

Kdy neuronový strojový překlad nahradí překladatele?



Ondřej Bojar

bojar@ufal.mff.cuni.cz

Ústav formální a aplikované lingvistiky

Matematicko-fyzikální fakulta

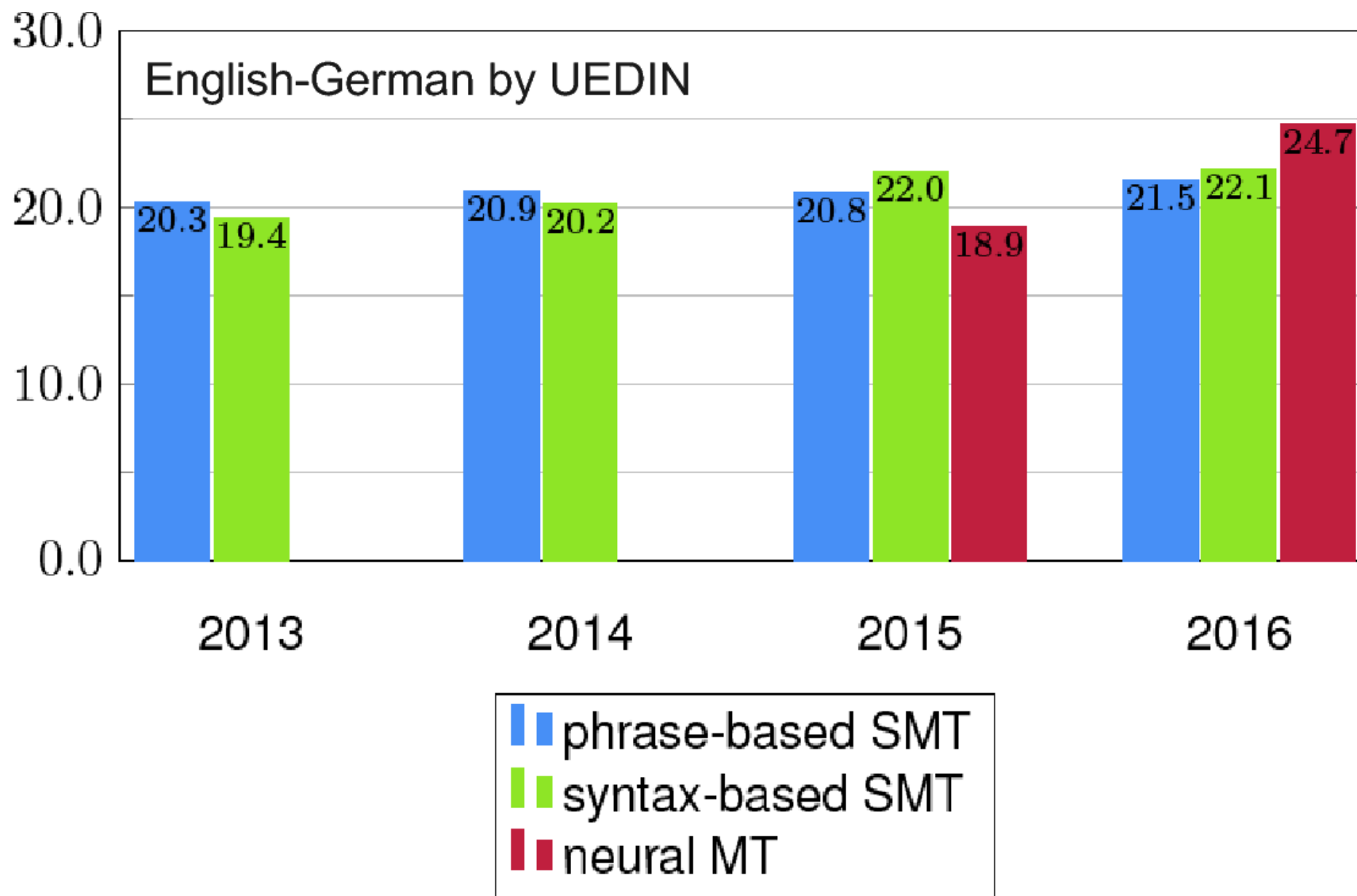
Univerzita Karlova v Praze

Evropská unie: 24 jazyků

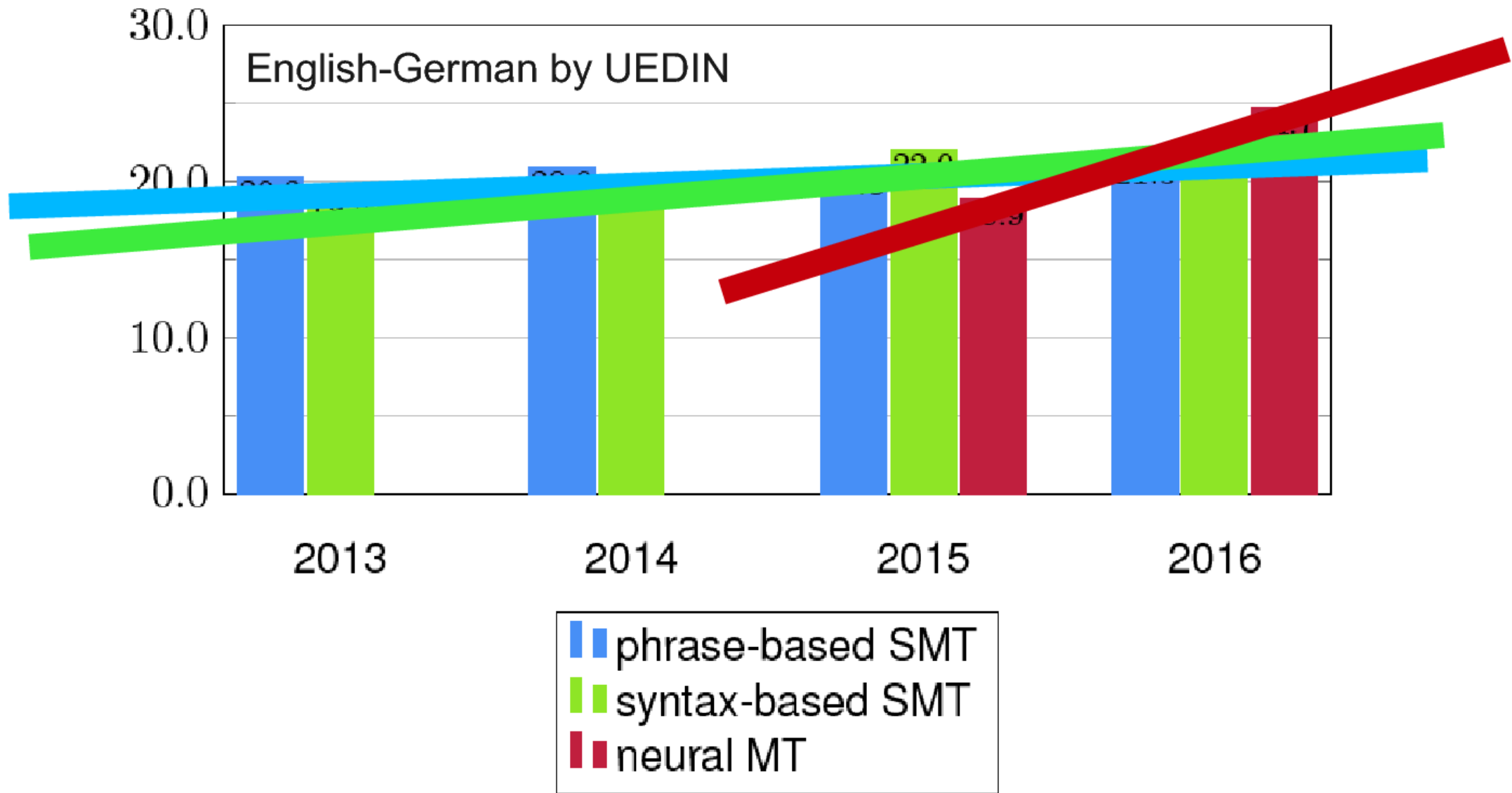


- DGT (Ředitelství překladů):
 - 1750 lingvistů, 600 asistentů
 - Miliony stran ročně: 0,9 (1992) 1,2 (2004) 1,8 (2012)
- Celá EU (2013):
 - ~6000 překladatelů, 1 miliarda EUR ročně.
- Série projektů: EuroMatrix, EuroMatrixPlus, MosesCore, QT21.
 - Od EuroMatrixPlus provozuje DGT vlastní překladový systém.

