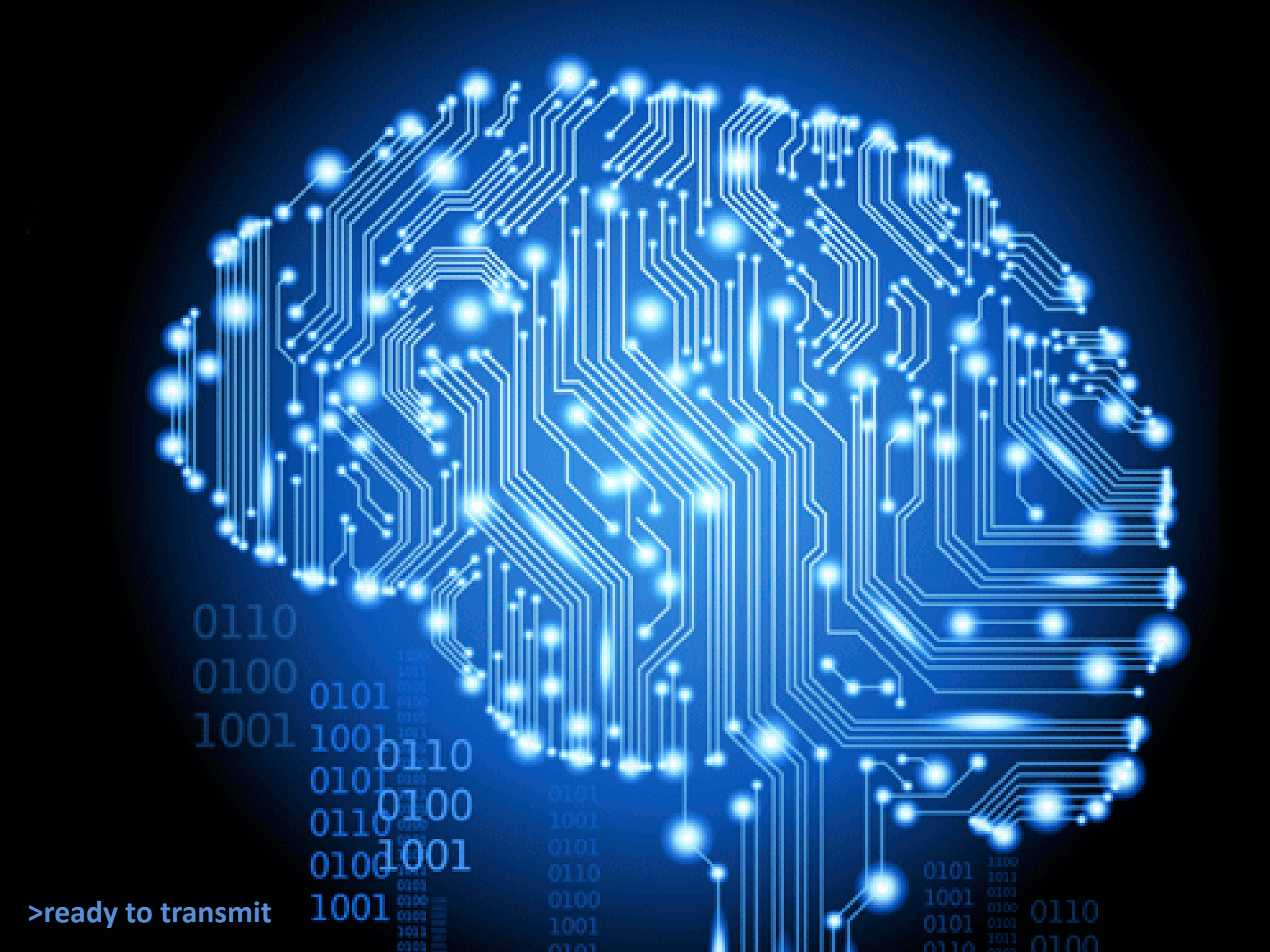

>ready to transmit

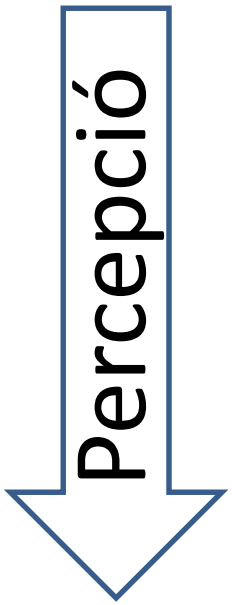
Rekurráló neurális hálózatok, beszédfelismerő programok

Szoldán Péter

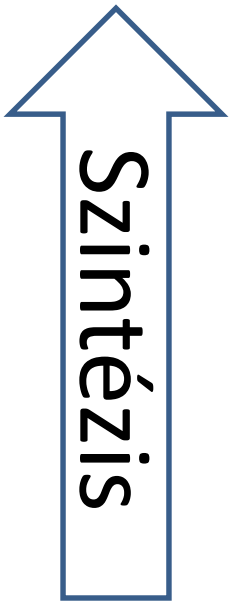
Új facebook csoport

- https://www.facebook.com/groups/496484064185906
- Ha kérdéseik vannak, ott is szívesen megválaszoljuk
- Érdekes újdonságokat, cikkeket is megosztunk

