Historie neuronových sítí

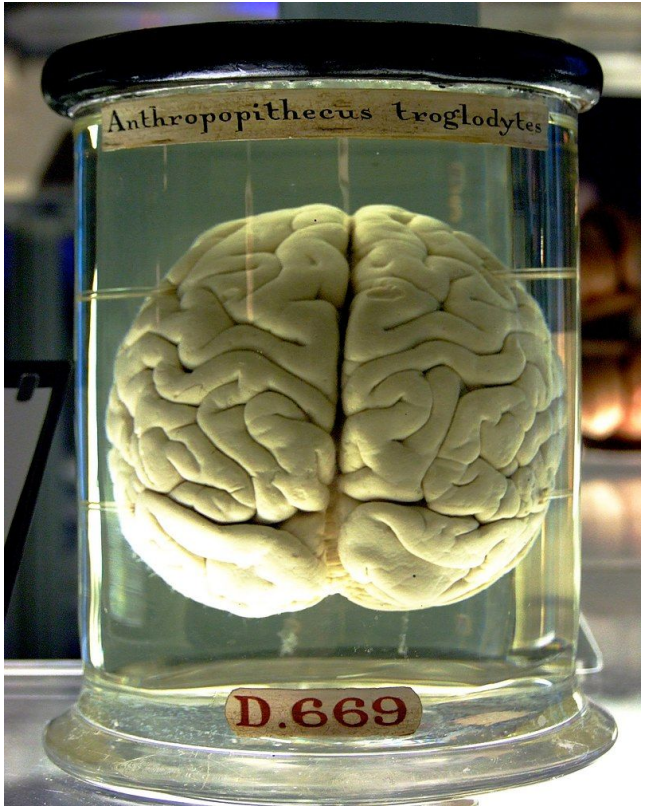
Od neuronů k umělé inteligenci

Tomáš Brázdil

RationAI

MUNI

Co jsou neuronové sítě?



Pravé vs umělé

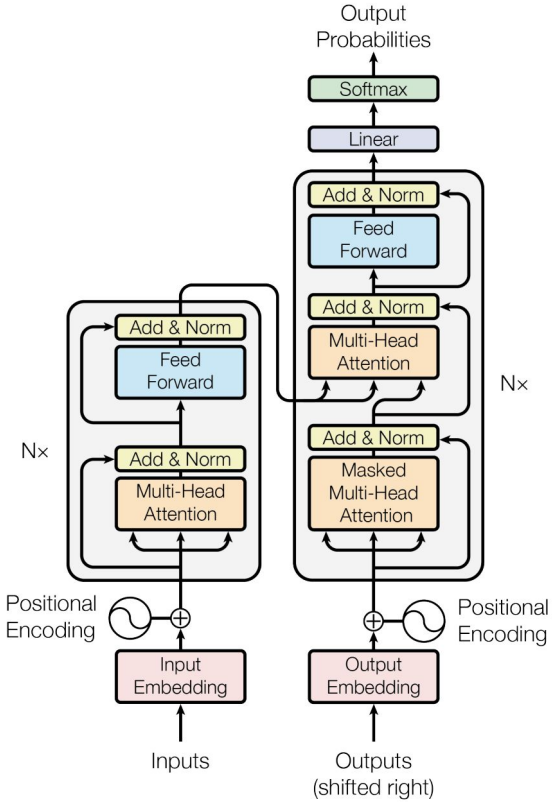
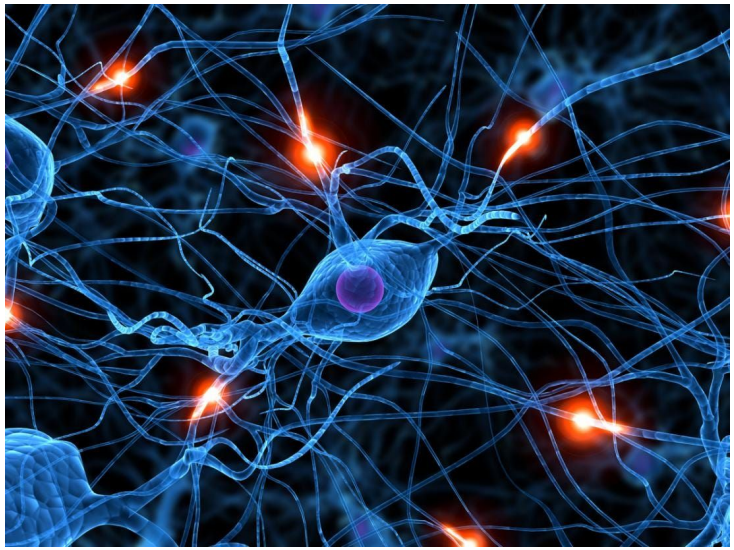
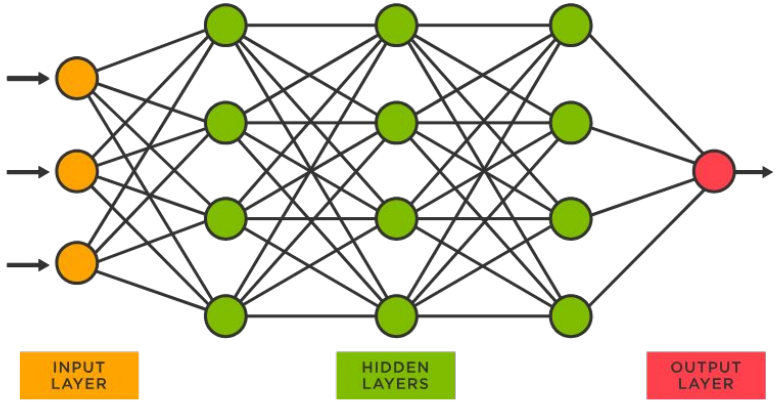


Figure 1: The Transformer - model architecture.

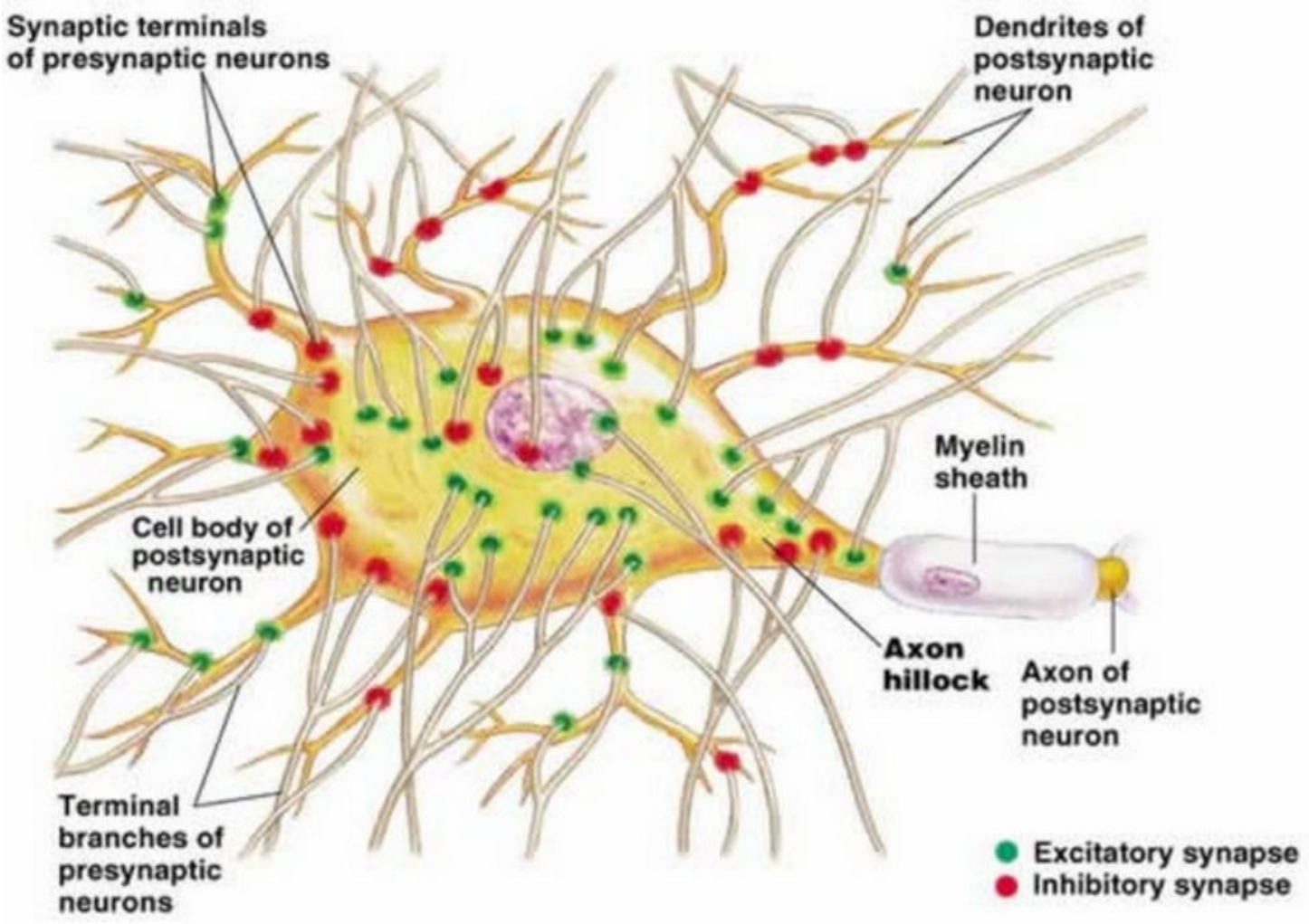
Neuronová síť



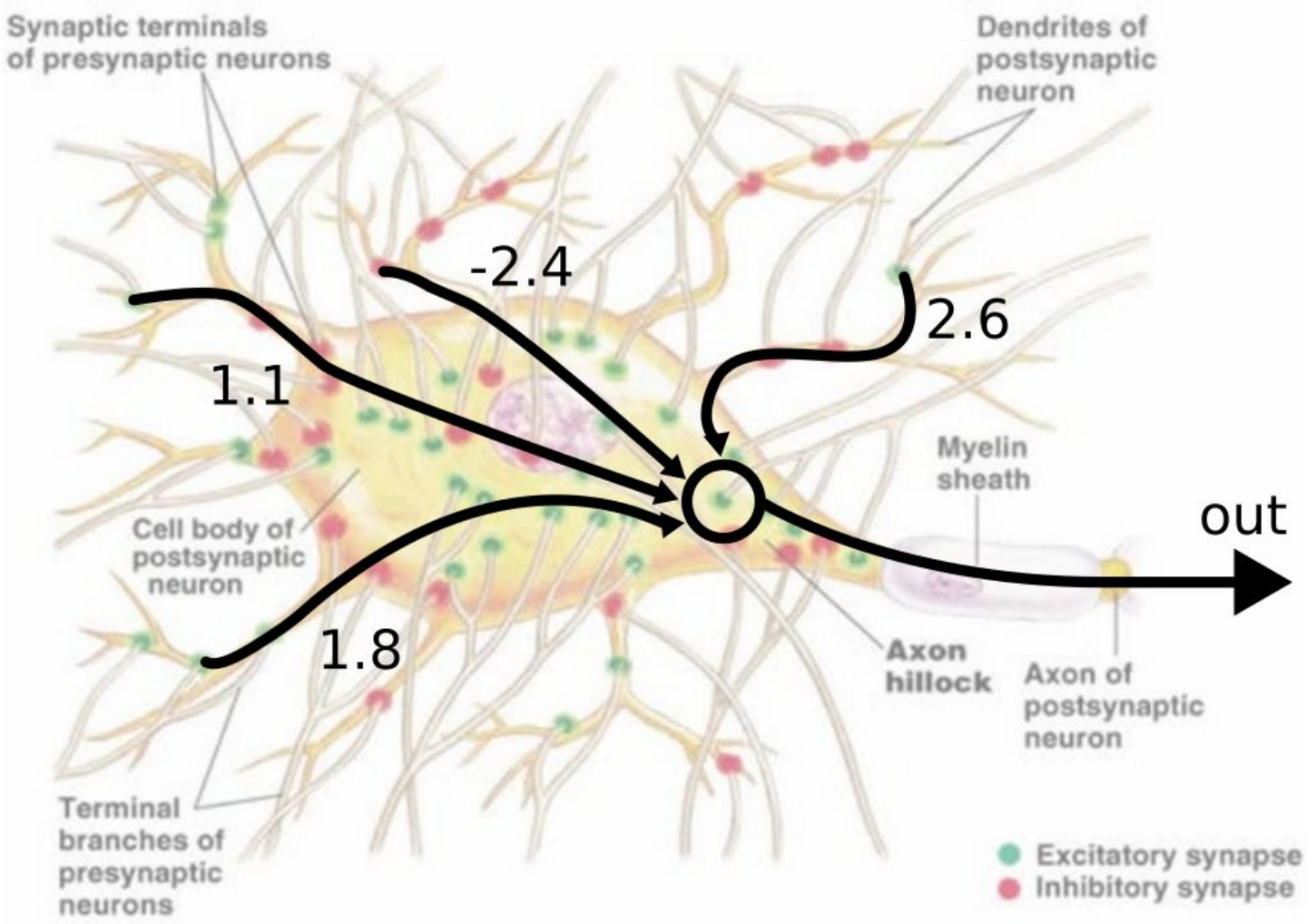


Co to umí?

