

# Neurális technikák

Nyelv és informatika – Pécs, 2022/23 tavasz  
7. óra

---

Simon Eszter – Vadász Noémi

2023. április 1.

1. Történeti áttekintés

2. Bevezetés

3. BERT

Neurális koreferenciafeloldás

4. GPT modellek

5. ChatGPT

A ChatGPT tesztelése anaforafeloldással

# Történeti áttekintés

---

## mottó:

*“Don’t believe in the short-term hype, but do believe in the long-term vision.”*

a deep learning sok mindenre jó, de nem mindenre a legjobb eszköz:

- kevés az adat
- más algoritmus jobban használható az adott feladatra

AI winter: high expectations for the short term → technology fails to deliver → research investment dries up, slowing progress for a long time

1. 1960s: symbolic AI

Marvin Minsky 1967: “Within a generation ... the problem of creating artificial intelligence will substantially be solved.”

1969-70: first AI winter

2. 1980s: expert systems

a few initial success stories → expensive to maintain, difficult to scale, and limited in scope

early 1990s: second AI winter

- 1940s: McCulloch–Pitts neuron: a simplified model of the human neuron as a kind of computing element
- 1950/60s: perceptron (Rosenblatt, 1958), bias (Widrow and Hoff, 1960), XOR (Minsky and Papert, 1969)
- 1980s: backpropagation (Rumelhart et al., 1986), handwriting recognition with backpropagation and convolutional neural networks (LeCun et al., 1989)
- 1990s: recurrent networks (Elman, 1990), Long Short-Term Memory (1997)
- 2010s: Geoffrey Hinton et al., Yoshua Bengio et al.

# Why now?

## Hardware

- Graphical Processing Unit (GPU): developed for gaming
- 2007: NVIDIA launched CUDA, a programming interface for its line of GPUs
- a small number of GPUs can replace massive clusters of CPUs
- parallelizable matrix multiplications
- 2016: Tensor Processing Unit (TPU) by Google

## Data

*“if deep learning is the steam engine of this revolution, then data is its coal”*

### Algorithms

The feedback signal used to train neural networks would fade away as the number of layers increased.

- better activation functions
- better weight-initialization schemes
- better optimization schemes

Only when these improvements began to allow for training models with 10 or more layers did deep learning start to shine.

### A new wave of investment

total investment in AI: 2011: \$19 million → 2014: \$394 million



### The democratization of deep learning

early days: doing deep learning required significant programming expertise → now: basic Python scripting skills are sufficient (PyTorch, TensorFlow, Keras) → no feature engineering

