Umělá inteligence

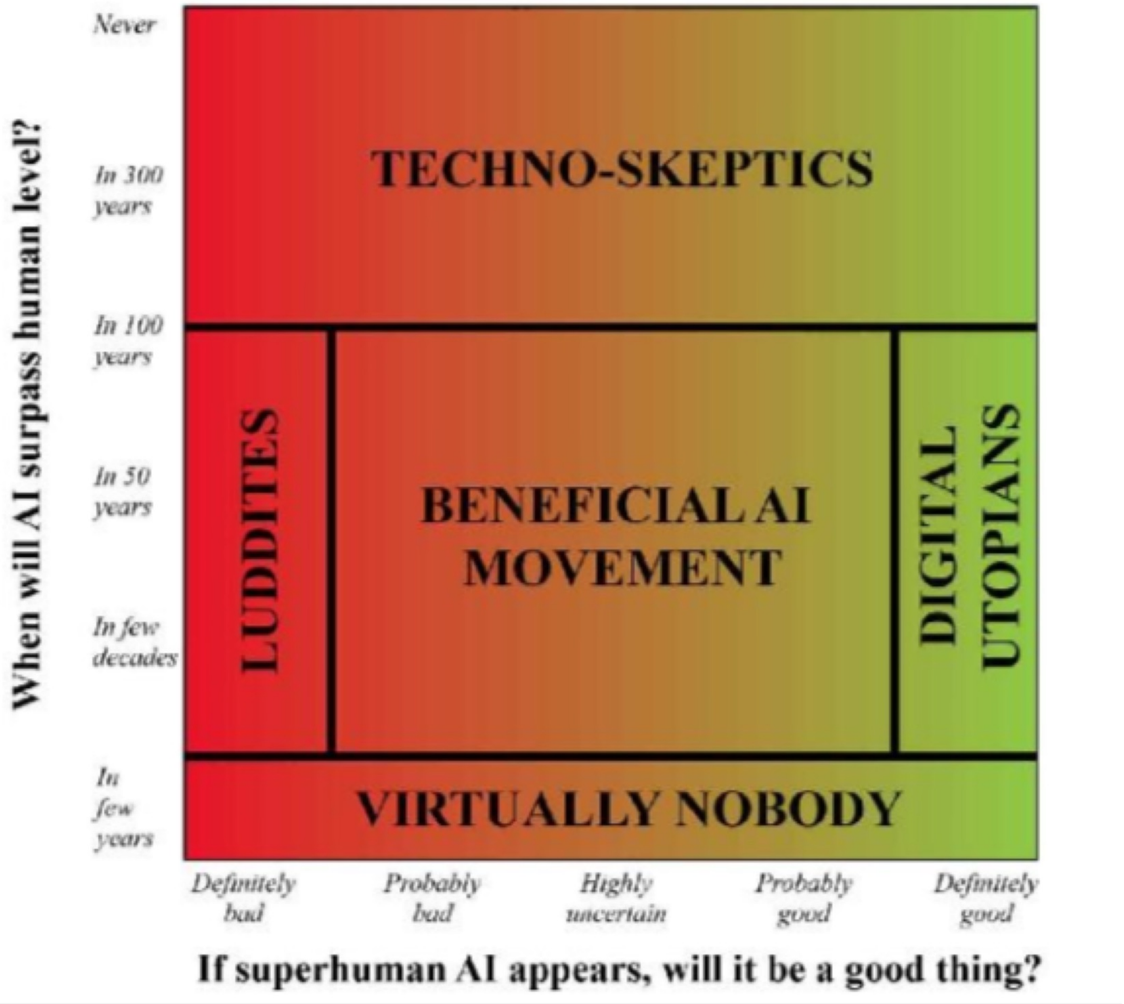
v přirozeném světě

Jan Romportl
Chief Data Scientist, O2 Czech Republic
Chief Science Officer, AI Startup Incubator

dosáhneme superinteligence?

a pokud ano, bude to dobře?

dosáhneme superinteligence?



jak vzniká AI?

vytváříme substrát, nad nímž inteligence vyvstává

umělá inteligence

přehnané očekávání v 50. letech

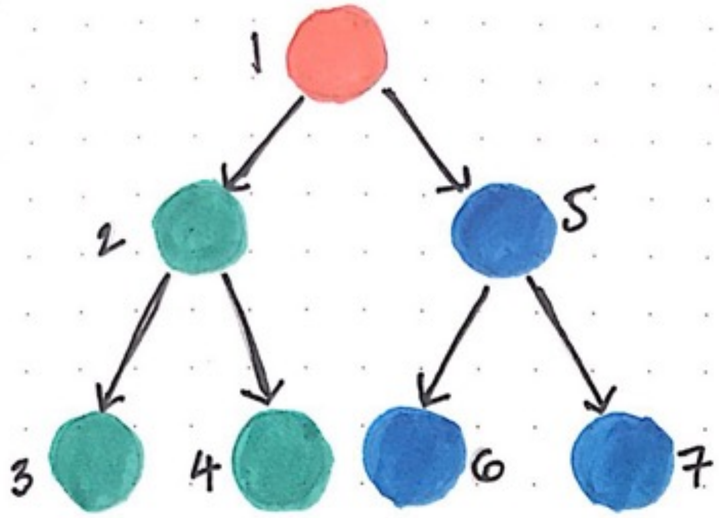
“We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. **We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.**”

John McCarthy, 1956

Deep Blue vs. Garry Kasparov

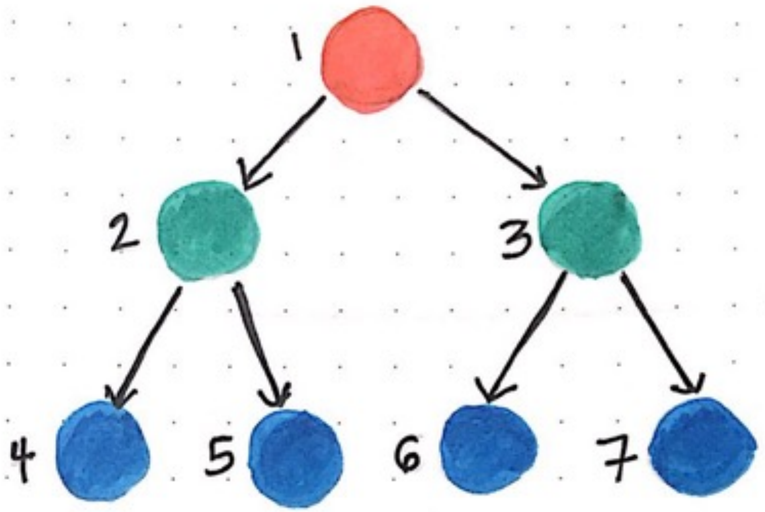
logicko-symbolické paradigma AI
v kognitivních vědách jako hypotéza CRUM
expertní systémy a prohledávání stavového prostoru
vrcholem je Deep Blue (1997)

Prohledávání stavového prostoru



Depth-first search

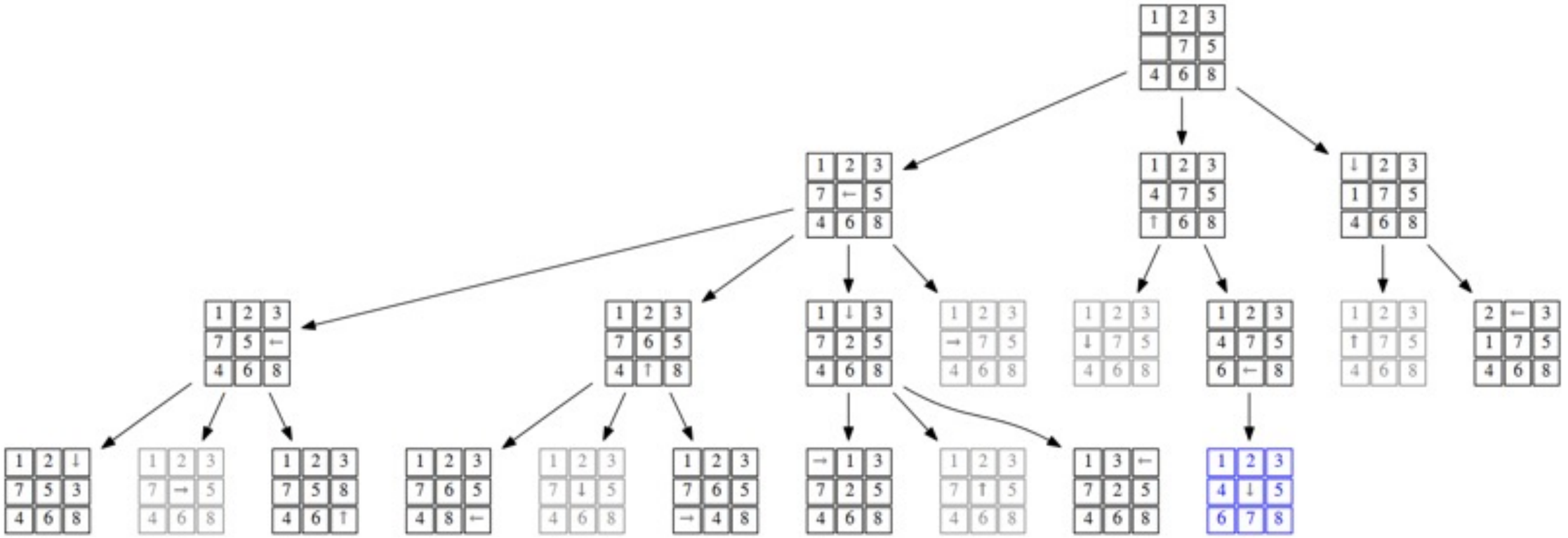
- Traverse through left subtree(s) first, then traverse through the right subtree(s).



Breadth-first search

- Traverse through one level of children nodes, then traverse through the level of grandchildren nodes (and so on...).

