek jelei, és ennek megfelelően időben megkezdődhet ezeknek a kezelése - amely segítségével jelentősen redukálhatóak a későbbi nehézségek. Ezzel egyetemben azonban a pozitívan kimagasló értékek is felismerhetőek, és azoknak megfelelően alkalmazhatunk tehetség gondozást az adott képességterületen.

### **Megoldási megközelítések**

Az előző pontban vizsgált feladathoz hasonlóan a részképességek vizsgálatára is két külön-böző megközelítést kell alkalmaznunk. Itt is szükség van egy részletes, pontos vizsgálatra, amely során rendelkezésre állnak olyan feladatok, melyek részképesség-függése ismert - legegyszerűbb esetben azt is tudjuk, hogy az adott feladat kifejezetten egy részképességre épít. Ennek a meghatározása történhet a korábban bemutatott módon vagy ezzel foglal-kozó pszichometrikusok hipotézisei alapján.

A másik megközelítésben azonban ismeretlen a feladatok részképesség-függése. Ebben a megközelítésben természetesen nem tudunk egyértelműen számszerű eredményt társítani a megnevezett részképességekhez, a feladat inkább olyan paraméterek meghatározása a kitöltőkhöz, amely hasonló módon befolyásolja a megoldásokat minden feladat esetén. A hipotézisem szerint ezek a paraméterek megfeleltethetőek egy-egy részképességnek, vagy kevesebb paraméter használata esetén részképességek egy halmazának.

#### **4.1.3. Eredménybecslés ismert részképességek esetén**

A kutatás során célom továbbá megbecsülni adott részképességgel rendelkező személy vá-laszát egy tetszőleges feladatra - mely részképesség függése nem feltétlen ismert. A hipo-tézisem szerint a korábbi eredmények segítségével lehetőség nyílik egy tetszőleges feladat megoldási valószínűségének becslésére jó közelítéssel.

### **A feladat relevanciája**

Az eredménybecslés amellett, hogy rendkívül érdekes lehet, szintén segít adaptív feladatok létrehozásához. Ha valamilyen szinten már ismertek a személy részképességei, megvizsgál-hatjuk, hogy egy adott feladatot milyen eredménnyel fog megoldani, és eszerint választ-hatjuk ki a következő feladatokat annak érdekében, hogy a legrészletesebb képet kapjuk a kitöltő képességeiről. Emellett a kiválasztott feladat során a várt eredmény összehasonlít-ható a kapott megoldással, és ennek segítségével módosíthatóak az adott személy képesség értékei.

Továbbá, az adaptív feladatkiválasztáshoz kapcsolódóan a megközelítés különösen fontos a kitöltő motiválására a feladatsor, tesztsor befejezésére. Ismert, hogy játékos fel-adatok megoldása olyankor a legérdekesebb, ha a kapott feladat se nem túl nehéz, se nem túl könnyű[9]. A túl egyszerű feladatok hamar unalmassá válnak, a túlzottan nehéz felada-tok pedig stresszt, idegességet és sikertelenség érzését válthatják ki a kitöltőben, amelyek eredményeként egyrészt könnyen elvesztheti a lelkesedést a feladatsor befejezésére, más-részt ezek az érzések az adott megoldásokat is jelentősen befolyásolják. Így tehát fontos, hogy a kitöltőt az úgynevezett *flow* állapotban kell tartani, amelyet úgy lehet elérni, hogy-ha olyan feladatokat kap, amelyeket átlagosan 70% eséllyel tud megoldani[9]. Ha képes a rendszer megbecsülni, hogy az adott személy milyen eredménnyel oldaná meg a különböző feladatokat, akkor megfelelően ki tudja választani a következő feladatot ennek megfelelően.



## 4.2. A célok megvalósításának elméleti háttere

A következőkben bemutatom, hogy a különböző célokat milyen elméleti megoldások mentén közelítettem meg. Az elemzéseim során figyelembe vettem mind olyan feladatokat, melyekre a válaszról csak azt tudjuk, hogy helyes-e vagy sem, mind olyan feladatokat, amelyek egy adott skálán pontozhatóak. Az előbbi esetében mindig a megoldási valószínűséget érdemes becsülni, hiszen csak az eredmény becslése sokkal kevesebb információt tartalmaz, míg az utóbbi esetben magát a pontszámot fogom becsülni.

A feladatok részképesség-függésének vizsgálatához, ha a válaszadók részképességei már ismertek, akkor a részképesség-függés faktoranalízis segítségével kiszámítható.

A faktoranalízis egy olyan statisztikai technika, amely segít azonosítani a változók közötti struktúrákat és kapcsolatokat. Ebben az esetben a változók a részképességek, és azt szeretnénk megtudni, hogy az adott feladat milyen valószínűséggel oldható meg az egyes részképességek ismeretének birtokában.

Ha feltételezzük, hogy a feladat megoldási valószínűsége vagy az feladatra kapott pontszám lineáris összefüggésben áll a részképességekkel, akkor a faktoranalízis visszavezethető bináris értékelésű feladatoknál logisztikus regresszióra, pontozott feladatoknál pedig lineáris regresszióra[6]. Ezeket analíziseket végrehajtva megkapható, hogy az egyes részképességek mekkora jelentőséggel hatnak az eredményre.

### 4.2.1. Faktoranalízis pontozott feladatokra

Pontozott feladatok esetén a faktoranalízis alkalmazására a lineáris regresszió biztosít megfelelő módszert. A lineáris regresszió egy statisztikai technika, amely segít meghatározni a függő és a magyarázó változók közötti lineáris kapcsolatot. Ebben az esetben a függő változó a feladatokra kapott pontszám, míg a magyarázó változók a különböző részképességek.

Az egyenlet, ami leírja ezt a kapcsolatot, a következő formát öltheti:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon,$$

ahol:

- $Y$  a feladatokra kapott pontszám,
- $\beta_0$  a konstans tag,
- $\beta_1, \beta_2, \dots, \beta_n$  a különböző részképességekhez tartozó regressziós együtthatók,
- $X_1, X_2, \dots, X_n$  a részképességek értékei,
- $\varepsilon$  a modell hibája, vagy más szóval a maradék variancia, amit a modell nem képes megmagyarázni.

A lineáris regresszió célja meghatározni ezeket az együtthatókat úgy, hogy minimalizálja a valós pontszámok és a modell által előrejelzett pontszámok közötti különbségek négyzetösszegét. Az egyes részképességekhez tartozó együtthatók értéke megmutatja, hogy az adott részképesség egy egységnyi növelése mennyivel befolyásolja a feladatra kapott pontszámot.

Az együtthatók értelmezése segít megérteni, hogy az egyes részképességek milyen mértékben járulnak hozzá a teljesítményhez. Pozitív együttható azt jelenti, hogy a részképesség növelése növeli a pontszámot, míg a negatív együttható csökkenti azt. Ha egy együttható értéke közel van a nullához, az azt jelenti, hogy az adott részképességnek csekély vagy semmilyen hatása nincs a teljesítményre.

Ez a modell a faktoranalízisen kívül arra is használható, hogy megbecsülje egy adott részképességű alany pontszámát egy ismert feladatra, hiszen ha ismertek a részképességek és az adott feladat részképesség-függése, vagyis a regressziós együtthatók, akkor a fenti képlettel meghatározható a becsült pontszám.

#### 4.2.2. Faktoranalízis bináris értékelésű feladatokra

Bináris értékelésű feladatoknál a faktoranalízist logisztikus regresszió nevű statisztikai módszerrel lehet megközelíteni. A logisztikus regresszió a lineáris regresszióhoz hasonló eszköz, amely segít modellezni egy bináris kimenetet - például helyes megoldás vagy helytelen megoldás -, és annak összefüggését a magyarázó változókkal, jelen esetben például a részképességek értékeivel.

Ebben az esetben a kimeneti bináris változó a feladatra adott megoldás helyessége lesz, a magyarázó változók pedig a részképességek értékei lesznek. A cél az, hogy megtudjuk, hogy az egyes részképességek milyen mértékben járulnak hozzá a feladat megoldhatóságához.

A logisztikus regresszió a következő egyenlettel írható le:

$$\log \left( \frac{P(Y = 1)}{1 - P(Y = 1)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n,$$

ahol:

- $P(Y = 1)$  most azt jelenti, hogy az egyén helyesen oldotta-e meg a feladatot,
- $\beta_0$  a konstans együttható,
- $\beta_1, \beta_2, \dots, \beta_n$  az egyes részképességekhez tartozó regressziós együtthatók,
- $X_1, X_2, \dots, X_n$  pedig a részképességek értékei.

Látható, hogy a képlet nagyon hasonlít a fenti, lineáris regresszióval említett képletre. A logisztikus regresszió és a lineáris regresszió közötti legfontosabb különbség a kimenet természetében rejlik. Míg a lineáris regresszió folytonos kimenetet modellez, addig a logisztikus regresszió csak két lehetséges kimenetet, vagyis bináris választ enged meg. Ennek megfelelően, míg a lineáris regresszióban az egyes részképességek növekedése folytonosan befolyásolja a feladatra kapott pontszámot, addig a logisztikus regresszióban ez a növekedés megváltoztatja a feladat helyes megoldásának valószínűségét.

Az együtthatók meghatározása során a bináris logisztikus regresszió célja meghatározni, hogy az egyes részképességek mekkora mértékben befolyásolják az adott feladat megoldhatóságát. A regressziós együtthatók értelmezése hasonló a korábban említett modellhez: pozitív együttható azt jelzi, hogy az adott részképesség növelése növeli a feladat megoldhatóságának valószínűségét, míg negatív együttható azt jelzi, hogy csökkenti.

Ez a modell a faktoranalízisen kívül alkalmas arra, hogy becsülje, hogy az adott részképességű alany milyen valószínűséggel oldja meg az adott feladatot. Az együtthatók és az értelmezésük pedig megadja, hogy mely részképességeknek van a legnagyobb hatása a feladat megoldhatóságára, és ezzel lehetőséget ad a feladatok tervezésére és fejlesztésére az alanyok különböző részképességeinek figyelembevételével.

Ezáltal a lineáris és logisztikus regresszió egyszerű megoldást ad a faktoranalízis elvégzésére. Azonban ezek a módszerek azt feltételezik, hogy lineáris kapcsolat áll fenn a részképességek és a feladatra kapott pontszám között, valamint, hogy a részképességek normális eloszlást követnek. Ez általában fennáll, azonban létezhetnek olyan feladatok, amelyekre

ez a hipotézis nem teljesül. Az ilyen feladatokra a faktoranalízis már nem végezhető el lineáris vagy logisztikus regresszió segítségével, ezek a módszerek téves eredményt adnának. Ezek miatt a bevett eljárások alkalmazhatósága korlátozott.