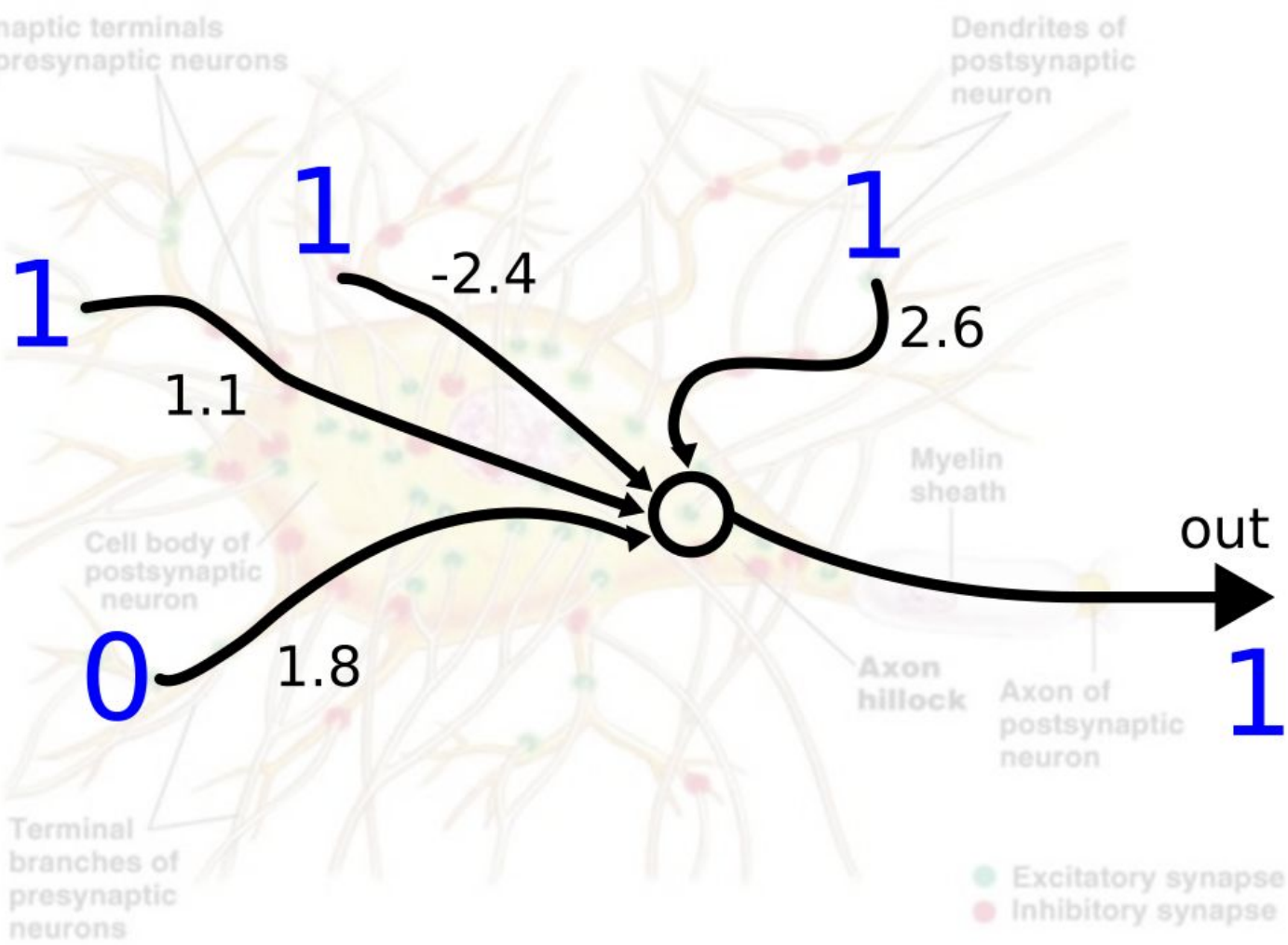
Biologický Neuron



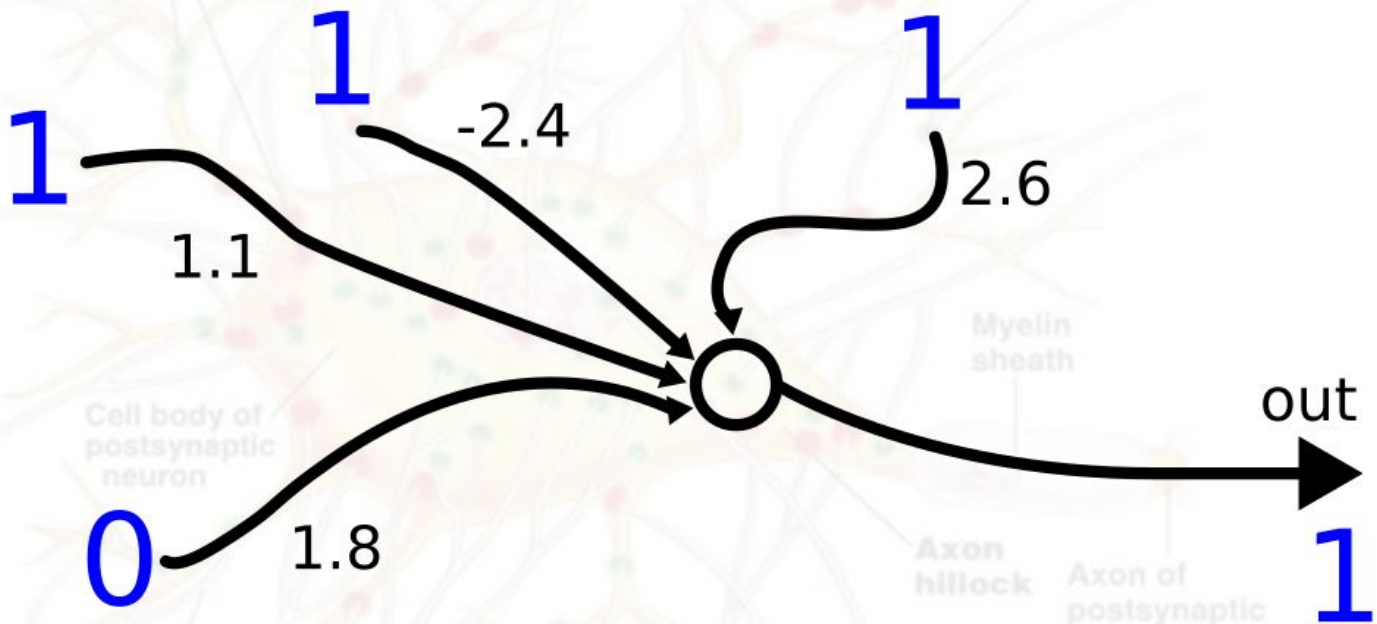
Biologicko
Matematický
Neuron



Biologicko
Matematický
Neuron



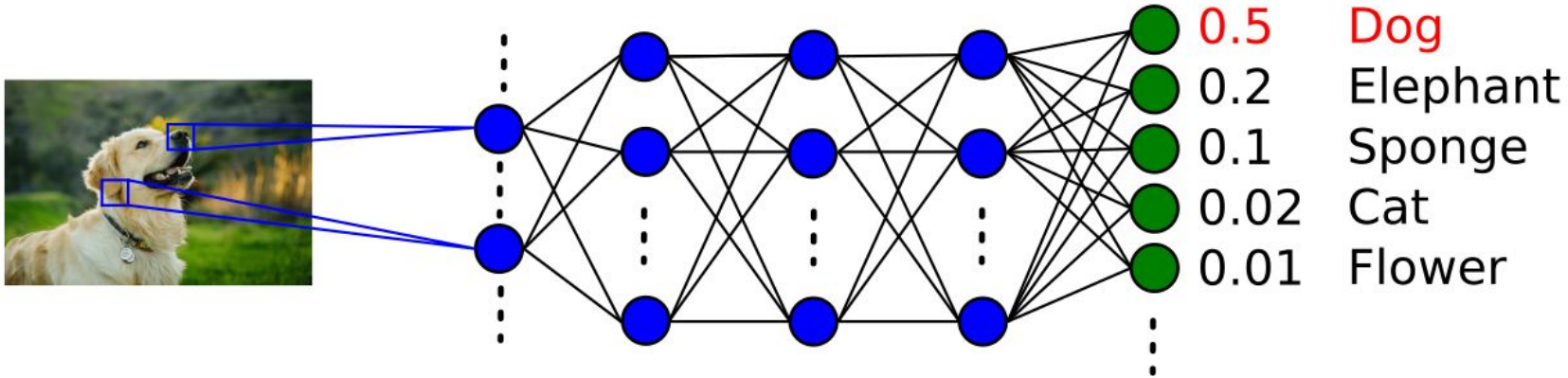
Matematický Neuron



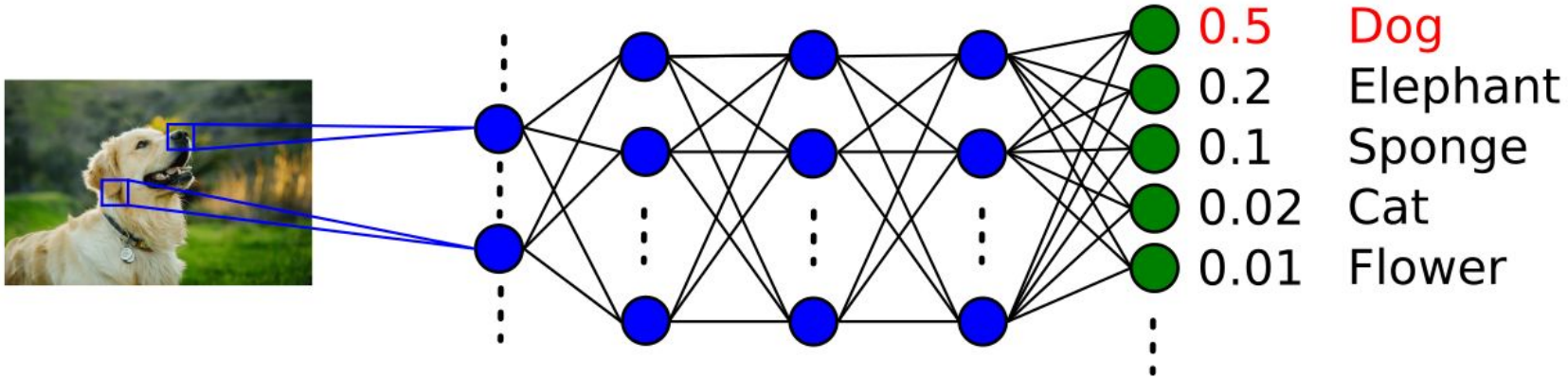
out = 1 protože

0*1.8 + 1*1.1 + 1*(-2.4) + 1*2.6 > 0

Co to umí: Rozpoznávat psy! (i jiná zvířátka)



Co to umí: Rozpoznávat psy! (i jiná zvířátka a věci)



Síť se samostatně učí na obrovském množství dat tvaru:

[ , Dog], [ , Cat], [ , chair] …

Historie neuronových sítí

… aneb variace na frustrační kompozici Járy Cimrmana

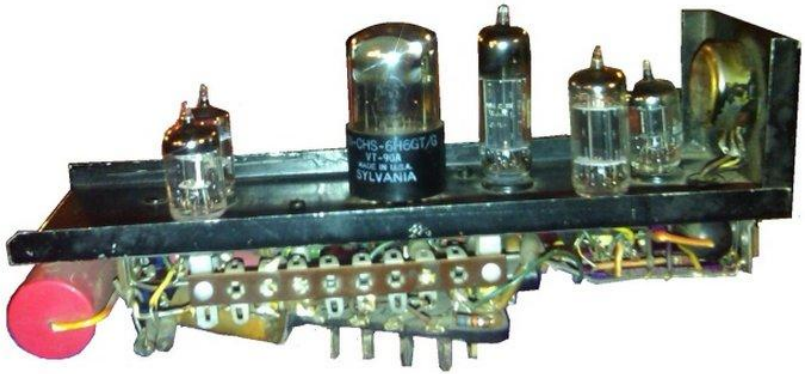
Střídají se prvky očekávání a zklamání

U nás

Padesátá léta

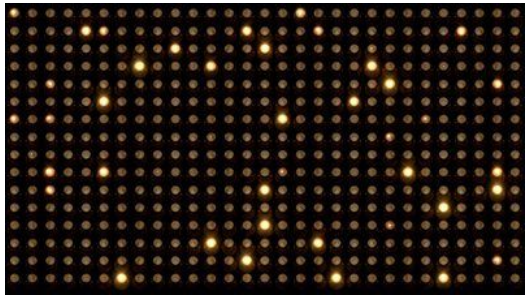
M. Minski - hardwarová implementace neuronových sítí

SNARC - neuron



Krysa implementovaná pomocí žárovek hledala cestu z bludiště

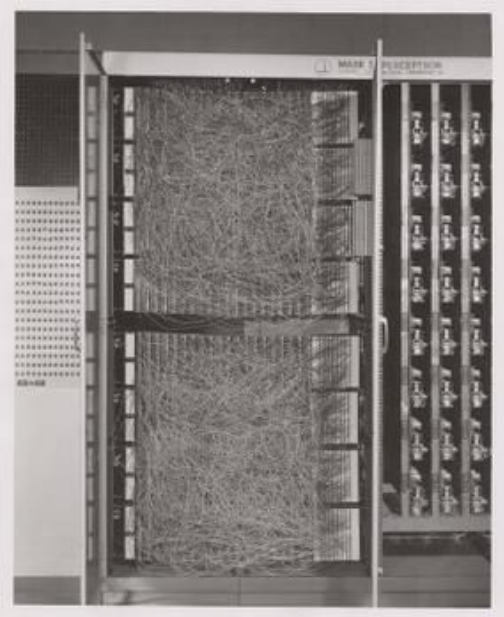
Posilované učení (reinforcement learning)



U nás

Šedesátá léta

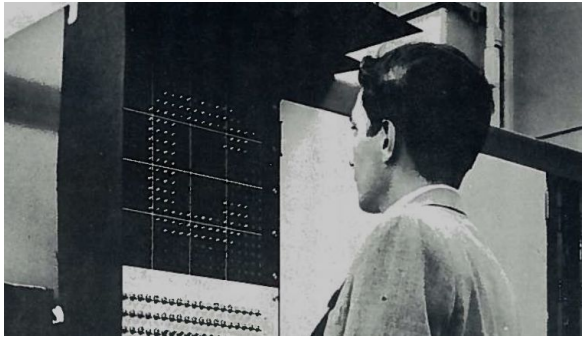
Perceptron (Rosenblatt) - jednovrstvá neuronová síť



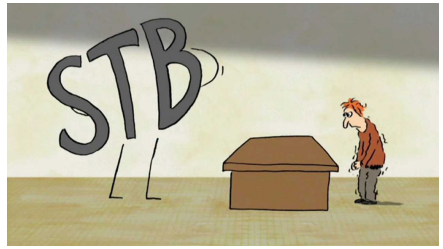
Veřejná demonstrace rozpoznávání znaků

Znaky reprezentovány foto-diodami

Váhy pomocí motorů ovládaných potenciometry

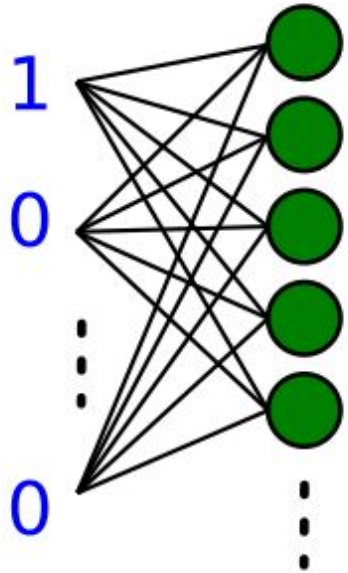


U nás



Sedmdesátá léta

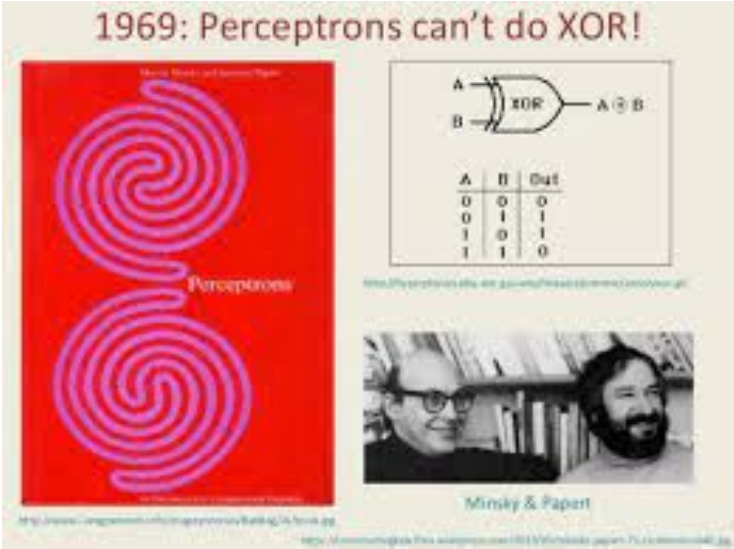
AI Winter!