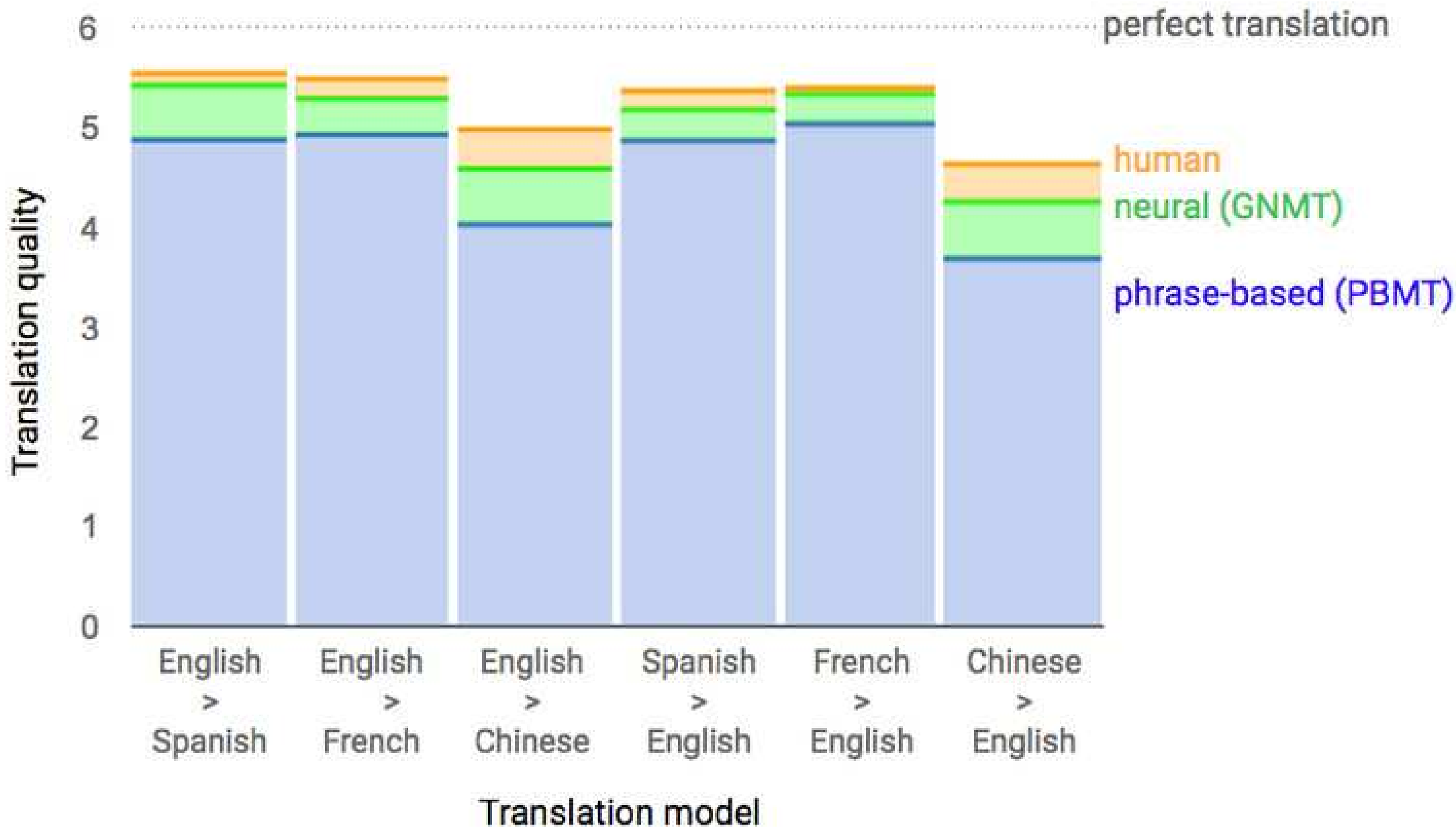
Nástup neuronového překladu



Nástup neuronového překladu



Reklama Google



English

Czech - detected



English

Czech

Translate

Máma mele maso?
Máma maso mele?
Mele máma maso?
Mele maso máma?
Maso mele máma?
Maso máma mele?



Mum is mincing meat?
Mommy meat?
My mom's meat?
My Flesh Mum?
My mom's meat?
My mom's meat?

Výstupy nejlepšího systému v letošní soutěži: <http://matrix.statmt.org/>

SRC A 28-year-old chef who had recently moved to San Francisco was found dead in the stairwell of a local mall this week.

Osmadvacetiletý kuchař, který se nedávno přestěhoval do San Francisca, byl tento týden nalezen mrtvý na schodišti místního obchodního centra.

Osmadvacetiletý šéfkuchař, který se nedávno přistěhoval do San Franciska, byl tento týden ∅ schodech místního obchodu.

SRC There were creative differences on the set and a disagreement.

Došlo ke vzniku kreativních rozdílů na scéně a k neshodám.

Na place byly tvůrčí rozdíly a neshody.

Výstupy nejlepšího systému v letošní soutěži: <http://matrix.statmt.org/>

SRC A 28-year-old chef who had recently moved to San Francisco was found dead in the stairwell of a local mall this week.

MT Osmadvacetiletý kuchař, který se nedávno přestěhoval do San Francisca, byl tento týden nalezen mrtvý na schodišti místního obchodního centra.

