

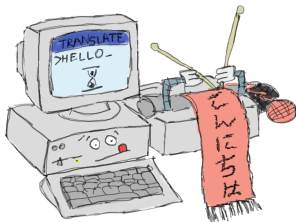
Překladač z angličtiny a ukrajinštiny a další aplikace AI

Mgr. Martin Popel, Ph.D.

Ústav formální a aplikované lingvistiky,
Matematicko-fyzikální fakulta, Univerzita Karlova

<https://ufal.mff.cuni.cz/martin-popel>

Informační technologie a jejich uplatnění ve vzdělávacím procesu
2022-10-05



source

Great talkers are little doers.

Yandex

Velké talkers jsou trochu činitelé.

Microsoft

Velcí vysílačky jsou malí činitelé.

Google

Velcí mluvčí jsou malí lidé.

Google 2022

Velcí řečníci jsou malí dělníci.

TectoMT

Velcí řečníci jsou malí vrazi.

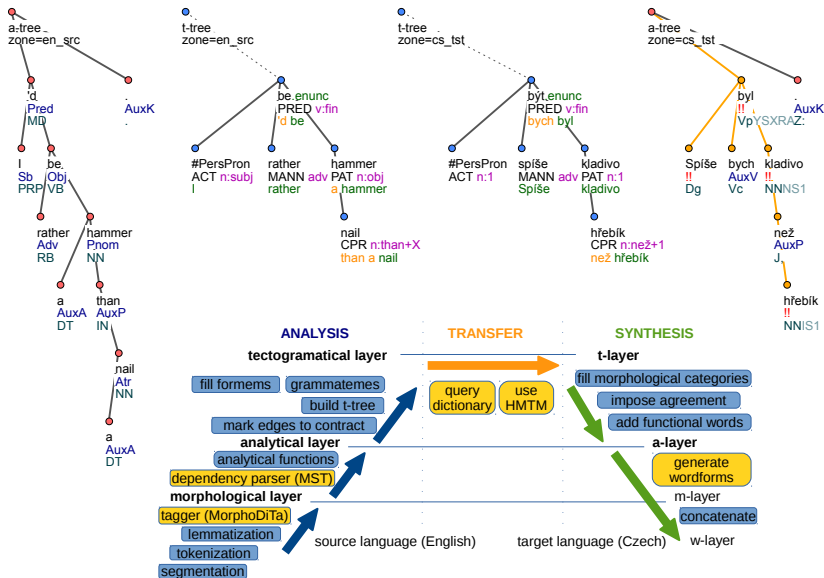
CUBBITT

Velcí mluvkové jsou malí dřiči.

Jak může strojový překlad pomoci výuce?

2





I'd rather be a hammer than a nail.

Spíše bych byl kladivo než hřebík/nehet.

output_label=hřebík#N	
feature	λ
child_formeme_n:in+X=1	1.64
is_member_of_coord=1	1.30
child_formeme_v:fin=1	1.04
next_lemma=down	0.84
is_capitalized=1	0.79
+precedes_parent=0	0.75
tense_g=post	0.74
+voice_g=active	0.66
prev_lemma=drive	0.66
parent_capitalized=1	0.62
formeme=n:from+X	0.60
+prev_lemma=hammer	0.59
child_lemma_few=1	0.55
child_lemma_remove=1	0.54
sempos=n.denot	0.50
next_lemma=and	0.50
formeme_g=v:until+fin	0.49
child_lemma_rusty=1	0.47
...	