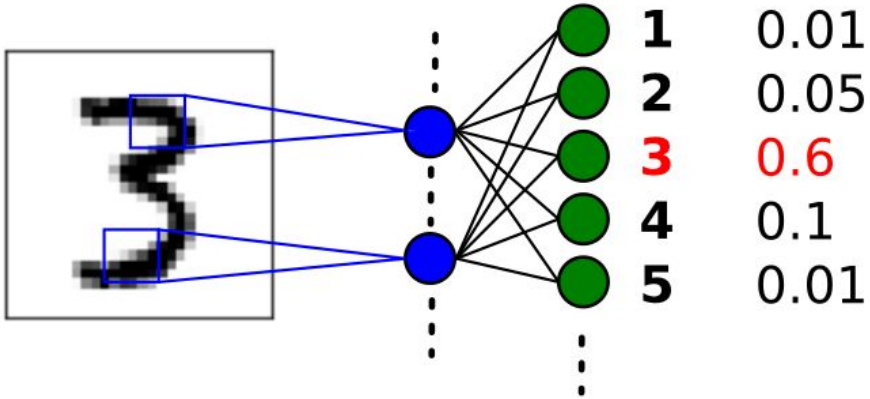
Jedna vrstva nestačí!

Víc se neumí trénovat!

Minski & Pappert



Osmdesátá léta



U nás

Umí se trénovat dvě vrstvy!

Gradientní sestup & backprop

Neuronové sítě se vyvíjejí v sw i hw verzích

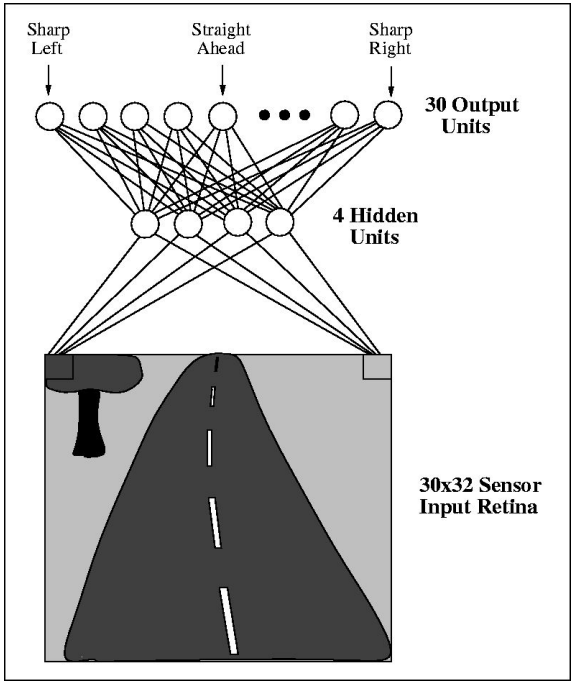
Končí AI zimou, krachem AI firem

Konec financování, akademici však statečně pokračují dál!

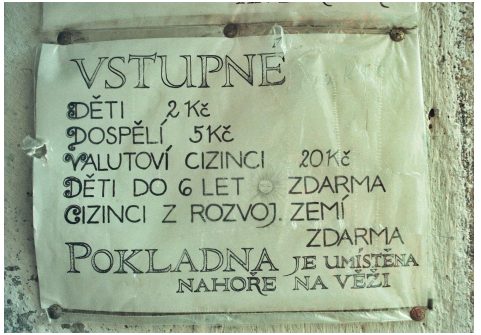
Konvoluční sítě (LeNet-1)

Devadesátá léta

ALVINN - self-driving car



U nás



Hardware se zlepšuje, oživení neuronek

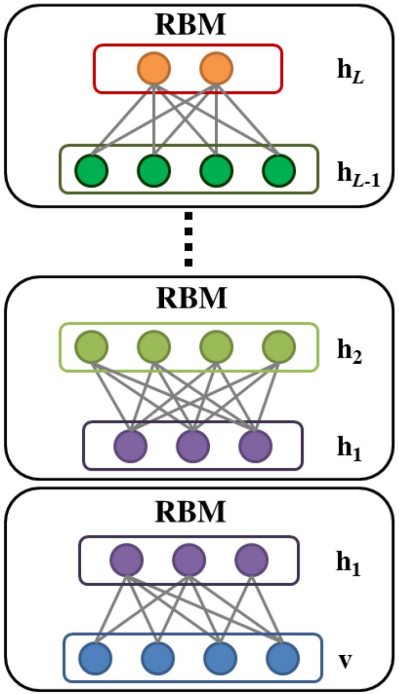
Stále se umí jen dvě vrstvy,
neuronky nejsou lepší než jiné metody!

Jak trénovat
***hluboké* sítě??**

ALVINN - self driving long long time ago

Nultá (???) léta

Hinton et al - Deep Belief Network - pokus o hlubokou síť



NN Winter!

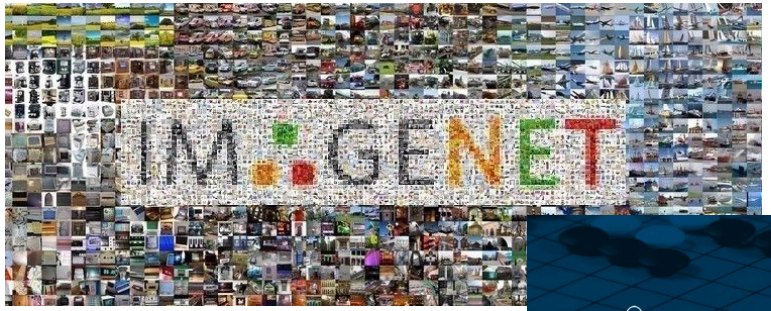
Snaha o více než dvě vrstvy!

Neuronové sítě udolány jinými metodami (SVM)

Málo peněz, pokračují převážně akademici
Ti pracují na všem …

Prudký rozvoj **grafických karet** v herním průmyslu!

Desátá léta



Google Translate

Konečně někdo zkusil použít

grafické karty pro trénink sítí!!

Obrovské sítě!

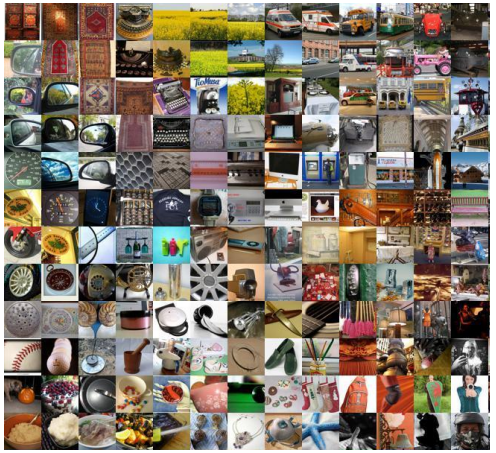
IT infrastruktura pro AI

Ztrhující výsledky v mnoha oblastech!

Prakticky použitelné je minimum, ale jsou prostě úžasné

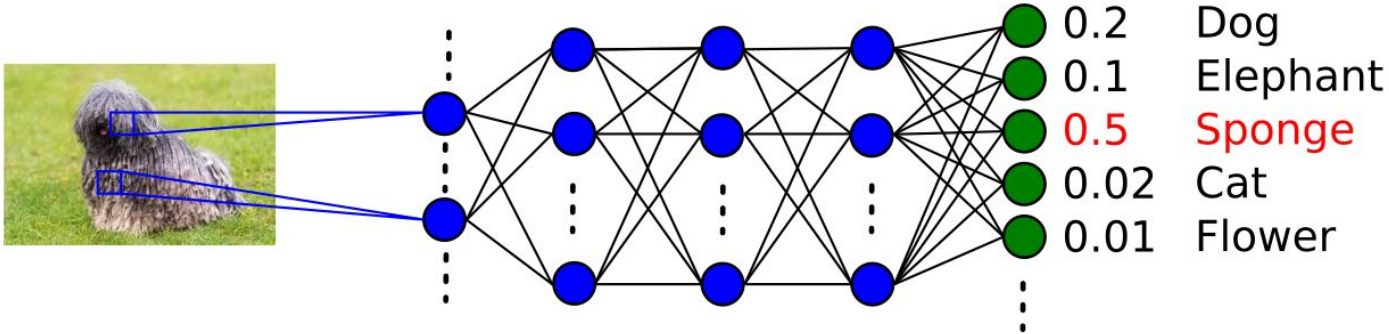
Šílenství kolem sítí začíná a roste!

Proč revoluce v AI? ILSVRC 2012



Soutěž v rozpoznávání obrázků

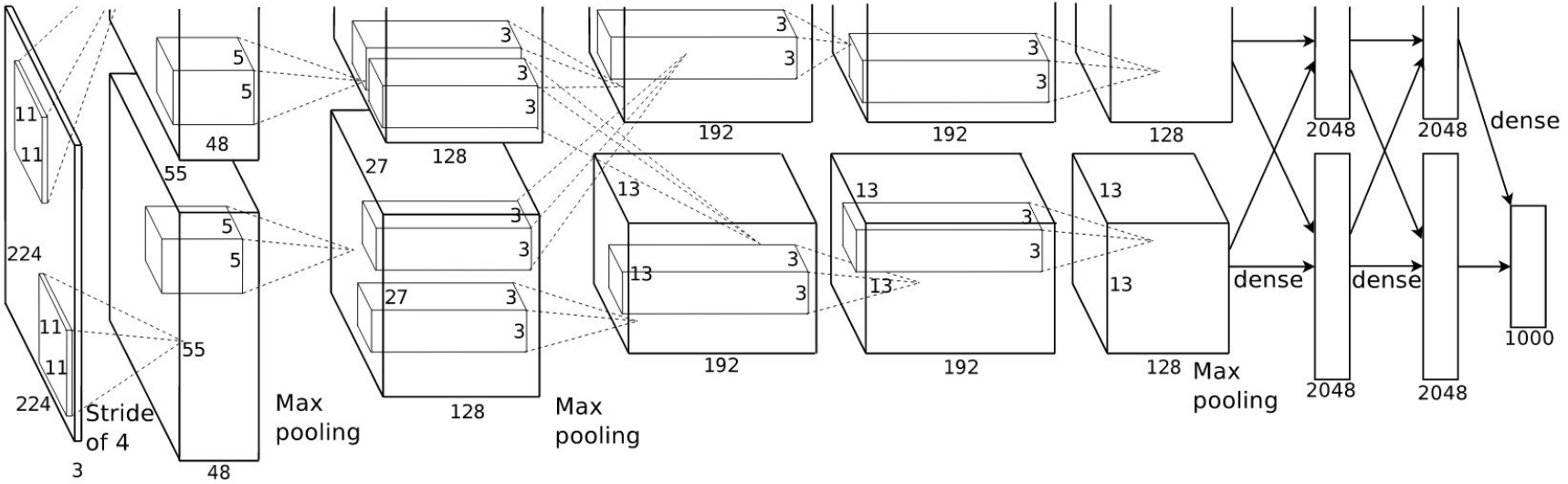
- 1 200 000 obrázků pro trénink
- 1 000 tříd (židle, pes, …)



Top 5 vyhodnocení = správná třída mezi prvními pěti

Vítěz roku 2012

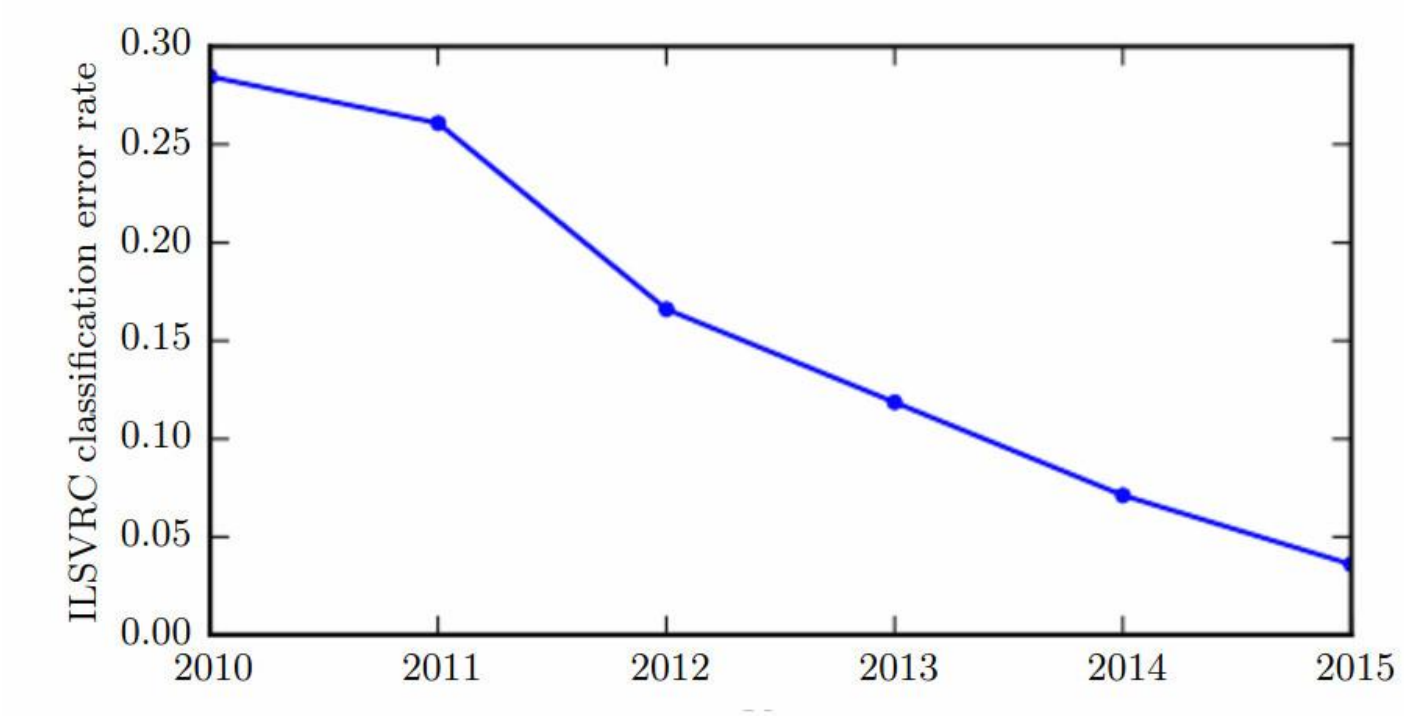
AlexNet



Chyboval v **15.3** procentech testovacích případů!