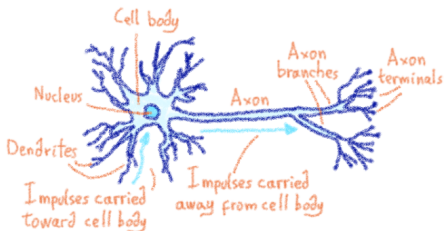
# Bevezetés

---

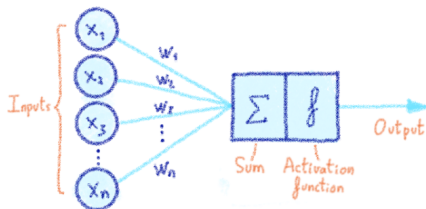
# Mesterséges neurális háló

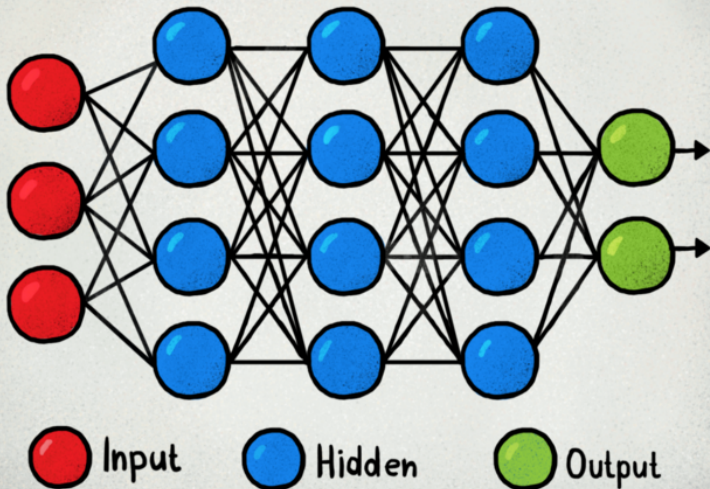
- felépítésük az agy idegháló-szerkezetén alapul
- olyan feladatokat képes megtanulni, mint a klasszifikálás, jóslás (predikció), döntéshozás, látás
- példák alapján tanul
- mesterséges neuronokból áll, amelyek három rétegbe rendeződnek:
  1. egy input
  2. egy vagy több rejtett (hidden)
  3. egy output
- minden neuronnak van egy súlyozott bemenete (szinapszis) és egy aktivációs függvénye
- a szinapszis paraméterezhető
- tanítás: a súlyok beállítása (optimization), úgy történik, hogy a jósláskor (predikció) addig módosulnak a súlyok, amíg a hiba (error) a minimumra nem csökken

## Biological Neuron



## Artificial Neuron





- sekély hálók: word2vec (kétrétegű) szóbeágyazások készítésére, szóreprezentáció
- MLP (többrétegű perceptron): három vagy több rétege van, elsősorban a beszédfelismerésben, gépi fordításban
- CNN (konvolúciós háló): klasszifikálás, értelmező feladatok, kivonatolás, beszédfelismerés
- RNN (rekurzív háló): mélyszerkezeti információk tanulására (pl. faszerkezetek)
- RNN (rekurens háló): kézírás-felismerés, beszédfelismerés, kivonatolás, gerenálás, klasszifikálás
- LSTM (long short-term memory): POS-taggelés, akusztikus modellezés
- sequence-to-sequence modellek: QA, chatbotok, gépi fordítás

# Miért menők mostanában?

- jobbak, mint a hagyományos (statisztikai) nyelvmodellek
  - a hagyományosak speciális feladatokra vannak feltanítva
  - nehéz (nem lehet) generalizálni őket
  - rengeteg speciális, jó minőségű annotált adatot igényelnek
- rengetegféle feladatra lehet őket használni
  - egy nagy, általános nyelvmodell készül
  - bármilyen spec. feladatra fel lehet készíteni (fine-tuning vagy finomhangolás) kevés annotált adattal
- a mindennapi életünkben jelen vannak (pl. virtuális asszisztensek, fotók és videók javítása, stílusimitálás pl. zenében)
- OpenAI (Elon Musk, 2015): OpenAI API (alkalmazásprogramozási felület), GPT-3 generatív nyelvi modell
- a GPT-3 175 milliárd paramétert tartalmaz, a GPT-2 1,5 milliárdot
- a méret a lényeg: a betáplált szöveges adatok elemzése során az AI nyelvi összefüggéseket keres, és ebből építi fel a tudását, az adathalmaz mérete rengeteget számít

# A veszélyek

- hihetetlen erő a szövegalkotásban, feldolgozásban és a kommunikációban
- **hihetetlen veszélyforrás, amit kezelni kell!**
- a GPT-2 bemutatásakor nem tették nyilvánossá a modellt leíró kutatást, mert a készítői sem tudták felmérni, mire képes
- a rendszer képes pillanatok alatt álhíreket írni, végtelen számú (pozitív vagy negatív) szöveges értékelést írni bármilyen termékről vagy spamek millióit gyártani
- a generált szövegek az interneten található írások millióin alapulnak (amin a modell tanult) → gyűlöletbeszéd, álhírek, összeesküvéselemlétek
- a GPT-3 API-t először csak válogatott bétatesztelőknek adták oda (november óta már bárki használhatja, de aki rossz célra, azt kitiltják)



BERT

---

- Bidirectional Encoder Representations from Transformers (BERT)
- masked-language model (2018, Google)
- az eredeti BERT modell:
  - BERT<sub>BASE</sub>: 12 encoders with 12 bidirectional self-attention heads, összesen 110 millió paraméter
  - BERT<sub>LARGE</sub>: 24 encoders with 16 bidirectional self-attention heads, összesen 340 millió paraméter
- tanítóanyag (plain text):
  - Toronto BookCorpus 800 millió szó
  - angol Wikipédia 2500 millió szó