Verbális kommunikáció



- Hang
- Betűk, szavak
- Értelem

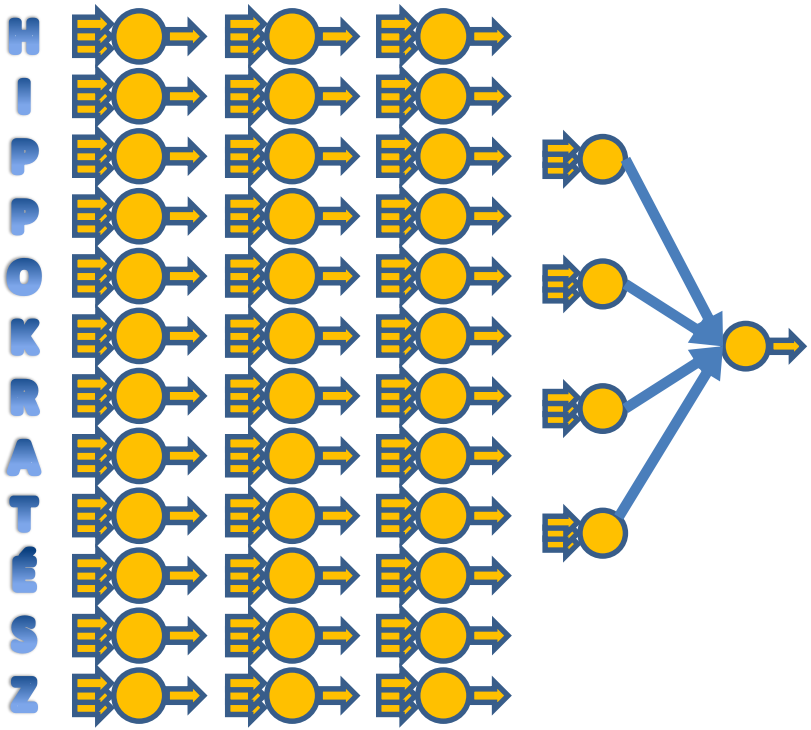


Szövegfeldolgozás

- A szöveg rengeteg karakter egymás utáni sorozata
- Minden karakternek van egy száma, például az „A” betű kódja 65. „B”-é 66, és így tovább. Ezt ASCII kódnak hívjuk („American Standard Code for Information Interchange”)
- Az ékezetes (magyar, és más nyelvek) karaktereknek speciális kódja van, és a programoknak ezzel néha nehézsége támad

Hálózat szövegre – naív megoldás

- Ha van például egy 12 karakteres szöveg, csinálhatnánk egy hálózatot 12 bemeneti neuronnal...



Karakterszámtól független megoldás kéne

- Ha több a karakter, mint a bemeneti neuron, mit csinálunk velük?

- Ha kevesebb a karakter, mint a bemeneti neuron, akkor ki kell tölteni a maradék helyet szóközzel, vagy valami mással, de ez rontja a hálózat teljesítményét, mert a szóközöket is megtanulja

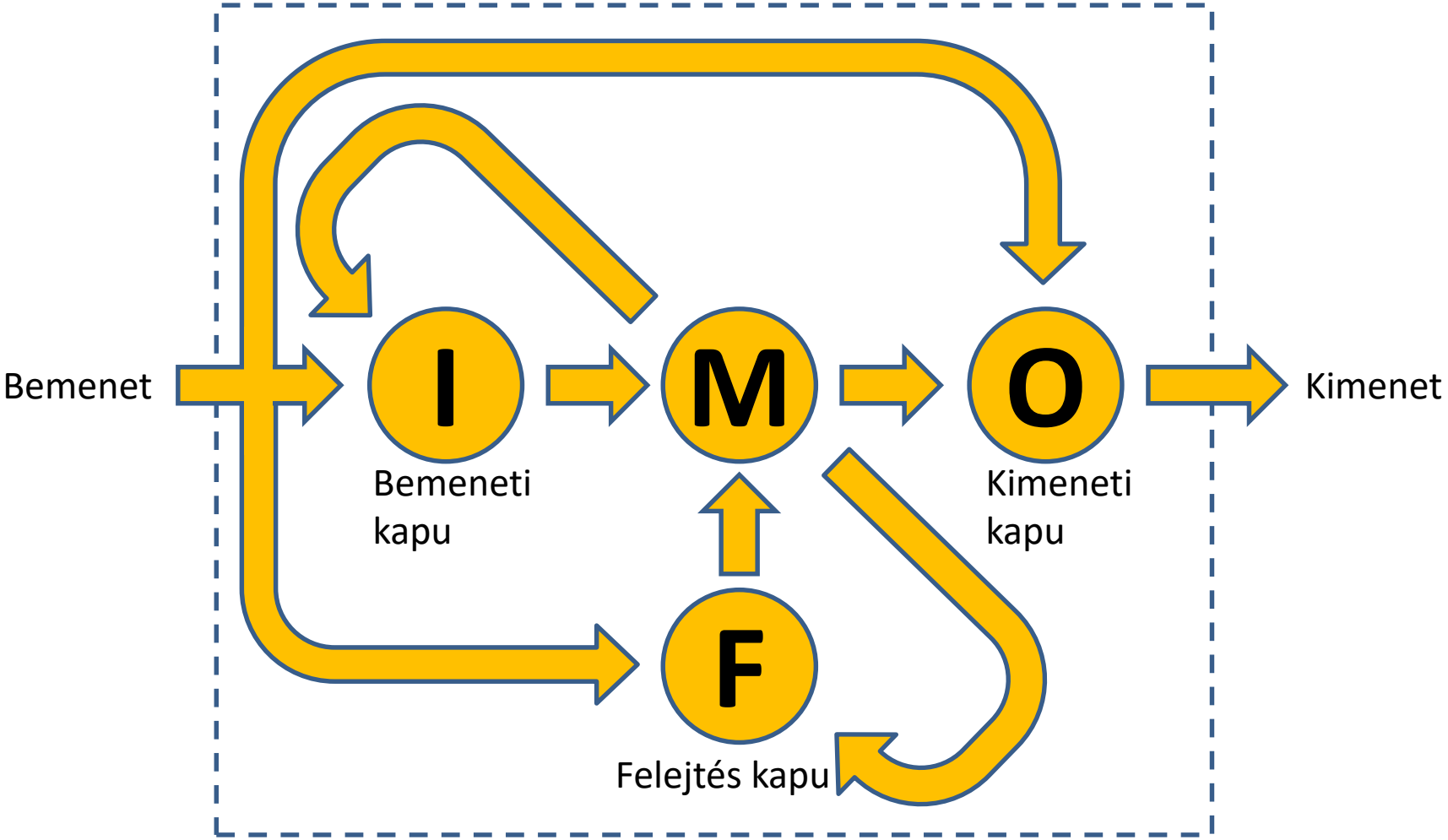
A rekurrens hálózatok (RNN)