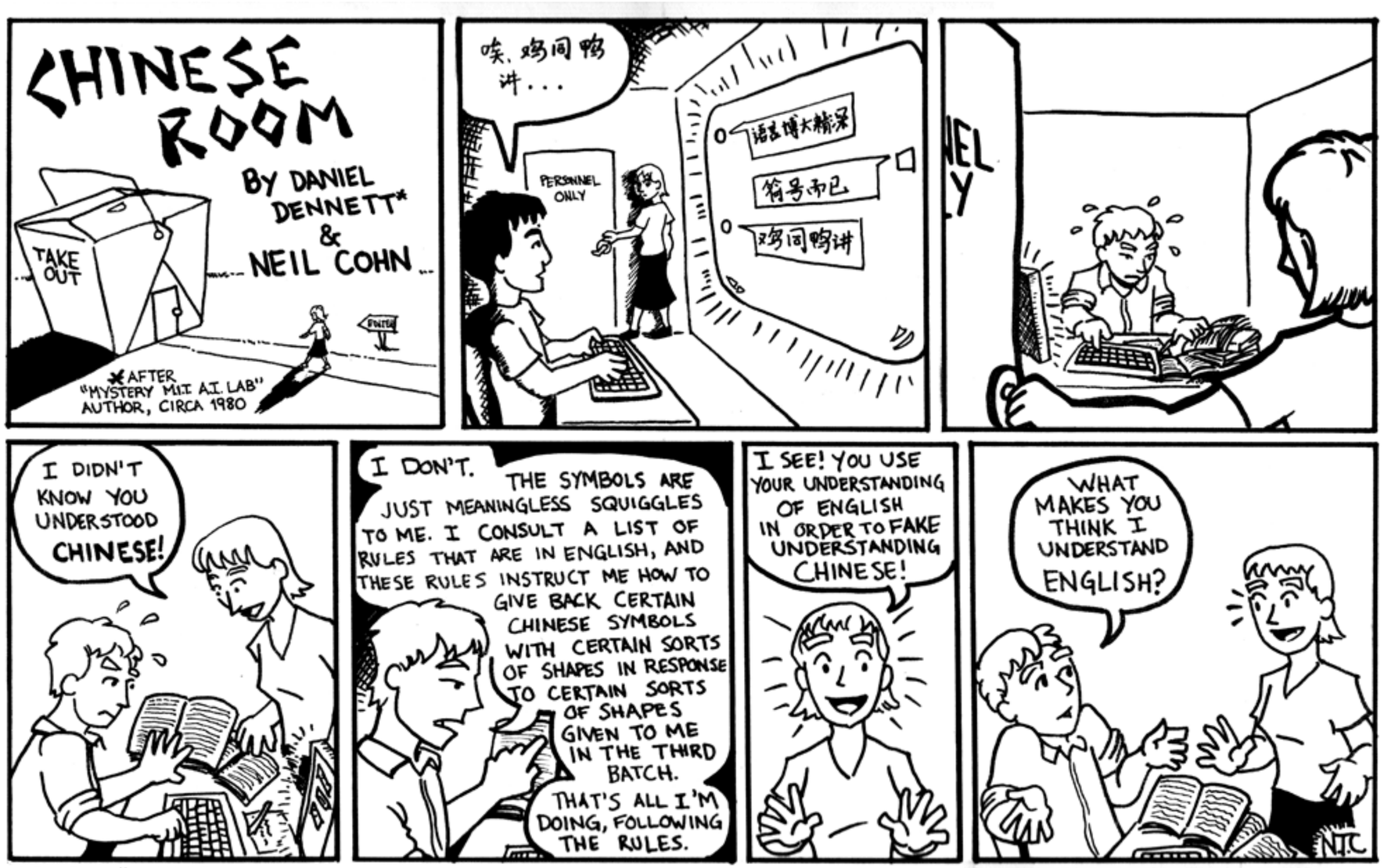
Prohledávání stavového prostoru



Prohledávání stavového prostoru

Searlův Čínský pokoj

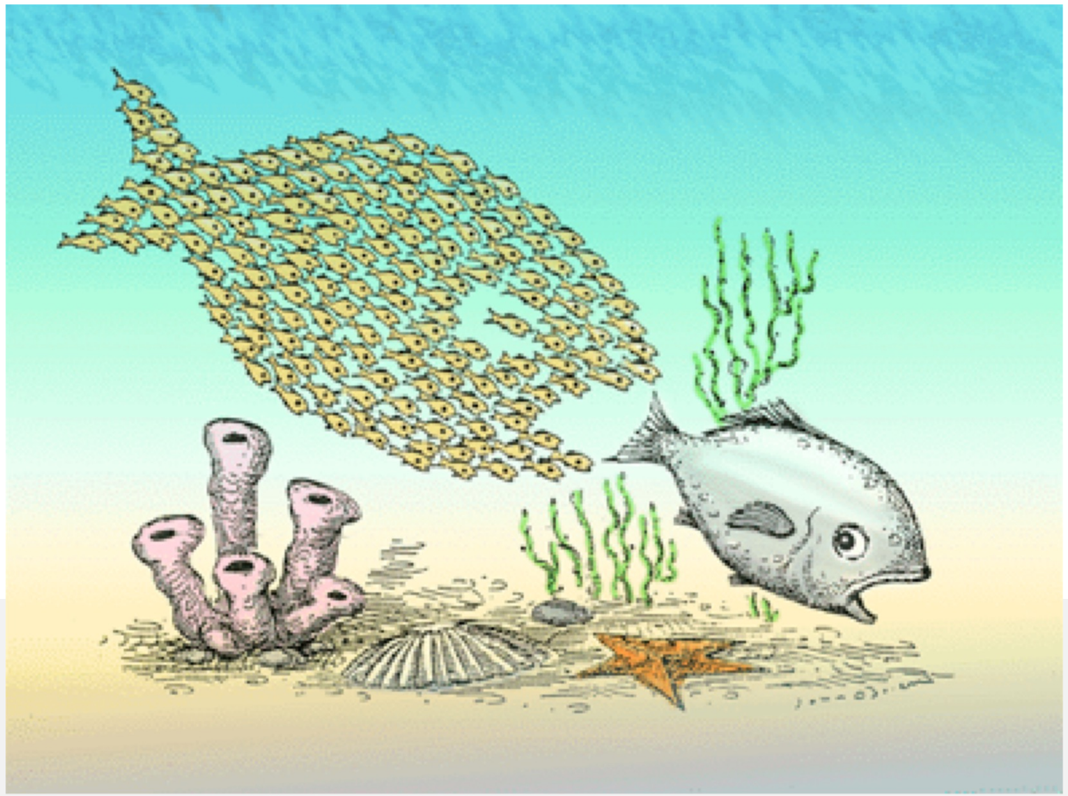
Searlův Čínský pokoj



© Dan Dennett and Neil Cohn

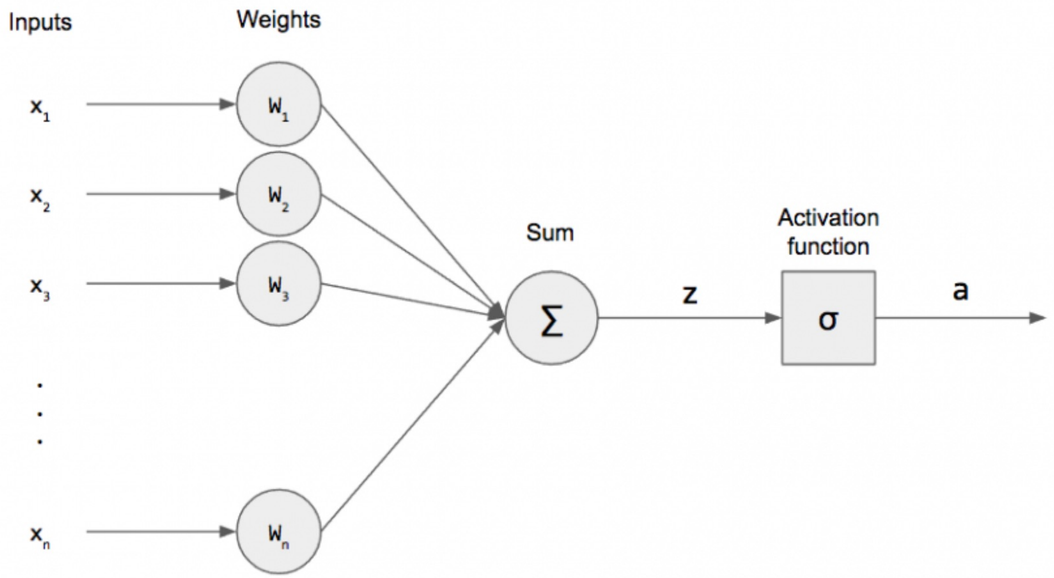
www.visuallanguagelab.com

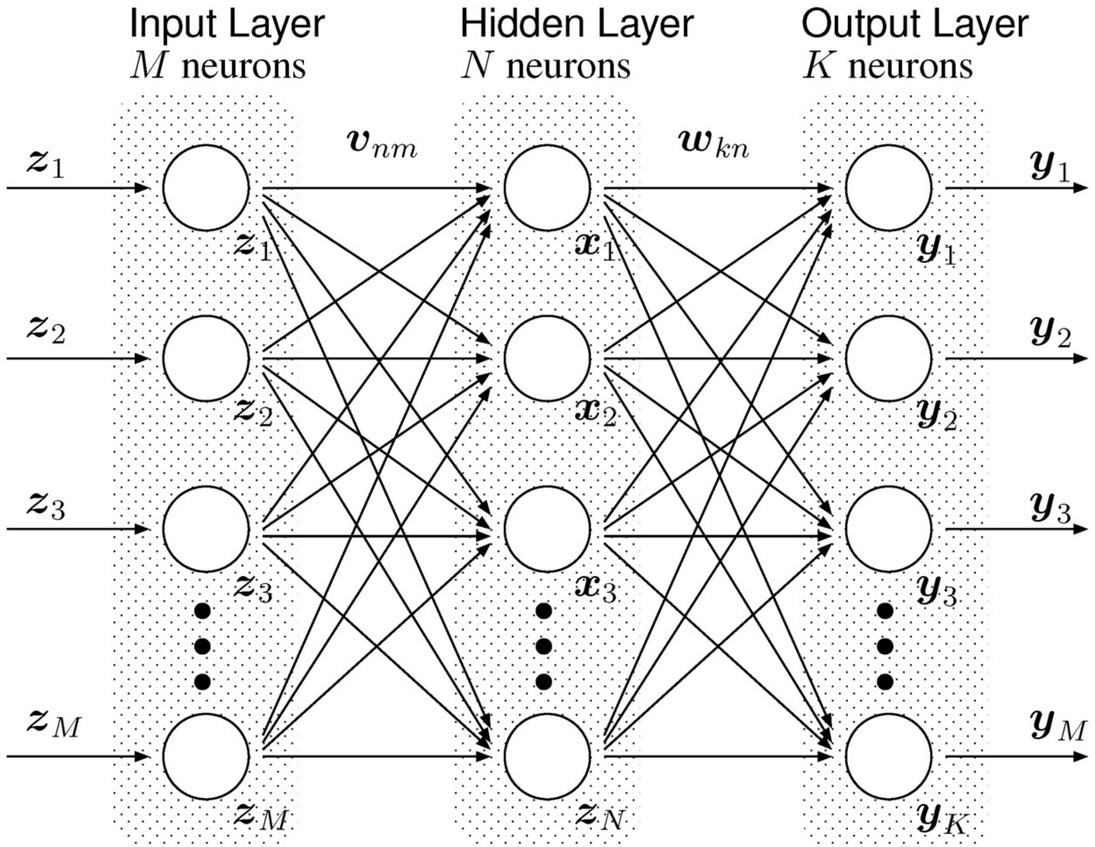
emergence v AI



kognitivní procesy (či jejich modely) nejsou reprezentovány top-down jako symboly a