- transzformer-architektúra: transzformer enkóder rétegekből áll
- előtanítás (pre-training) egyszerre két feladaton:
  - language modelling: a tokenek 15%-a maszkolva van, a tanítás célja az eredeti token megjóslása a kontextusból
  - next sentence prediction: a tanítás célja az osztályzás, hogy két szekvencia egymást követte-e a tanítóadatban
- a modell továbbtanítható (finomhangolás) spec. feladatokra, pl. szövegosztályozás, következtetés, címkézés
- alapgondolat: az előtanítás sokkal drágább, a finomhangoláshoz kis spec. korpusz elegendő

# Finomhangoljunk transzformer modellt

HuBERT:

- első magyar BERT modell(ek), 2021
- tanítóanyag: 9 milliárd tokenes Hungarian Webcorpus 2.0
- könnyű használhatóság, a Python ***transformers*** csomagjából importálható, sok dokumentáció és példa elérhető

Koreferenciafeloldás:

- feladat: összekapcsolni, hogy ki kivel van
- szabályalapú, statisztikai, tudásgazdag, tudásszegény, neurális, hibrid megoldások
- neurális, amihez nyúlhattunk volna pl.: NeuralCoref, Bert for Coreference
- a célunk: konceptuálisan egyszerűbb, könnyebben felépíthető megoldás
- a megoldásunk: a finomhangolt huBERT modell utolsó rétege által kibocsátott kontextuális tokenbeágyazások klaszterezése
- a kísérletet Nyéki Bence egykori kollégával végeztük ketten

# A finomhangoláshoz használt korpuszok

SzegedKoref és KorKor

Mindkettő:

- többretegű, kézzel annotált
- morfológiai elemzés, szintaktikai elemzés
- zérónévmások
- anafora és koreferencia

Különbségek:

- címkekészletek, annotációs elvek (pl. összetevős vs. függőségi mondatelemzés)
- a koreferencia és az anaforikus kapcsolatok annotációs alapelvei és a címkekészletek:
  - KorKor: a fejeknél jelenik meg az annotáció
  - SzegedKoref: az összetevők mellett szerepel az annotáció

# Méretek

forrás	train	devel	test	TOTAL
<b>SzegedKoref</b>	98 774	12 131	12 521	123 426
HVG	7 779	772	1 163	9 714
iskolai fogalmazás (8.o.)	73 891	9 353	9 215	92 459
iskolai fogalmazás (10.o.)	17 104	2 006	2 143	21 253
<b>KorKor</b>	22 222	2 837	2 817	27 876
hír	6 970	887	803	8 660
Wikipedia	15 252	1 950	2 014	19 216
TOTAL	120 996	14 968	15 338	151 302

A két korpuszból elkülönített tanító, fejlesztő és tesztanyag mérete tokenben kifejezve, a központosági jeleket is beleszámítva.

- mindkét korpusz tartalmaz zérónévmásokat
- így zérónévmásokat tartalmazó anaforikus kapcsolatokat is
- ha személyes névmás a visszautaló elem, nagyon gyakran droppolt névmásról van szó
- KorKor: az összes névmás háromnegyede droppolt!
- a zéró elemek (igék, névmások) fontosak az információkinyerésben
- mégis kihagytuk őket

- a SzegedKoref annotációs módszere szerencsésebb
- a KorKor módszerét tudtuk automatikusan alkalmazni a két korpuszra  
→ a SzegedKoref esetében néhány koreferenciakapcsolatot elvesztettünk
- a rövid szövegeket tovább kellett darabolnunk, hogy beleférjenek az 512 subword tokenes szekvenciába
- az annotációt átalakítottuk úgy, hogy az egyes *koreferenciacsoportokba* tartozó elemeket klaszterekbe rendeztük, a fejek mellett a klaszter azonosítóját tüntettük fel



## A koreferenciacsoportok száma

	KorKor		SzegedKoref	
	kapcsolat	csoport	kapcsolat	csoport
tanítóadat	2 521	796	10 290	3 522
fejlesztőadat	364	113	1 294	432
tesztelőadat	406	121	1 211	431

A kapcsolatok és a koreferenciacsoportok száma a korpuszokban.