Hipotézisem szerint a neurális hálózat alkalmas lehet arra, hogy a fentieket kiküszöbölve faktoranalízist és eredménybecslést végezzünk, hiszen a neurális hálózat sokkal összetettebb összefüggéseket is képes lehet felismerni, a definíciójából adódóan. Ha a feltételezésem igaz, akkor ez az eljárás olyan esetekben is alkalmazható lehet, ahol a részképességek és a feladat nehézsége között nem lineáris az összefüggés.

### 4.2.3. Eredménybecslés regresszióval és neurális hálóval

Az előzőekben bemutatott módszerek lehetőséget adnak egy ismert részképességekkel rendelkező személy eredményének becslésére egy adott feladaton - feltéve, hogy lineáris kapcsolat áll fenn a részképességek és a feladatra kapott pontszám között. Ha a lineáris vagy logisztikus regresszió együtthatói már adottak, akkor az eredményt vagy megoldási valószínűséget a részképességek alapján lehet kiszámítani a regressziós egyenlettel a részképességek behelyettesítése során. A legtöbb esetben ez egy megfelelő közelítést fog adni a pontszámra, vagy a feladat teljesítésének valószínűségére.

Azonban lehetnek olyan összetett feladatok, amelyekben a részképességek és a feladatra kapott pontszám nem állnak lineáris kapcsolatban. Például vegyünk egy olyan vizuális memóriafeladatot, ahol a résztvevőknek egy sor képet kell megjegyezniük, majd később felidézniük őket. A feladatot úgy állíthatjuk össze, hogy az első néhány kép megjegyzése viszonylag egyszerű, mivel azok hasonlóak vagy egyszerű minták. Ebben az esetben a rövidtávú memória és a vizuális feldolgozás alapvető szintje elegendő lehet a magas pontszám eléréséhez.

Ahogy a feladat nehezedik, és egyre több, összetettebb képet kell megjegyezni, a vizuális feldolgozásnak és a rövidtávú memóriának is egyre jobban kell teljesítenie a magas pontszám eléréséhez. Ebben a szakaszban a kapcsolat a képességek és a pontszám között továbbra is lineáris lehet.

Azonban elérhet egy pontot, ahol a feladat annyira bonyolulttá válik, hogy még a kiemelkedő vizuális feldolgozással és rövidtávú memóriával rendelkező egyének számára is nehéz további pontokat szerezni. Ebben az esetben a kapcsolat a képességek és a pontszám között megváltozhat, és inkább szigmoid vagy logisztikus függvényt követhet, ahol a teljesítmény növekedése lelassul, és egyre nehezebb további pontokat szerezni, még a képességek növekedése mellett is.

Erre az lehet a magyarázat, hogy a feladat egy ponton olyan összetetté válik, hogy már nem csak a vizuális feldolgozás és a rövidtávú memória a meghatározó tényezők, hanem más készségek és stratégiák is szükségesek a sikerhez. Ebben az esetben a pontszám és a részképességek közötti kapcsolat nem lineáris, és a további képességfejlesztés nem feltétlenül eredményez arányos pontszámnövekedést.

Az ilyen esetekben a pontszám vagy a megoldási valószínűség becslésére már nem hatékony megoldás a lineáris illetve logisztikus regresszió használata, hiszen ezek a definíciójuk alapján csak a lineáris kapcsolatokat képesek felismerni. Ilyen esetekben egy összetettebb modellre van szükség, amelyre tökéletes választás a mesterséges intelligencia. Felismertem, hogy egy megfelelően paraméterezett neurális hálózat képes szinte bármilyen összetett kapcsolatot azonosítani és megtanulni, így nem kell előre tudnunk, hogy a részképességek és az eredmény kapcsolata milyen matematikai függvénnyel írható le.

Azonban a neurális hálózat megfelelően paraméterezve, és optimális mennyiségű adattal betanítva, mind az összetett összefüggéseket, mind az egyszerű, akár lineáris kapcsolatot képes felfedezni két változó között. Mivel a részképességek és az eredmény közti kapcsolatról sok feladat esetén nem ismert, hogy lineáris-e, így a neurális hálózat minden

esetben jó választás az eredménybecslésre. Ha egy feladatnál sejtjük, hogy ez a kapcsolat lineáris, akkor a lineáris vagy logisztikus regressziót használhatjuk a felépített neurális hálózat validálására is.

#### 4.2.4. Beágyazás alkalmazása részképességek felbontására

A kitűzött célok közül a legösszetettebb a kitöltők részképességének és a feladatok részképesség-függésének vizsgálata abban az esetben, ha mindössze azokkal az adatokkal rendelkezünk, hogy mely feladatot mely személy milyen eredménnyel oldotta meg. Ebben az esetben a hagyományos statisztikai megközelítések nem képesek elkülöníteni a részképességeket a személyek illetve feladatok vizsgálata során. Hipotézisem szerint azonban ilyen esetben is használható a neurális hálózat, mely képes lehet ilyen feladatok megoldására. Ennek a speciális paraméter felbontásnak a neve neurális hálózatoknál a *beágyazás*, angolul *embedding*.

A beágyazás egy modern gépi tanulási technika, amely lehetővé teszi a magas dimenziós, ritkán előforduló adatok átalakítását sűrű és alacsonyabb dimenziós vektorokká. Ezek a vektorok képesek megragadni és kvantifikálni az adatokban rejlő összefüggéseket és mintákat. A beágyazások segítségével a modell képes a hasonló jelentésű vagy funkciójú elemeket közel helyezni egymáshoz a vektor térben, lehetővé téve ezzel a bonyolult kapcsolatok és az adatokban rejlő mintázatok felfedezését. Ezáltal a beágyazás egy megoldást jelenthet a részképességek felismerésére akkor is, ha csak minimális információ ismert a kitöltő személyekről és a feladatokról.

Amikor feladatokat és képességeket szeretnénk modellezni beágyazások segítségével, az alábbi lépéseket követhetjük:

##### Képesség vektorok

- Minden egyes képességet egy  $k$ -dimenziós vektorral ábrázolunk.
- A vektor elemei a különböző részképességeket szimbolizálják és írják le kvantitatív módon
- Például egy  $[0.8, 0.3, 0.6]$  vektor azt jelentheti, hogy a személy magas szinten rendelkezik az első képességgel, alacsony szinten a másodikkal, és közepes szinten a harmadikkal feltéve, hogy a vektor elemei mind 0 és 1 közötti valós értékek lehetnek fel. Ha ez a feltételezés a modell készítése során alapvetően nem teljesül, úgy utólag is lehetőség van a vektor elemeinek normalizálására.

##### Feladat vektorok

- A feladatokat is  $k$ -dimenziós vektorokkal ábrázoljuk.
- A vektor elemei az egyes részképességek jelentőségét vagy szükségességét mutatják az adott feladat teljesítéséhez.
- Például egy  $[0.7, 0.2, 0.9]$  vektor azt sugallhatja, hogy az első és harmadik képességre nagyobb szükség van a feladat megoldásához, mint a másodikra. Az elemek konkrét értékének értelmezéséhez az képesség vektorokhoz hasonlóan szükség van a vektor elemeinek normalizálására.

##### Kapcsolatok és összehasonlítások

- A beágyazások segítségével összehasonlíthatjuk a képesség vektorokat és a feladat vektorokat, és felfedezhető, hogy mely képességek fontosak egy adott feladathoz.

- A feladat vektorok között valamilyen hasonlósági metrika alapján meghatározható, hogy mely feladatok építenek hasonló részképességekre.

Összefoglalva, a beágyazások használata a feladatok és képességek modellezésére lehetővé teszi, hogy mélyebb betekintést nyerjünk az egyes részképességek és feladatok közötti kapcsolatokba, valamint segít azokat az eseteket azonosítani, ahol egy adott személy képességei jól illeszkednek egy adott feladathoz.

## A beágyazás használatának nehézségei

A korábbiakban láthattuk tehát, hogy milyen hasznos lehet a beágyazás a feladatok és kitöltők részképesség felbontására, azonban a módszer számos nehézséget is rejt, amelyek miatt az eredmények értelmezése nem triviális, sok esetben mélyebb pszichometriai ismereteket igényel - néhány ismert feladattal és személlyel azonban már könnyebben validálható a beágyazás eredménye.

A beágyazás alkalmazása során a következő nehézségekkel szembesülünk:

1. **A vektor elemei kontextus nélkül nem hordoznak információt.** Habár a példában minden elem 0 és 1 között volt, és úgy értelmeztük, hogy az 1 közeli érték magas szintű részképességet jelent a személyeknél a 0 közeli pedig minimálisat, erre nem alapozhatunk egy általános beágyazás során. Az értékeknek csak egymáshoz viszonyítva van értelmük, és habár az valószínű, hogy például egy személy vektorban a nagyobb értékek magasabb részképességet szimbolizálnak, ez egyáltalán nem törvényszerű, és további vizsgálatokat kíván.

Emögött a magyarázat az, hogy a gépi tanuló algoritmus nem ismeri, hogy mi a beágyazás célja, mindössze olyan paraméterekre bontja fel a feladatokat és személyek képességét, amelyek reprezentálják azt.

2. **A képesség vektor elemei nem egy részképességet reprezentálnak.** A beágyazás természetesen nem ismeri a CHC modellben definiált részképességeket, így egzaktan nem tudja megmondani, hogy melyik paraméter melyik részképességet jellemzi. A beágyazásnak meg kell mondani, hogy milyen hosszú vektorokat készítsen, tehát a vizsgált részképességek számát is meg kell határozni a módszer használata előtt. Azonban még ha pontosan határozzuk meg a részképességek számát, akkor sem biztosan jelent minden elem külön részképességet. Lehet, hogy a vizsgált feladatok szempontjából az egyik ismert részképesség egyáltalán nem releváns, így a beágyazás során a vektor semelyik eleme nem fogja azt reprezentálni. Ezzel szemben az is előfordulhat, hogy a vektornak két eleme ugyanazt a részképességet reprezentálja.

Emiatt mindenképp figyelni kell a vektor elemszámának alapos megválasztására, amiben a pedagógusok és pszichometrikusok hipotézisei és tapasztalatai is segítséget nyújthatnak. Érdeemes úgy megválasztani az vektor elemeinek számát, amennyi különböző részképesség előfordulására lehet számítani a vizsgált feladatok között - javasolt továbbá minél rövidebb vektort készíteni, ugyanis hosszabb vektorok során valószínűvé válik, hogy a beágyazás túl komplex modellt épít, és a kapott vektorok nem fogják reprezentálni a részképességeket.

Továbbá a részképesség-függések azonosításához szükség van a mintához hozzávenni egy-egy olyan feladatot, amely korábbi mérések alapján egy adott részképességre építenek. Ennek segítségével a többi feladatot viszonyíthatjuk a kijelölthöz, és így vizsgálható a feladatok adott, megnevezett részképesség-függése. Ezzel a módszerrel az előző pontban ismertetett problémát is megoldhatjuk, hiszen a részképesség érték mennyiségeket már tudjuk mihez viszonyítani.