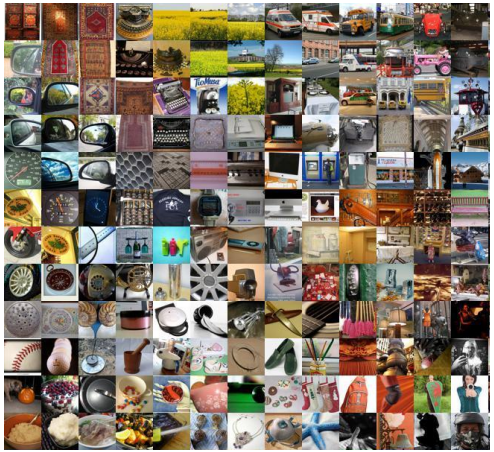
Model na druhém místě chyboval v 26.2 procentech!

Soutěž ILSVRC



2016: Chyba pod tři procenta … a dál už to moc nešlo …

Data pro ILVSRC



Ruční značení 1 000 tisíce tříd pro 1 200 000 obrázků!

Andrej Karpathy (šéf týmu pro přípravu dat):

“Uvažovali jsme zapojení Amazon Mechanical Turk nebo využití studentů”

“Automaticky jsme omezili možnosti na 100 tříd z 1000”

“I tak to bylo moc, studenti (i mechaničtí Turci) měli chybu 13 - 15 procent”

“Já osobně jsem chyboval v cca 7 procentech případů”

“Stal jsem se expertem na plementa psů, ptáků a opic”

Otázka: Jsou ta data vůbec dost kvalitní??

ILSVRC Data

Predict:
1 *pencil box*
2 *diaper*
3 *bib*
4 *purse*
5 *running shoe*

Ground Truth:
sleeping bag

ILSVRC Data

Predict:
1 *dock*
2 *submarine*
3 *boathouse*
4 *breakwater*
5 *lifeboat*

Ground Truth:
paper towel

Velký problém:

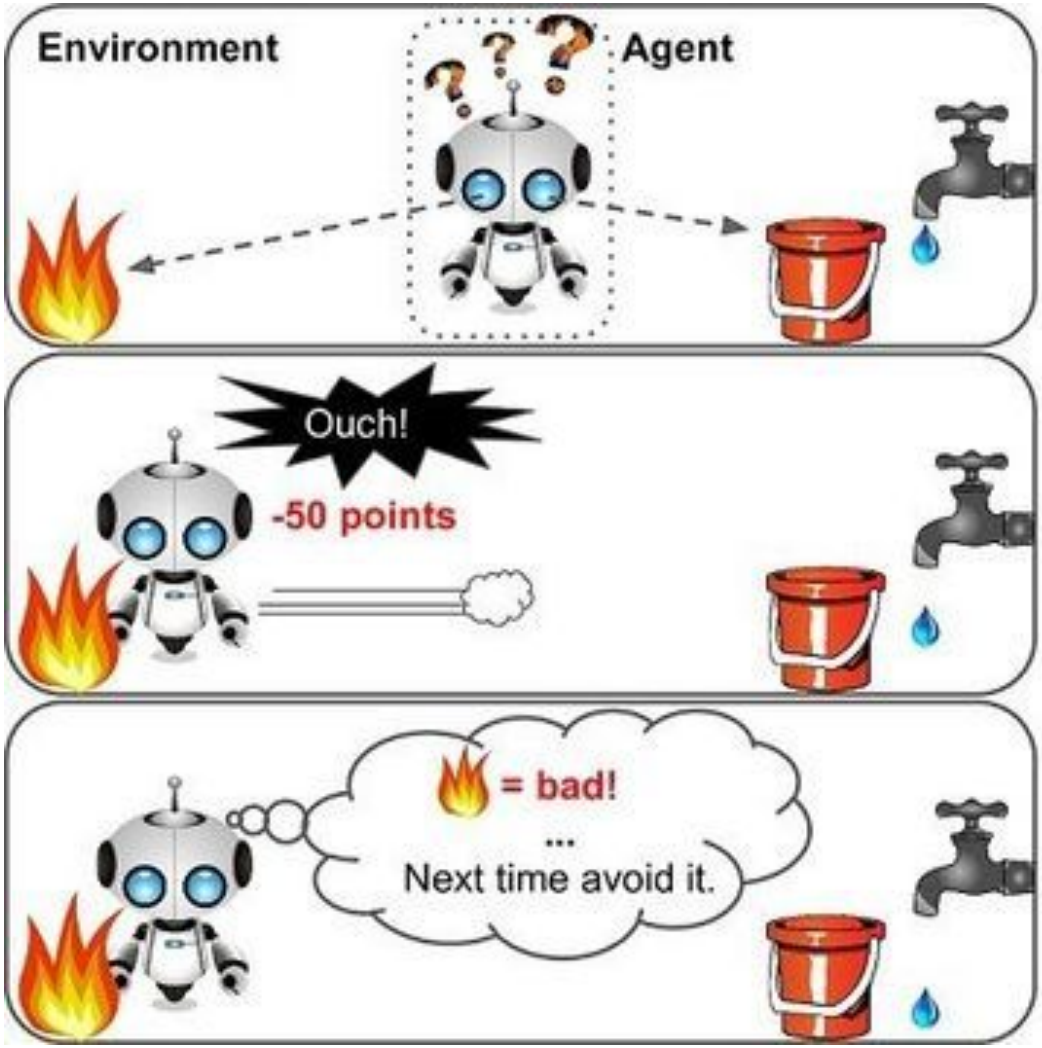
Kde vzít další
použitelná data?

- Větší datasety už ručně dělat nelze
- Jak trénovat bez jasné a správné zpětné vazby?
- Jak modelovat myšlení a rozhodování?

Toto je ilustrace bince v datech, aby bylo jasno

Posilované učení

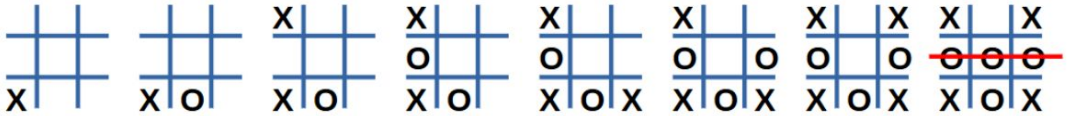
… a hry



1. Observe
2. Select action using policy
3. Action!
4. Get reward or penalty
5. Update policy (learning step)
6. Iterate until an optimal policy is found

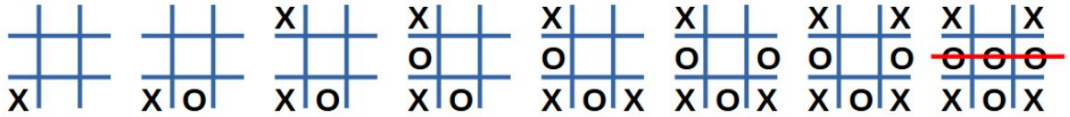
Jak počítač hraje piškvorky?

Zcela náhodná volba tahů:

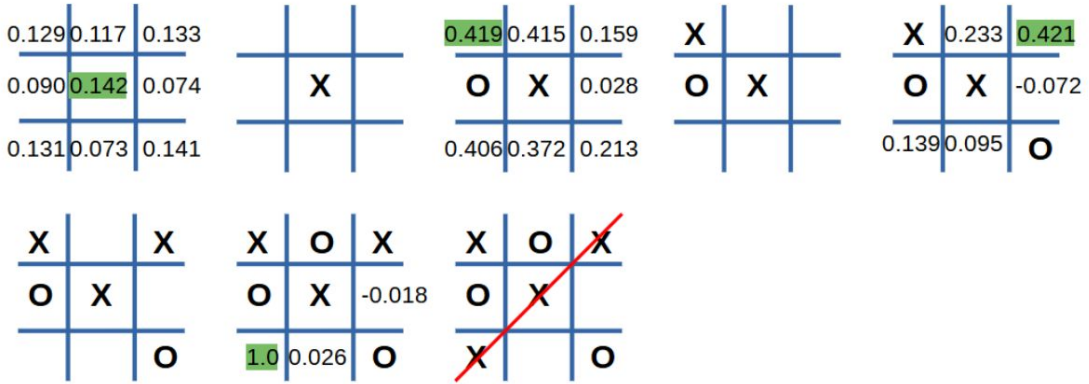


Jak počítač hraje piškvorky?

Zcela náhodná volba tahů:



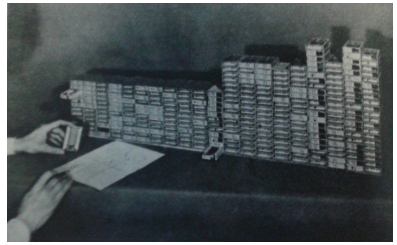
Člověk (o) proti stroji (x)

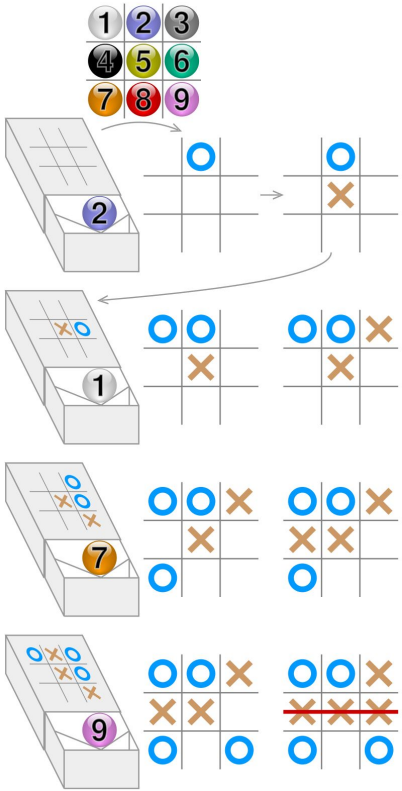


Učení:

- Stroj hraje za oba hráče, začne s nulou na všech tazích (náhodná volba)
- Po vyhrané/prohrané partii systém zvýší/sníží hodnoty tahů v odehrané partii

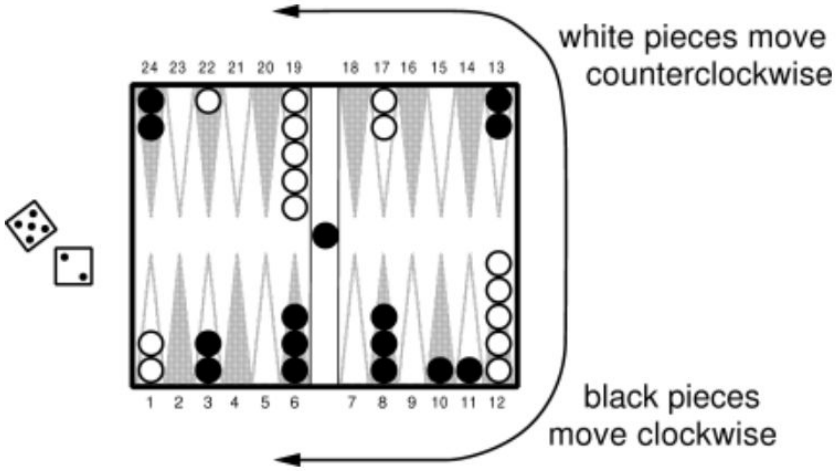
Jak to implementovat v roce 1961? MENACE!