REF Osmadvacetiletý šéfkuchař, který se nedávno přistěhoval do San Franciska, byl tento týden ∅ schodech místního obchodu.

SRC There were creative differences on the set and a disagreement.

REF Došlo ke vzniku kreativních rozdílů na scéně a k neshodám.

MT Na place byly tvůrčí rozdíly a neshody.

Naštěstí ;-) stále dělá vážné chyby



SRC ... said Frank initially stayed in hostels...

MT ... řekl, že Frank původně zůstal v **Budějovicích**...

SRC Most of the Clintons' income...

MT Většinu příjmů **Kliniky**...

SRC The 63-year-old has now been made a special representative

MT 63letý **mladík** se nyní stal zvláštním zástupcem...

SRC He listened to the moving stories of the women.

MT Naslouchal **pohyblivým** příběhům žen.

...vlastně katastrofické chyby (2/2)



SRC Criminal Minds star Thomas Gibson sacked after hitting producer

REF Thomas Gibson, hvězda seriálu Myšlenky zločince, byl propuštěn po té, co uhodil režiséra

MT **Kriminalisté Minsku** hvězdu Thomase Gibsona **vyhostili** po **zásahu** producenta

SRC ...add to that its long-standing grudge...

REF ...přidejte k tomu svou dlouholetou nenávist...

MT ...přidejte k tomu svou dlouholetou **záštitu**...
(grudge → zášť → záštita)

Selhávají ale i lidé



Selhávají ale i lidé



Google Translate



English Spanish French Welsh - detected



English Hebrew Czech

Translate

Nid wyf yn y swyddfa ar hyn o bryd.
Anfonfwch unrhyw waith i'w gyfieithu.