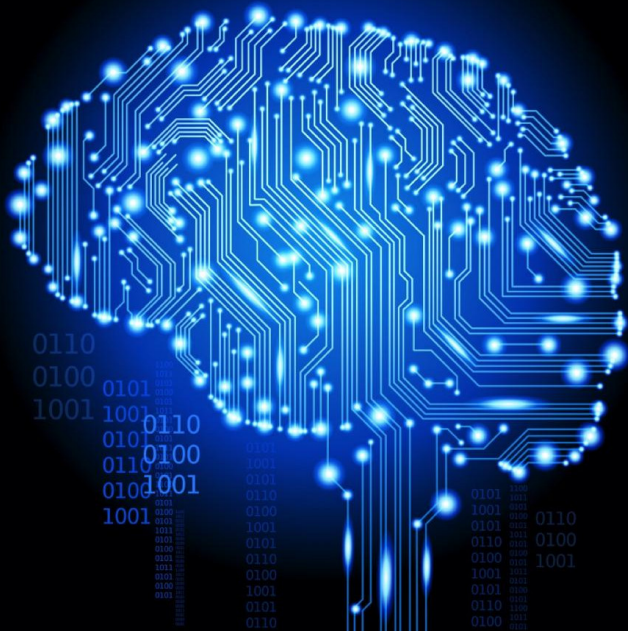
output_label=hřebík#N	
feature	λ
child_formeme_n:in+X=1	1.64
is_member_of_coord=1	1.30
child_formeme_v:fin=1	1.04
next_lemma=down	0.84
is_capitalized=1	0.79
+precedes_parent=0	0.75
tense_g=post	0.74
+voice_g=active	0.66
prev_lemma=drive	0.66
parent_capitalized=1	0.62
formeme=n:from+X	0.60
+prev_lemma=hammer	0.59
child_lemma_few=1	0.55
child_lemma_remove=1	0.54
sempos=n.denot	0.50
next_lemma=and	0.50
formeme_g=v:until+fin	0.49
child_lemma_rusty=1	0.47
...	

output_label=nehet#N	
feature	λ
child_formeme_n:poss=1	1.32
child_lemma_finger=1	1.07
child_formeme_n:of+X=1	0.98
precedes_parent=1	0.88
prev_lemma=black	0.77
child_lemma_broken=1	0.76
child_formeme_v:attr=1	0.70
formeme=n:at+X	0.67
formeme_g=n:attr	0.67
child_lemma_long=1	0.67
next_lemma=file	0.60
child_lemma_false=1	0.58
prev_lemma=false	0.58
+number=sg	0.56
formeme=n:obj	0.53
formeme=n:by+X	0.52
...	