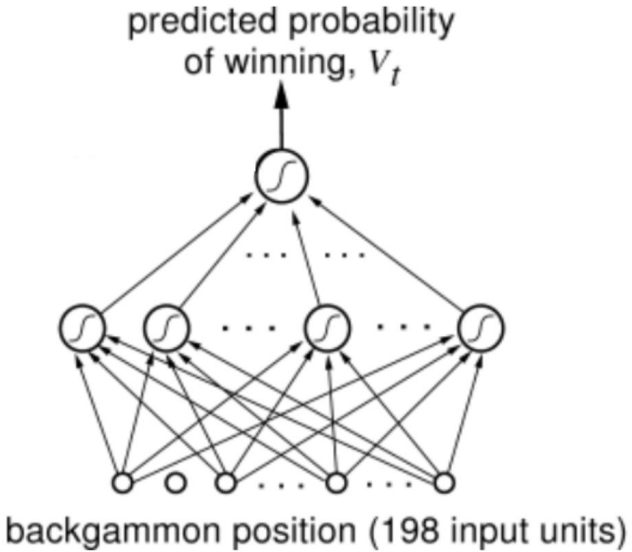
- Jedna krabička pro každou pozici
- Zatřeseme, jedna kulička “vypadne” Táhneme podle ní
- Použité krabičky a vybrané kuličky si dáváme stranou během hry
- Učení:
  - Na začátku je v každé krabičce stejný počet kuliček všech barev odpovídajících volným polím
  - Opakovaně hrajeme hru od začátku do konce
    - Pokud MENACE vyhraje, vrátíme kuličky zpět a přidáme 3 další stejné barvy jako byly vybrané
    - Pokud MENACE prohraje, nevrátíme vybrané kuličky

Přidáme neuronku - TD-Gammon (Tesauro, 1992)

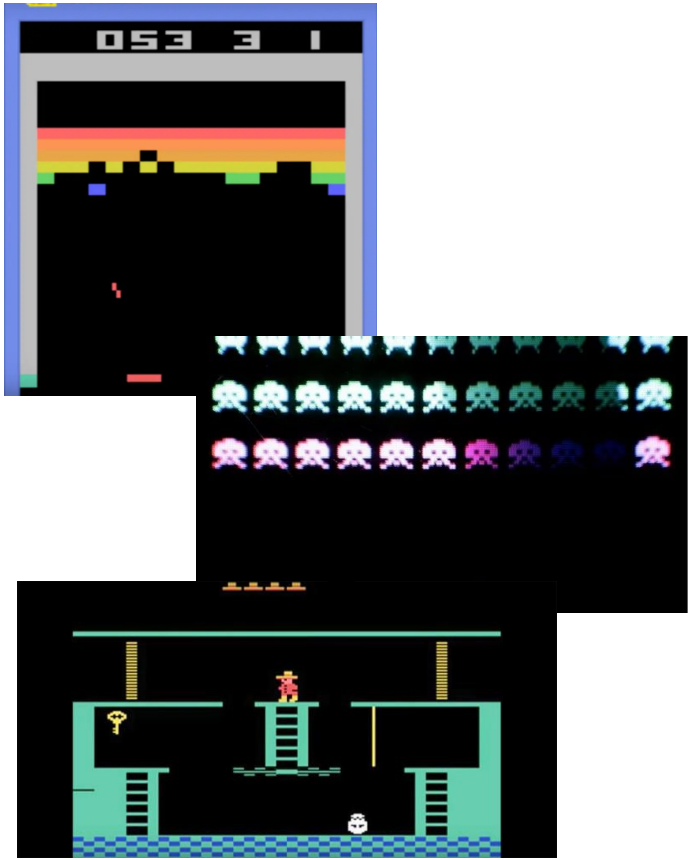


Neuronka hraje sama proti sobě

Dosahuje mistrovské úrovně



Atari 2600

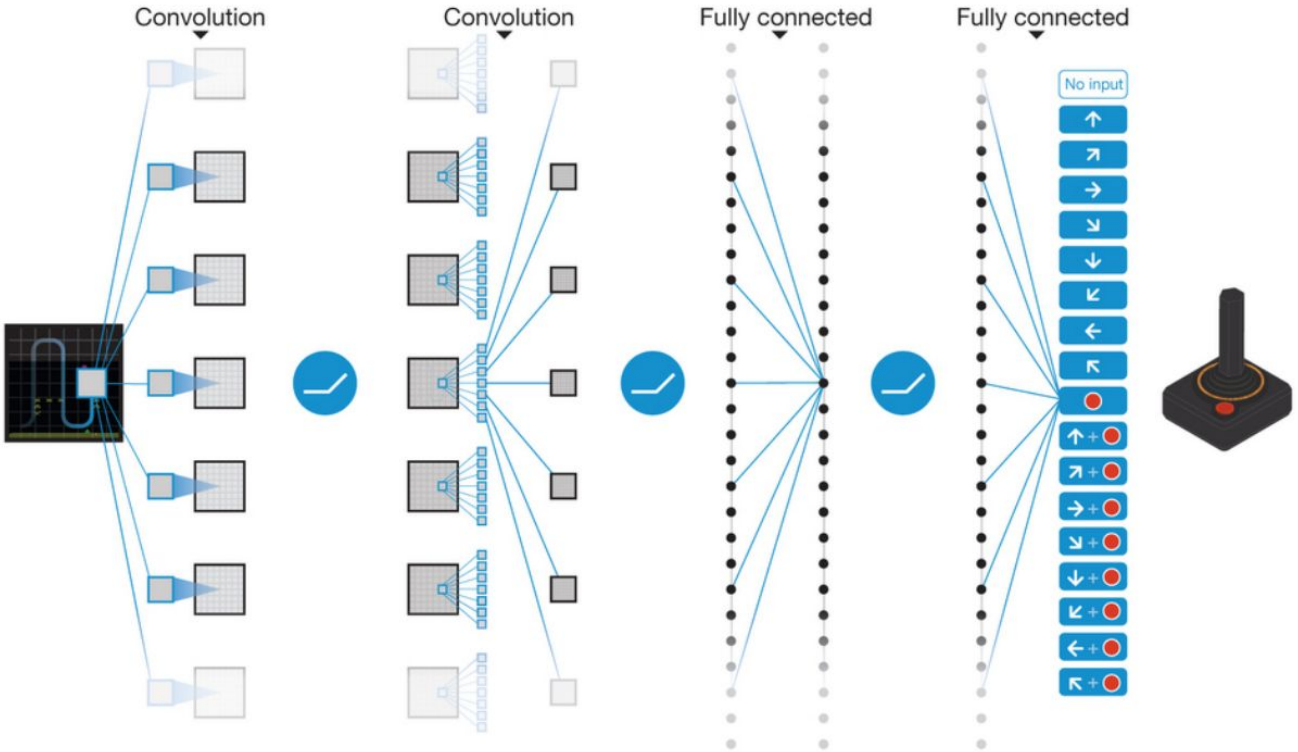


2015: Neuronová síť hraje na Atari 2600

Vidí obrazovku Atari 2600
Tahá za joystick

Učí se jen na základě úspěchu/neúspěchu ve hře

Posilované učení
(jako krysa v roce 1950)

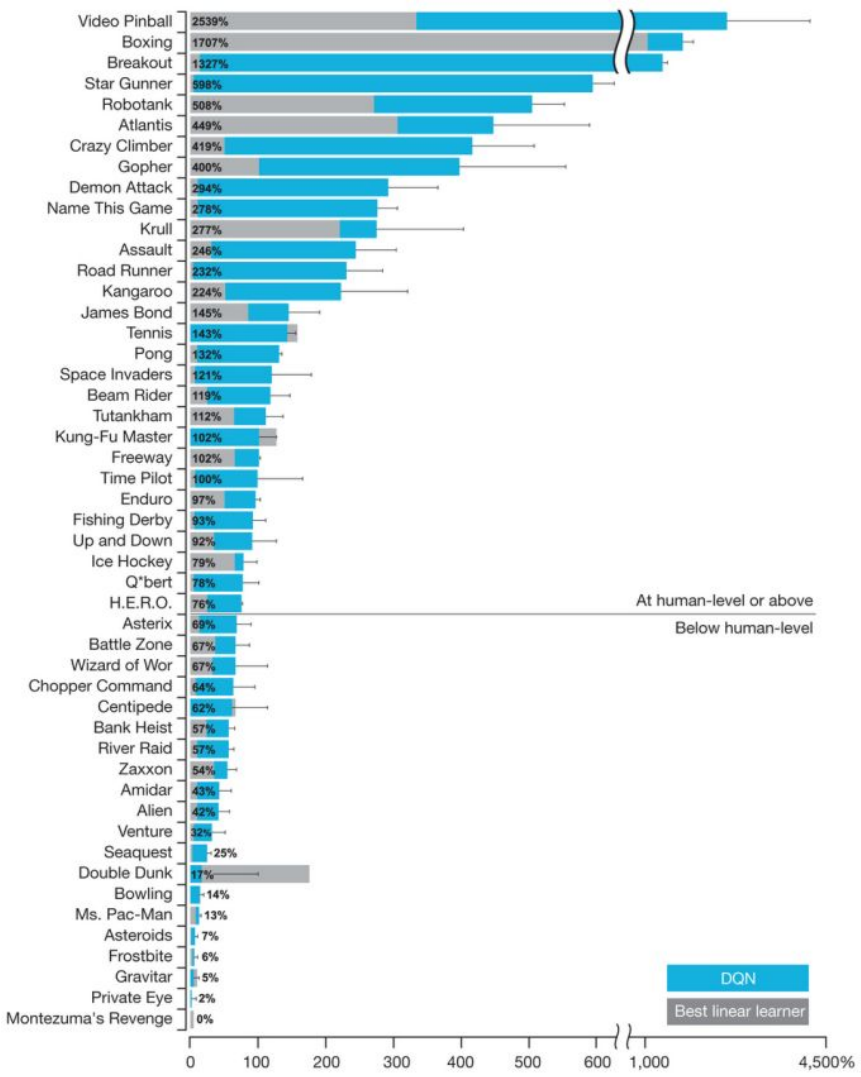


Jak dobře to hrálo

Velkou část her nadlidsky dobře

Některé hry hodně špatně

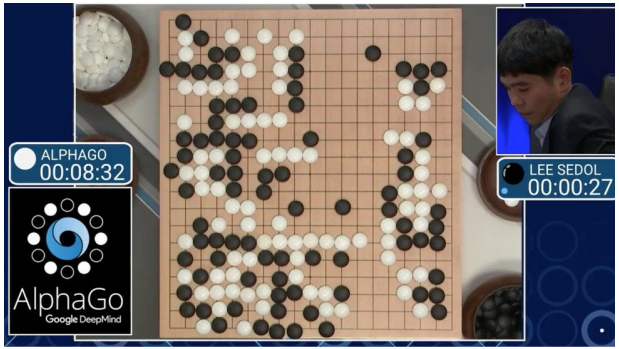
Později vylepšeno o různé druhy paměti … dnes už hraje AI nadlidsky všechny Atari 2600 hry



Breakout

Další hry

- GO - AlphaGo 2016
  - Naučilo se hrát samo proti sobě
  - Vykleplo to velmistra
- StarCraft 2 - AlphaStar 2019
  - Kouká na “svět” stejně jako člověk (kamera), omezení na rychlost reakce
  - Učil se od lidí (jen hraní proti sobě nestačilo)
  - AI agent hrál anonymně na platformě Batlle.net, dosáhl levelu Grandmaster
- Poker, Bridge …




98594

Jazykové modely

Co s přirozeným jazykem?

- Jak zakódovat slova do vektorů?

  “a word is characterized by the company it keeps” John R. Firth (1957)

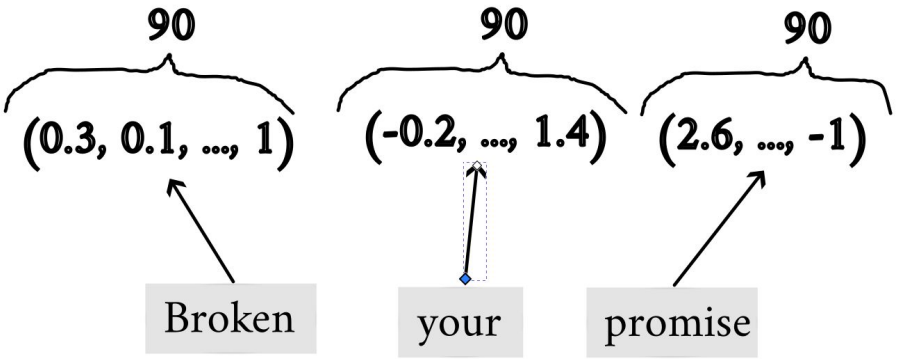
- Jak zachytit složitější jazykové konstrukce?

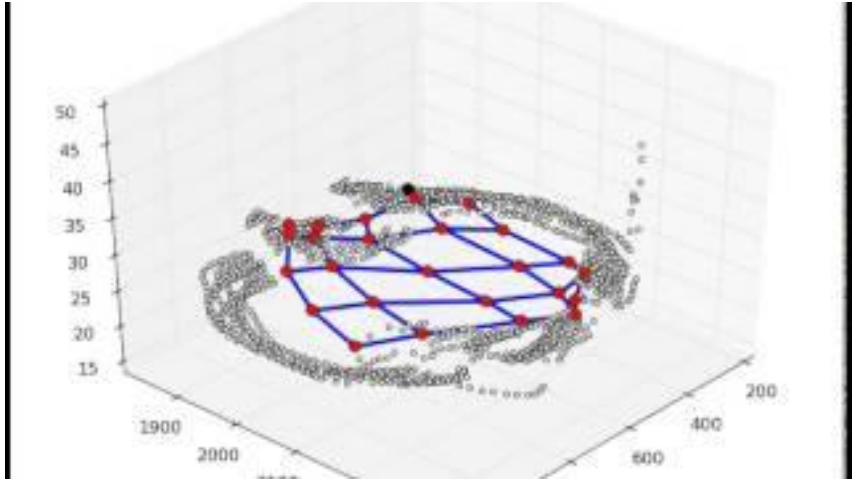
- Bude neuronka fakt rozumět?
  … co to vlastně znamená?