- Adjuk be egyesével a karaktereket, és futtassuk ugyanazt a hálózatot újra és újra, amíg ki nem fogyunk a karakterből!
- Ez így még nem tökéletes, mert csak külön fogja értelmezni az egyes karaktereket
- Kéne valami memóriát gyártani neki, hogy megjegyezze, mi volt korábban

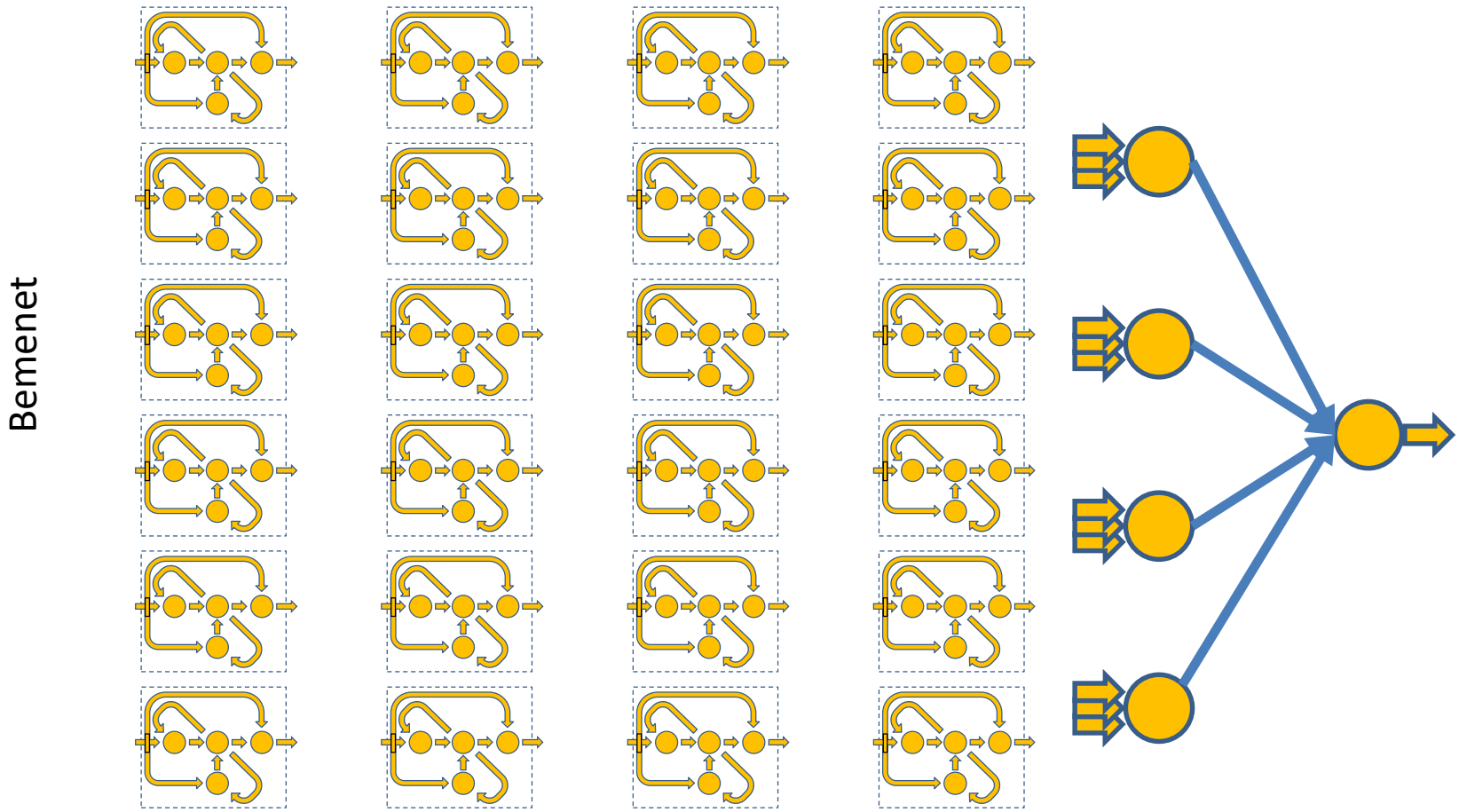
Ötlet memória egységre

- Építsünk egy több neuronból álló memória-egységet!
- Legyen benne középen egy tároló, ami egy értéket tárol hosszabb távon
- Csináljunk neuronokat, amik eldöntik, hogy mit jegyezzünk meg, mikor felejtsük el, illetve mi legyen a kimenet: a tárolt érték, vagy a bemenet

Long Short-Term Memory (LSTM)



Ugyanúgy lehet ebből is építkezni



Ez már tud valamit kezdeni a szöveggel

- A memória meg tudja jegyezni a korábbi részeit a szövegnek, és kontextusba tudja helyezni
- Nagyon jó a szintaktika (helyesírási szabályok, zárójelezési szabályok) felismerésében
- De nehéz neki nagyobb struktúrákat áttekinteni, mert a memória nem tudja, hogy milyen régen volt ott az adat

Szöveg hangulatelemzése

- Automatikus hírolvasó tőzsdei felhasználáshoz:
  - Kideríti, melyik tőzsdei cégről (vagy országról) van szó
  - Milyen a szöveg hangulata? Bizakodó, semleges, vagy negatív?
- Ez alapján lehet kereskedni: részvényt, kötvényt venni vagy eladni

Szöveg generálás

- Egy érdekes applikáció a szöveg generálás egy adott szerző stílusában
- Andrej Karpathy 2015-ös cikke[1] ad erre néhány érdekes példát
  - Próza Paul Graham stílusában
  - Dráma Shakespeare stílusában
  - Wikipedia szócikk

[1]Andrej Karpathy: The Unreasonable Effectiveness of Recurrent Neural Networks http://karpathy.github.io/2015/05/21/rnn-effectiveness/

AI Wikipedia: szintaktika pontos

'''See also''': [[List of ethical consent processing]]

== See also ==
*[[Iender dome of the ED]]
*[[Anti-autism]]

===[[Religion|Religion]]===
*[[French Writings]]
*[[Maria]]
*[[Revelation]]
*[[Mount Agamul]]

AI Shakespeare: értelem hiányzik

CASSIUS:
And in desire,
And call'd me ballant Cassius.