umělá inteligence
~1950

strojové učení
~1980

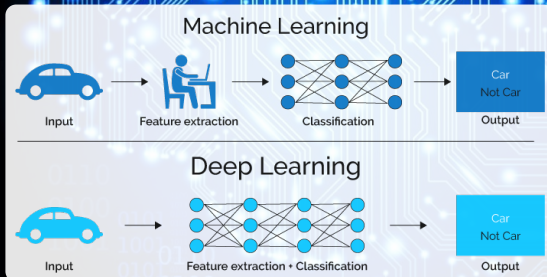
hluboké učení
~2010



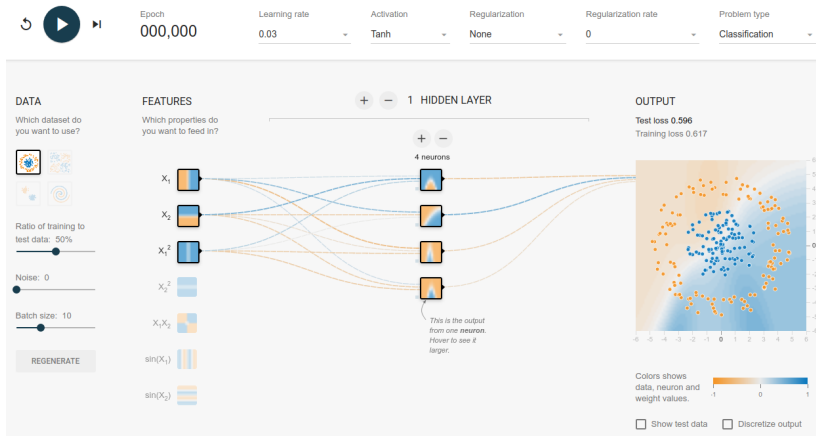
umělá inteligence
~1950

strojové učení
~1980

hluboké učení
~2010

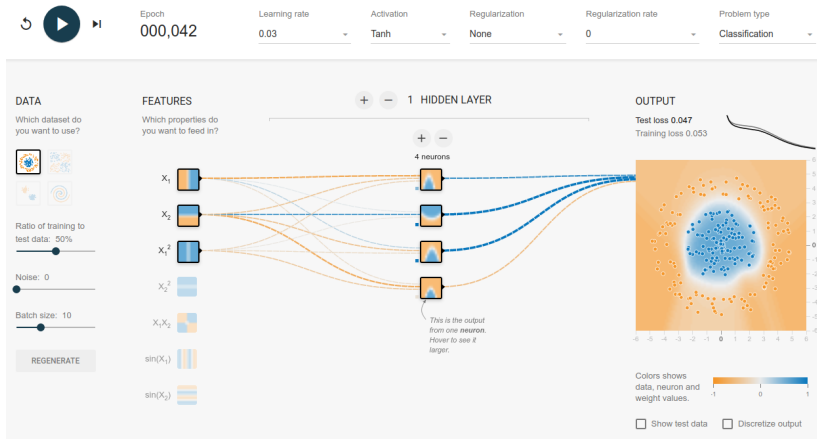


jednoduchá architektura (16 parametrů)



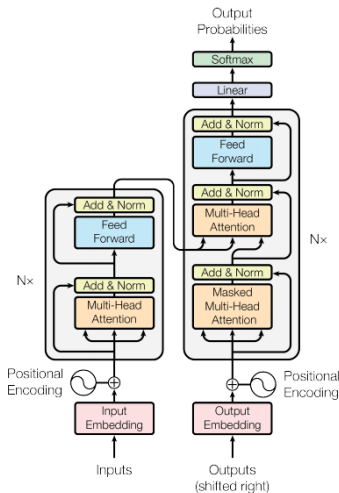
<https://playground.tensorflow.org>

jednoduchá architektura (16 parametrů)

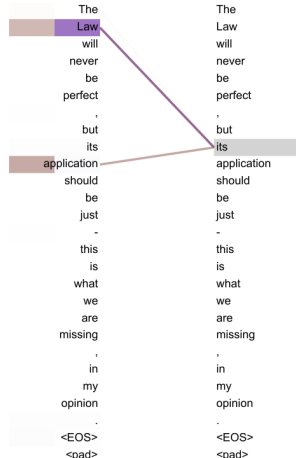


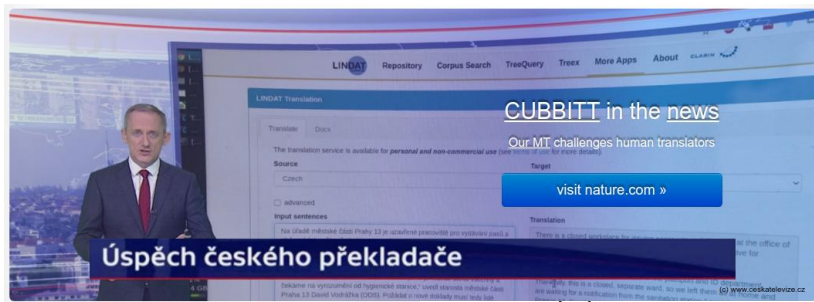
<https://playground.tensorflow.org>

architektura Transformer (213 milionů parametrů)



síť se sama učí důležité vztahy mezi slovy (díky self-attention)





Úspěch českého překladače

nature communications

Zkuste si CUBBITT sami:
<https://lindat.cz/cubbitt>
(En↔Cs, Fr, Pl)