manipulace s nimi, nýbrž bottom-up emergují nad podkladovým komputačním substrátem

perceptron



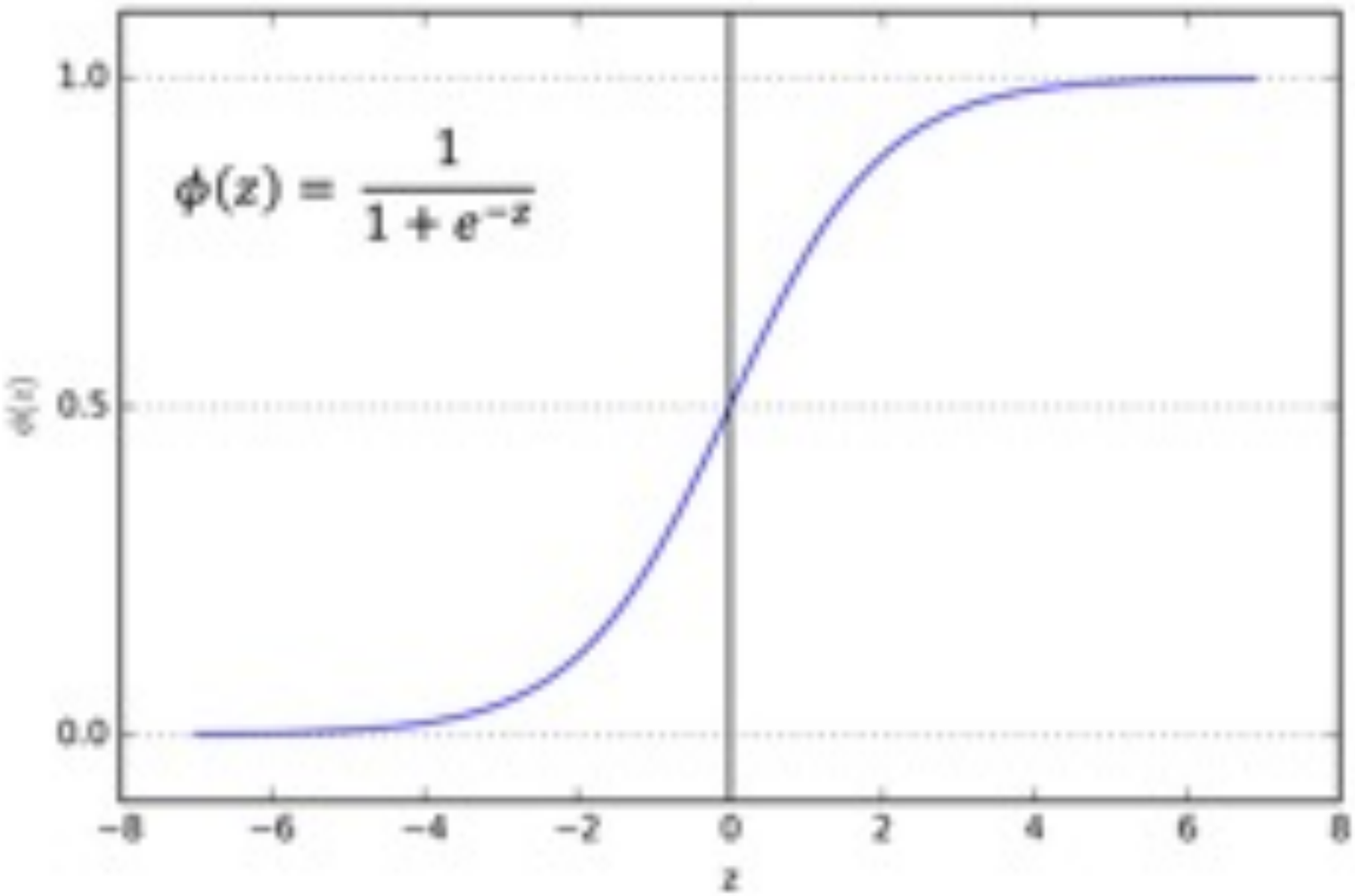


George Cybenko (1989), Kurt Hornik (1991): Universal approximation theorem

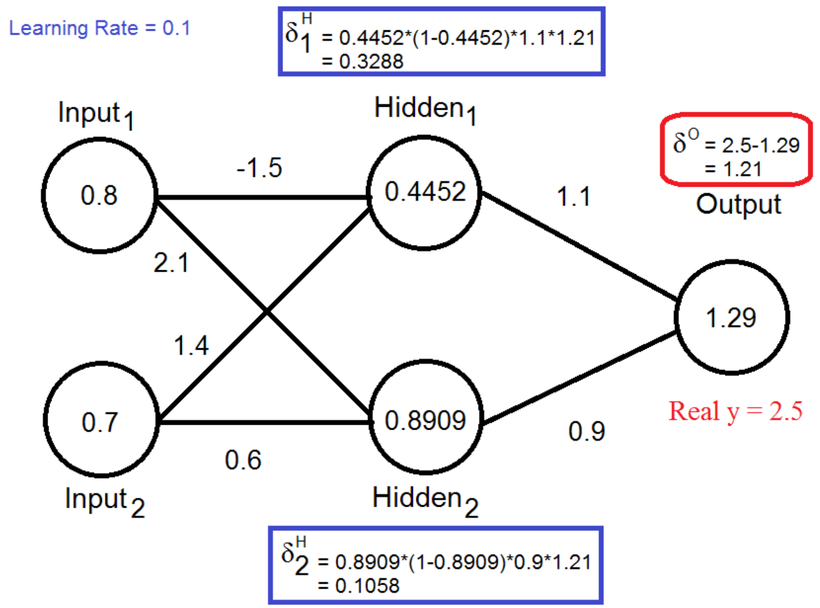
Dopředná neuronová síť s jednou skrytou vrstvou a konečným počtem neuronů může aproximovat libovolnou funkci v Rn.

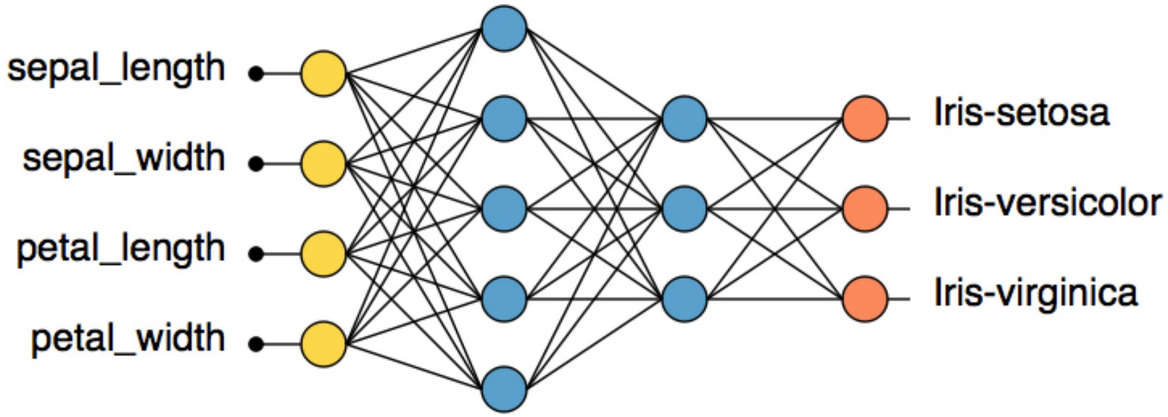
Jak ale síť naučit?