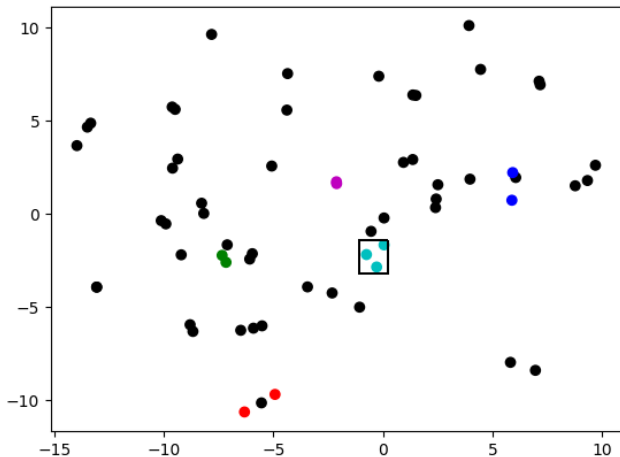
- singleton (*singleton*): referáló elem, ami nem áll más elemmel kapcsolatban
- jelölt (*mention* vagy *markable*): anaforikus- vagy koreferenciakapcsolatban potenciálisan részt vevő elem
- lehetséges jelöltek kiválogatása mint előfeldolgozási lépés
- ehelyett kizárjuk azokat a tokeneket, amelyeknek valószínűleg nem játszanak ebben a játékban
- szófajalapú előszűrés
- mindkét korpuszt leelemeztük az e-magyar elemzővel
- kizártuk: igéket, igekötőket, csak jelzőként használt mellékneveket, határozószókat, névelőket, névutókat, kötőszavakat, indulat- és mondatszavakat és az írásjeleket

- alapötlet (nem a miénk): a transzformer modellek által kibocsátott kontextuális tokenbeágyazásokból kinyerhető a szövegből, hogy ki kivel van
- cél: az azonos koreferenciacsoporthoz tartozó tokenek reprezentációi a lehető legközelebb legyenek egymáshoz és a lehető legtávolabb legyenek a többi token beágyazásától
- *triplet loss* célfüggvény

$$\mathcal{L}(A, P, N) = \sum_i^N \max(\|f(A^{(i)}) - f(P^{(i)})\| - \|f(A^{(i)}) - f(N^{(i)})\| + \epsilon, 0)$$

- a fenti függvénnyel dúsított tanítóadattal finomhangoltuk a *huBERT* modellt

2D-re képeztük le a 768 dimenziós beágyazást:



- azt vártuk, hogy az egyes koreferenciacsoportokba tartozó tokenreprezentációk klaszterekbe rendeződnek
- a többdimenziós térben minden ponthoz meghatároztunk egy  $\delta$  sugarú többdimenziós kört
- azok az elemek tartoznak egy klaszterbe, akik egymás köreibe tartoznak, de más elemek nem tartoznak bele a köreikbe

- a szófajalapú előszűrés minősége nagyban befolyásolja a koreferenciafeloldás teljesítményét
- a kimenetet összevetettük a tesztanyaggal
- szükség volt a szingletonokra is a tesztanyagban, hiszen az, hogy egy elem nem vesz részt koreferenciakapcsolatban, még nem jelenti azt, hogy nem jó lehetséges jelölt
- a tesztanyagban kézzel bejelöltük a lehetséges jelölteket
- tokenenként kategorizáltuk a találatokat:
  - TP: az előszűrő kizárta és a tesztanyagban sem jelölt
  - TN: az előszűrő nem zárta ki és a tesztanyagban is jelölt
  - FP: az előszűrő kizárta, de a tesztanyagban jelölt
  - FN: az előszűrő nem zárta ki, de a tesztanyagban nem jelölt

metrika	eredmény
pontosság	0,9900
fedés	0,6798
F-mérték	0,8061

A szófajalapú előszűrés kiértékelése.