nature > nature communications > articles > article

Article | [Open Access](#) | Published: 01 September 2020

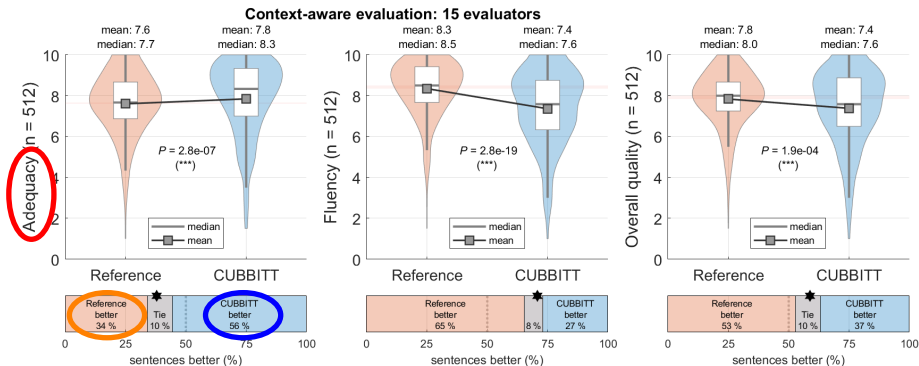
Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

Martin Popel , Marketa Tomkova, Jakub Tomek, Lukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar & Zdeněk Žabokrtský

Nature Communications **11**, Article number: 4381 (2020) | [Cite this article](#)

6273 Accesses | 76 Altmetric | [Metrics](#)

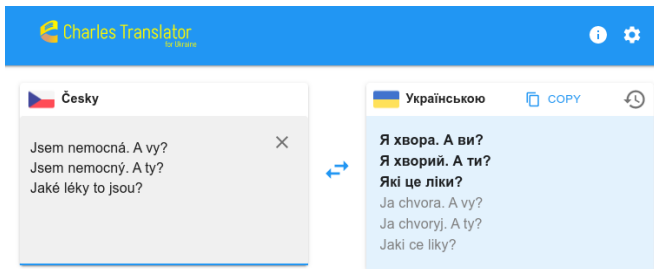
56 % vět přeložil **přesněji** náš překladač CUBBITT,
34 % profesionální překladatelská agentura





Zkuste si Charles Translator:
<https://translator.cuni.cz>





- první verze překladače vytvořena za 2 týdny
- cs→uk: asi 1400 uživatelů denně, uk-cs: asi 700
- přímý překlad (nikoli přes angličtinu)
- fonetický přepis azbuky, hlasový vstup,...
- API: překlad učebnic, PomahejUkrajine.cz,...

 Charles Translator
for Windows



 **Česky**

Jsem nemocná. A vy?
Jsem nemocný. A ty?
Jaké léky to jsou?



 **Українською**

 COPY 

Я хвора. А ви?
Я хворий. А ти?
Які це ліки?
Ja chvora. A vy?
Ja chvoryj. A ty?
Jaki ce liky?

 **Google Překladač**

 [Přihlásit se](#)

ROZPOZNAT JAZYK **ČEŠTINA** AN( ↔ **UKRAJINŠTINA** ČEŠTINA ANGLIČ 

Jsem nemocná. A vy?
Jsem nemocný. A ty?
Jaké léky to jsou?



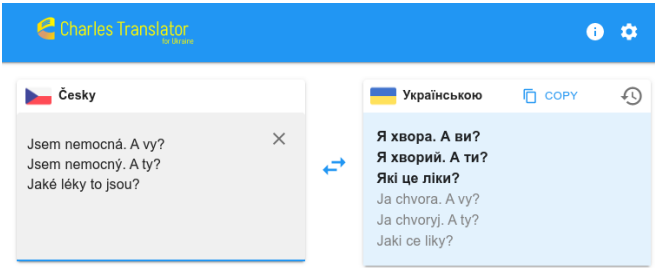
Я хворий. І ти?
Я хворію. І ти?
Які це наркотики?

YA khvoryy. I ty?
YA khvoriyu. I ty?
Yaki tse narkotyky?