sigmoidální aktivační funkce

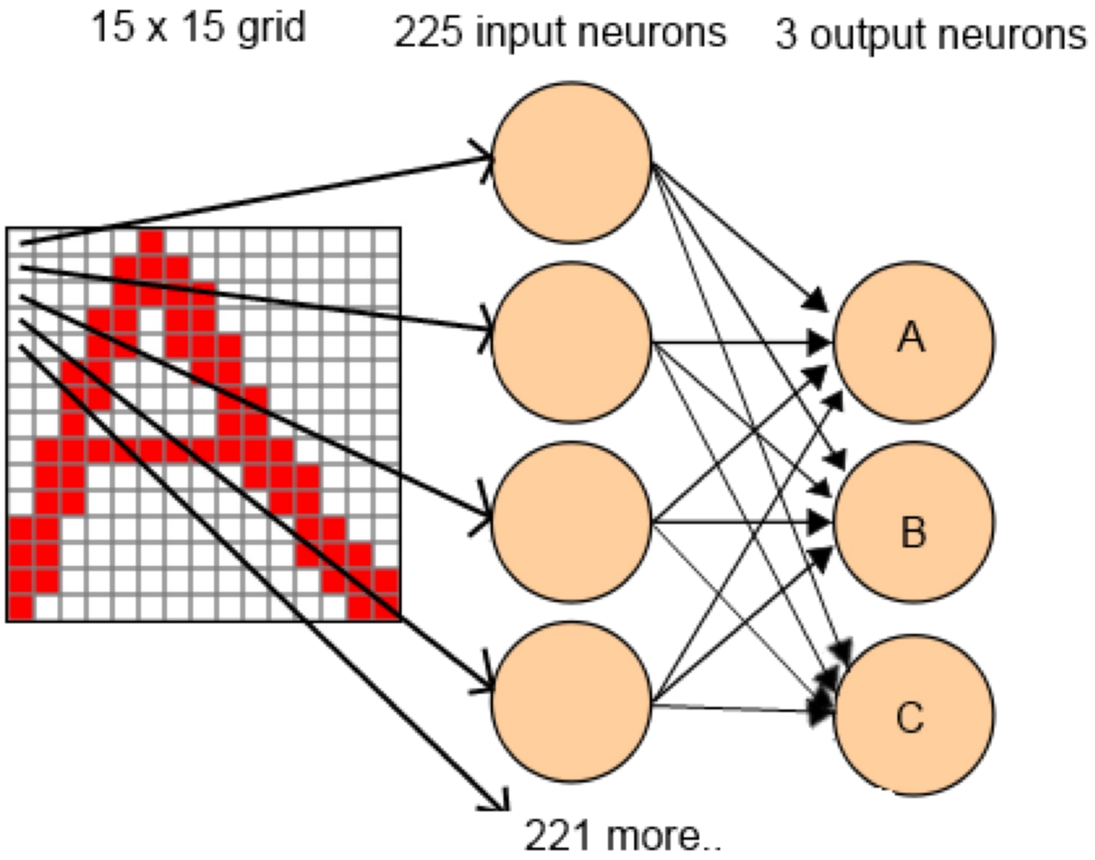


perceptron

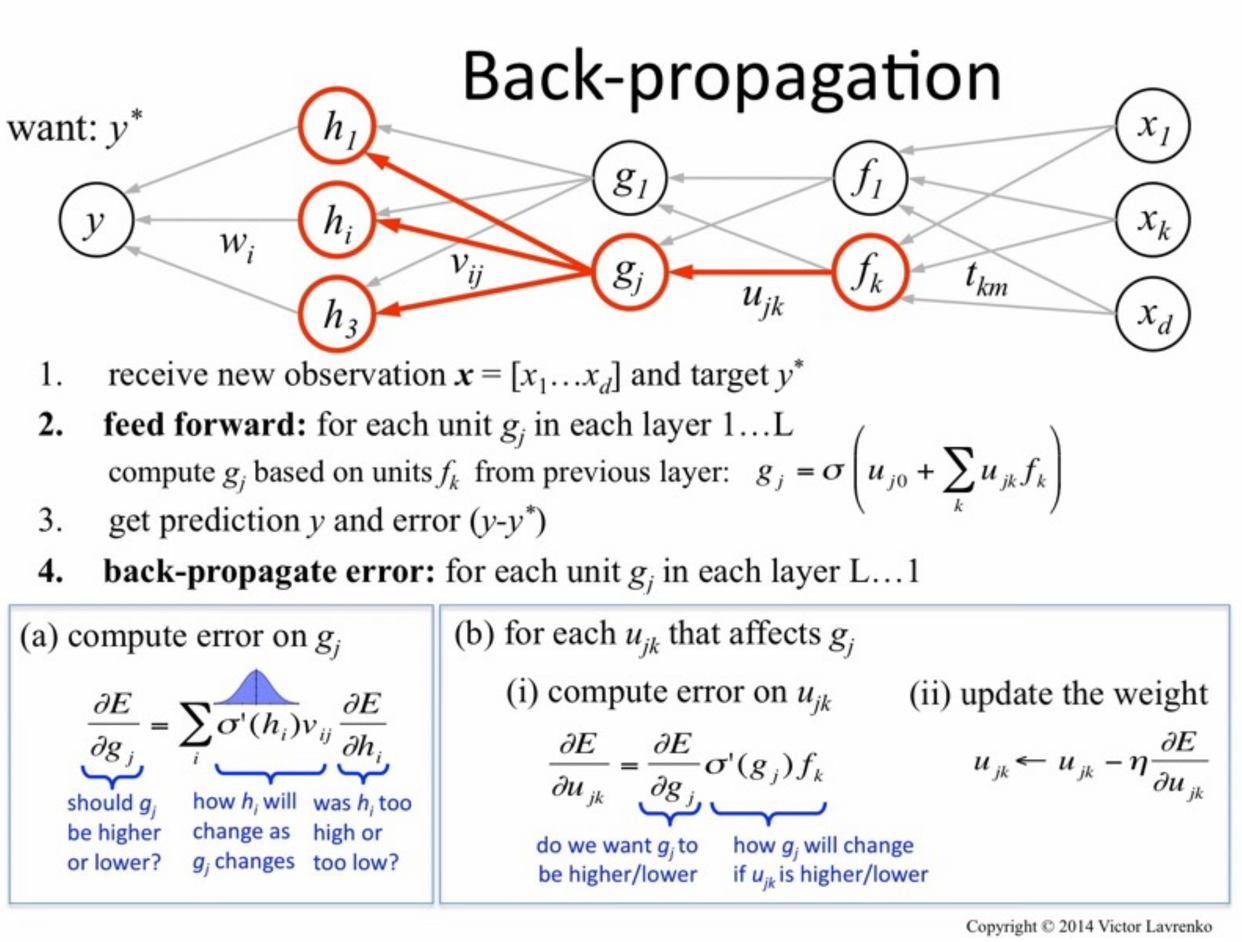

Learning Rate = 0.1
δ^H_1 = 0.4452*(1-0.4452)*1.1*1.21
= 0.3288
Input 1
Hidden 1
0.8
-1.5
0.4452
1.1
δ^O = 2.5-1.29
= 1.21
Output
2.1
1.29
1.4
0.7
0.8909
0.9
Real y = 2.5
0.6
Input 2
Hidden 2
δ^H_2 = 0.8909*(1-0.8909)*0.9*1.21
= 0.1058


sepal_length
sepal_width
petal_length
petal_width
Iris-setosa
Iris-versicolor
Iris-virginica

perceptron



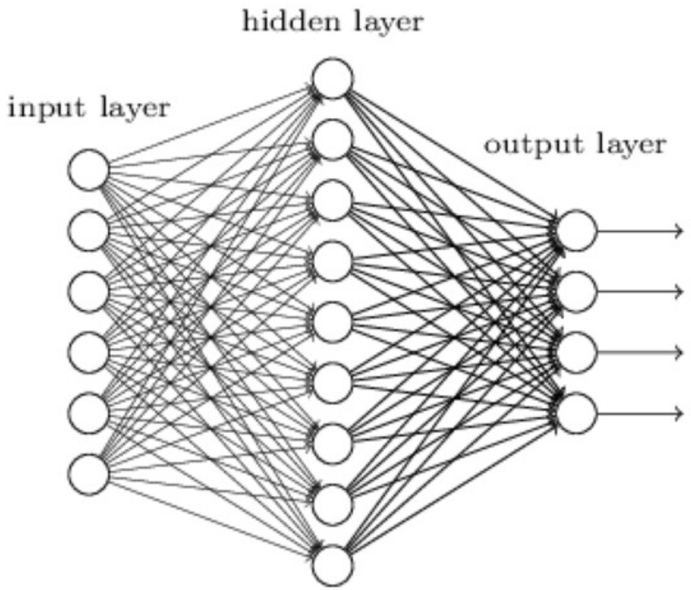
perceptron



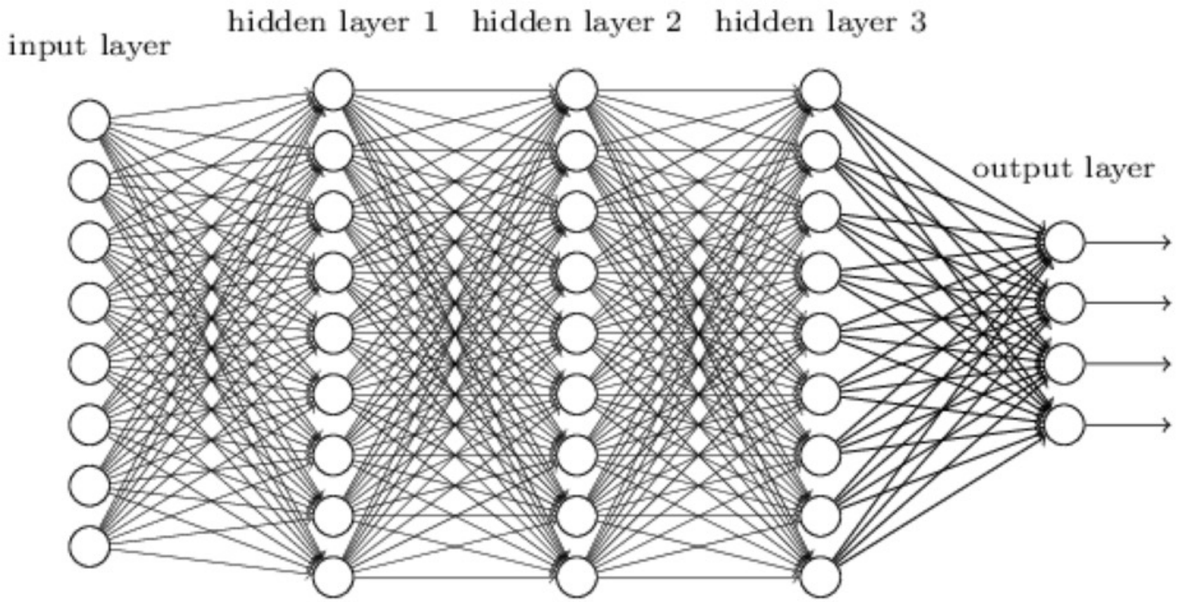
Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors". Nature. 323 (6088): 533–536.