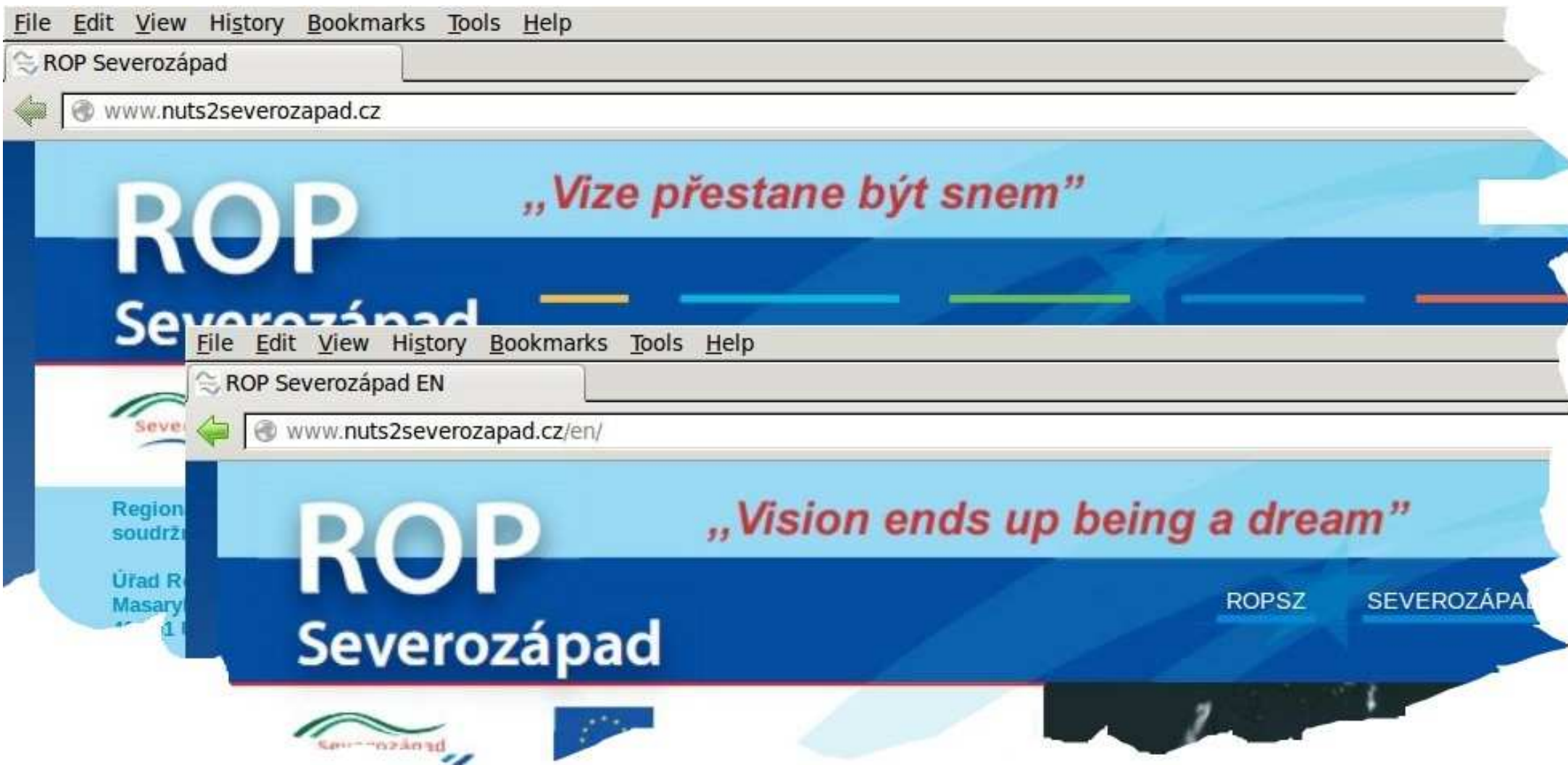


73/5000

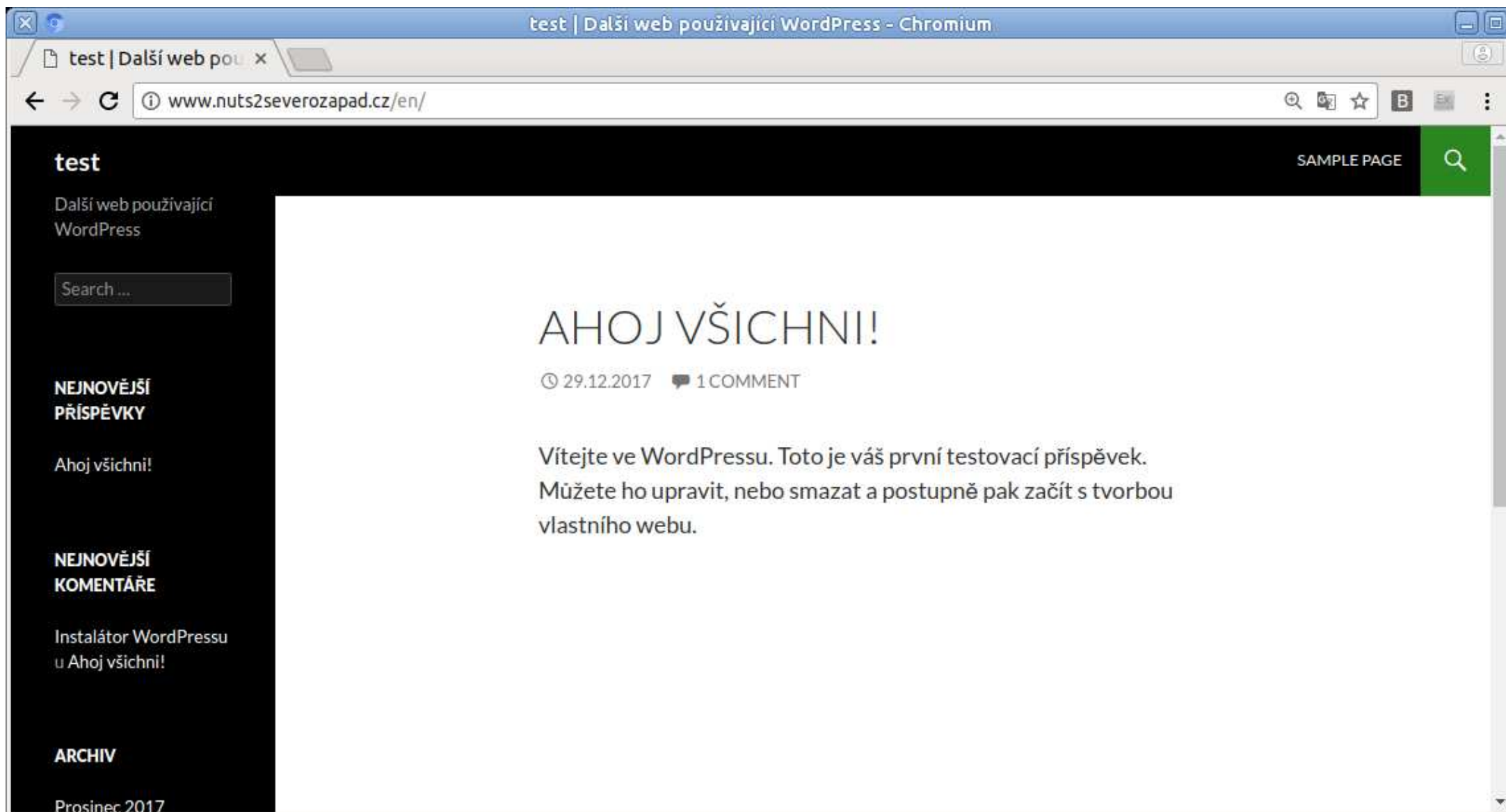
V současné době nejsem v kanceláři. Zašlete prosím jakoukoli práci, kterou chcete přeložit.



Selhávají ale i lidé



... nuts2severozapad.cz ... Vize skončí jako sen.



test | Další web používající WordPress - Chromium

test | Další web pou x

www.nuts2severozapad.cz/en/

test

Další web používající WordPress

Search ...

NEJNOVĚJŠÍ PŘÍSPĚVKY

Ahoj všichni!

NEJNOVĚJŠÍ KOMENTÁŘE

Instalátor WordPressu u Ahoj všichni!