BARDOLPH:
Tost in it, what then take your madder?

DUKE OF YORK:
She would be ready, this advice, say you a chaste.

Second Neris:
Now, blessed France, and with thy speech can know?

AI film: Sunspring

- Youtube-on megtalálják: https://youtu.be/LY7x2Ihqjmc
- Érezhető rajta, hogy bizonyos kliséket megtanult
- De átfogóan egy érthetetlen zűrzavar az egész
- Aki szereti a groteszket, jól fog rajta szórakozni

Fordító AI („bábel hal”)

- Sokkal nehezebb feladat, a szöveg értelmét kell átadni
- Szintaktika kell, de az még édeskevés
- Azzal kísérleteznek, hogy nem betűket, hanem szavakat jelölnek kóddal
- Egész pontosan vektorokkal, „word2vec”

Wikipedia több nyelven elérhető

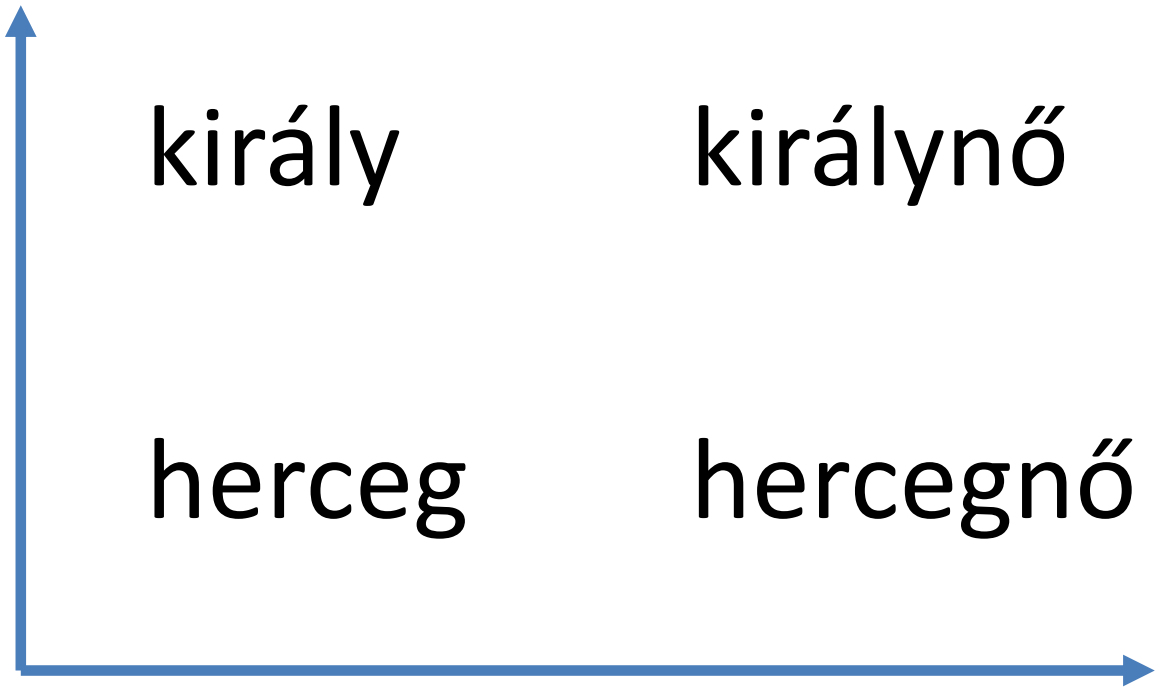
- Stefan Jansen 2017. májusban közzétett cikke[1]
- Szócikkeken lehet tanítani a hálózatot (persze ellenőrizni kell, hogy a hossza és szerkezete a cikkeknek hasonlít)
- Először beágyazta a szavakat minden nyelven
- Majd a beágyazáson tanította a hálózatot fordítani: 30-40%-os pontosságot ért el

[1]Stefan Jansen: Word and Phrase Translation with word2vec, https://arxiv.org/pdf/1705.03127.pdf