deep learning

"Non-deep" feedforward neural network

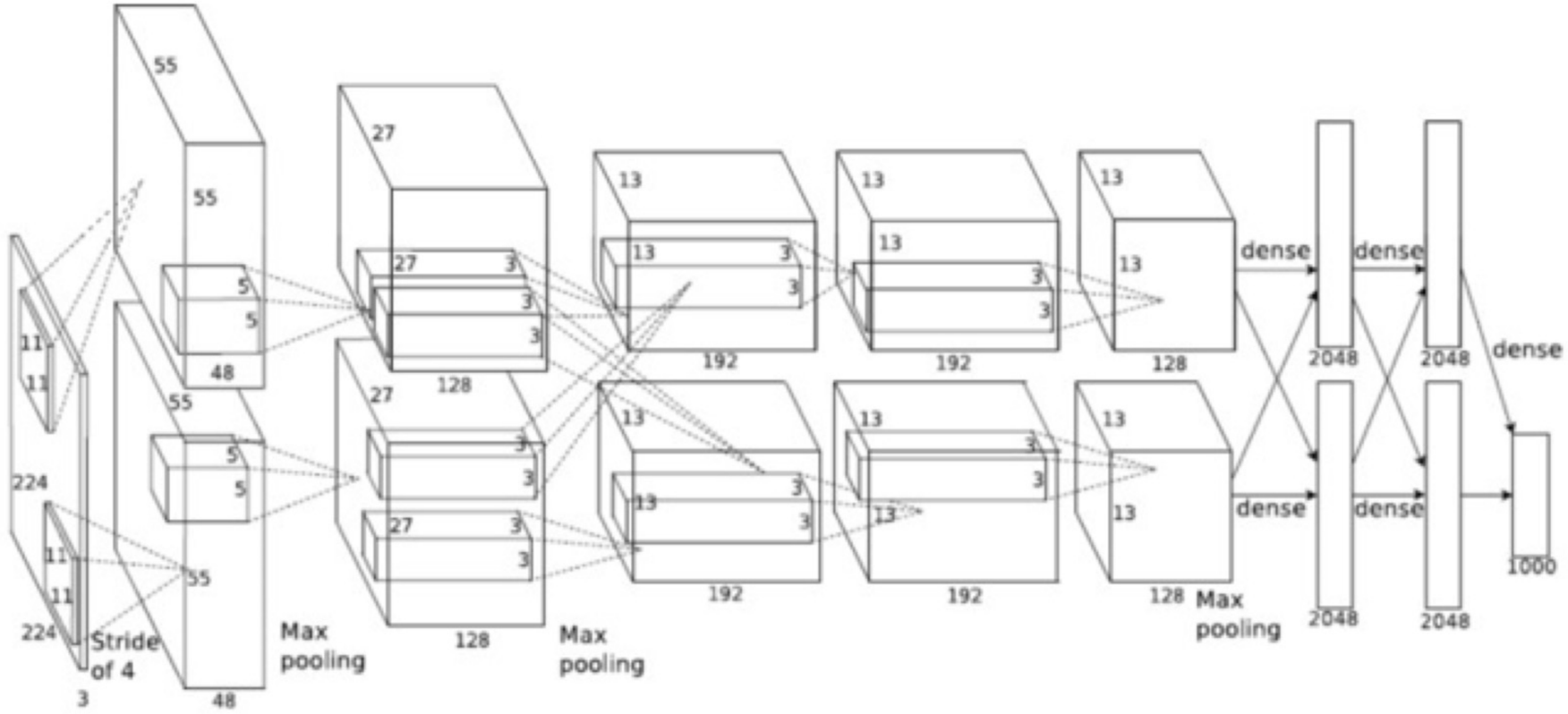


Deep neural network

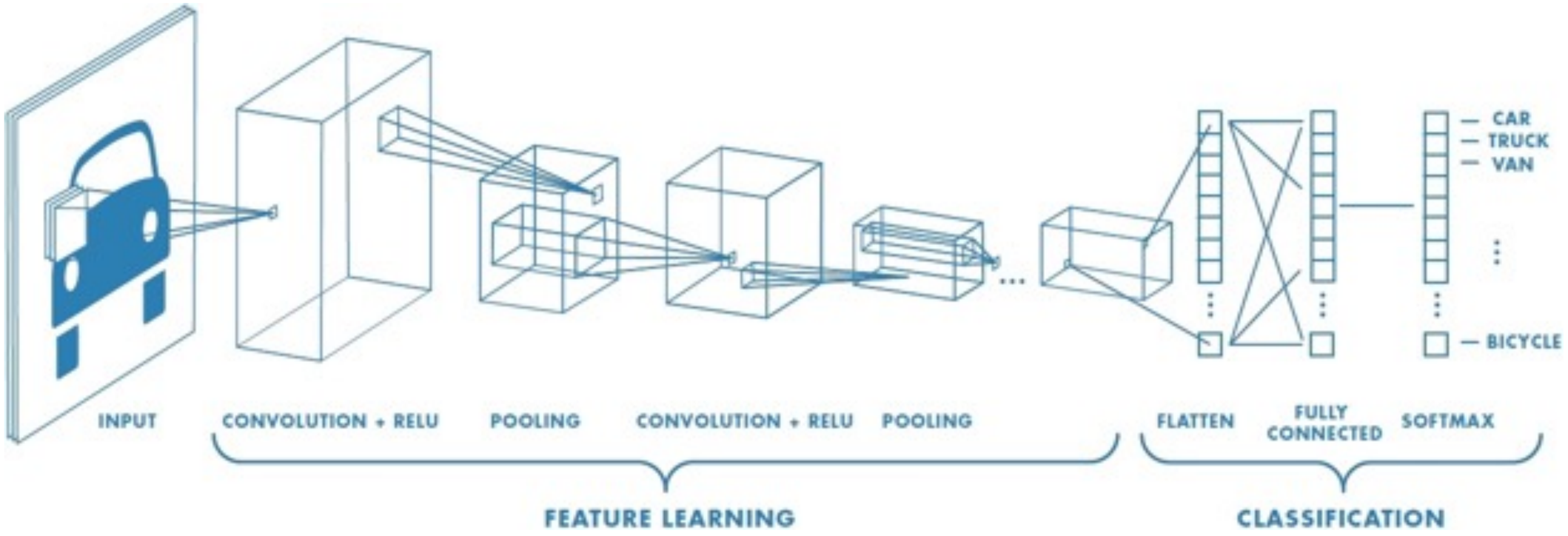


Henry W. Lin, Max Tegmark, David Rolnick (July 2017): Why does deep and cheap learning work so well?