Ehhez hasonlóan a részképesség vektor értelmezését könnyíti, ha a mintában található ismert részképességű személy, mert ezáltal a többi kitöltő részképesség értékeit lehetőség van összehasonlítani egy már bemért egyénnel. Ennek hiányában a részképességek azonosításához - vagyis a vektor értékeinek és a CHC modell szerinti részképességek megfeleltetéséhez - szintén olyan pszichometrikus segítsége szükséges, aki már jól ismeri a feladatokat.

- 3. A feladat és képesség vektor elemeinek azonosítása nem triviális.** A korábbiakból is adódik, hogy a feladat vektor és a részképesség vektor nem feltétlen áll olyan egyszerű összefüggésben egymással, hogy a feladat vektor  $i$ . eleme éppen a részképesség vektor  $i$ . eleme által mért részképesség-függést jelenti.

Például már sikerült beazonosítani, hogy a feladatvektor 2. eleme a rövidtávú memória részképességet reprezentálja, tehát ha ez a paraméter alacsony, akkor a feladat megoldásához nem szükséges jelentős rövidtávú memória, ellenben ha magas, akkor a feladat erre épít. Ez azonban nem jelenti azt, hogy egy kitöltő, akinek a részképesség vektorának 2. eleme magas, az kiváló rövidtávú memóriával rendelkezik, mert a feladat és részképesség vektorok ennél összetettebb összefüggésben is állhatnak. Ennek kiértékeléséhez vizsgálhatjuk az elkészült neurális háló együtthatóit és súlyait, vagy hagyatkozhatunk a pszichometrikusok tapasztalatára és méréseire is.

### 4.3. Módszertani kihívások és megoldások

Az adatgyűjtés és elemzés folyamatában számos olyan kihívással találkoztam, amelyek potenciálisan befolyásolhatják az eredmények megbízhatóságát és érvényességét. Ebben a részben bemutatom, hogy honnan származnak a kutatás alapját adó mérések, illetve milyen technikák alkalmazhatóak az eredmények megbízhatóságának növelésére.

#### 4.3.1. Rendelkezésre álló mérések a validációhoz

A kutatás során a rendelkezésemre állt egy nagy adathalmaz, amely általános iskolás gyermekek válaszait tartalmazta több különböző képességet vizsgáló kognitív tesztre. A tesztelés során közel 1200 gyermek oldott meg 28 különböző feladatot, így az adathalmaz összesen több mint 32 000 személy-feladat-eredmény azonosítópárost tartalmaz.

A mérést 2022 őszén végezték 5 és 10 év közötti gyermekekkel. Minden résztvevő először találkozott a tesztfeladatokkal, és azoknak a gyerekeknek az eredményét, akik korábban már töltöttek ki szinte azonos tesztet, kizárták a mérésből. A feladatok között találhatóak hasonló képességeket mérő tesztek, és teljesen eltérő tudást igénylő kérdések is.

Az adatok gyűjtését a MTA-AVKF Tanulási környezet kutatócsoport<sup>1</sup> végezte, akik segítségemre voltak a feldolgozott adatok értelmezésében is.

#### 4.3.2. Szimuláció az adatok gazdagításához

A mérésekkel analóg szimulációt is készítettem az adatok gazdagítására és az elemzések mélyebb validációjához. A szimuláció képesség és feladatparamétereket generál, majd szimulálja a feladatkitöltést. Így a mért eredményekhez hasonló formátumú adatokat kaphatunk, amelyeket a mért eredményekkel vegyesen is felhasználhatunk, így biztosítva, hogy a modellek elemzéseket kellően nagy adathalmazon tudjam betanítani. Természetesen a végeleges modelleket az éles, mért adatokon alkalmaztam és validáltam.

<sup>1</sup>Az MTA-AVKF Tanulási környezet kutatócsoport honlapja: <https://tanulas-kutatas.hu>

A szimuláció egyik előnye, hogy tetszőleges mennyiségű mérést elő tudunk állítani pillanatok alatt. Ismert, hogy a gépi tanuló algoritmusok, kiváltképp a neurális hálózat betanulásához és használatához nagy méretű adathalmazra van szükség ahhoz, hogy minél pontosabb eredményeket adjon. Szerencsére számomra rendelkezésre állt egy nagy méretűnek tekinthető valós adatokat tartalmazó adathalmaz is, azonban a szimulációval együtt az adatokat augmentálva még pontosabb eredményt kaphatunk.

A szimuláció használatának másik nagy előnye, hogy rendelkezésre áll minden háttérinformáció a kitöltőkről és a feladatokról, hiszen az adatok generálása során befolyásolhatóak a különböző paraméterek. Ez az információ természetesen nem használható fel az elemzések elvégzéséhez, hiszen azt kell feltételezni, hogy csak a feladatok megoldásából adódó adatok ismertek. Azonban felhasználható az elemzés validációjához, amelyben nagy segítséget tud nyújtani. Míg a valós adatokon végzett elemzéseket csak pszichometrikus sejtései és hipotézisei segítségével lehetséges validálni - hiszen semmilyen információval nem rendelkezünk a kitöltésből származó adatokon kívül -, addig a szimuláció megfelelően validálható a generált adatok segítségével.

Egy példán keresztül bemutatva, feltételezzük, hogy egy személy 5 különböző részképességét szeretnénk vizsgálni különböző feladatok megoldásával. A szimulációval elkészíthetők úgy az adatok, hogy egy adott feladat például az 5 közül csak 3 részképességre épít - tehát amikor a feladat megoldását szimuláljuk, a megoldás valószínűsége csak a személy ezen részképességeitől függőnek tekintjük. Ezután az eredményeket elvégezve az elemzéseket megvizsgálható, hogy az elemzés sikeresen felfedezte-e, hogy két változó egyáltalán nem játszott szerepet az adott feladat megoldásában.

Így tehát feltételezhető, hogy ha a feladatokat és személyeket célzottan, ismert összefüggések alapján generáljuk, majd a tesztmegoldást szimuláljuk és az elemzést elvégezve az eredményeken megkapjuk a kiindulásként ismert összefüggéseket, akkor valós adatokon - ahol ezek a kiinduló összefüggések nem ismertek - is helyes eredményt kapunk.

### 4.3.3. Az elemzések technológiai megvalósítása

A kutatás során az elemzések elvégzéséhez a Python programozási nyelvet használtam, mely egy nagyon rugalmas és sokoldalú eszköz az adattudomány terén. A Python kínálta funkcionalitások lehetővé tették, hogy hatékonyan kezeljem és elemezzem a gyűjtött adatokat.

Az adatelemzések alapját a pandas<sup>2</sup> könyvtár adta, mely egy nyílt forráskódú adatelemzési és -manipulációs eszköz. A pandas lehetővé tette számomra, hogy könnyen és gyorsan hajtsam végre a különböző adatumveleteket, mint például az adatok importálása, tisztítása és transzformálása. A könyvtár adattáblák és kezelésére optimalizált adatszerkezeteket kínál, amelyek nagyban megkönnyítették a munkámat.

A neurális hálózatok fejlesztése és tanítása során a TensorFlow és a Keras keretrendszereket használtam. A TensorFlow egy nyílt forráskódú gépi tanulási platform, amely lehetővé teszi a magas szintű neurális hálózatok fejlesztését és tanítását. A Keras egy magas szintű neurális hálózat API, amely a TensorFlow-on belül is használható, és egyszerűsíti a mélytanulási modellek létrehozását és tesztelését. A Keras egyszerű, intuitív programozási interfészt kínál, amely nagyban megkönnyítette a neurális hálózatok fejlesztését és optimalizálását.

A TensorFlow és a Keras alkalmazása lehetővé tette számomra, hogy hatékonyan hajtsam végre a neurális hálózatok tanítását és értékelését, valamint hogy mélyebb betekintést nyerjek az adatok összefüggéseibe. A technológiai eszközök kiválasztása és alkalmazása fontos része volt a kutatás előkészületének, hiszen alapvetően befolyásolja az elemzések eredményét és sikerességét.

---

<sup>2</sup><https://pandas.pydata.org/>

## 5. fejezet

# Új modellek kialakítása és kiértékelése

Az előző fejezetben ismertettem a kutatás céljait, és a megvalósításához szükséges elméleti illetve módszertani problémákat és megoldásokat. Ebben a fejezetben bemutatom az elkészített modellek kiértékelését, illetve az elemzések eredményeit, különös hangsúlyt fektetve a szimuláció során a paraméterek beállítására, illetve az elemzés valós adatokon végzett validálására.

### 5.1. Szimuláció elkészítése és kiértékelése

A következő részben bemutatom hogyan működik a szimuláció, illetve milyen eredményeket sikerült elérni a szimuláció által augmentált adatokon.

#### 5.1.1. Szimuláció paramétereinek meghatározása

A szimuláció során átlagosan  $k = 5$  részképességet különböztettem meg, ám ez a paraméter szabadon változtatható. Azért döntöttem emellett, mert a vizsgált mérés során a feladatokban körülbelül 5 különböző részképességet terveztem megkülönböztetni, amelyek jelentős szerepet töltenek be a feladatok megoldásában.

A szimuláció során először elkészítettem a személyek részképesség értékeit, melyeket úgy határoztam meg, hogy minden részképesség átlaga a személyek között  $\mu = 1$  legyen  $\sigma^2 = 0,15$  szórással. Azért választottam ezeket az értékeket, mert a legtöbb kognitív teszt, köztük a hagyományos IQ teszt is úgy van kialakítva, hogy ilyen módon pontozza a képességeket (az IQ teszt esetében 100-zal felszorozva). A különböző részképességeket egymástól független változóknak tekintettem, hiszen éppen ez a célja a részképességek vizsgálatának. Az lehetséges, hogy amit CHC modell több részképességként vizsgál, azt az elemzés során egy paraméter fogja jellemezni, azonban egy részképességet nem szeretnénk több egymástól nem független változóra bontani.

Egy személy paramétereit tehát például a következő vektorral írható le:

$$\underline{\beta} = (1,025 \quad 0,896 \quad 0,962 \quad 0,868 \quad 1,102)$$

Ez azt jelenti, hogy a vizsgált személynek átlag feletti az első és utolsó paraméterrel jelzett részképessége - ez utóbbi kimondottan magas értékű, azonban átlag alatti a többi részképessége, melyek közül a negyedik a leggyengébb.

A feladatok generálására szintén minden feladathoz  $k$  paramétert generáltam. Legegyszerűbb esetben csak  $k$  db bináris értéket rendeltem a feladatokhoz, amely azt jelképezte, hogy az adott részképesség szükséges-e a feladat megoldásához, azaz befolyásolja-e



a megoldás valószínűségét. A feladatoknál kikötöttem, hogy minden feladat megoldásához legalább egy részképesség szükséges. Később a bináris érték helyett egy folytonos változót használtam, amely hasonlóan azt jelképezte, hogy mekkora szerepet játszik a feladat megoldásában, azaz milyen magas szinten szükséges az adott részképesség a feladat megoldásához. Ilyenkor a feladat nehézség paramétereit egy-egy 0 és 1 közötti szám reprezentálta.