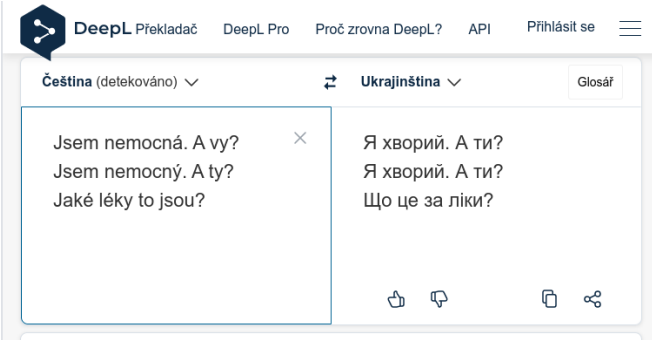


  58 / 5 000 

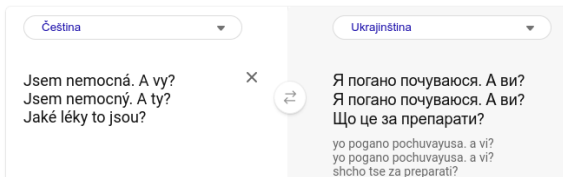
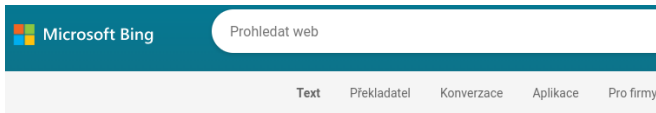
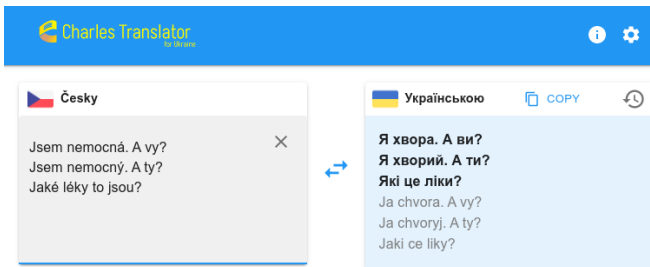
   

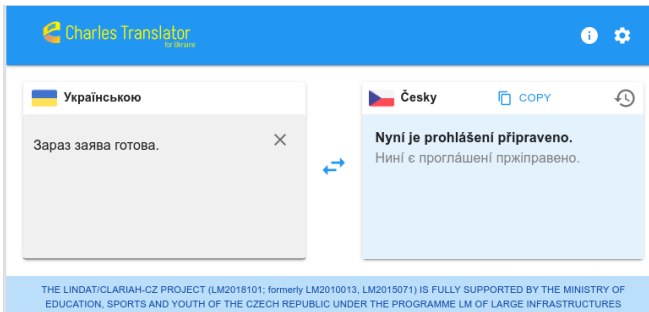


The Charles Translator interface features a blue header with the logo and name "Charles Translator" on the left, and information and settings icons on the right. Below the header, there are two main text boxes. The left box, titled "Česky" with a Czech flag icon, contains the text: "Jsem nemocná. A vy?", "Jsem nemocný. A ty?", and "Jaké léky to jsou?". The right box, titled "Українською" with a Ukrainian flag icon, contains the text: "Я хвора. А ви?", "Я хворий. А ти?", "Які це ліки?", "Ja chvora. A vy?", "Ja chvoryj. A ty?", and "Jaki ce liky?". A double-headed arrow is positioned between the two boxes. Above the right box, there are icons for "COPY" and a refresh symbol.

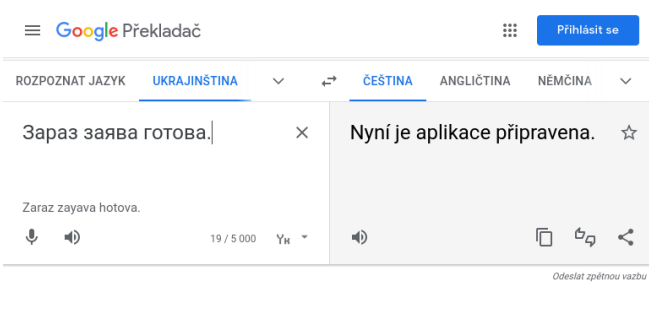


The DeepL Translator interface has a dark blue header with the DeepL logo and navigation links: "DeepL Překladač", "DeepL Pro", "Proč zrovna DeepL?", "API", and "Přihlásit se". Below the header, the source language is set to "Čeština (detekováno)" and the target language is "Ukrajínština". A "Glosář" button is on the right. The left text box contains: "Jsem nemocná. A vy?", "Jsem nemocný. A ty?", and "Jaké léky to jsou?". The right text box contains: "Я хворий. А ти?", "Я хворий. А ти?", and "Що це за ліки?". At the bottom of the interface, there are icons for thumbs up, thumbs down, copy, and share.