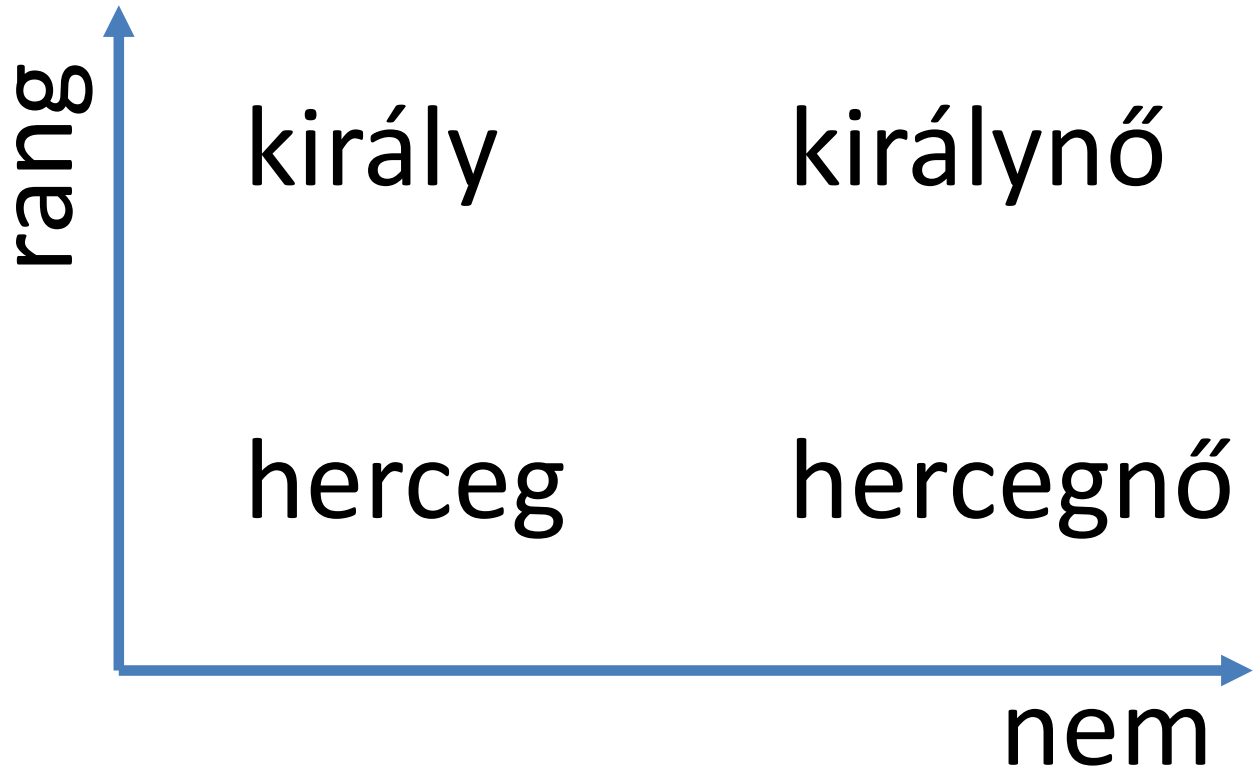
Szó jóslás a környezetéből

- A hálózat feladata az, hogy egy (kitakart) szót kitaláljon a szövegkörnyezetből
- „Elizabeth's early years were marked by constant fluctuations of fortune.”
- Meglepően pontos struktúráját építi fel a nyelvnek ez az egyszerű feladat

Beágyazás (embedding)



50-100 jelentés dimenzió

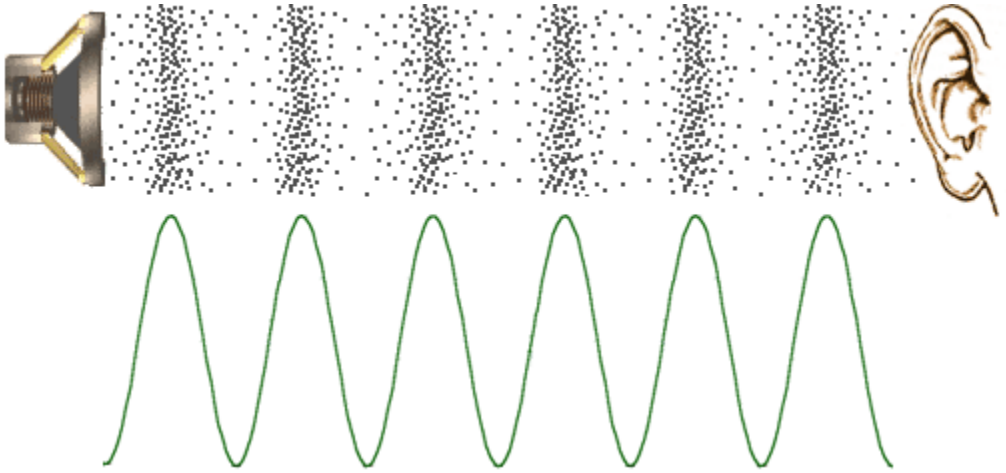


Egyenlet formában is működik

Királynő = király – férfi + nő

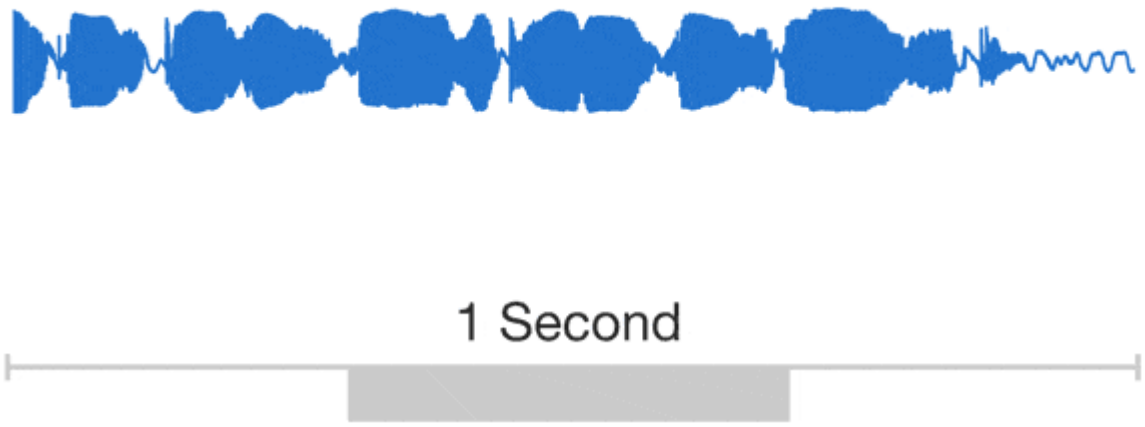
- Hasonló struktúrák alakultak ki mind a négy nyelven
- Ez nagyon jó kiindulópont volt a fordító program megalkotására
- 30-40%-os pontosságot értek el