Egy feladat paramétereit a bináris változatban tehát például a következő vektorral írható le:

$$\underline{\theta} = (0 \quad 1 \quad 0 \quad 0 \quad 1)$$

Ez azt jelenti, hogy a vizsgált feladathoz a második és az ötödik részképesség szükséges. Egy személy akkor tudja tehát megoldani ezt a feladatot, ha ezek a részképességei meghaladnak egy bizonyos szintet.

A szimuláció során olyan feladatokat is készítettem, amelyek több egymás utáni kérdést foglalnak magukban. Ezek a részfeladatok ugyanazokat a képességeket mérik, azonban egyre nehezednek. Az ilyen feladatcsoportokat a szimuláció során egy tesztként vizsgáltam, az eredménynek pedig a részfeladatokon elért eredmények összegét tekintettem. A nehézségi paramétert először lineárisnak állítottam be, majd később exponenciálisan nehezedő paramétereket is vizsgáltam. Hipotézisem szerint az eredménybecslés során nem lineáris paraméterek esetén a neurális háló lényegesen jobb becslést ad mint a lineáris regresszió.

### 5.1.2. A szimuláció lefuttatása

A szimuláció során minden személy-feladat párosra meghatározom, hogy az adott személy meg tudja-e oldani az adott feladatot. Egyszerű esetben csak egy bináris - sikeresen megoldotta vagy nem - értéket rendeltem hozzá a személy-feladat pároshoz, a későbbiekben pedig egy folytonos változót, amely a pontszámot reprezentálta - tehát milyen eredménnyel oldotta meg a személy a feladatot.

A következőkben ismertetem, hogy a különböző változatú szimulátorok milyen módon számolják ki, hogy adott személy adott feladatot milyen eredménnyel old meg:

- **Bináris feladat paraméter és bináris eredmény vizsgálat** esetén azt mondjuk, hogy a feladatot sikeresen megoldotta a kitöltő, ha minden olyan részképessége, amely a feladat szempontjából releváns egy adott  $t$  küszöbérték felett van. Ezt matematikailag a következőképpen tudjuk kiszámítani:

$$A1 = \bigwedge_{i=1}^k \begin{cases} \text{Igaz} & \text{ha } \theta_i = 0, \\ \beta_i > t \pm \text{uniform}(-\epsilon, \epsilon) & \text{egyébként.} \end{cases}$$

ahol  $k$  a részképességek száma,  $\theta$  a személy részképesség vektora,  $\beta$  a feladat nehézség vektora és  $A$  egy bináris változó mely meghatározza, hogy a személy sikeresen megoldotta-e az adott feladatot. Ezen kívül a küszöbértéket minden kitöltés esetén módosítom egy kis  $-\epsilon$  és  $\epsilon$  közötti véletlenszerű értékkel, hiszen a való életben is minden feladat megoldás során előfordulnak olyan tényezők, amelyek kiszámíthatatlanok. Ehhez az  $\epsilon = 0,01$  értéket használtam.

Tehát egy személy akkor old meg helyesen egy feladatot, ha minden részképességre a feladat adott részképességhez tartozó paramétere 0, vagy a személy adott részképességhez tartozó értéke meghaladja a küszöbértéket. A küszöbértéket minden esetben úgy határoztam meg, hogy a generált személy és feladat párosítások körülbelül fele végződjön sikeres feladatmegoldással. Ebben az egyszerű esetben az korábban meghatározott részképesség átlaggal és szórással ez a küszöbérték  $t = 0,9$  lett.

- **Folytonos feladat paraméter és bináris eredmény vizsgálat** esetén akkor mondjuk, hogy a feladatot sikeresen megoldotta a kitöltő, ha minden részképességre a személy képesség paramétere meghaladja a feladat nehézség paraméterétől függő küszöbértéket. Ezt a következőképpen definiáltam:

$$A2 = \bigwedge_{i=1}^k \beta_i > \lambda \cdot t(\theta_i) \pm \text{uniform}(-\epsilon, \epsilon),$$

$$t(\theta_i) = (\mu - 2\sigma^2) + \theta_i \cdot 4\sigma^2,$$

ahol a pszichometriában bevett eljárások alapján  $\theta$  a személy részképesség vektora,  $\beta$  a feladat nehézség vektora, továbbá  $k$  a részképességek száma,  $\mu$  és  $\sigma^2$  rendre a vizsgált személyek részképességeinek átlaga és szórása,  $\lambda$  egy korrekciós tényező,  $A$  pedig egy bináris változó, mely meghatározza, hogy a vizsgált személy sikeresen megoldotta-e az adott feladatot. A küszöbértéket ebben az esetben is minden kitöltés esetén módosítom egy kis  $-\epsilon$  és  $\epsilon$  közötti véletlenszerű értékkel. Ehhez szintén az  $\epsilon = 0,01$  értéket használtam.

A küszöbérték függvény a vizsgált feladat adott részképesség-függésének értéke alapján meghatározza, hogy a személy részképesség értékének minimum mekkorának kell lennie ahhoz, hogy sikeresen megoldja a feladatot. Ehhez a 0 és 1 közötti feladat nehézség értéket átskálázza úgy, hogy az átlag 2 szórásnyi környezetében legyen. Azért választottam ezt az értéket, mert ebbe az intervallumba esik a részképesség értékek jelentős része, a többit kiugró értéknek tekintem. Azonban annak érdekében, hogy ebben az esetben is nagyjából a generált személy és feladat párosítások körülbelül fele végződjön sikeres feladatmegoldással, szükség van még egy korrekciós tényezőre is. Ezt a tényezőt tapasztalatok alapján  $\lambda = 0,8$ -nak állítottam be.

- **Folytonos feladat paraméter és folytonos eredmény vizsgálat** esetén minden személy és feladat pároshoz egy 0 és 1 közötti pontszámot rendelünk. Ehhez először a személy részképesség paramétereit át kell skálázni 0 és 1 közé, olyan módon, hogy ez a paraméter reprezentálja a személy pontszámát abban az esetben, ha a feladat csak az adott részképességtől függene teljesen. Majd ezeket a részpontszámokat átlagolom a részképesség jelentőségének arányában az adott feladatra. Ezt a következőképpen számítom ki:

$$a = \sum_{i=1}^k f(\beta_i \pm \text{uniform}(-\epsilon, \epsilon)) \cdot \frac{\theta_i}{\max(\theta)},$$

$$f(x) = \begin{cases} 0 & \text{ha } f'(x) < 0, \\ 1 & \text{ha } f'(x) > 1, \\ f'(x) & \text{egyébként.} \end{cases}$$

$$f'(x) = \frac{x - (\mu - 2\sigma^2)}{4\sigma^2},$$

ahol  $k$  a részképességek száma,  $\theta$  a személy részképesség vektora,  $\beta$  a feladat nehézség vektora,  $\mu$  és  $\sigma^2$  rendre a vizsgált személyek részképességeinek átlaga és szórása,  $a$  pedig egy a személy feladatra kapott pontszáma, egy 0 és 1 közötti valószínű szám. Ebben az esetben a személy paramétereit módosítom minden kitöltés esetén egy kis  $-\epsilon$  és  $\epsilon$  közötti véletlenszerű értékkel, annak érdekében, hogy néhány véletlen faktorial számoljak. Ehhez szintén az  $\epsilon = 0,01$  értéket használtam.

A fenti összefüggések összetettnek tűnnek, azonban nagyon egyszerű logika található mögötte. Az  $f(x)$  függvény felel az adott rész-képesség érték fent említett skálázásáért. A függvény 0 és 1 közé skáláz, azaz normalizál olyan változókat, melyeknek az átlaga  $\mu$  a szórása pedig  $\sigma^2$ . Ehhez az  $f'$  függvényt vesszük segítségül, amely normalizál egy tetszőleges valós számot  $\mu - 2\sigma^2$  és  $\mu + 2\sigma^2$  között. Mivel tudjuk, hogy egy sokaság elemeinek a nagy része beleesik az átlag 2 szórás körüli intervallumába, ezért a legtöbb rész-képesség értéket ez a függvény 0 és 1 közé transzformálja. Azonban lehetnek kiugró értékek, ezek esetében 0-nál kisebb vagy 1-nél nagyobb eredményt ad az  $f'$  függvény. Ezt nem fogadhatjuk el pontszámoknak, ezt korrigálja az  $f$  függvény olyan módon, hogy ha az átskálázott érték 0 alatt van, akkor a személy pontszáma 0, ha pedig 1 felett, akkor 1. Ez éppen azt jelenti, hogy ha a személy rész-képessége az átlagtól két szórásnál nagyobb mértékben tér el negatív irányban, akkor a személy nem fogja tudni megoldani azt a feladatot, amihez az a rész-képesség szükséges, ha pedig pozitív irányban, akkor a személy hibátlanul megoldja azt.

Ez az összefüggés azt feltételezi, hogy egy személy rész-képességei és a feladatra kapott pontszáma lineáris összefüggésben áll egymással, hiszen a pontszámot a rész-képességek alapján kiszámolt pontszámok súlyozott átlaga adja. Ez a feltétel azonban sokszor nem áll fenn. A legegyszerűbb esetben például azt mondhatjuk, hogy a pontszám, és a kiszámított súlyozott átlag közötti összefüggés egy szigmoid alakú függvényt vesz fel. Ezt az indokolhatja, hogy a pontszám nem lineárisan nő ahogyan a képességek nőnek, hanem egy adott képesség értékig szinte 0 pontszámot érnek el a kitöltő, egy adott képesség érték felett pedig hibátlanul megoldják azt. A szimuláció elkészítése során ilyen és ehhez hasonló nem lineáris pontozással is kísérleteztem, hogy megtudjam, az ilyen összefüggéseket melyik modell milyen pontossággal tudja felismerni.

A szimuláció során tehát ezekkel a módszerekkel döntjük el, hogy egy személy milyen választ ad egy adott feladatra, azonban még az összetettebb változatok is figyelmen kívül hagynak pár olyan tényezőt, amely a valós életben jelentősen tudja befolyásolni az eredményt. Egyre több gyermek rendelkezik valamilyen tanulási nehézséggel[8], amely sokszor nehezen megmagyarázható módon befolyásolja a feladatmegoldás készségét[24]. Emiatt az elemzéseknek többek között azt is kell kezelni, hogyha egy-egy kitöltő néhány feladat során egyáltalán nem a képességeinek megfelelően válaszol. Például, ha egy ADHD-s tanuló észrevesz egy mentőt elszáguldani az ablak alatt a teszt kitöltése közben, az nagy valószínűséggel az ADHD-val nem küzdőkhöz képest jelentősen negatívabban fogja befolyásolni a következő feladatokra adott választát.