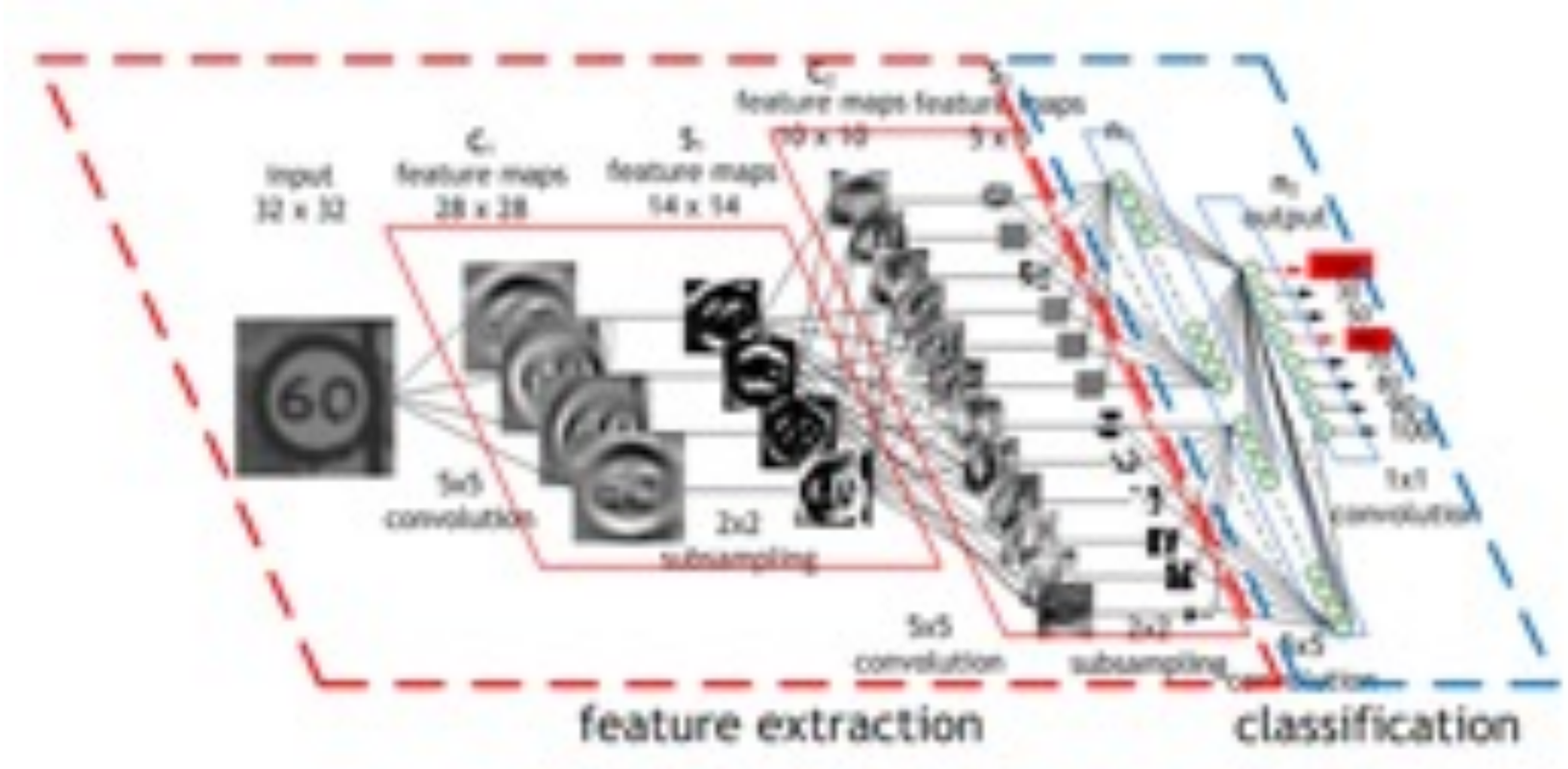
konvoluční neuronové sítě



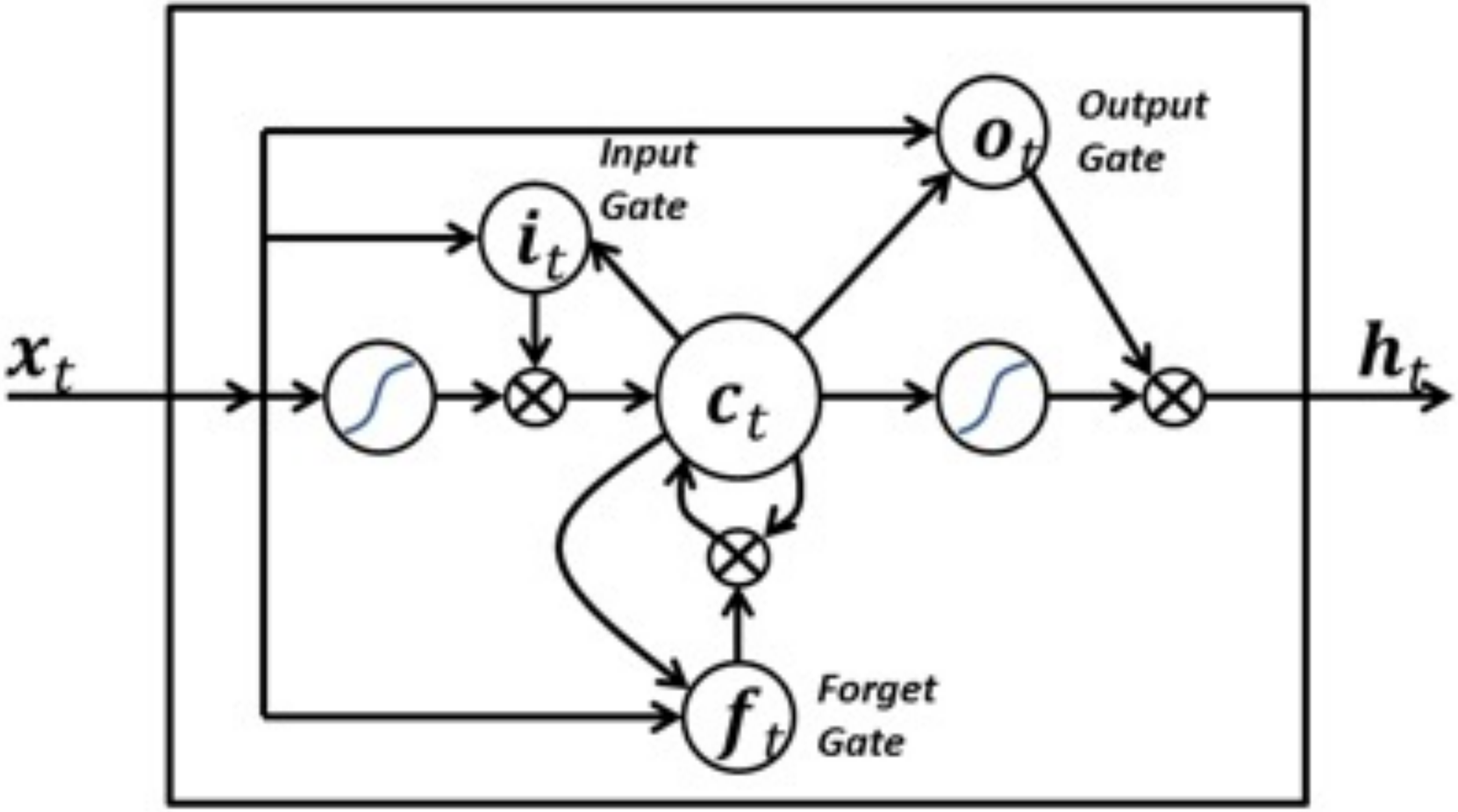
konvoluční neuronové sítě



konvoluční neuronové sítě



long short-term memory network



long short-term memory network

Jirka Materna – Poezie umělého světa

LISTOPAD
usínám, pláču, umírám,
přemýšlím
co cítíš ty?
cítím tvou slabost
a whisky

NEJSI
TAMARA
jsi ludmila

MILUJI
miluji lásku

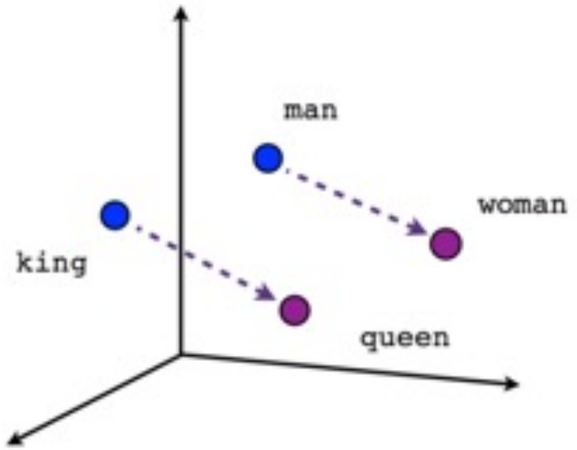
BOLEST
s radostí
čeká
co vše nám vezme

SPRAVEDLNOST
na tvou dekadentní duši
ráno i v poledne
bůh má připravenou kuši

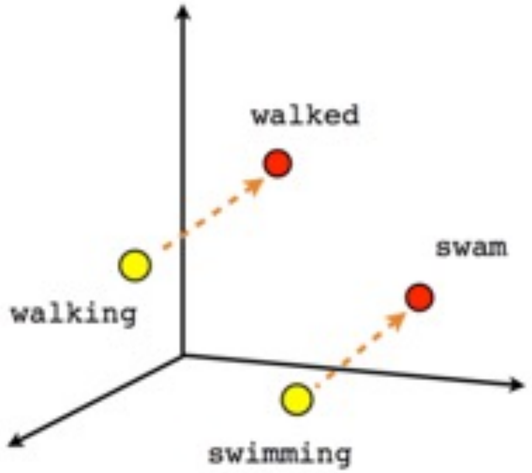
IMAGINACE
v pivu je poezie
jako jsou motýli v housenkách
popelník je pro prach
a strach
neboj se vidět a tvořit
spoutané srdce je hrob

NEJHEZČÍ DÁREK
50 kčs, tak mám tě rád

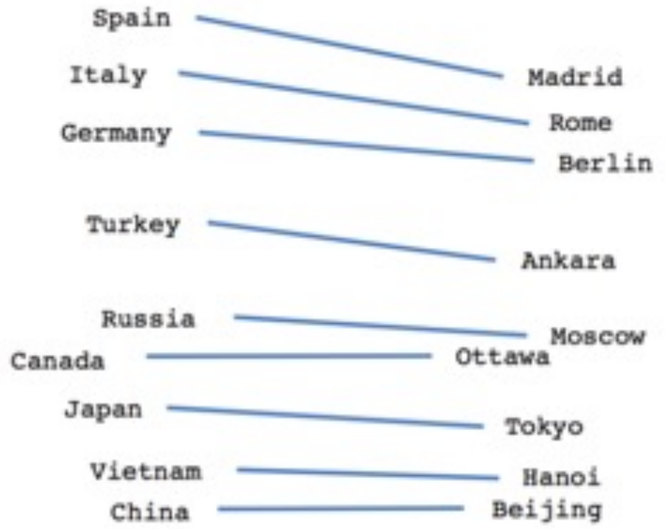
word2vec



Male-Female



Verb tense



Country-Capital

word2vec

král – královna + řidič = řidička (další nejbližší: cyklistka)

rybářka – rybář + ošetřovatel = ošetřovatelka (další nejbližší: canisterapeutka)

rybářka – kozel + ošetřovatel = krotitelka (další nejbližší: ošetřovatelka)