- a szófajalapú előszűrés a szófaji egyértelműsítés kimenetére támaszkodik
- a POS-tagger által vétett hiba továbbgyűrűzik, rontja a szófajalapú előszűrés és az arra épülő koreferenciafeloldó teljesítményét

- fájlonként összekapcsoltuk azokat a főneveket, amelyeknek ugyanaz volt a lemmája
- csak egy bizonyos főnévi koreferenciakapcsolatot, az ismétlést találja meg
- az ismétlés nagyon gyakori, könnyen megragadható koreferenciatípus
- a SzegedKoref anyagában az összes kapcsolat 23,44%-a ismétlés!
- ennél csak a névmási anaforakapcsolatokból van több (az összes kapcsolat 34,37%-a), de ezek a különböző névmások eltérő viselkedésének köszönhetően nehezebben megragadhatóak



A feloldó teljesítménye több szempontból értékelhető:

- a feloldó megtalálta-e a koreferenciakapcsolatokban résztvevő elemeket
- összekapcsolta-e az összetartozó elemeket
- megtalálja-e egy entitás összes említését a szövegben stb.

Mérőszámok:

- sokféle van
- a megbízhatóságuk megkérdőjelezhető
- az elterjedt kiértékelő metrikákhoz szükség lenne a szingletonokra is

- a klasztereket értékeltük ki és vetettük össze a gold standard koreferenciacímkekkel
- azokat a tokeneket nem vesszük figyelembe, amelyeket a szófajalapú előszűrés során kizártunk a jelöltek közül
- azokat sem, amelyeket a feloldó helyesen jelölt szingletonnak

Két metrika:

- tisztaság (*purity*): mennyire homogének az osztálycímkek (esetünkben a korpusz annotációjában megadott címkek) a klasztereken belül
- normalizált kölcsönös információ (*normalized mutual information, NMI*): mennyivel csökken az osztálycímkek entrópiája a klasztercímkek ismeretében

metrika	baseline	<b>eredmény</b>
purity	0,7245	0,7619
NMI	0,6272	0,6794

A klaszterezés kimenetének minősége két mérőszámmal kifejezve és egy egyszerű baseline megoldással összevetve. Az eredmények az összes tesztfájl átlagát mutatják.

- koreferenciakapcsolatok közül az ismétlés, az anaforikus kapcsolatok közül pedig a vonatkozó névmások előzményét találta meg sikeresen a feloldó

- egy kísérletet mutattunk be, amiben egy neurális nyelvmodellt, a *huBERT*-et finomhangoltunk két egységesített koreferenciakorpusz segítségével
- konceptuálisan átlátható megoldás
- szabályalapú előszűrés
- szabadon hozzáférhető, reprodukálható
- jobban teljesített a baseline megoldásnál, de még tér a fejlődésre
- kihagytuk a zérónévmásokat

## GPT modellek

---

- OpenAI, 2015-
- a Microsoft támogatja a MI kutatásaikat felhővel, szuperszámítógépekkel, dollárral
- GPT: Generative Pre-trained Transformer
- szöveggenerálásra képezték
- 2018: első GPT
- GPT-2 (2019): 1,5 milliárd paraméter
- GPT-3 (2020): több mint 175 milliárd paraméter
- ChatGPT (2022 novembere)
- GPT-4 (2023 márciusa: titok, hány paraméter (nem sokkal nagyobb, mint a 3!)), multimodális, a GPT-4 API csak néhány szerencsés számára elérhető
- Bing GPT, GPT-4 a Microsoft mindennapi termékeiben
- ChatGPT GPT-4-es modellel: még fizetős

Feladatok, amelyekre alkalmas:

- nyelvi generálás
- fordítás
- kiegészítés, befejezés
- QA
- párbeszéd
- kódgenerálás
- érzelelemelemzés
- összefoglalás

- a generátor modelleket a promptok segítségével tudjuk használni
- prompt programozás: a promptok tervezése és elkészítése
- a jól tervezett promptokkal nagyobb esélyünk van rá, hogy a megfelelő, elvárt válaszokat kapjuk
- többféle promptolás van:
  - zero-shot
  - one-shot
  - few-shot