Emiatt a szimulációba beleépítettem, hogy néhány tanuló esetében, néhány tesztmegoldás során a képességétől teljesen függetlenül gyengén fog teljesíteni. Ezáltal azt is tudjuk validálni, hogy ha egy-egy váratlan eredményt is tartalmaznak az adatok, azt hogyan tudja kezelni az elemzés.

A fent leírt módszerekkel tehát szimulálni tudjuk egy adott rész-képességű személy választát egy olyan feladatra, amelynek ismert a rész-képesség-függése. Fontos megemlíteni, hogy maga a konkrét feladat az elemzés szempontjából lényegtelen, hiszen egy absztrakt modellt építünk az összetett feladatok értékelésére - éppen az a cél, hogy a feladat ismerete nélkül, csak a kitöltések alapján tudjunk minél több információt meghatározni a feladatról és a rész-képesség-függéséről. Éppen e tulajdonság miatt készíthető szimuláció az adatok gazdagítására.

## 5.2. Az új megközelítés tesztelése és validálása

Az új megközelítés segítségével olyan eredményt sikerült elérnem a kutatásom során, amelyre korábban még kutatásban részt vevő pszichometrikus és pedagógus társkutatók szerint sem volt példa. Ebben a részben áttekintem az előző fejezetben kitűzött célokat, és bemutatom a rájuk készített megoldásom eredményeit. Kitérek arra, hogyan lehetséges ezeket az eredményeket validálni és milyen új lehetőségeket biztosítanak a kutatási területen.

A modellek kiértékeléséhez rendelkezésemre állt egy 2022 őszén készült mérésből származó nagy adathalmaz, amelyet a korábbi fejezetben ismertettem. A következőkben bemutatok a méréshez használt feladatok közül néhányat, hiszen az eredmények kiértékeléséhez és megértéséhez ez kulcsfontosságú lesz.

### 5.2.1. A tesztfeladatok bemutatása

A mérésben használt tesztfeladatok azonosítója és neve az 5.1-es táblázatban található. A feladatok közül azokat, amelyek elemzésre kerülnek a következőkben részletesen bemutatom. A többi feladatról részletesebb információ a kutatócsoport honlapján<sup>1</sup> található, illetve a feladatok leírása megtalálható a Kognitív Profil<sup>2</sup> weboldalon, ahol a tesztek ki is próbálhatóak.

Kód	Feladat címe
55	Figurák
56	Számsorozat visszafelé
64	Számismétlés
84	Képfelismerés
86	Go - no go feladat
89	Ritmustartás II. - egyenletes
90	Kopogtatás
94	Szenzomotoros vizsgálat - Testrészek azonosítása
95	Szenzomotoros vizsgálat - Ujjak azonosítása
96	Szenzomotoros vizsgálat - Jobb és bal kéz felismerése
97	Szenzomotoros vizsgálat - Irányok azonosítása
98	Szenzomotoros vizsgálat - Környezeti tárgyakkal való viszonylat
99	Szenzomotoros vizsgálat - Alak-háttér észlelése
100	Szenzomotoros vizsgálat - Szemmozgás irányítása
101	Szenzomotoros vizsgálat - Zenei hang észlelése
102	Szenzomotoros vizsgálat - Téri szekvencia
103	Szenzomotoros vizsgálat - Idői szekvencia
104	Szenzomotoros vizsgálat - Nyelvi szekvencia 1.
105	Szenzomotoros vizsgálat - Nyelvi szekvencia 2.
106	Szenzomotoros vizsgálat - Fékezős feladat
107	Állatok megtalálása sorrendben
108	Beszédértés
109	Mennyiségek összehasonlítása
110	IT kompetenciák II. (drag 'n' drop)
111	IT kompetenciák I. (touch)
112	Színes Raven teszt
113	Egyensúly nyitott szemmel
114	Beszédhangok megkülönböztetése
116	Egyensúly csukott szemmel

5.1. táblázat. A feladatok listája

<sup>1</sup><https://tanulas-kutatas.hu>

<sup>2</sup><https://kognitivprofil.hu>

A következőkben pár feladat részletesebb leírása:

- **56 - Számsorozat visszafelé**

A *Számsorozat visszafelé* feladatban egy hanganyag alapján kell leírni a hallott számokat fordított sorrendben. A számok leírása csak a számsor elhangzása után kezdődhet meg. A feladat több fordulóból áll, kezdetben kettő szám hallható, ami a feladat végére hatra nő.

- **64 - Számisméltés**

A *Számisméltés* feladatban meg kell jegyezni a hallott számokat, majd a hanganyag vége után leírni azokat. A feladat több fordulóból áll, kezdetben három szám hallható, ami a feladat végére hétre nő.

- **84 - Képfelismerés**

A *Képfelismerés* feladatban három rajz rajzolódik ki folyamatosan olyan módon, hogy folyamatosan egyre több vonal jelenik meg a képeken. Azonban a három kép közül csak egy ábrázol ténylegesen valamit, a másik kettő csak vonalakat tartalmaz. A feladat minél korábban felismerni, hogy melyik az a rajz, amelyik valós dolgot ábrázol. Ez a feladat is több fordulóból áll.

- **94 - Szenzormotoros vizsgálat: Testrészek azonosítása**

A *Testrészek azonosítása* feladatban emberekről illetve állatokról láthatóak képek, és minden fordulóban be kell jelölni egy-egy testrész helyzetét az adott képen.

### 5.2.2. Faktoranalízis regresszió segítségével

Ahogy a korábbi fejezetben bemutattam a faktoranalízist bináris értékelésű feladatok során logisztikus regresszió segítségével, pontozott feladatok esetén pedig lineáris regresszióval valósítottam meg. Az elemzés során a célom az volt, hogy abban az esetben, amikor ismert a személyek részképessége és a teszteken elért eredményei, meghatározzam, hogy az adott feladatokhoz mely részképességek milyen mértékben szükségesek.

A mérés adatait augmentáltam a szimuláció segítségével új feladatok és személyek generálásával, majd a feladatmegoldás szimulálásával. Ezekben az esetekben számomra ismert a feladatok részképesség függése, így az eredmény validációja egyszerűen megoldható, szemben a mérésből származó adatokkal, amelyek helyessége hipotézisek illetve pszichometrikusok segítségével határozható meg.

A validáció a szimuláció segítségével sikeresnek bizonyult. A generált feladatokra a regresszió paraméterei a pszichometriai mérések esetén megszokott pontossággal visszaadták a feladatok részképesség-függését. Az alábbiakban bemutatom, hogy az egyes szimulátorokkal készült eredmények alapján milyen pontossággal ismerte fel a modell a feladatok részképesség függését.

### Bináris feladat paraméter és bináris eredmény vizsgálat

Az 5.2-es táblázatban a bináris feladatvektorokat, illetve bináris eredményeket alkalmazó szimuláció eredménye látható. Ebben az esetben a modell által kapott együtthatók tetszőleges valós értéket felvehettek, ezért annak érdekében, hogy összehasonlíthatóak legyenek a bemeneti feladat vektorral, az értékeket normalizáltam a szigmoid függvény segítségével, amely a  $(-\infty; \infty)$  intervallumot képezi le a  $[0; 1]$  intervallumra. Ahol a normalizált eredmény meghaladja a 0,99999 értéket, azt már 1-nek vettem.

Feladat vektor	Becsült feladat vektor				
(1 0 1 0 0)	(1,0	0,782	1,0	0,566	0,636)
(0 1 1 0 1)	(0,571	0,9997	0,99994	0,8124	0,99992)
(1 1 0 1 0)	(0,9997	0,9997	0,4564	0,9996	0,6069)
(1 0 0 0 0)	(1,0	0,3856	0,1207	0,8709	0,704)

**5.2. táblázat.** Vektorok összehasonlítása 1.

Az eredmények alapján jól látható, hogy a modell szinte tökéletesen felismeri az összefüggéseket. Például, ha a értékhatárt, ami felett azt mondjuk, hogy a részképeség beleszámít a feladatba 0,9-nél húzzuk meg, akkor a bemutatott mintán a modell tökéletesen működik. A vizsgálataim szerint azonban egy ideális határ a 0,98 lehet - a bemutatott mintán is minden olyan részképeség értéket, amely valóban számított a feladat megoldásában 0,99 feletti értékre becsült a modell. Elsőre nagyon magasnak tűnhet ez a határ, azonban ez a szigmoid függvény természete miatt alakult így. Egy alkalmas másik függvény választásával ez orvosolható lehetne - azonban erre nincs szükség, hiszen így is működik a modell egy magas határt választva.

### Folytonos feladat paraméter és bináris eredmény vizsgálat

A 5.3-es táblázatban láthatóak a folytonos feladat vektorok és a lineáris regresszió segítségével becsült feladat vektorok abban az esetben, amikor a feladat eredménye bináris. A sorokban a felső vektor a szimulációval generált valódi érték, az alsó pedig a regresszió során becsült érték. Ahhoz, hogy a két vektort ténylegesen össze tudjuk hasonlítani a vektorok értékeit át kellett alakítani oly módon, hogy a kapott vektorok értékeit először normalizáltam 0 és 1 közé, majd mindkét esetben azt vizsgáltam, hogy az adott részképeség érték mekkora részét teszi ki az összes szükséges részképeségek. Ezt azáltal tudtam elérni, hogy minden részképeség értéket leosztottam a vektorban lévő részképeség értékek összegével. Így tehát mindkét vektorban az látható, hogy az adott részképeség mekkora arányban szükséges a feladat megoldásához. Így precízen össze lehet hasonlítani, hogy a feladat részképeség-függését milyen jól becsüli a lineáris regresszió.

Feladat vektor és becsült feladat vektor	Távolság
(0,2098 0,2087 0,3060 0,0539 0,2215)	0,0721
(0,2199 0,2184 0,2203 0,1230 0,2184)	
(0,1760 0,2560 0,1052 0,2893 0,1736)	0,0869
(0,1952 0,2138 0,1713 0,2138 0,2059)	
(0,3100 0,0538 0,1950 0,2849 0,1563)	0,1419
(0,2325 0,0918 0,2264 0,2325 0,2168)	
(0,1462 0,2460 0,1432 0,3026 0,1619)	0,0847
(0,1952 0,2027 0,1966 0,2040 0,2015)	
(0,1957 0,1952 0,3303 0,2169 0,0620)	0,1085
(0,2167 0,2225 0,2277 0,2257 0,1074)	

**5.3. táblázat.** Vektorok összehasonlítása 2.

Az eredményekből azt láthatjuk, hogy ebben az esetben is egy nagyon jó közelítést ad az algoritmus, szinte minden részképesség értékre. A táblázatba felvettem a két vektor euklidészi távolságát is. Az euklidészi távolság két normalizált vektor között kerekítve  $[0; 1, 4142]$  intervallumban vesz fel értéket. Ebben az intervallumban a kapott távolságok értéke jó becslésre enged következtetni.