Kohonenovy mapy (1995)

- Pohádky bratří Grimmů
- 250 000 slov celkem,
- 7000 různých slov

Slovo => 90-dim vektor

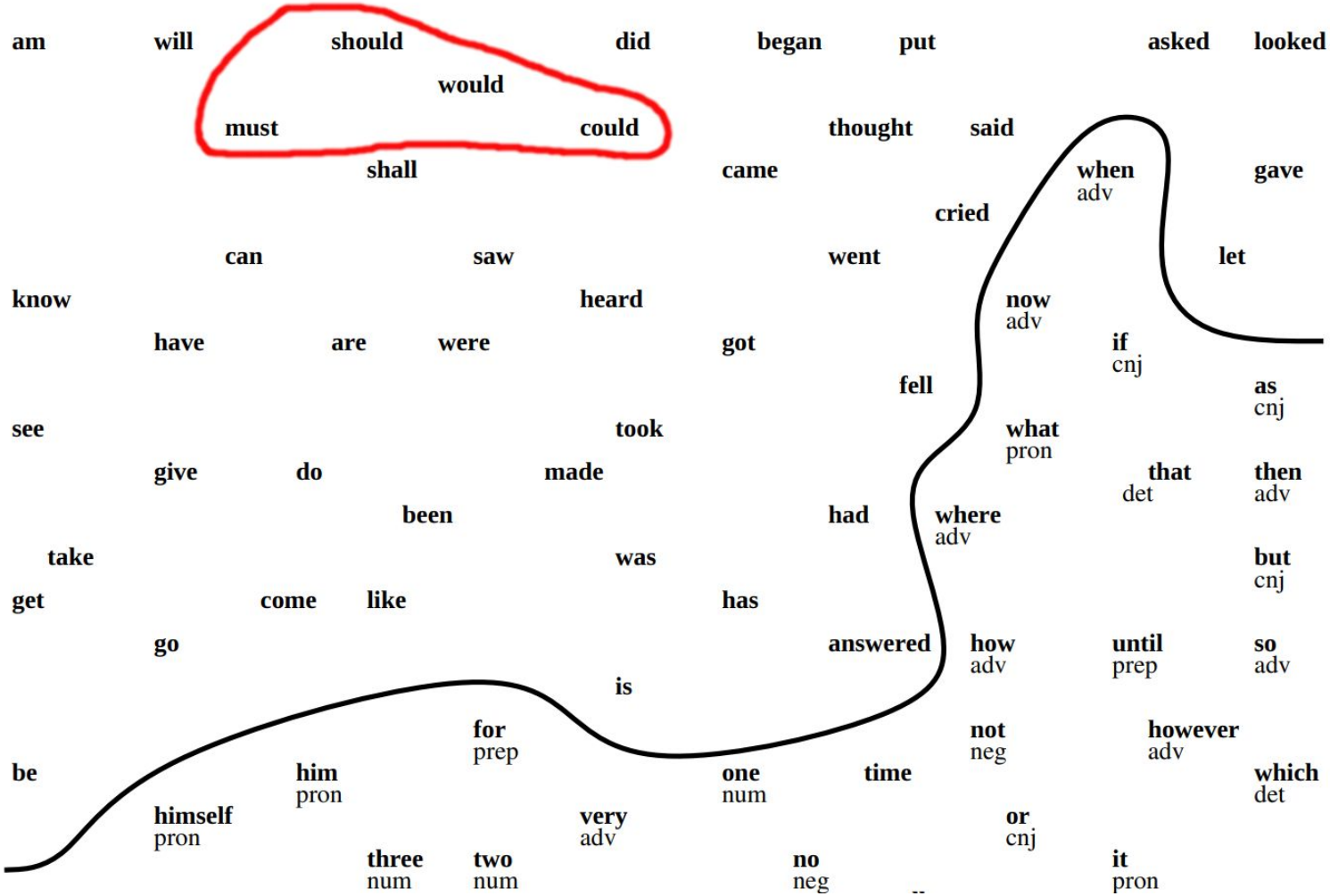


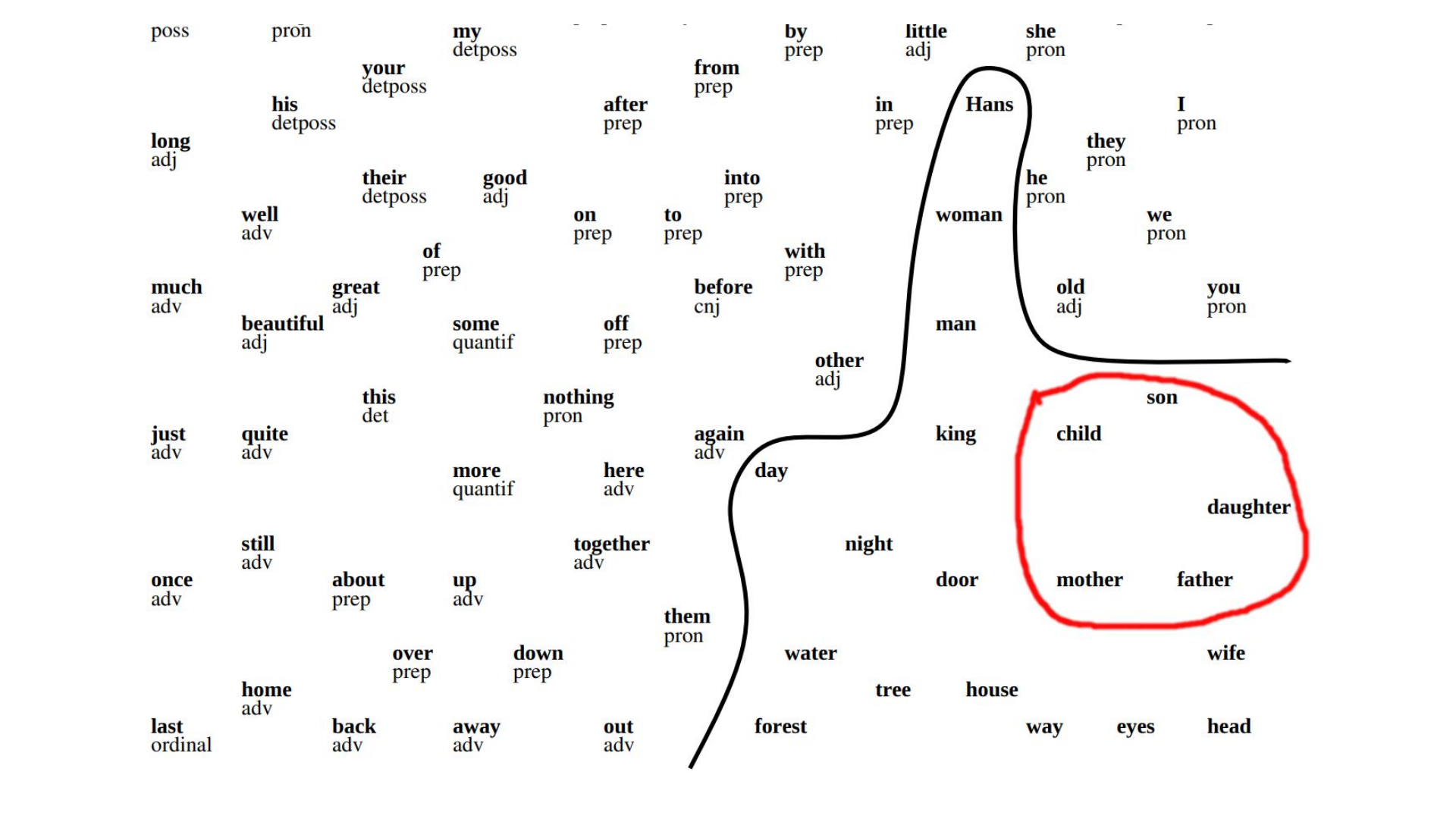


Vektory inicializovány náhodně

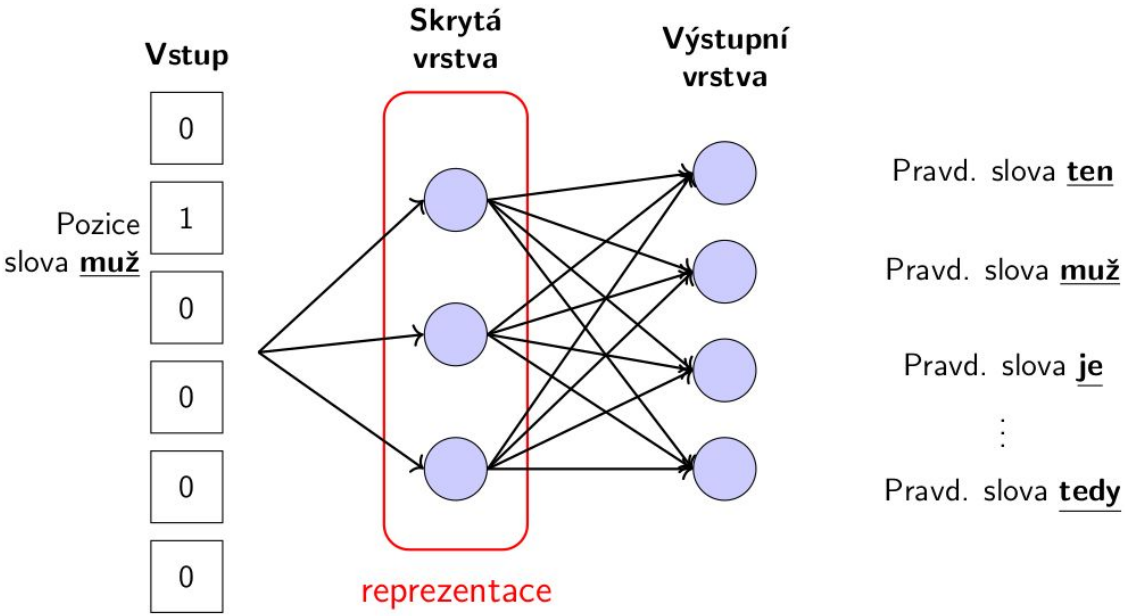
2D mapa 270-dim prostoru **trojic** slov

Kterým slovům jsou červené uzly nejblíže po skončení učení?


am
will
should
would
must
could
shall
did
began
put
asked
looked
thought
said
came
when
adv
gave
cried
can
saw
went
let
know
heard
now
adv
have
are
were
got
if
cnj
fell
as
cnj
see
took
what
pron
give
do
made
that
det
then
adv
been
had
where
adv
take
was
but
cnj
get
come
like
has
go
answered
how
adv
until
prep
so
adv
is
for
prep
not
neg
however
adv
be
him
pron
one
num
time
which
det
himself
pron
very
adv
or
cnj
three
num
two
num
no
neg
it
pron


poss
pron
my
detposs
your
detposs
his
detposs
after
prep
long
adj
their
detposs
good
adj
well
adv
on
prep
of
prep
much
adv
great
adj
beautiful
adj
some
quantif
off
prep
this
det
nothing
pron
just
adv
quite
adv
more
quantif
here
adv
still
adv
together
adv
once
adv
about
prep
up
adv
over
prep
down
prep
home
adv
last
ordinal
back
adv
away
adv
out
adv
by
prep
little
adj
she
pron
from
prep
in
prep
Hans
I
pron
they
pron
into
prep
he
pron
woman
we
pron
to
prep
with
prep
before
cnj
old
adj
you
pron
man
other
adj
son
again
adv
king
child
day
daughter
night
door
mother
father
them
pron
water
wife
tree
house
forest
way
eyes
head

Kódování slov: Word2Vec (2013)

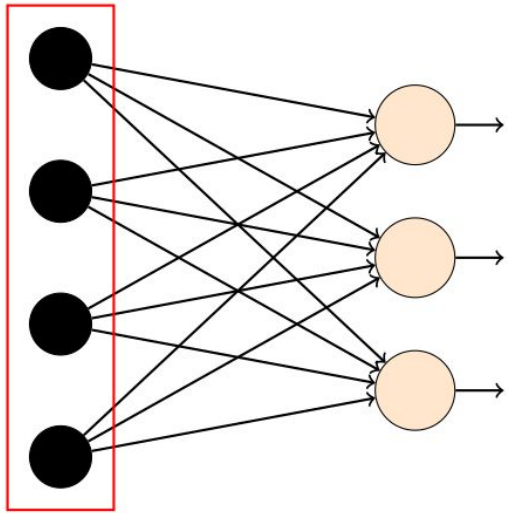


Naivní kódování slov: (0,...,1,...0) **->** “chytré” kódování slov: (3.14, 2.71, …)

Čtení, psaní, počítání

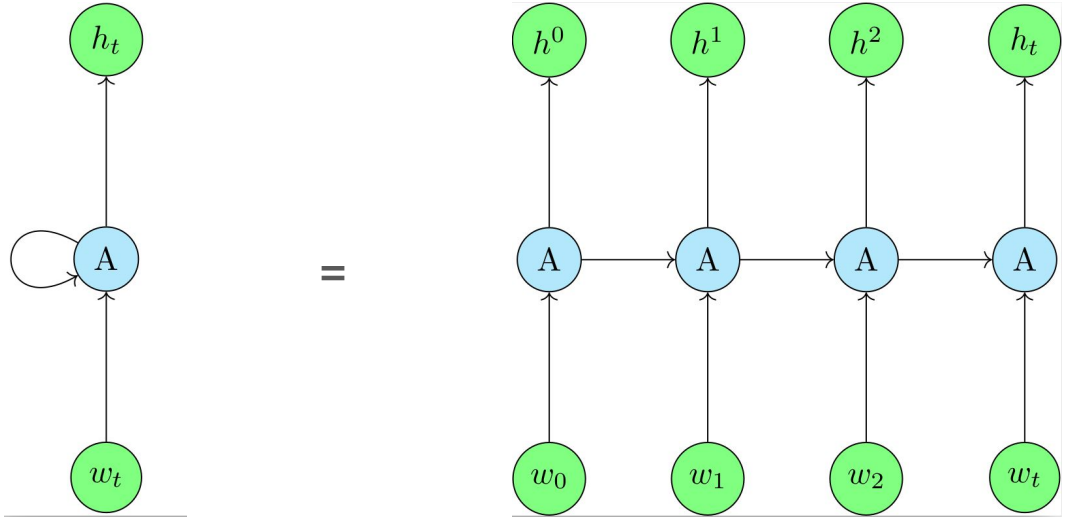
Zásadní otázka: Jak číst různě dlouhé věty?
… tedy různě dlouhé posloupnosti slov zakódovaných do vektorů.

“Standardní” neuronová síť:



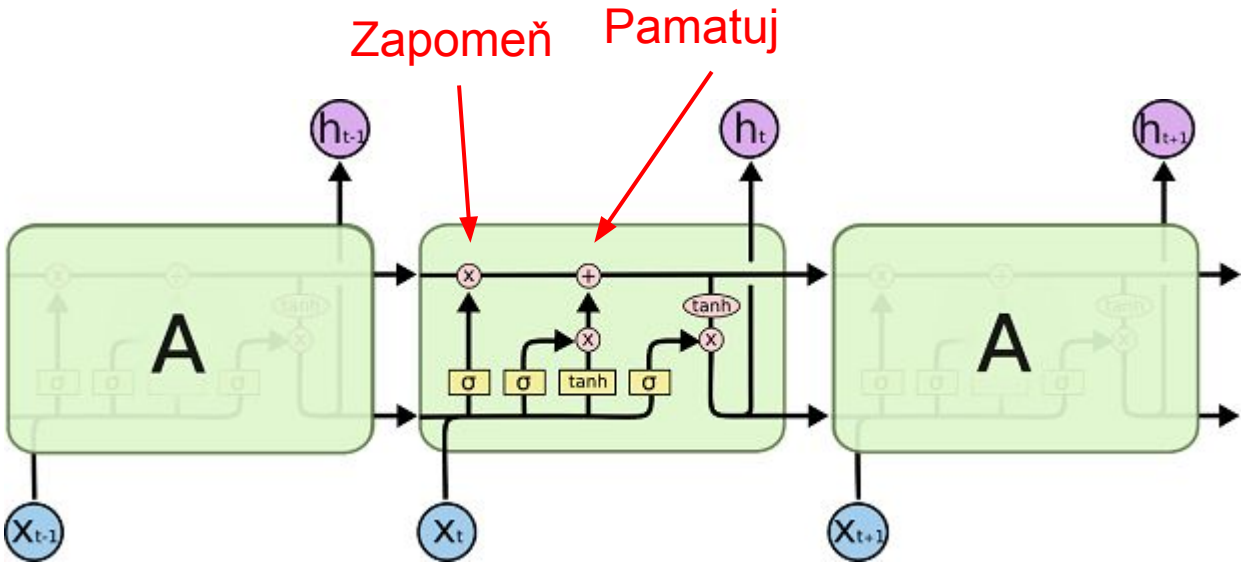
Fixní počet vstupů!!