Hanghullámok

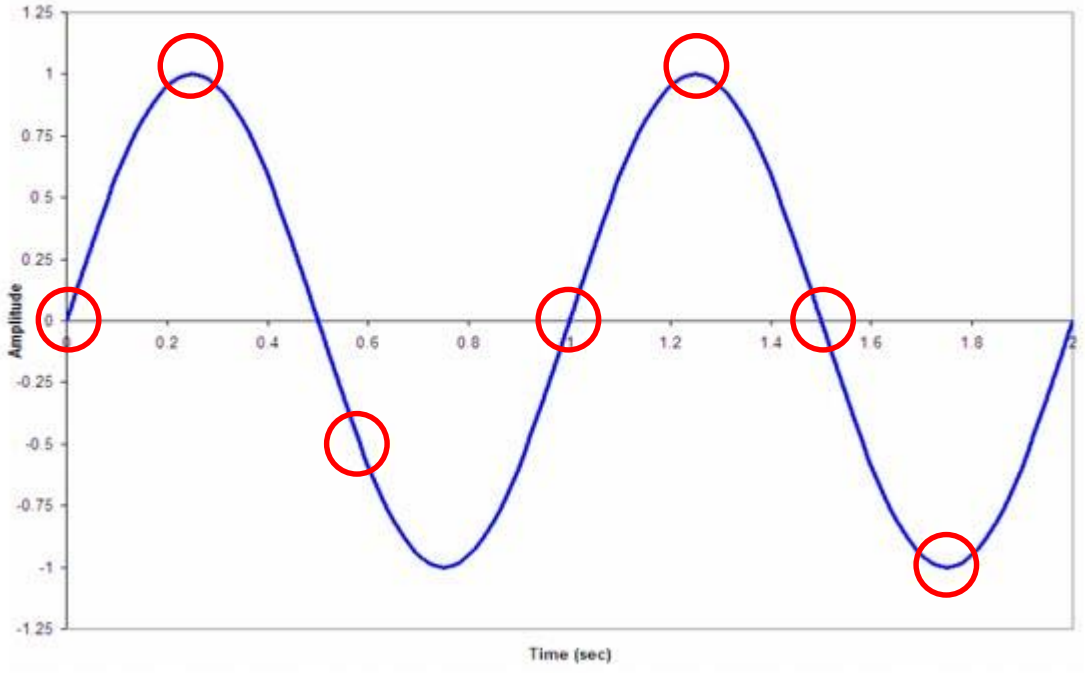


- Sűrűsödő-ritkuló levegő, a változás terjed
- Minél gyorsabban sűrűsödik-ritkul, annál nagyobb a frekvencia, magasabb a hang
- https://youtu.be/qNf9nzvnd1k

Különböző zoom szinteken



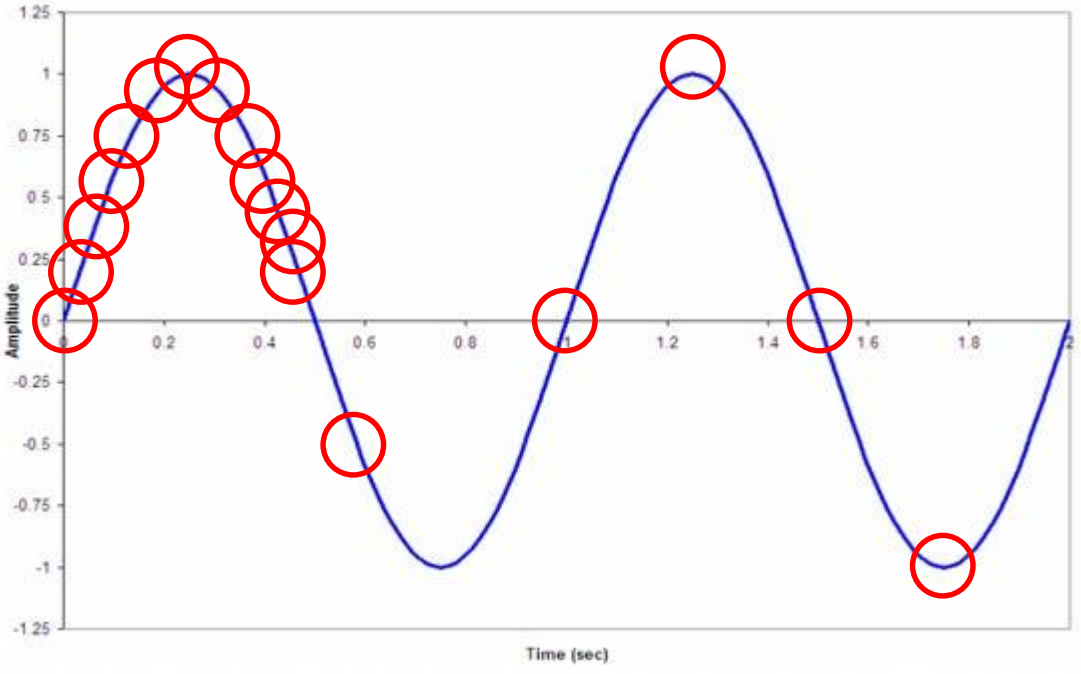
Konvertáljuk számokká



0 1 -0,5 0 1 0 -1

Mintavételezési frekvencia

- Minél több mintát veszünk, annál hűségesebb a hang, de több adatot kell rögzíteni