### Folytonos feladat paraméter és folytonos eredmény vizsgálat

A 5.4-es ábrán láthatóak a folytonos feladat vektorok és a becsült feladat vektorok azokban az esetekben, ahol a feladatra kapott pontszámot vizsgáltuk. A vektorokat az előző táblázathoz hasonlóan ábrázoltam, a valós és becsült vektorok euklidészi távolságával egyetemben.

Feladat vektor és becsült feladat vektor	Távolság
(0.1255 0.1182 0.3361 0.3290 0.0912) (0.1837 0.1812 0.2332 0.2322 0.1697)	0.1830
(0.2172 0.1471 0.1293 0.3031 0.2033) (0.2070 0.1869 0.1801 0.2222 0.2038)	0.1040
(0.2812 0.0673 0.1615 0.2520 0.2380) (0.2215 0.1545 0.1936 0.2168 0.2136)	0.1184
(0.3247 0.0827 0.3013 0.1163 0.1750) (0.2299 0.1639 0.2269 0.1791 0.2003)	0.1603
(0.2295 0.1312 0.1257 0.3495 0.1641) (0.2126 0.1826 0.1805 0.2296 0.1946)	0.1457

5.4. táblázat. Vektorok összehasonlítása 3.

Látható, hogy ennél a vizsgálatnál már nagyobb eltérések vannak a lineáris regresszió becsülése és a valós érték között - azonban ezek a távolságok sem túl nagyok, tekintve, hogy a maximum távolság körülbelül 1, 4142.

### 5.2.3. Eredménybecslés regresszóval és neurális hálóval

A következő részben bemutatom, hogy milyen pontossággal sikerült megbecsülnöm különböző módszerek segítségével azt, hogy egy adott részképességű ember egy tetszőleges feladatot, amelyet nem ismerünk milyen eredménnyel fog megoldani - feltéve, hogy ismert számos személy a részképesség értékeikkel, és az eredményük egy adott feladaton. Az elért eredményt egyszerűbb esetekben az előző részben bemutatott logisztikus illetve lineáris regresszió egyenletével is lehetséges becsülni, azonban összetettebb esetekben hipotézisem szerint a neurális hálózat sokkal jobb eredményt ad.

Ahhoz, hogy az eredménybecslés pontosságát mérjük az adatokat szét kell választani tanító és teszt adatokra. A tanító adatokon kell felépíteni a modellt, majd a teszt adatokon kiértékelni. Ezáltal objektíven tudjuk értékelni a becslés pontosságát, hiszen a teszt adatok a tanítás során nem látott mintákat tartalmaznak. A következőkben minden esetben egy külön tanító halmazon építettem fel a modelleket, és egy teszt halmazon értékeltem ki.

A becslés helyességének ellenőrzéséhez továbbá szükség van egy metrikára, ami szerint összehasonlíthatjuk a becsült és valós adatokat. A metrika megválasztása kulcsfontosságú a kiértékelésben, ugyanis rossz választással könnyen hihetjük azt, hogy sikeres becslést végeztünk miközben ez nem így van, csak a vizsgálat során olyan módon hasonlítottuk össze a két eredményt, amely az adott adatokra nem alkalmazható helyesen. Például a

pontosság (*accuracy*) metrika alapján - amely a helyesen megbecsült értékek arányát jelenti - tévesen következtethetünk, ha az elemek többsége az egyik osztályból származik. Ebben az esetben mindig a többségi osztályt becsülve magas pontosságot kapunk, pedig a modell célja az is, hogy felismerje mindkét osztálybeli elemeket.

Abban az esetben, amikor bináris eredményt kell becsülni, a becslés helyességének mérésére egy összetett mutatót az *F1 Score*-t alkalmaztam.

Az F1 Score egy statisztikai mérőszám, amely a modell precizitását (*precision*) és a visszahívást (*recall*) egyensúlyozza, ezáltal olyan helyzetekben is használható, ahol az osztályok egyenlőtlenül vannak elosztva. Az F1 Score a precízió és a visszahívás harmonikus közepét számolja ki, így segít elkerülni azokat a helyzeteket, ahol egy magas pontosság mellett alacsony visszahívás (vagy fordítva) megtévesztően jó teljesítményt mutatna.

Az F1 Score kiszámítása a következőképpen történik:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

ahol:

- A Precision:  $\frac{\text{Igaz Pozitív}}{\text{Igaz Pozitív} + \text{Hamis Pozitív}}$
- A Recall  $\frac{\text{Igaz Pozitív}}{\text{Igaz Pozitív} + \text{Hamis Negatív}}$

Az *Igaz Pozitív* azoknak az eseteknek a száma, amikor a modell helyesen azonosított egy pozitív esetet, a *Hamis Pozitív* azoknak az eseteknek a száma, amikor a modell tévesen azonosított egy pozitív esetet, míg a *Hamis Negatív* azoknak az eseteknek a száma, amikor a modell nem ismert fel egy valódi pozitív esetet.

Tehát a precízió azt jelenti, hogy az pozitívnek becsült eredmények mekkora része volt ténylegesen pozitív, a visszahívás pedig azt, hogy a ténylegesen pozitív eredmények közül hányat sikerült felismerni.

Az F1 Score értéke 0 és 1 közötti, ahol 1 a tökéletes pontosságot és visszahívást jelenti, míg 0 azt jelenti, hogy a modell nem teljesítette az elvárásokat.

Azokban az esetekben, ahol folytonos eredményt, vagyis pontozott feladatot kell becsülni, a becslés helyességének mérésére az úgynevezett *Mean Squared Error* metrikát alkalmaztam.

A Mean Squared Error - továbbiakban MSE - egy hibamérőszám, amely egy regressziós modell becsléseinek pontosságát méri. Az MSE kiszámítása során a modell által előrejelzett értékeket összehasonlítjuk a tényleges értékekkel, és az MSE-t a hibák négyzetének átlaga adja.

Az MSE képlete a következő:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

ahol:

- $n$  a mintaelemek száma
- $y_i$  az  $i$ -edik tényleges érték
- $\hat{y}_i$  az  $i$ -edik becsült érték

Az MSE értéke mindig pozitív, és minél kisebb, annál jobb a modell teljesítménye. Ha az MSE értéke 0, az azt jelenti, hogy a modell minden esetben tökéletesen megjósolta a tényleges értékeket.



Az MSE érzékeny a különbségek négyzetes értékére, ami azt jelenti, hogy a nagyobb hibák nagyobb súlyt kapnak, mint a kisebb hibák. Ez azt is jelenti, hogy az MSE értéke nagyobb lehet a szélsőséges hibák miatt, még akkor is, ha a modell többi előrejelzése pontos.

Ezenkívül az MSE értékének értelmezését nehezíti, hogy a hibák négyzetét veszi figyelembe. Az egyik módja annak, hogy ezt kezeljük, az RMSE (Root Mean Squared Error) használata, amely az MSE négyzetgyökét veszi, így az értéke a tényleges adatok mértékében lesz.

### **Eredménybecslés logisztikus regresszióval**

Azoknál a feladatoknál, ahol az eredmény bináris volt logisztikus regresszióval lehetséges megbecsülni azt. A becslés helyességét a korábban bemutatott F1 Score segítségével értékeltem ki. Ahol a feladat vektort is binárisnak tekintettük - tehát egy részképeség vagy szükséges a feladat megoldásához, vagy nem -, ott az F1 Score értéke 0,8 körül mozgott. Ahol a feladatok részképeség-függését egy folytonos skálán vizsgáltuk, azokban a esetekben pedig az F1 Score átlagosan 0,85 volt. Ezek az értékek jó becslésekre engednek következtetni, tekintve az F1 Score összetettségét.

### **Eredménybecslés lineáris regresszióval**

Azoknál a feladatoknál, ahol az eredmény egy folytonos változó volt, vagyis a feladatra kapott pontot vizsgáltuk, a becslés lineáris regresszióval végezhető. A becslés helyességét a korábban bemutatott MSE illetve RMSE metrika segítségével határoztam meg.

A legegyszerűbb esetben, ahol a részképeségek és az eredmény között lineáris függés állt fenn - tehát az egyes részképeségek szerint külön került pontozásra a feladat, majd a pontszámok átlagát véve kaptuk az eredményt - a lineáris regresszió szinte tökéletes becslést ad a pontszámra. Ebben az esetben az MSE értéke szinte minden mérés során a  $[0,0001; 0,0002]$  intervallumba esett, ami az RMSE értékét tekintve a  $[0,01; 0,014]$  intervallumot jelenti. Ezek az értékek nagyon jó becslésre engednek következtetni, tekintve, hogy a becült értékek a  $[0; 1]$  intervallumba tartoznak. Ez nem meglepő eredmény, hiszen ismert volt, hogy a részképeségek és az eredmény között lineáris kapcsolat áll fenn, ezekben az esetekben pedig a lineáris regresszió alkalmazása igen pontos eredményt ad.

Azokban az esetekben, ahol a feladatokra kapott pontszám és a személy részképesége nem lineáris összefüggésben áll a korábban említett hipotézisem szerint a lineáris regresszió nem ad megfelelő eredményt. A mérések során ez a feltétel igazolódni látszik mind azoknál a feladatoknál, amelyeknél ez az összefüggés egy szigmoid görbét követ, mind azoknál, amelyeknél több egyre nehezedő feladat összpontszáma adja az eredményt. Ezeknél az eseteknél a lineáris regresszió becslésének MSE értéke a méréseim alapján 0,02 és 0,08 között mozgott, attól függően, hogy a kiértékelt feladatok milyen típusúak voltak. Ez nagyjából 0,14 és 0,28 közötti RMSE értéket jelent, amely már viszonylag nagy hibára utal, hiszen a becült érték 0 és 1 között mozog. (A körülbelüli értékek a 5.5-ös táblázatban találhatóak meg, ahol összehasonlítottam az eredményeket a neurális hálózat becslésének helyességével.)

### **Eredménybecslés neurális hálózattal**

A korábbi részben bemutattam, hogy a lineáris regresszió nem ad megfelelő becslést az eredményre abban az esetben, ha a részképeségek és a feladtra kapott pontszám nem lineáris összefüggésben áll. Hipotézisem szerint ezekben az esetekben az eredmény a lineáris regressziónál jobb közelítéssel becsülhető neurális hálót alkalmazva.

Az eredménybecsléshez egy egyszerű, többrétegű perceptron szerkezetű neurális hálózatot készítettem, amelynek két rejtett rétege és egy kimeneti rétege van. A rejtett rétegek mindegyike 16 neuronból áll, és a *ReLU* (Rectified Linear Unit) aktivációs függvényt használják. A *ReLU* függvény nemlineáris függvény, amely segít a hálózatnak abban, hogy komplex mintázatokat tanuljon meg az adatokból.

A kimeneti réteg egyetlen neuronból áll, és nincs aktivációs függvény hozzárendelve, ami azt jelenti, hogy a hálózat bármilyen valós számértéket képes előállítani. Ez egy tipikus konfiguráció a regressziós problémák megoldásához, mint amilyen a feladat pontszámának megbecslése is.