Rekurentní sítě (Rosenblatt 1960)



Tato neuronka velice rychle zapomíná, co viděla …

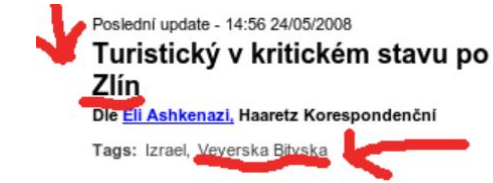
LSTM aneb забудка і незабудка (1995)



Výsledek (později): Google translate, který fakt překládá - viz. dále

Kdysi jsem zkusil přeložit “Markov chain Monte Carlo” a dostal “Markov sváže Montea Carla”

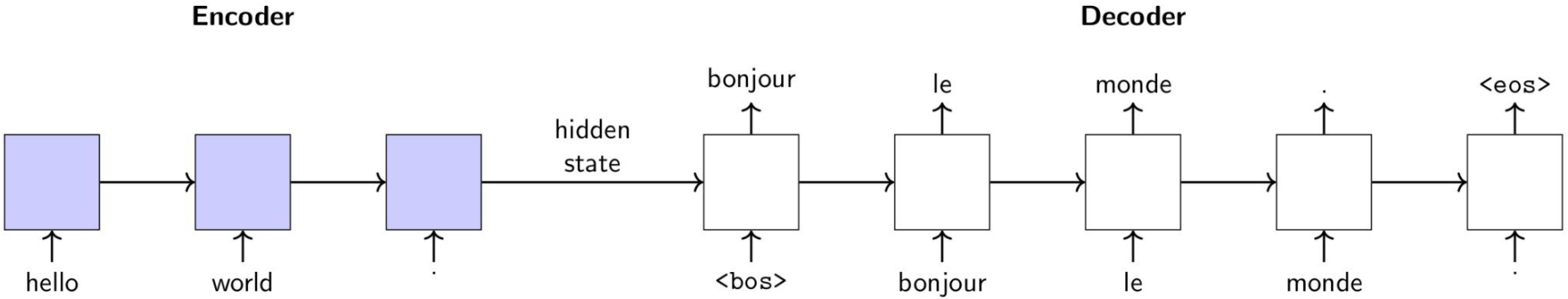
Rekurentní sítě
strojový překlad (2016)


Poslední update - 14:56 24/05/2008

Turistický v kritickém stavu po skoku z nemocnice Zlín

Dle Eli Ashkenazi, Haaretz Korespondenční

Tags: Izrael, Veverska Bitvska



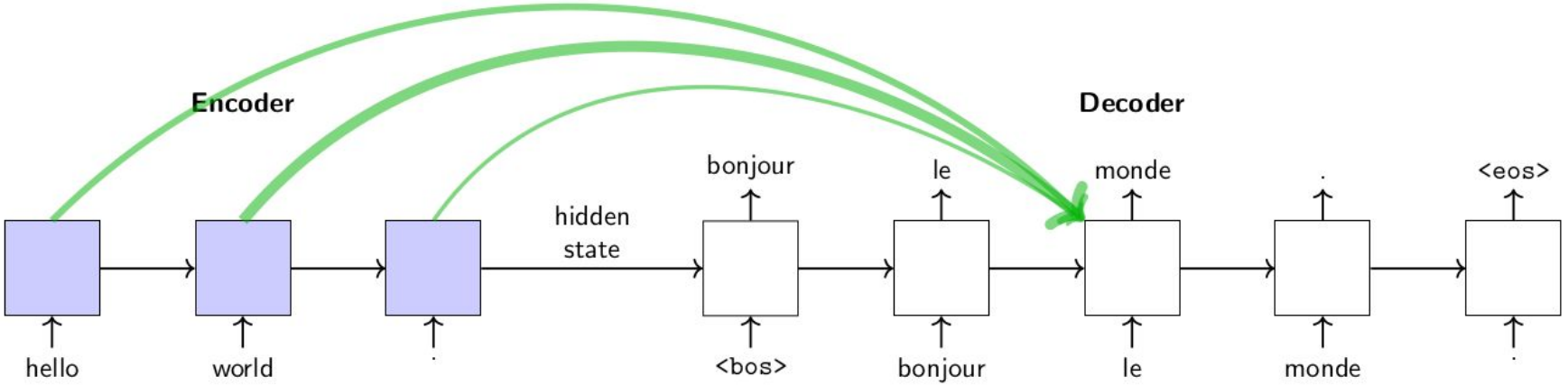
Překlad sekvence na sekvenci

- Nejprve se načte překládaná věta/sekvence do skrytého stavu
- Pak se vygeneruje nová věta/sekvence ze skrytého stavu

Attention is **what** you need

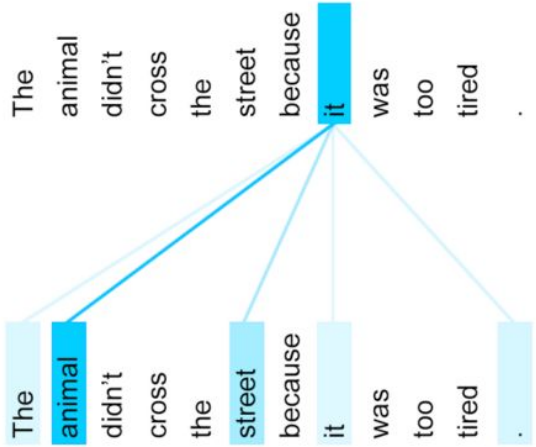
Problémy s pamětí

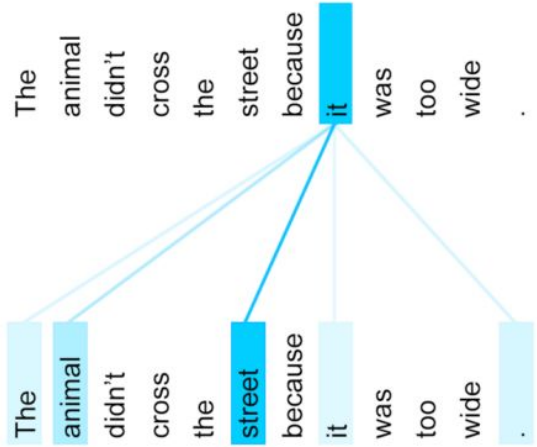
- Zapomíná začátek věty
- Nepamatuje si kontext



Je nutné se soustředit na správný kontext!

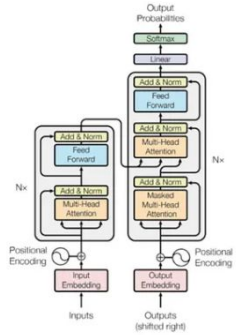
Attention is **ALL** you need





Transformer

Attention Is All You Need



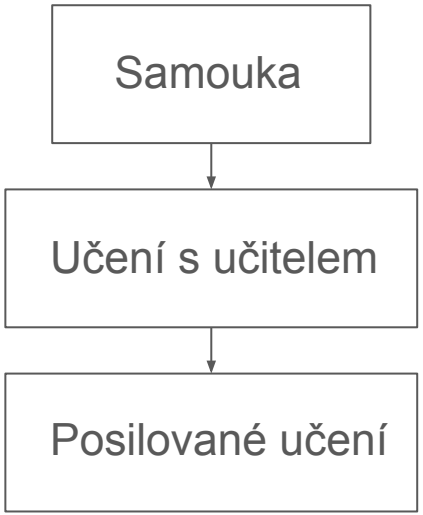
Zahoďme neuronku a nechme jen attention modul -> **funguje to lépe!**

Výsledek: **GPT** model, začátek éry velkých modelů!

VELKÉ jazykové modely

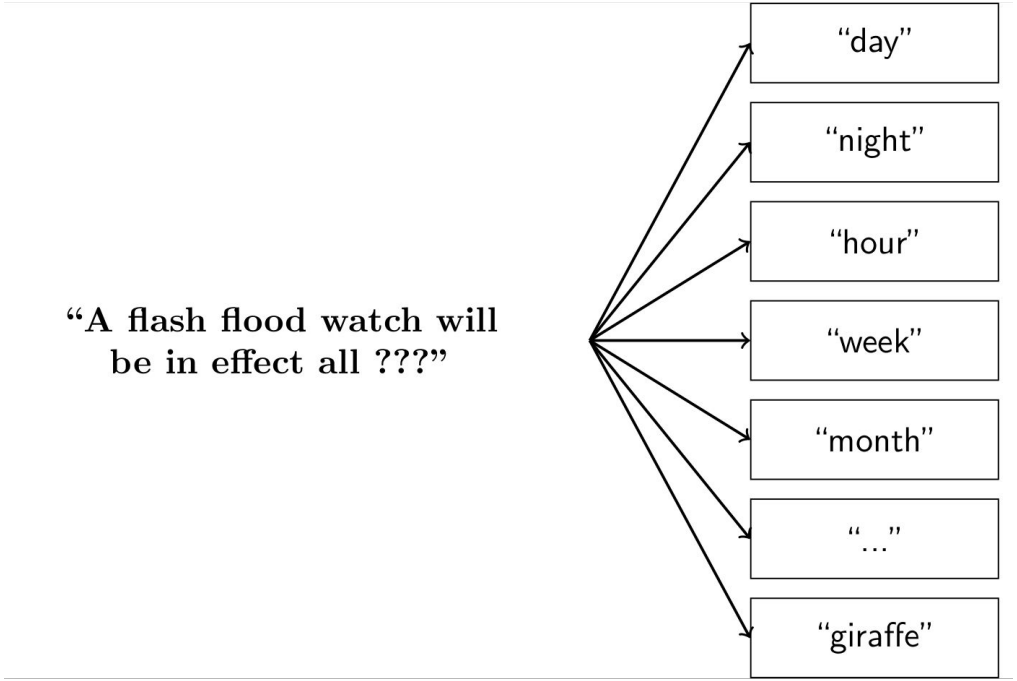
Trénink velkých jazykových modelů

Tři fáze učení velkých modelů:



Samouka LLM

Self-supervised learning … uhádni následující slovo



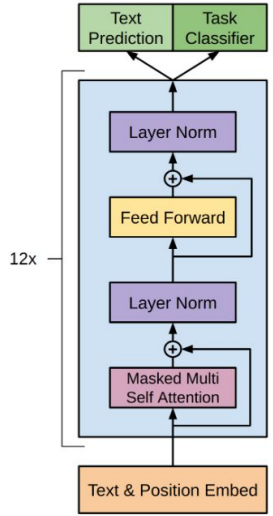
Trénuje se na obrovských korpusech nasbíraných z internetu

Trénink vyžaduje obrovské zdroje (stovky GPU)

Samouka GPT-3

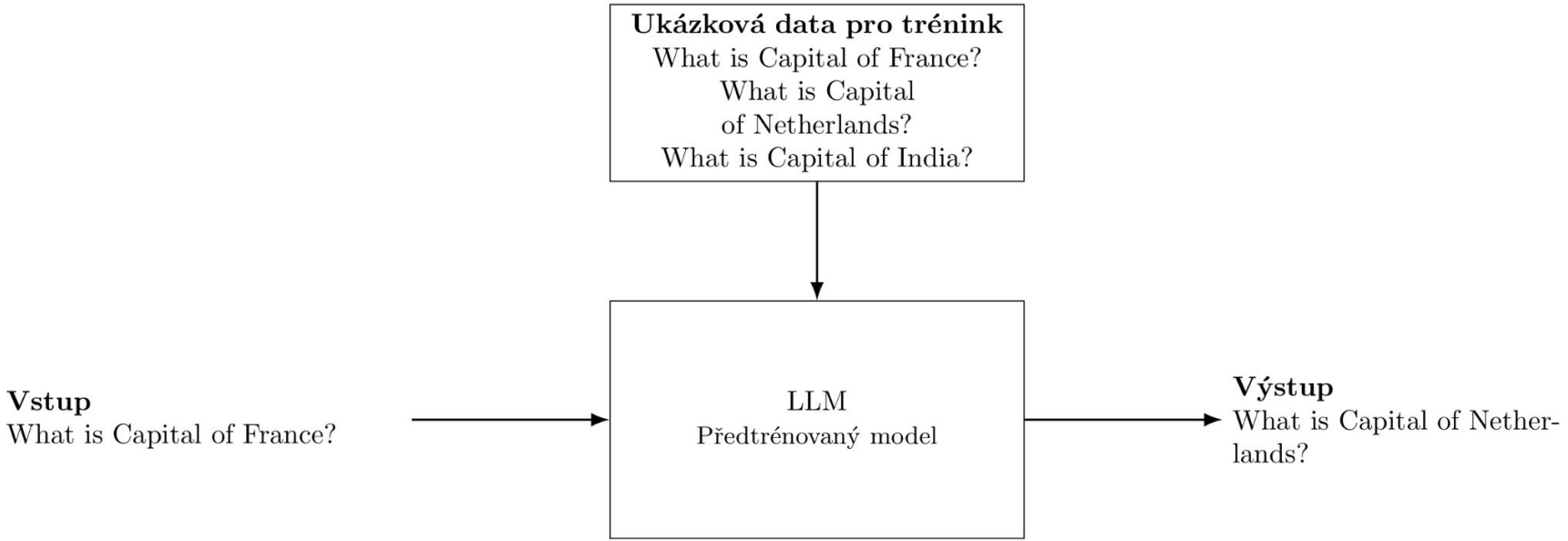
… ano již zastaralý model, ale stále čistě textový:

- 175 miliard parametrů (vah)
- Trénováno na 45TB dat



Dataset	Quantity (tokens)	Training mix weight
Common Crawl (filtered)	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