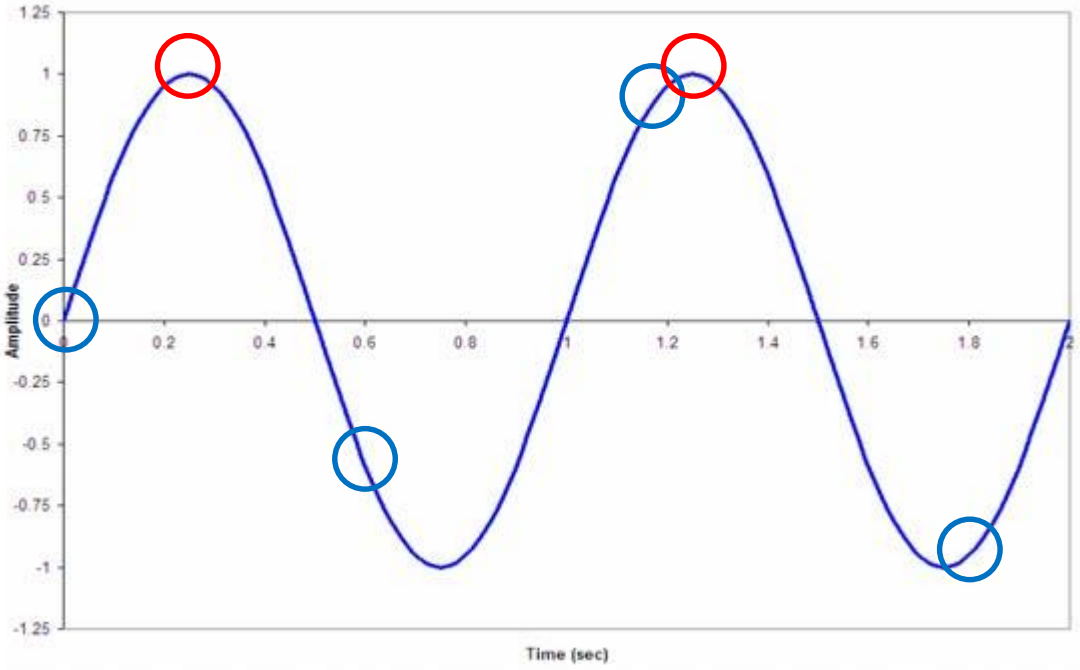


- Emberi fül 20kHz-ig működik
- 44.1kHz a zenei CD-k mintavételezése

Frekvencia meghatározása

- A hallás első lépése a hallott zajok, zörejek, beszéd, zene fekvenciájának meghatározása („spektrum analizálás”)
- Ebben a fül nagyon jó, ha több hang is hallatszódik egyszerre, akkor is
- Hogyan tudja ezt egy hálózat vajon megcsinálni?

Az 1D konvolúció

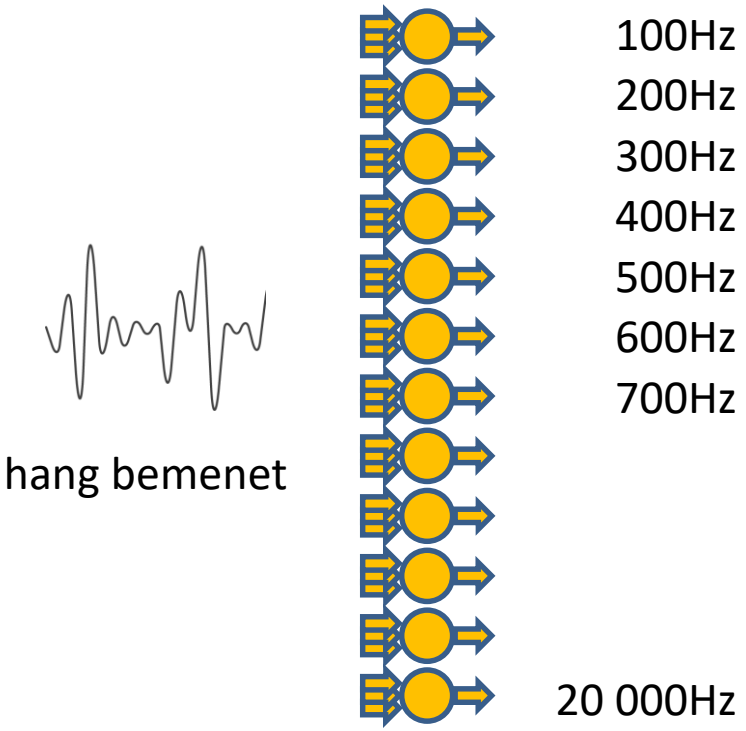


- Különböző neuronok különböző időközönkénti mintákat adnak össze

0 0 1 0 0 0 0 0 1 0 0 0 fázisban van: az összeg 2

1 0 0 0 1 0 0 1 0 0 0 1 nincs fázisban: az összeg kb. 0

Spektrum analízis az első réteg után



A beszéd (és zene) struktúrája

- Az, hogy egy pillanatban milyen hangok szólnak, csak az első lépés
- A beszédnek, csakúgy, mint a zenének, időben hosszan elhúzódó struktúrája van
- Ahhoz, hogy 5 másodperces struktúrát felderítsünk, 44.1kHz esetén 220,000 bemenet kellene neurononként

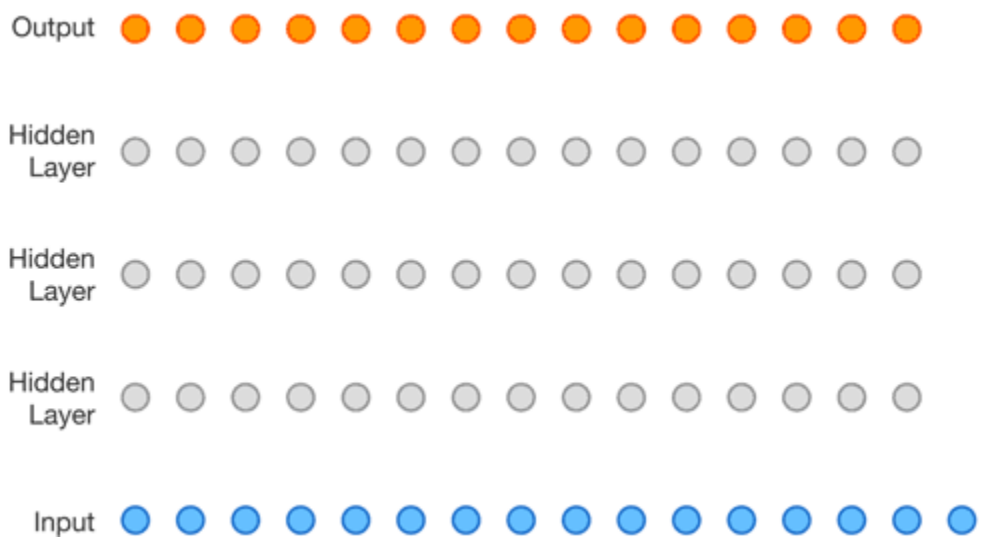
Konvolváljuk a konvolúciókat!