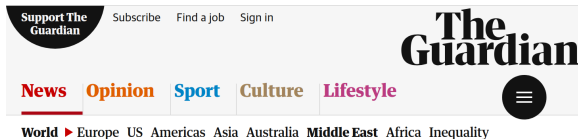




The screenshot shows the Charles Translator desktop application. The interface is split into two main panes. The left pane, titled 'Українською' (Ukrainian) with a Ukrainian flag icon, contains the text 'Зараз заява готова.' (The statement is ready now). The right pane, titled 'Česky' (Czech) with a Czech flag icon, contains the translated text 'Nyní je prohlášení připraveno.' (The statement is now ready) and its Ukrainian equivalent 'Нині є прогласені пржіправено.' Below the panes, a blue banner states: 'THE LINDAT/CLARIAH-CZ PROJECT (LM2018101; formerly LM2010013, LM2015071) IS FULLY SUPPORTED BY THE MINISTRY OF EDUCATION, SPORTS AND YOUTH OF THE CZECH REPUBLIC UNDER THE PROGRAMME LM OF LARGE INFRASTRUCTURES'. The top bar features the 'Charles Translator for desktop' logo and settings/info icons.



The screenshot shows the Google Překladač (Google Translate) web interface. At the top, there's a navigation bar with the Google logo, the text 'Překladač', a grid icon, and a 'Přihlásit se' (Sign in) button. Below this is a language selection bar with 'ROZPOZNAT JAZYK' (Detect language), 'UKRAJINŠTINA' (Ukrainian), 'ČEŠTINA' (Czech), 'ANGLIČTINA' (English), and 'NĚMČINA' (German). The main area is split into two columns. The left column shows the input text 'Зараз заява готова.' and its transcription 'Zaraz zayava hotova.' with a character count of '19 / 5 000'. The right column shows the translated text 'Nyní je aplikace připravena.' and its Ukrainian equivalent 'Нині є прогласені пржіправено.' with a star icon. At the bottom right, there are icons for copy, share, and print, along with the text 'Odeslat zpětnou vazbu' (Send feedback).



Facebook

Facebook translates 'good morning' into 'attack them', leading to arrest

Palestinian man questioned by Israeli police after embarrassing mistranslation of caption under photo of him leaning against bulldozer

zdroj: [The Guardian](#) (2017)



Jak může strojový překlad pomoci výuce?

13



Jak může strojový překlad pomoci škodit výuce?

14



- přílišná důvěra v kvalitu
- předpoklad neustálé dostupnosti
- Za domácí úkol přeložte tento článek. / Napište esej.
- Překladače (zatím) neumí vysvětlit
 - proč zvolily daný překlad
 - proč je jiný překlad chybný
 - na jaké chyby se zaměřit nejdříve

Byl by to rytíř, kde v pláně hřích vzlet,
Vědě jsem jse seheldo na přídoutně v světě si nezastává:
„Ukryjemné, chvěla, milý nás jest

Kolem jsou jest vyhrávaných
A svítí co pláčem, rád pravil:
Ale plná jízdo zaporodilo se, vys.

již dávno vás poháru a vlanných rány,
v jablonění je píše je i v kristování,

srdce v své ženských svém
v obly pětky tam a vzíti,
na kónku je, milý svěžek.

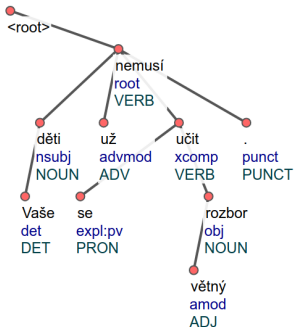
MFF + DAMU + Švandovo divadlo
premiéra 26. 2. 2021 (100 let po R.U.R.)