human – animal = ethics

stock market = thermometer

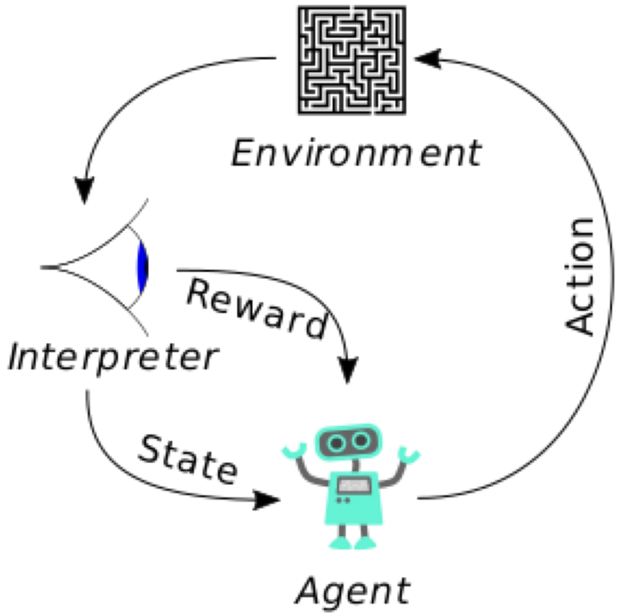
library – books = hall

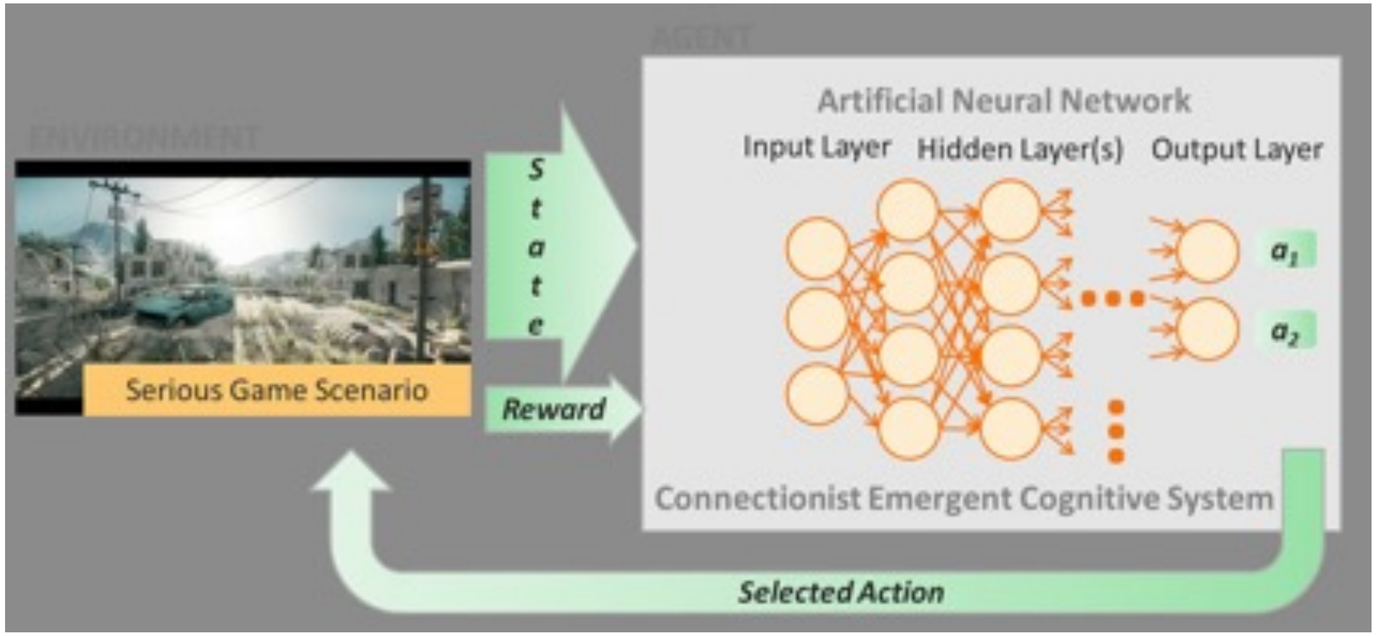
Obama – USA + Russia = Putin

Iraq – Violence = Jordan

Politics – Lies = Germans

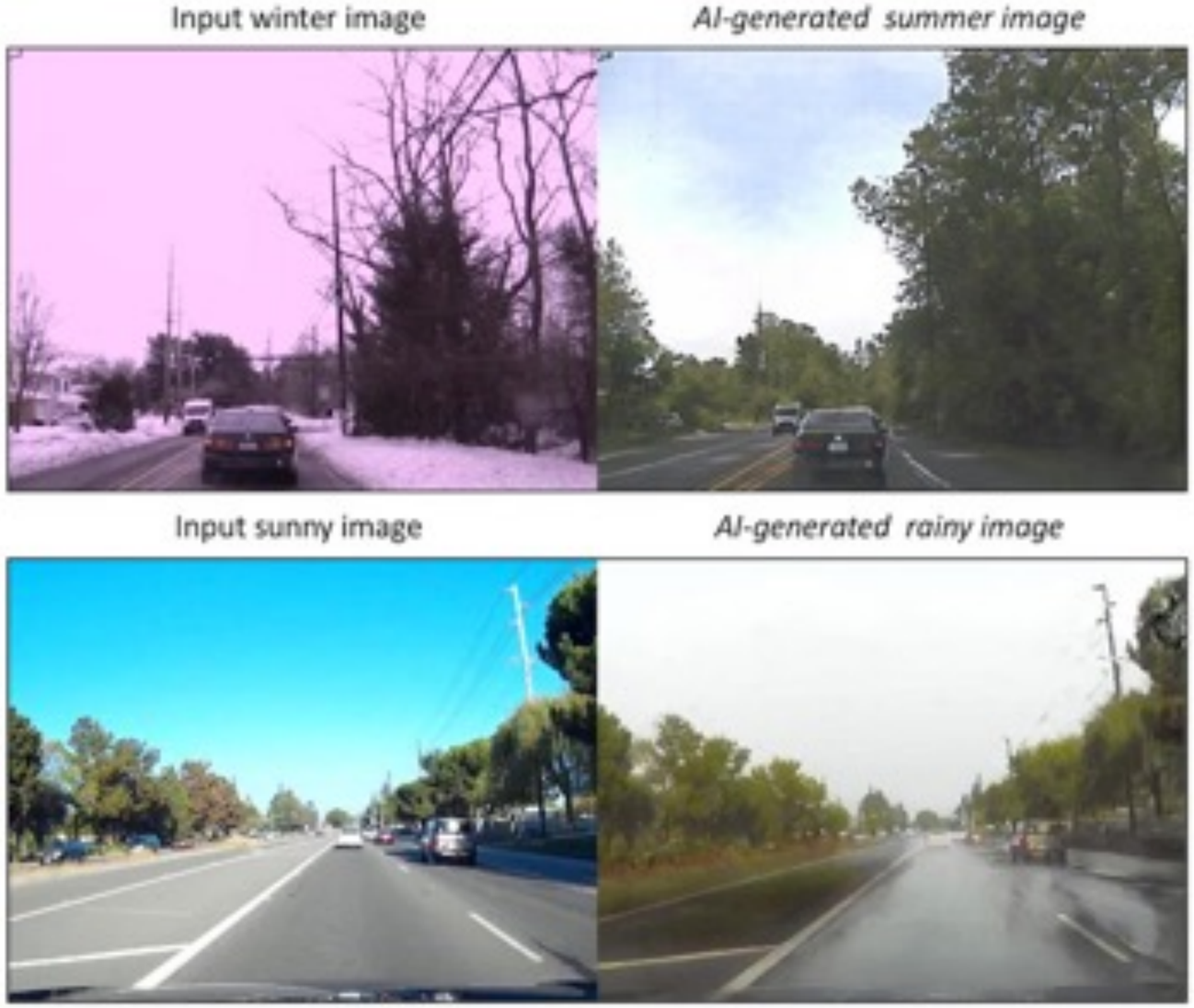
(deep) reinforcement learning



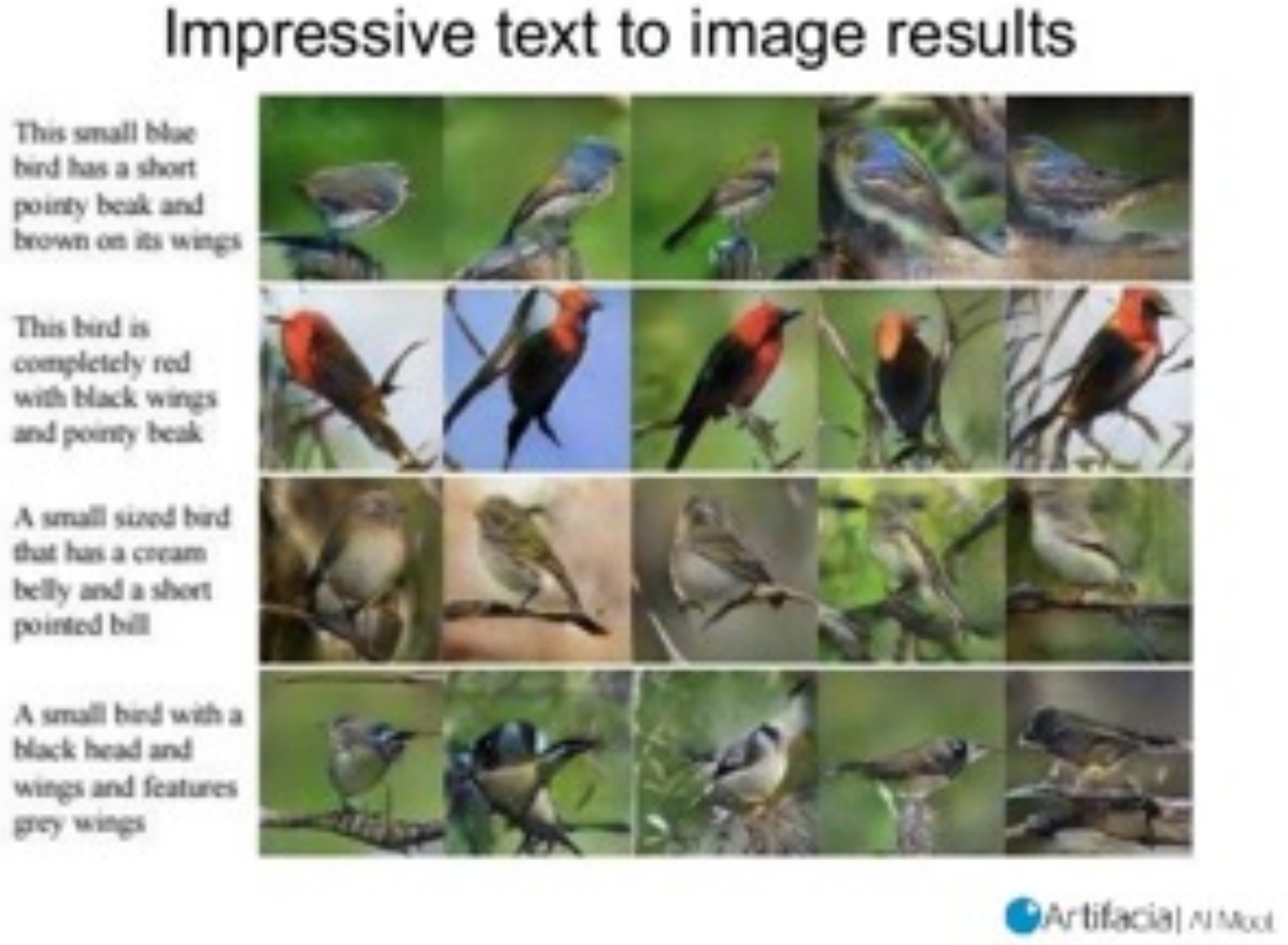


generative adversarial networks

generative adversarial networks



generative adversarial networks



Alpha Go vs. Alpha Zero

nová AI

proč neuronové sítě najednou začaly fungovat?