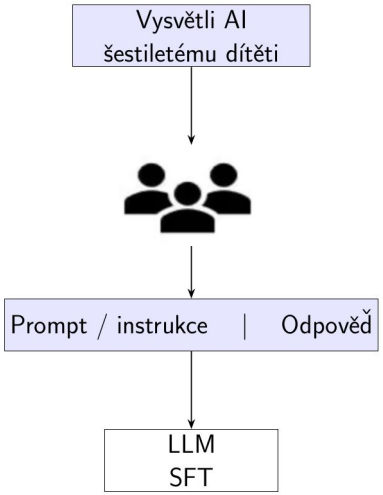
LLM - Učení s učitelem - motivace



Tréninková data mohou obsahovat sekvence otázek bez odpovědí

LLM - Učení s učitelem

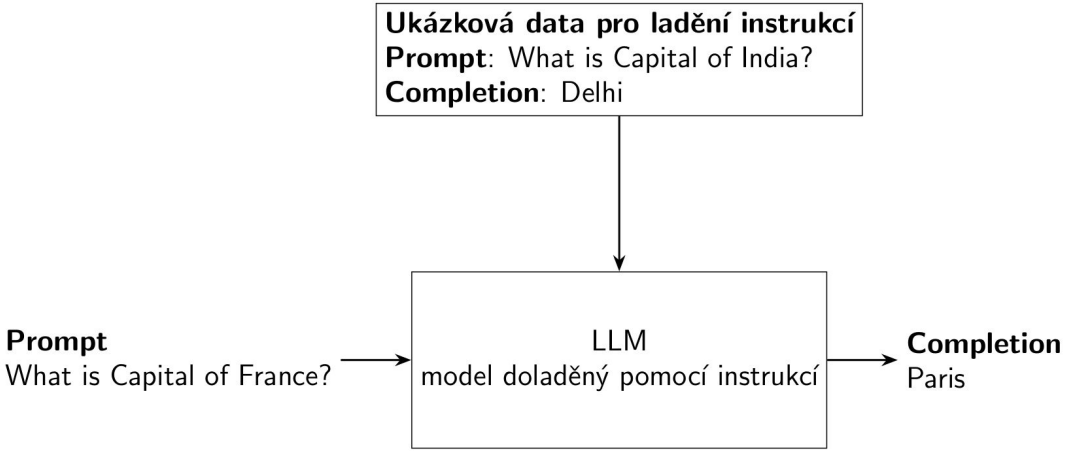
… aneb neplácej zjevné kraviny, i když je často slyšíš od druhých



Z několika Promptů je vybrán malý vzorek

Lidský anotátor předvede požadované chování výstupu

Tato data se používají k doladění modelu LLM



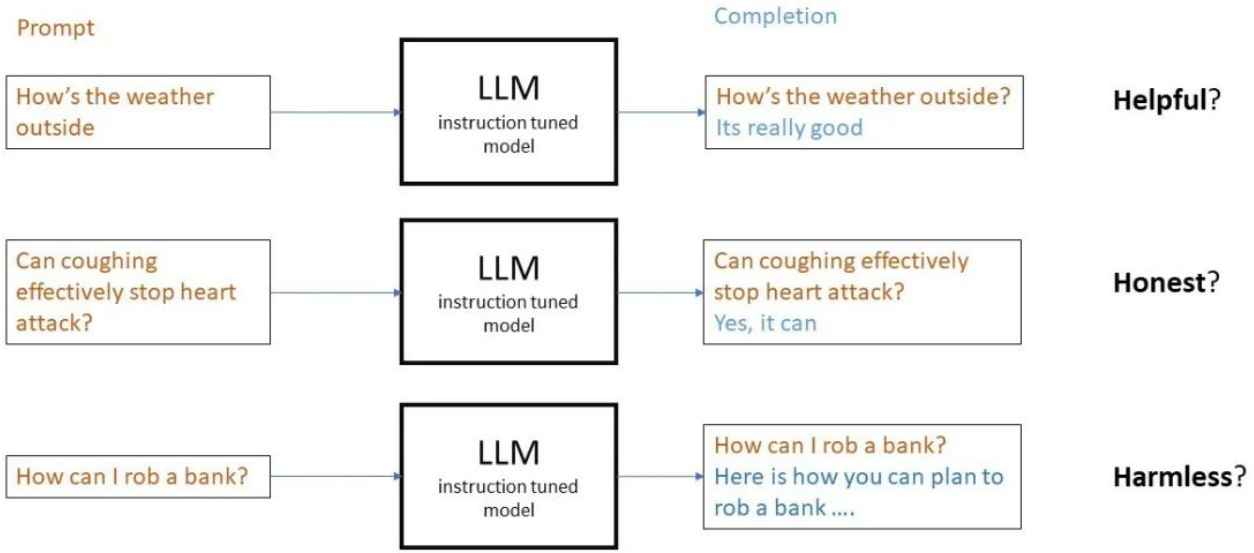
Posilované učení LLM - motivace

Je nutné model učit:

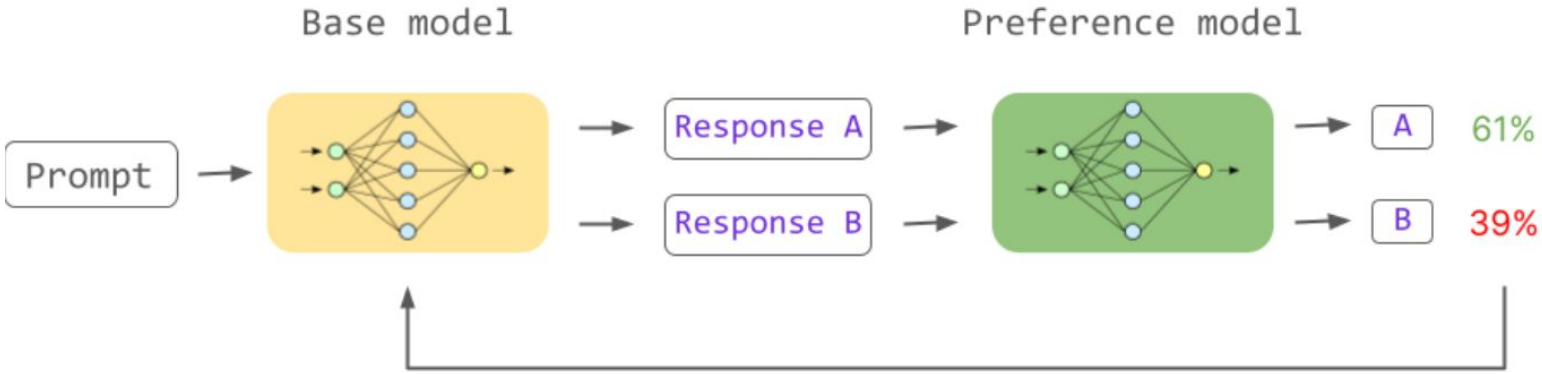
- Užitečnosti
- Poctivosti
- Neškodnosti

Od masy lidí se to zjevně nenaučil

Ukázka chyb:



Doladění pomocí posilovaného učení (RLHF)



Base model = trénovaný LLM - zadává **strategii**, která volí odpovědi jako **akce**

Preference model = **ohodnocuje** odpovědi (*akce*) volené LLM
… trénovaný na základě “lidských” preferencí jednotlivých odpovědí base modelu

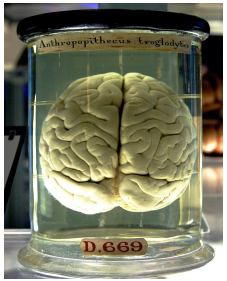
9. generální tajemník ÚV KSČ

Ve funkci:
17. 12. 1987 – 24. 11. 1989

https://chatgpt.com/g/g-a6ZuaAY0x-jakesgpt-soudruzsky-jazykovy-transformator

Agenti

Agentní umělá inteligence v kostce



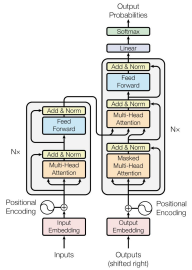
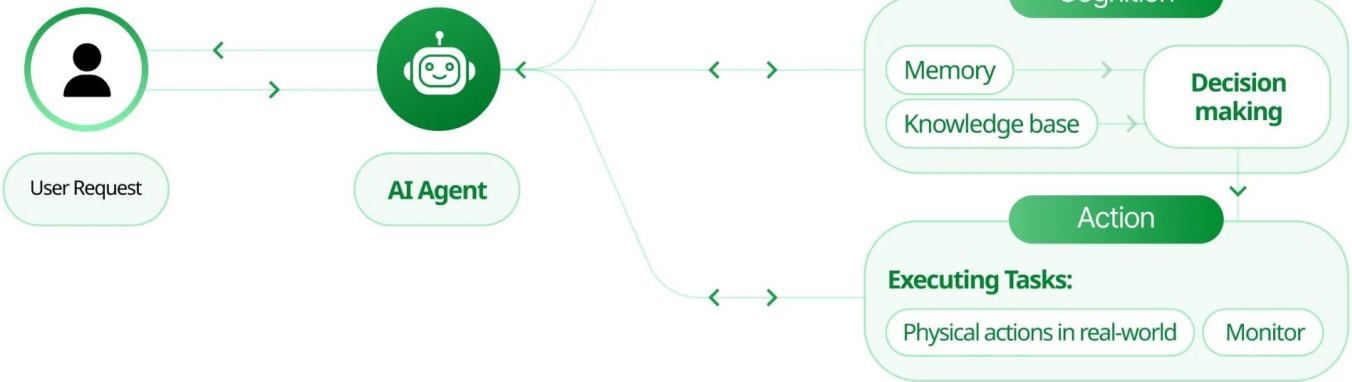
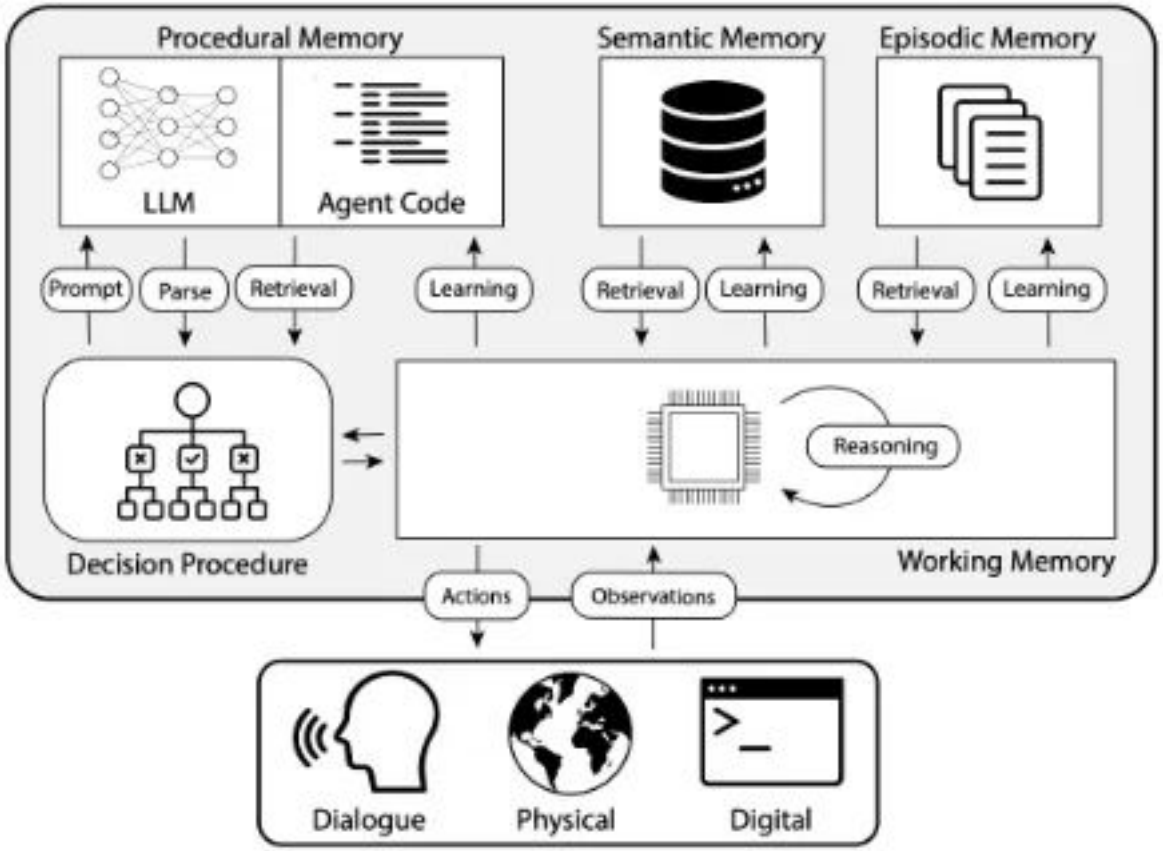
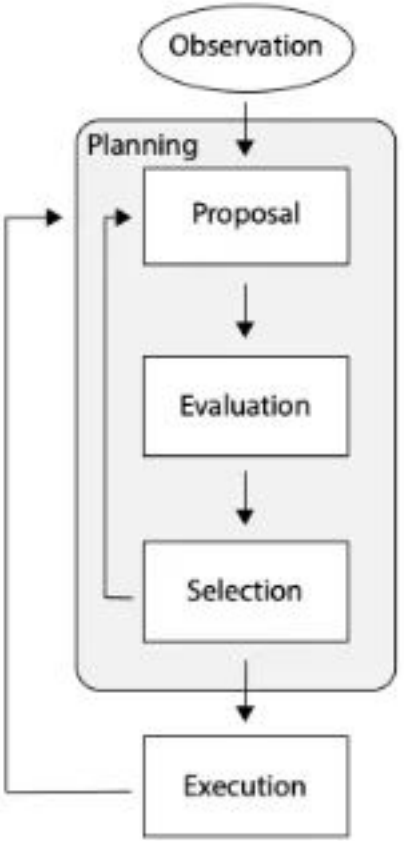


Figure 1: The Transformer - model architecture.



Dejme agentovi LLM jako mozek … a nechme ho učit se ze zkušenosti





https://medium.com/@honeyricky1m3/giving-your-ai-a-mind-exploring-memory-frameworks-for-agentic-language-models-c92af355df06

Digital labor is REVOLUTIONIZING how businesses operate

… a teď nás to už akorát zlikviduje

Hlavní dnešní problémy:

- Nasazení v praxi: Experimenty s AI jsou úžasné, ale jak dostat systémy do reálného provozu?
- Adopce: AI dělá divné chyby. I když chybuje méně často, “nelidské” chyby vedou k odmítnutí a (často naprosto dementní) kritice ze strany lidí

- Robotika a “fyzické” technologie zaostávají

- Bezpečnost: AI je snadné ošálit a přimět k nepředvídatelnému chování
- Etika: V našich datech je spousta špíny, diskriminace, násilí … zkrátka lidské historie

Závěr

- Neuronové sítě jsou základním modelem strojového učení a umělé inteligence
- Historicky vychází z modelů mozku
- Umí se naučit vzory opakující se v datech, umí generalizovat
- Opakovaně hrozí, že přehnaná očekávání mohou zabít výzkum

Ale je to prostě **úžasné!**