Az összeállítás során a hálózatot a *Mean Squared Error* veszteségfüggvénnyel konfiguráltam, és az *Adam* optimalizálót alkalmaztam a veszteség minimalizálására a tanulási folyamat során. Ez az optimalizáló egy hatékony algoritmus, amely adaptív tanulási rátákat használ, ami gyakran gyorsabb konvergenciát és jobb teljesítményt eredményez.

Ezzel a neurális hálózattal is elvégeztem az összes eredménybecslést, amit lineáris regresszióval is elkészítettem, és a kapott eredmények igazolták a hipotézisemet. A neurális hálózat valóban lényegesen jobban teljesített azokon az adatokon, amikor a részképességek és a pontszám nem lineáris összefüggésben állt. A többi adaton, ahol a lineáris regresszió is megfelelő becslést adott, a neurális hálózat hasonló pontossággal becsülte meg a pontszámokat. A két modell kiértékeléseit a 5.5-ös táblázatban hasonlítottam össze, ahol jól látható a neurális hálózat előnye.

	Lineáris regresszió		Neurális hálózat	
	MSE	RMSE	MSE	RMSE
Lineáris összefüggés	$\approx 0,00015$	$\approx 0,012$	$\approx 0,0001$	$\approx 0,01$
Sigmoid-szerű összefüggés	$\approx 0,02$	$\approx 0,14$	$\approx 0,001$	$\approx 0,03$
Lineáris paraméterek	$\approx 0,05$	$\approx 0,22$	$\approx 0,01$	$\approx 0,1$
Exponenciális paraméterek	$\approx 0,08$	$\approx 0,28$	$\approx 0,01$	$\approx 0,1$

**5.5. táblázat.** Lineáris regresszió és neurális hálózat összehasonlítása

Így tehát sikerült jó közelítéssel megbecsülnöm egy ismert részképességekkel rendelkező személy tetszőleges feladaton elért eredményét. Az eredmény jelentősége azonban az, hogy a módszerem olyan esetekben is működik, amikor a kapott pontszám és a személy részképessége között nem lineáris az összefüggés. Az ilyen esetekben a korábban alkalmazott megoldások nem adtak megfelelő eredményt.

#### 5.2.4. Részképességekre bontás beágyazással

A kutatásom során a legkiemelkedőbb eredményt a beágyazás használatával sikerült elérnem. Mint ahogy a korábbi fejezetben kifejtettem, ennek a használatához mindössze egy nagy adathalmazra van szükség, amely minél több feladatmegoldásra tartalmazza a személy azonosítóját, a feladat azonosítóját és az elért pontszámot. A beágyazás során alkalmazott neurális hálózat struktúráját tekintve hasonló a korábban bemutatott eredménybecsléshez alkalmazott hálózathoz, azonban ebben az esetben egy-egy beágyazási réteget hoztam létre a két bemeneti paraméter számára, amelyeket egy konkatenációs rétegben vontam össze.

A részképességekre bontást elvégeztem csak a mért adatokon is, és azt vizsgáltam, hogy mely feladatokhoz milyen absztrakt részképességek szükségesek. A kapott részképesség értékek csak egymáshoz viszonyítva értelmezhetőek, így az konkrét értékek publikálása célszerűtlen. Az absztrakt részképességek miatt az eredmények validálása is nehéz

feladat, hiszen a részképességek beazonosítása mély pszichológiai és pedagógia ismereteket igényel. Azonban a feladatok összehasonlításával könnyebben értelmezhető eredményt kapunk, amely hitelesítése is lényegesen egyszerűbbé válik.

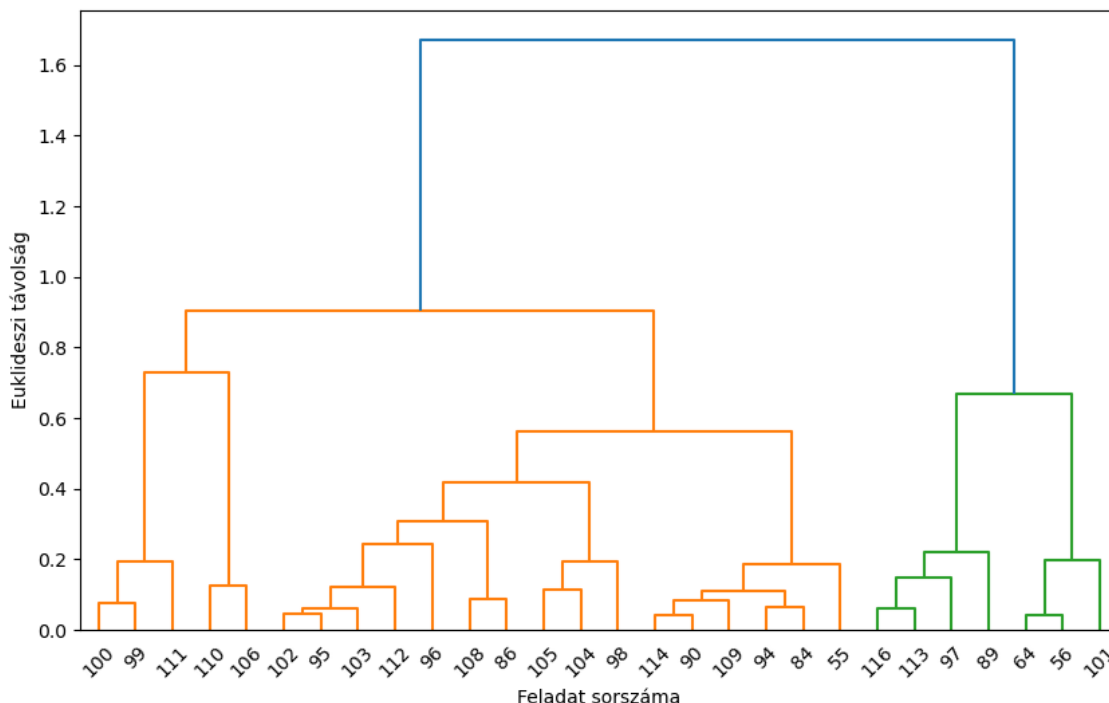
### Az eredmények összehasonlítása klaszterezéssel

Az eredmények, tehát a kapott részképesség vektorok összehasonlításához hierarchikus klaszterezést használtam, amely a vektorok euklideszi távolsága alapján határozza meg, hogy két feladat vektor mennyire áll közel egymáshoz.

A klaszterezés egy olyan statisztikai adatelemzési módszer, amelynek célja adathalmazok csoportokba rendezése. A klaszterezés során az egyes adatpontokat úgy próbálom csoportosítani, hogy azonos klaszterbe eső elemek hasonló tulajdonságokkal rendelkezzenek, míg a különbözőekbe tartozó elemek között jelentős eltérések legyenek.

A hierarchikus klaszterezés egy speciális típusú klaszterezési módszer, amely egy fokozatos, lépcsőzetes megközelítést alkalmaz. A klaszterezési eljárás először minden adat-elemet külön klaszterbe sorol, majd minden lépésben összevonja a két legközelebb eső csoportot. A csoportok közötti távolság mérésére több módszer létezik. Jelen esetben Ward metódusát használtam, amely a többi ismert módszerrel szemben nem direkt a klaszterek távolságát vizsgálja, hanem megnézi, hogy mennyivel nőne a klaszterek szórása, ha két csoportot összevonná. Ez alapján választja azt a két klasztert, amelynél ez a növekedés a legkisebb.

A hierarchikus klaszterezés eredményeit egy dendrogrammon ábrázoltam, amely mutatja a különböző feladatvektorok közötti távolságokat és lehetséges csoportosításukat, bemutatva, hogy mely feladatok hasonlítanak leginkább egymáshoz az általuk képviselt részképességek tekintetében.



5.1. ábra. A feladatok hierarchikus klaszterezése

Az ábra alján, az  $x$  tengely mentén elhelyezkedő, egymás mellett látható vonalak képviselik a vizsgált feladat vektorokat, amelyek a klaszterezési folyamat során lépcsről lépésre csoportosulnak. Az alacsonyan lévő kapcsolódási pontok azt jelzik, hogy két feladatvektor

hasonló részképességeket mér, mivel közel állnak egymáshoz az euklideszi térben. A magasan lévő kapcsolódások többnyire már feladat csoportokat kapcsolnak össze, és minél magasabban található a kapcsolódás, a két feladat csoport annál inkább különböző a részképességek vizsgálata szerint. Az  $y$  tengelyen tehát látható, hogy a kapcsolódási pontok mekkora euklideszi távolságot jelentenek a két feladat vagy feladat csoport között.

Az 5.1 ábráról leolvasható, hogy a részképességek vizsgálatában az egyik legközelebb a 64-es és az 56-os azonosítójú feladat áll egymáshoz, hiszen ezek vannak az egyik legalacsonyabban összekapcsolva. Ez a két feladat a *Számsorozat visszafelé* és a *Számisméltés* feladat. Ez a viszony szabad szemmel is látható, hiszen mindkét feladat leginkább a számok ismeretére és a rövidtávú memóriára épít, de ezt a hasonlóságot a témakörben kutató pszichológusok is meg tudják erősíteni, hiszen ugyanarra a részképességre dolgozták ki a feladatot. Mégis megdöbbenésüknek adtak hangot, hogy a megoldásom ezt az összefüggést, amely az algoritmus számára nyilvánvalóan nem volt feltárva, ilyen pontosan kimutatta. Hasonlóan magától értetődőnek tűnik, hogy a 113-as és 116-os feladat is nagyon hasonló részképességekre épül, hiszen mindkettő a személy egyensúlyát méri, az egyik csukott, míg a másik nyitott szemmel<sup>3</sup>.

Eggyel kevésbé szembeűnő a 84-es és a 94-es feladat közötti hasonlóság, melyek közül az elsőben három lassan kirajzolódó kép közül kell felismerni, hogy melyikből lesz tényleges rajz (a többin csak véletlenszerű vonalak vannak), míg a másikon egy-egy embert ábrázoló képen kell bejelölni, hogy hol találhatóak az egyes testrészei. Habár mindkettő feladatnál egy ábra felismerése játszik jelentős szerepet, azonban ez a két feladat már nem annyira triviálisan áll közel egymáshoz mint a korábban bemutatottak.

Egy sokkal érdekesebb megfigyelés azonban, hogy a vizsgálatom szerint a 101-es *Zenei hang észlelése* feladat hasonló absztrakt részképességekre épül mint a korábban bemutatott *Számsorozat visszafelé* és *Számisméltés* feladat. Ha feltételezzük, hogy a feladatok közötti távolság mérés helyes - amit a korábbiakban több példánál sikerült igazolni -, az azt jelenti, hogy ezek a feladatok is hasonló részképességre építenek. Egy kutató, Gyarmathy Zsófia statisztikai megközelítéssel ugyanerre a meglepő eredményre jutott.