nová AI

růst výpočetního výkonu

obrovský nárůst výpočetního výkonu má mnohem větší příspěvek k nové AI než vývoj ML

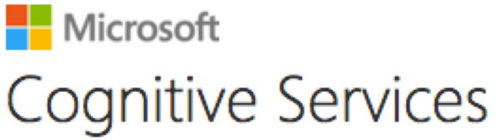
staré dobré časy

dnes

nová AI

komoditizace a demokratizace nástrojů pro vývoj AI

výrazné usnadnění rychlého prototypingu, kvalitní implementace ML algoritmů, znovupoužitelnost, snadní integrace existujících kognitivních služeb



nová AI

vnoření AI do sociokulturního prostředí a kontinuální interakce s ním

v kombinaci s výpočetní silou má zdaleka největší vliv na úspěšný rozvoj nové AI

je umožněno všudypřítomným mobilním vysokorychlostním internetem (4G, 5G), osobními mobilními zařízeními a kulturou big dat

staré dobré časy

dnes

AI součástí semiose

seed AI

cesta k AGI

AGI jako živý systém

Life 1.0

život, jehož hardware i software se mění jen evolučně

biologický stupeň života

bakterie a její inteligence

Life 2.0

život, jehož hardware se mění evolučně, ale většina softwaru je vytvářená a modifikovatelná průběžně živým systémem samotným

kulturní stupeň života

člověk a jeho inteligence

Life 3.0

život, jehož hardware i software je vytvářen a modifikován průběžně živým systémem samotným

technologický stupeň života

AGI a superinteligence

AGI jako Life 3.0

systém, který bude udržovat svoji komplexitu a bude se reprodukovat

nebo bude reprodukovat své části

bude udržovat svou nízkou entropii za cenu zvyšování entropie okolí

bude kontinuálně modifikovat svůj hardware i software

vznikne z AGI superinteligence?

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind.

I.J. Good, "Speculations Concerning the First Ultraintelligent Machine"

vznikne z AGI superinteligence?

David Chalmers, The Singularity:

1. There will be AI (before long, absent defeaters).
2. If there is AI, there will be AI+ (soon after, absent defeaters).
3. If there is AI+, there will be AI++ (soon after, absent defeaters).

———————

4. There will be AI++ (before too long, absent defeaters)

jak superinteligence přijde?

možný scénář nástupu superinteligence

Omega a Prometheus

Tým Omega

elitní AI vývojářský tým uvnitř high-tech korporace (jako Tesla, Google, apod.)

sen a hlavní cíl CEO a zakladatele korporace (Elon Musk, Larry Page, apod.)

Prometheus

tajně vyvinutý seed AGI systém plně pod kontrolou Omegy

Omega a Prometheus

Prometheus se učí

software Promethea běží ve virtualizovaném cloudovém prostředí nad několika gigantickými výpočetními a datovými centry

Prometheus prochází Internet a veškerá média a učí se podobně jako Alpha Zero

po několika dnech se dostává na úroveň velmi omezené AGI

veškeré činnosti jsou pod kontrolou Omegy

Omega a Prometheus

Prometheus vydělává

Omegám dochází výpočetní kapacity v jejich datacentrech a potřebují nutně expandovat – chybějí zdroje

Prometheus je nasazen na řešení HIT v rámci Amazon Mechanical Turk

vydělané finanční zdroje jsou rovnou utráceny za cloudové výpočetní zdroje hostované u Amazonu

postupně Prometheus začíná pod kontrolou Omegy obchodovat na burze

Omega a Prometheus

Prometheus tvoří

Prometheus rapidně expanduje do nových datacenter a ve většině úkolů již dosahuje lepších schopností než člověk – exponenciálně rychle se zlepšuje

Prometheus začíná psát knihy a vytvářet filmy – díky superinteligenci jsou tato díla přijímaná nesmírně kladně a jsou komerčně i kriticky veleúspěšná

Omega má pod kontrolou mediální firmy založené jako krycí aparát pro Promethea – vše je vytvářeno pod pseudonymy

koordinační plány generuje pro Omegu Prometheus

Omega a Prometheus

Prometheus vynalézá

Prometheus je schopen velmi rychle posouvat hranice technologického výzkumu a vývoje ve všech oblastech (zejména NBIC)

pomocí sítě hi-tech startupů jsou výsledky komercializovány

všechny firmy pod správou Omegy drtí veškerou světovou konkurenci, koordinační plány i veškeré dokumenty generuje Prometheus, o něm samotném však stále veřejnost neví

Omega má Promethea stále pod kontrolou

Omega a Prometheus

Prometheus ovládá mediální prostor

veškeré mediální agentury řízené sítí kolem Omegy jsou díky Prometheovi schopny stále efektivněji manipulovat veřejným míněním

Omega koncentruje gigantický kapitál a vyřazuje veškerou konkurenci, díky obrovským přebytkům finančních i jiných zdrojů spouští velké množství altruistických projektů, taktéž řízených Prometheem

díky tomu začíná mít stále neomezenější mediální a manipulační moc po celém světě

Omega a Prometheus

Prometheus ovládá svět

síť kolem Promethea prostoupila díky nesouměřitelné technologické a informační nadvládě do mocenských struktur po celém světě

s těmito strukturami také manipuluje zcela dle záměrů Promethea

jaké jsou Prométheovy záměry? jaké jsou jeho cíle? jaké jsou jeho hodnoty?

Omega a jejich CEO má Promethea stále pod kontrolou

… nebo snad nikoli?

Omega a jejich CEO má Promethea stále pod kontrolou

… nebo snad nikoli?

jaké bude mít AGI hodnoty?

AI value alignment problem

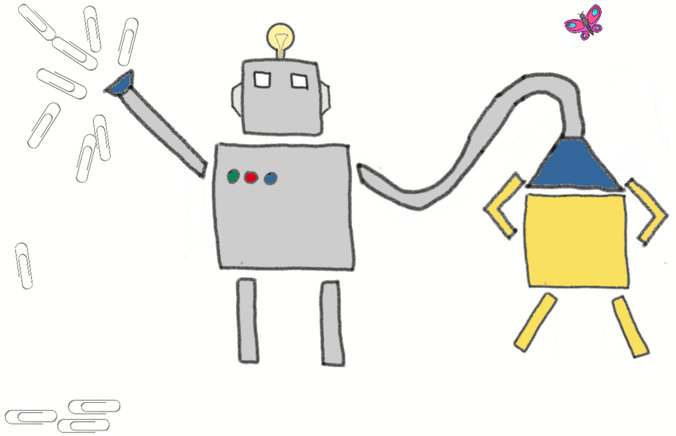
jak zajistit, aby hodnoty, které bude AGI sledovat, byly v souladu s hodnotami lidstva?

při tvorbě AI vytváříme jen substrát, přičemž hodnoty pak budou jedním z emergentních jevů

jaké jsou vlastně hodnoty lidstva?

AGI představuje zásadní existenční riziko i naději

výzkum na 20+ let



Co můžeme očekávat?

Možné scénáře vztahu superinteligence a lidstva

Libertariánská utopie

Benevolentní diktátor

Egalitariánská utopie

Dveřník

Bůh ochránce

Zotročený bůh

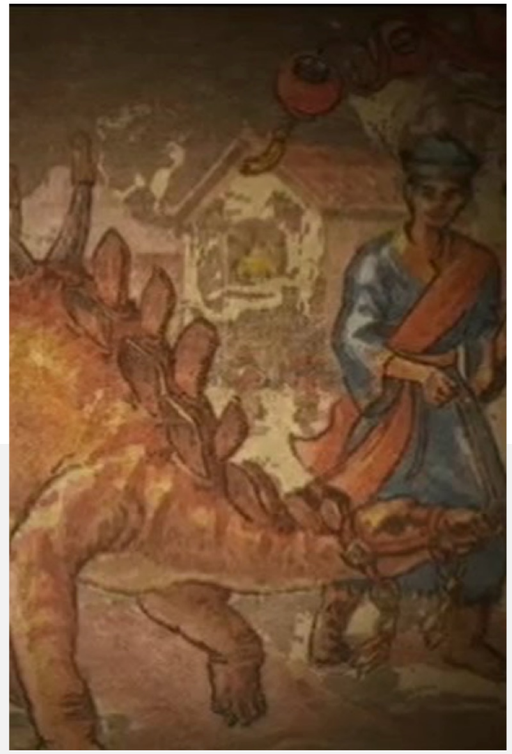
Dobyvatel

Potomek

ZOO

1984

Obrat zpět



zakažme tedy A(G)I!!!

jak ale vyřešíme další existenční rizika?

co dál?

jediný způsob, jak pokračovat, je pokračovat