- Az egymás után fűzött konvolúciós rétegekkel keressünk egyre magasabb struktúrákat a hangban
- Például a felharmonikusok határozzák meg a hangszínt: a második réteg már fel tudja ismerni, hogy ki beszél
- Vagy milyen hangszer(ek) játszik(-anak)

DeepMind WaveNet

- Keressünk egyre magasabb struktúrákat a hangban
- Például a felharmonikusok határozzák meg a hangszínt: a második réteg már fel tudja ismerni, hogy ki beszél
- Vagy milyen hangszer(ek) játszik(-anak)

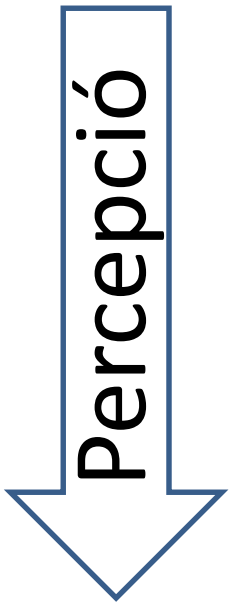
Kevés művelet, nagy átlátóképesség



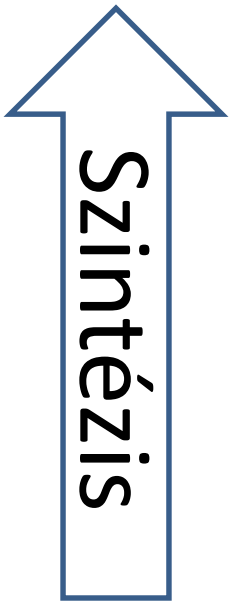
Beszéd szintetizálás

- Hagyományos beszédszintetizálás: 
- WaveNet beszédszintetizálás: 
- Zongoramű: 

Verbális kommunikáció

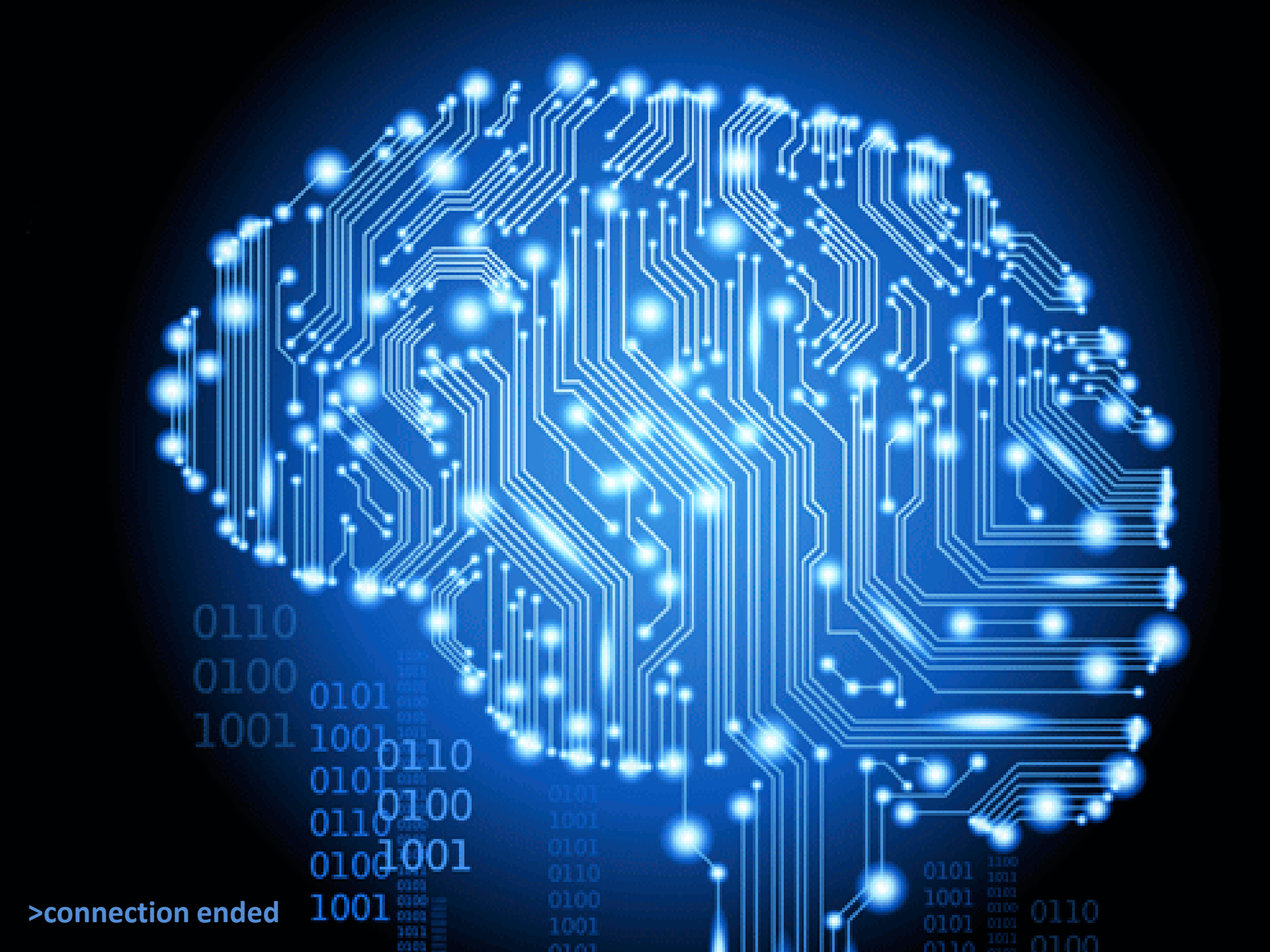


- Hang
- Betűk, szavak
- Értelem



Google Assistant

https://youtu.be/D5VN56jQMWM?t=70


>connection ended