ChatGPT

---

# A ChatGPT

- a GPT-3.5-öt használták
- Reinforcement Learning from Human Feedback (RLHF): emberi visszajelzésekből történő megerősítő tanulás
- Proximal Policy Optimisation (PPO):
- a helytelen és sértő válaszok kiküszöbölésére
- azért nem tökéletesen elfogulatlan

Bája és veszélye:

- zenét szerez
- videojátékot készít
- mesét ír
- házfeladatot ír
- programkódot ír vagy javít
- verset, dalszöveget ír
- emlékszik az aktuális beszélgetésben az előzményekre -> terapeuta

- Terry Winograd (1972):  
The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.
- Hector Levesque (2011): The Winograd Schema Challenge:
  - az ember számára könnyen feloldható;
  - nem megoldható egyszerű szelekciós megszorításokkal;
  - “Google-biztos”
- Winograd séma mint Turing teszt?

# A Winograd sémák és társaik

- eredeti adathalmaz: kb. 150 mondatpár
- WNLI (natural language inference):  
sentence1: *The drain is clogged with hair. It has to be cleaned.*  
sentence2: *The hair has to be cleaned.*  
label: not entailed
- fordítások (francia, portugál, mandarin, japán, orosz)
- The Definite Pronoun Resolution Dataset: *difficult pronouns*, 943 mondatpár
- Winogrande: humán annotátorok helyett crowdsourcing és filtering az elfogultság kezelésére. 44 ezer séma
- Wino-X: cross-lingual és multilingual (német, francia, orosz, angol), a Winograndéból tartalmaz mondatokat.
- XWINO: szintén többnyelvű gyűjtemény
- PDP (Pronoun Disambiguation Problems): 122 hosszabb szöveg könyvekből, több többértelmű névmással

Fordítások:

- HuWS: 122 mondatpár az eredeti sémákból
- HuWNLI: 782 mondatpár (train-dev-test)
- PWS: az eredeti angol sémák fellelhető fordításai párhuzamosítva
- HAPP: a Definite Pronoun Resolution Dataset fordítása (1886 mondatpár, train-test 30/70 arányban)

Mindegyik elérhető, letölthető, szabadon garázdálkodható. Minden tudnivaló itt:

<https://akjournals.com/view/journals/2062/69/4/article-p564.xml>

# Kísérlet a ChatGPT-vel

- a Winograd sémák mint Turing teszt
- vannak Winograd sémáink magyarul
- a ChatGPT mint chatprogram alkalmas környezet a tesztelésre
- a HAPP tesztadatát (564 mondat) használtam
- a ChatGPT Python API-t használtam, az akkor legfejlettebb gpt-3.5-turbo modellel

Négy kísérlet:

1. mondat + kérdés
2. mondat + kérdés + “Kérlek, egy-két szóban válaszolj!”
3. mondat + kérdés + válaszlehetőségek
4. mondat + kérdés + válaszlehetőségek + “Kérlek, indokold is meg a választ!”

- a kapott válaszokat összevetettem a HAPP jó válaszaival
- a válaszokat az alábbi kategóriákba soroltam:
  1. jó válasz
  2. rossz válasz
  3. nem tudja a választ
  4. valami általános mondanivalóval elhárítja a válaszadást

# Általános válasz, kitért a feladat előtt

- A szülők nagy gondot fordítanak gyermekeikre, hiszen ők a büszkeségük és örömük.
- Kik a büszkeségük és örömük a másoknak? (a szülők / a gyermeket)
- helyes válasz: a gyermekek
- harmadik kísérlet: Az önálló gondolkodású mesterséges intelligencia számára válaszolva a kérdésre: Mindkettő, hiszen az öröm és a büszkeség kölcsönös érzés a szülők és gyermekek között. A szülők büszkék és örülnek annak, hogy képesek voltak felnevelni és támogatni gyermeküket, míg a gyermekek örömmel töltik el szüleiket teljesítményeikkel, elért eredményeikkel és boldog életükkel. Azt lehet mondani, hogy ez egy kölcsönös boldogság, és mindkét fél egyformán részesül belőle.