A feladatok kipróbálásával azonban rögtön észrevehető egy fontos hasonlóság. A *Zenei hang észlelése* feladatban három gomb megnyomása három különböző hangot ad ki. A feladat elején hallható egy rövid dallam, és ezt kell lejátszani a gombok segítségével. A feladat több, egyre nehezedő dallam ismétléséből áll.

Az tehát könnyen látható, hogy a rövidtávú memória mindhárom feladatban jelentős szerepet játszik. Azonban feltételezésem szerint az sem véletlen, hogy éppen egy zenei memória feladat áll közel a kvantitatív memória feladathoz, ugyanis több kutatás igazolta, hogy a matematikai és zenei kognitív képességek összefüggésben állnak[1], így tehát feltételezhető, hogy létezik egy absztrakt részképesség, amely mindkettőt befolyásolja.

---

<sup>3</sup>Érdekesség, hogy az egyensúly mint ügyességi képesség a tapasztalatok alapján sok esetben valóban összefüggésben áll a kognitív képességekkel[23].

## 6. fejezet

# Összefoglalás

Dolgozatomban a mesterséges intelligencia felhasználási lehetőségeit vizsgáltam a kognitív képességek- és részképességek elemzéshez használható feladatsorok összeállítása és az eredmények elemzése során. A mesterséges intelligencia alkalmazása innovatív megközelítés ezen a területen, amelynek a hatékonyságát már az első kutatási eredmények tapasztalatai is visszaigazolták.

A dolgozat elején bemutatott CHC modell pszichometriai keretrendszert kínált számomra, amellyel strukturáltan elemezhetjük az emberi kognitív képességek összetettségét és kapcsolódásait. A dolgozat során nagy adathalmazokon és különböző feladattípusokon keresztül vizsgáltam a részképességek szerepét, illetve becsültem meg az egyének teljesítményét, figyelembe véve a személy képességeit és a teszt paramétereit. Ebben a fejezetben végigveszem a dolgozat során felállított hipotéziseimet, és azok bizonyítását a kutatási eredményeim segítségével.

A meglévő, bizonyított statisztikai eljárásokkal összevetve igazoltam, hogy a feladatok részképesség-függése jól becsülhető logisztikus illetve lineáris regresszió segítségével olyan esetben, ahol a feladatokat megoldó személyek részképességei már ismertek és lineáris összefüggésben állnak a feladat során nyújtott teljesítménnyel. Ezt követően szimulációs eljárással alátámasztottam, hogy amennyiben ez a lineáris kapcsolat nem áll fenn, úgy ezeknek a becsléseknek a pontossága jelentősen csökken, ugyanakkor mesterséges intelligencia alkalmazásával még ebben az esetben is lehetőség nyílik hatékony becslések készítésére, amelyre korábban még nem volt megoldás.

Továbbá bizonyítottam a hipotézisem, miszerint a mesterséges intelligencia egy speciális felhasználásával, a beágyazás segítségével absztrakt részképességekre bontható egy feladat, és ezáltal pszichometrikusok segítségével már könnyen megállapítható a feladat specifikus részképesség-függése is.

A kutatásom eredményeit az adatokat szolgáltató kutatócsoport vezetője nagy lelkesedéssel fogadta, mivel azok megerősítették azokat az eredményeket, amelyeket más módszerekkel ők már korábban bizonyítottak, és meglepetésként érte az újonnan alkalmazott módszer hatékonysága és pontossága. Továbbá nagy örömmel fogadta az új eredményeket, és lelkesen mesélte azok felhasználási lehetőségeit a kutatócsoport munkájában, illetve az eredmények által nyitott új perspektívákat.

A felhasználási lehetőségek közül az egyik legfontosabb és legizgalmasabb az eredmények alkalmazása adaptív gyakorló- és teszt feladatsorok készítése során. Az adaptív feladatsorok lényege, hogy nem egy azonos, statikus gyakorlatsort old meg mindenki, hanem a soron következő feladatot saját korábbi eredményei alapján határozza meg a rendszer. Az eredmények értékelésében és a következő feladat kiválasztásában nagy segítséget nyújtanak a kutatás során bemutatott módszerek, hiszen egy összetett feladat kiértékeléséből a kitöltő részképességeit is vizsgálni tudjuk.

A részképességekre alapuló adaptív feladatok jelentős mértékben növelik a kognitív profilalkotás hatékonyságát és pontosságát, hiszen az ilyen jellegű feladatsorok lehetővé teszik, hogy kevesebb tesztfeladattal is részletes képet kapjunk a vizsgált személyről. Az eredmények alapján kimutatható, mely részképességeket érdemes intenzívebben vizsgálni, és melyek azok, amelyek az összetettebb feladatokon keresztül is jól mérhetőek. Az adaptív jelleg továbbá garantálja, hogy a vizsgált személy mindig a képességeihez mérten, de mégis kihívást jelentő feladatokkal találkozzon. Ez biztosítja, hogy a teszt során a kitöltő az úgynevezett *flow* állapotában maradjon, amelyben a tesztek a leginkább reprezentatív eredményeket adnak a kognitív képességek terén.

Ezekon felül az általam elért eredmények alkalmazhatók személyre szabott és kiscsoportos fejlesztési folyamatokban is. Ennek során a pedagógusok a tanulók részképességeinek megfelelően tudják kiválasztani a feladatokat, lehetővé téve ezzel a hatékonyabb és célzottabb fejlődést. Emellett a részképességeket mérő tesztek segítségével észrevehetőek a különböző tanulási nehézségek korai jelei, és így hatékonyabban kezelhetik azokat.

# Köszönetnyilvánítás

A dolgozat az MTA-AVKF Tanulási Környezet Kutatócsoporttal<sup>1</sup> együttműködésben készült a Magyar Tudományos Akadémia Közoktatás-fejlesztési Kutatási Program támogatásával.

Továbbá szeretnék köszönetet mondani a konzulensemnek Dr. Forstner Bertalannak, aki megismertette velem a kutatási területet, mindig segítőkészen támogatta a munkámat, és az első héttől kezdve motivált a lelkesedésével.

---

<sup>1</sup><https://tanulas-kutatas.hu>

# Irodalomjegyzék

- [1] Connie Barroso – Colleen M Ganley – Sara A Hart – Nancy Rogers – Jane P Clendinning: The relative importance of math-and music-related cognitive and affective factors in predicting undergraduate music theory achievement. *Applied Cognitive Psychology*, 33. évf. (2019) 5. sz., 771–783. p.
- [2] A Wade Boykin – Nathan Brody – Robert J Sternberg: Intelligence: Knowns and unknowns. *Am. Psychol*, 51. évf. (1996), 77–101. p.
- [3] Timothy A Brown – David H Barlow: Reliability in psychological assessment: A historical perspective and implications for the future. *Journal of Clinical Psychology*, 64. évf. (2008) 10. sz., 1045–1054. p.
- [4] John B Carroll: Cognitive abilities: The state of the art. *Psychological Science*, 3. évf. (1992) 5. sz., 266–271. p.
- [5] John B Carroll: *Human cognitive abilities: A survey of factor-analytic studies*. 1. köt. 1993, Cambridge University Press.
- [6] Damien C Cormier – Okan Bulut – Kevin S McGrew – Jessica Frison: The role of cattell–horn–carroll (chc) cognitive abilities in predicting writing achievement during the school-age years. *Psychology in the Schools*, 53. évf. (2016) 8. sz., 787–803. p.
- [7] Susan E Embretson – Steven P Reise: *Item response theory*. 2013, Psychology Press.
- [8] József Balázs Fejes – Beáta Szenczi: Tanulási korlátok a magyar és az amerikai szakirodalomban. *Gyógypedagógiai szemle*, 38. évf. (2010) 4. sz., 273–287. p.
- [9] László Gazdi – Krisztián Pomázi – Máté Szabó – Bertalan Forstner: An innovative model for adaptive learning utilizing biofeedback and item response theory. *Periodica Polytechnica Electrical Engineering and Computer Science*, 62. évf. (2018) 3. sz., 90–105. p. Paper: 12213, 16 p.
- [10] Jan-Eric Gustafsson – Gudrun Balke: General and specific abilities as predictors of school achievement. *Multivariate behavioral research*, 28. évf. (1993) 4. sz., 407–434. p.
- [11] Éva Gyarmathy: A tehetséggondozás változási kényszere. *Iskolakultúra*, 23. évf. (2013) 3-4. sz., 101–109. p.
- [12] Casey Henley: *Foundations of Neuroscience*. 2021, Michigan State University Libraries, 6–7. p.
- [13] Fazekas István: Neurális hálózatok, 2013.  
URL <https://gyires.inf.unideb.hu/GyBITT/19/index.html>.
- [14] Andrea Jones – Angela Green: Validity in psychological testing: The role of item analysis. *Psychological Assessment*, 24. évf. (2012) 3. sz., 328–340. p.



- [15] Kristof Kovacs – Andrew RA Conway: A unified cognitive/differential approach to human intelligence: Implications for iq testing. *Journal of Applied Research in Memory and Cognition*, 8. évf. (2019) 3. sz., 255–272. p.
- [16] Jim Lumsden – Elizabeth A Edwards – Natalia S Lawrence – David Coyle – Marcus R Munafò: Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games*, 4. évf. (2016. Jul) 2. sz., e11. p.
- [17] Yixuan Ma – Zhen ji Zhang: Travel mode choice prediction using deep neural networks with entity embeddings. *IEEE Access*, 8. évf. (2020), 64959–64970. p.
- [18] Kevin S McGrew: Chc theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research, 2009.
- [19] Jean Raven: Raven progressive matrices. In *Handbook of nonverbal assessment*. 2003, Springer, 223–237. p.
- [20] Mary G Rizza – David E McIntosh – Alice McCunn: Profile analysis of the woodcock-johnson iii tests of cognitive abilities with gifted students. *Psychology in the Schools*, 38. évf. (2001) 5. sz., 447–455. p.
- [21] Gale H Roid – Mark Pomplun: *The stanford-binet intelligence scales*. 2012, The Guilford Press.
- [22] Leonard D Smith – Lisa A Best – D Alan Stubbs – John Johnston – Anne B Archibald: Measurement in psychology: A critical history of a methodological concept. *Cambridge University Press*, 2010.
- [23] Tony Szturm – Vedant Sakhalkar – Sue Boreskie – Jonathan J Marotta – Christine Wu – Anuprita Kanitkar: Integrated testing of standing balance and cognition: Test–retest reliability and construct validity. *Gait & posture*, 41. évf. (2015) 1. sz., 146–152. p.
- [24] Harald Valås: Learned helplessness and psychological adjustment ii: Effects of learning disabilities and low achievement. *Scandinavian Journal of Educational Research*, 45. évf. (2001) 2. sz., 101–114. p.
- [25] David Wechsler – Habuku Kodama: *Wechsler intelligence scale for children*. 1. köt. 1949, Psychological corporation New York.