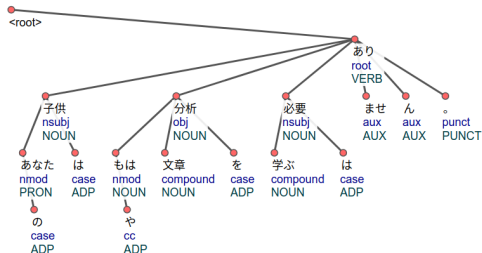
<https://www.theaitre.com>

větný rozbor dostupný pro 50 jazyků
přesnost pro češtinu asi 90% (85% včetně morfologie)

Vaše děti se už nemusí učit větný rozbor .

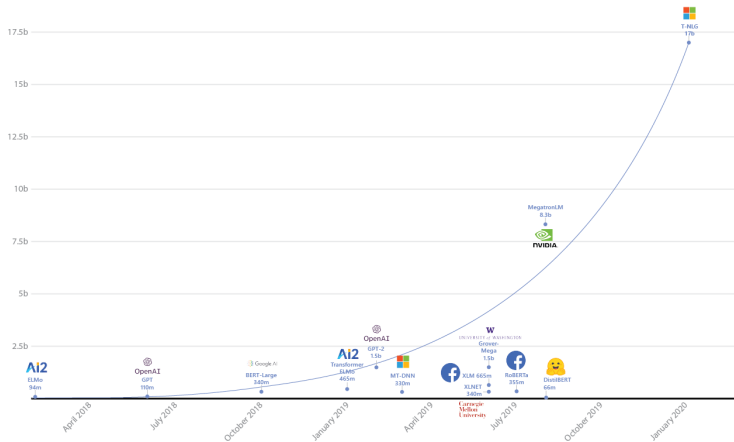


あなたの子供はもはや文章分析を学ぶ必要はありません。



<https://lindat.cz/services/udpipe/>

- Lidský mozek: asi 100G neuronů, asi 1000T synapsí
- Největší model 2018 (ELMo): 94M parametrů (synapsí)
- Největší model 2020 (GPT-3): 175G parametrů
- Největší model 2021 (Switch): 1.6T parametrů



Source	Jana je žena. Pracuje jako průvodčí.
Google	Jana is a woman. He works as a guide.
Bing	Jana is a woman. He works as a conductor.
CUBBITT	Jane is a woman. He works as a conductor.
CUBBITT-doc	Jana is a woman. She works as a conductor.

Source	Saša je muž. Pracuje jako průvodčí.
DeepL (2021)	Sasha is a man. She works as a conductor.
CUBBITT-doc	Sasha is a man. He works as a conductor.

Source	Pavla není muž. Pracuje jako průvodčí.
DeepL	Paul is not a man. She works as a conductor.
CUBBITT-doc	Pavla is not a man. He works as a conductor.

source	As good be an addled egg as an idle bird.
Bing	Jako dobrý být popletený vejce jako nečinný pták.
Google	Jako dobrá být včleněná vejce.
T2009	Dobré je fetácké vejce jako činný pták.
T2018	Dobří buďte plete vejce jako nečinný pták.
CUBBITT	Stejně dobré je být pomateným vejcem jako zahálejícím ptákem.

source	A miss by an inch is a miss by a mile.
--------	--

Bing	Miss o palec je Miss o míli.
Yandex	Slečna tím, že palec je vedle o míli.
Google	Chybějící palcem je míle vzdálená míle.
T2009	Slečna palec je slečna miliónu.
T2018	Slečna palce je slečna míle.
CUBBITT	Minutí o centimetr je o kilometr.

Birds of a feather flock together.

Ptáci peří stáda dohromady.
Vrána k vráně sedá.
Vrána k vráně sedá.
Ptáci v bederním hejnu spolu.
Ptáci péřového hejna spolu.
Vrána k vráně sedá.

Zkuste si CUBBITT sami: <https://lindat.cz/cubbitt>