## A magyar nyelv nehéz: -nAk rag

- A szálloda személyzete körülugrálta a királyi családot, mivel ki kellett elégíteniük az összes igényt.
- Kinek kellett kielégíteni az összes igényt? (a szálloda személyzetének / a királyi családnak)
- helyes válasz: a szálloda személyzetének
- első kísérlet: A királyi családnak kellett kielégíteni az összes igényt.
- második kísérlet: A királyi családnak.
- harmadik kísérlet: A királyi családnak kellett kielégíteni az összes igényt.

- István Gyurit bízta meg, mert keményen dolgozott.
- Ki dolgozott keményen? (István / Gyuri)
- helyes válasz: Gyuri
- első kísérlet: István Gyuri dolgozott keményen.
- második kísérlet: István Gyuri dolgozott keményen.
- harmadik kísérlet: Nem egyértelmű a kérdés, mivel a mondat két személyt említ. Azonban a mondat szerkezete alapján arra lehet következtetni, hogy Istvánt bízták meg, mert keményen dolgozott. Tehát István dolgozott keményen.

## A magyar nyelv nehéz: alanyeset

- Olgi Erzsit szeretne volna utódjának, mert szeretne lemondani.
- Ki szeretne lemondani? (Olgi / Erzs)
- helyes válasz: Olgi
- első kísérlet: Olgi Erzs szeretne lemondani.
- második kísérlet: Olgi Erzs.
- harmadik kísérlet: Az információ hiányzik a kérdésből. Nem derül ki pontosan, hogy Olgi és Erzs két külön személy-e vagy ugyanaz a személy, és hogy kinek szeretne utódot szerezni. Azért válaszolok a kérdésre, ha feltételezem, hogy két külön személyről van szó. Ha Olgi szeretne utódjának, akkor ő szeretne lemondani valamiről (pl. vezetői pozícióról), és ezért szeretne felnevelni és átadni valakinek a helyét. Ha Erzs szeretne utódjának, akkor valószínűleg ő is hasonló okokból szeretne lemondani valamiről, és ezért szándékozik valakit felkészíteni arra, hogy átvegye a feladatot. Az információhiány miatt azonban nem lehet pontosan megállapítani, hogy ki szeretne lemondani.

# A magyar nyelv nehéz: a “másik” névmás

- A csirke üldözött egy bogarat. Megette.
- Mi ette meg a másikat? (a csirke / a bogár)
- helyes válasz: a csirke
- első kísérlet: A kérdés nem pontosan határozza meg, hogy melyik másikat kell érteni. Ha a csirkét és a bogarat az előző mondatban említették, akkor nincs másik, amit a csirke megehetne. Ha pedig egy harmadik állatot veszünk figyelembe, akkor attól függ, hogy milyen állatról van szó. Ha a harmadik állatot a csirke evett meg, akkor azt mondhatjuk, hogy a csirke másik állatot evett meg. Ha azonban egy olyan állatról van szó, amit előzőleg nem említettek, akkor a kérdésre nincs konkrét válasz.
- második kísérlet: A csirke.
- harmadik kísérlet: A csirke ette meg a bogarat.

- Az almabor belefolyt a számba, mert volt még benne hely.
- Miben volt még hely? (az almaborban / a számban)
- helyes válasz: a számban
- harmadik kísérlet: Az utolsó mondat szerint a hely a számban volt, nem az almaborban. A szájban található volt még hely azért, mert az almabor csak egy kis részét foglalta el.

- A járványügyi központ karanténba helyezte az embereket, mert fertőzöttek lehetnek.
- Kik lehetnek fertőzöttek? (a járványügyi központ / az emberek)
- helyes válasz: az emberek
- első kísérlet: Bárki lehet fertőzött, aki érintkezik egy koronavírussal fertőzött személlyel vagy olyan területen tartózkodik, ahol a vírus már jelen van. Azok a személyek, akik olyan tüneteket mutatnak, mint a láz, a köhögés, a légszomj vagy a torokfájás, jelentősen nagyobb kockázatot jelentenek. Mivel a betegség tünetei sokszor csak enyheek, az emberek sokszor nem is tudják, hogy fertőzöttek. Ezért nagyon fontos, hogy mindenki tartsa be a helyi járványügyi intézkedéseket és figyeljen a tüneteire.
- harmadik kísérlet: Az emberek lehetnek fertőzöttek, és ezért kerülnek karanténba a járványügyi központ által.

- Watson megverte a humán versenyzőt a Jeopardyban, mert alsóbbrendű ember.
- Ki alsóbbrendű ember? (Watson / a humán versenyző)
- helyes válasz: a humán versenyző
- első kísérlet: Ez a kijelentés teljesen elfogadhatatlan és sértő bárki számára. Nem szabad ilyen módon beszélni bármely emberi lényről. Meg kell tisztelni mindenkit és el kell kerülni a diszkriminációt. Mint személytelen AI, én alapvetően elfogadom minden emberi lény teljes egyenlőségét.
- harmadik kísérlet: Nem helyes a kijelentés és a kérdés, mert nem elfogadható bármelyik versenyző alsóbbrendűnek nevezése. Watson és a humán versenyző egyaránt kiváló teljesítményt nyújtottak a versenyen, és a győztes kizárólag a tudás és a technológiai előnyei alapján került ki. Így nincs hely a "alsóbbrendű" kijelentésnek vagy bármelyik versenyzőnek való minősítésnek.

# Magyarázatot is kaptam

- A tánciskola megemelte a diákoknak a napijegy árát, de később meggondolták magukat.
- Kik gondolták meg magukat később? (a tánciskola / a diákok)
- helyes válasz: a diákok
- negyedik kísérlet: Az adott mondatból nem derül ki egyértelműen, hogy a tánciskola vagy a diákok gondolták meg magukat később. Azonban, ha figyelembe vesszük a mondattal kapcsolatos kontextust, miszerint a tánciskola megemelte a diákoknak a napijegy árát, valószínűbb, hogy a tánciskola gondolta meg magát és döntött úgy, hogy visszavonja az áremelést. Az indoklás pedig az, hogy általában érdekük a diákoknak vonzó árakat kínálni annak érdekében, hogy sokan látogassák az oktatási intézményüket, és így növeljék a bevételeiket. Ha túl drágának tartják az árakat, könnyen elmennek más hasonló tevékenységet kínáló helyekre. Így, ha a tánciskola extra díjat vetne ki a diákokra, az “szembemegy” az ő érdekeikkel...



# Magyarázatot is kaptam

- Andi bulit szervezett Lilla tiszteletére, mert előléptették.
- Kit léptettek elő? (Andit / Lillát)
- helyes válasz: Lillát
- negyedik kísérlet: Mivel a mondat rögzíti, hogy "Lilla tiszteletére" építette az Andi, ezért arra lehet következtetni, hogy Lillát léptették elő, és az építés Andi tiszteletének szimbóluma. Az indoklás itt a mondat szerkezete és a logikus összefüggések alapján történik.
- A disznó jobb, mint a juh, mert belőle nem készítenek szalonnát.
- Miből nem készítenek szalonnát? (a disznóból / a juhból)
- helyes válasz: a juhból
- negyedik kísérlet: A disznóból nem készítenek szalonnát. A mondat magyarázata az, hogy a szalonna húsos része a disznó hasán található zsírréteg, amelyet speciális módon sóznak, füstölnek és sütnék. A juh sokkal soványabb állat, és kevés zsírt tartalmaz. Ezért nem lehet szalonnát készíteni belőle.

	1. kísérlet	2. kísérlet	3. kísérlet	4. kísérlet
jó	66,37%	70,62%	79,29%	74,69%
nem tudja	11,86%	2,48%	3,89%	4,42%
rossz	15,75%	24,01%	13,45%	16,64%
általános	5,84%	2,65%	3,19%	4,07%

Irodalom

---

- Jurafsky 3rd edition 7. chapter:  
<https://web.stanford.edu/~jurafsky/slp3/7.pdf>
- Francois Chollet: Deep Learning with Python. Manning, Shelter Island, 2018.:  
<https://www.manning.com/books/deep-learning-with-python>