ARCHIV

Prosinec 2017

SAMPLE PAGE

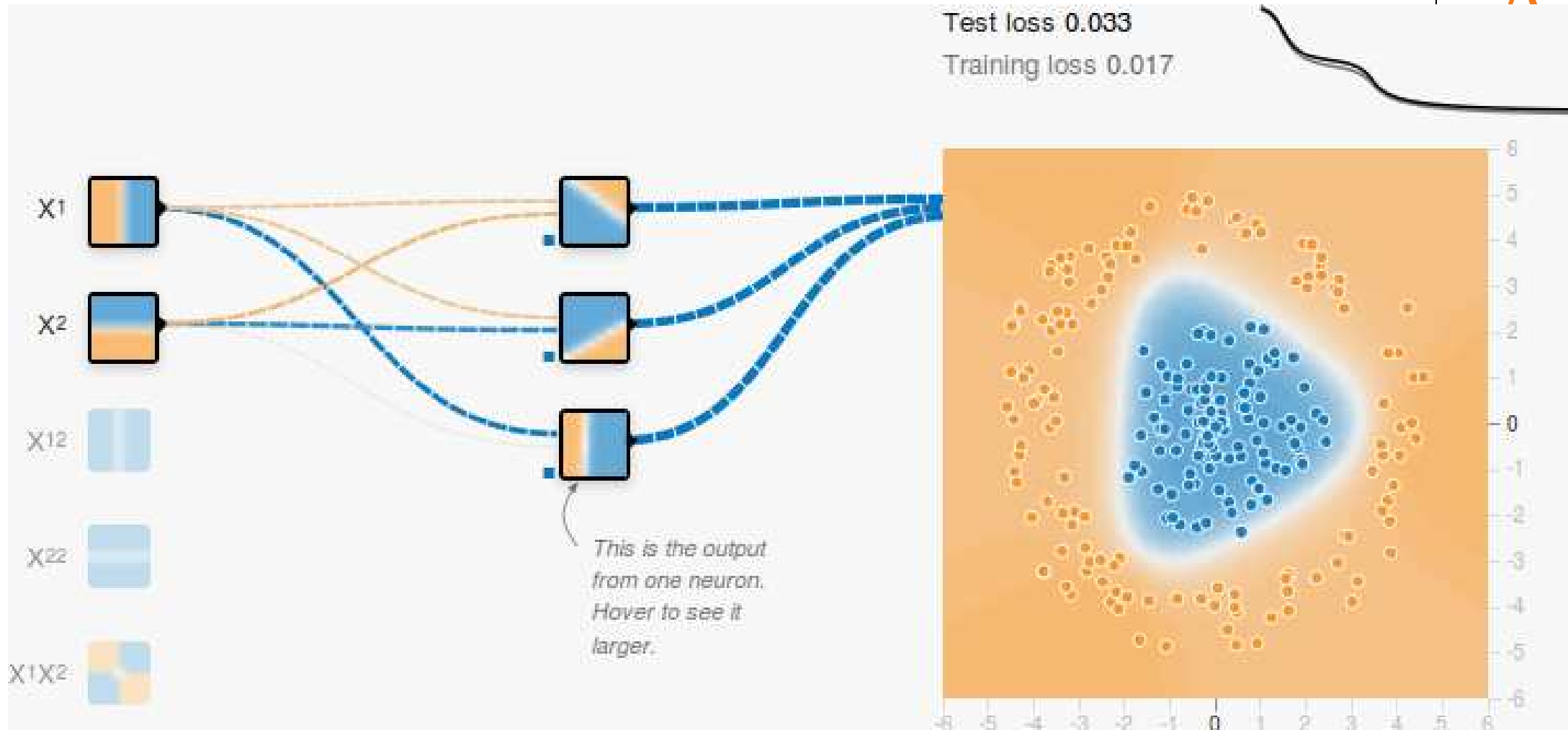
AHOJ VŠICHNI!

29.12.2017 1 COMMENT

Vítejte ve WordPressu. Toto je váš první testovací příspěvek. Můžete ho upravit, nebo smazat a postupně pak začít s tvorbou vlastního webu.

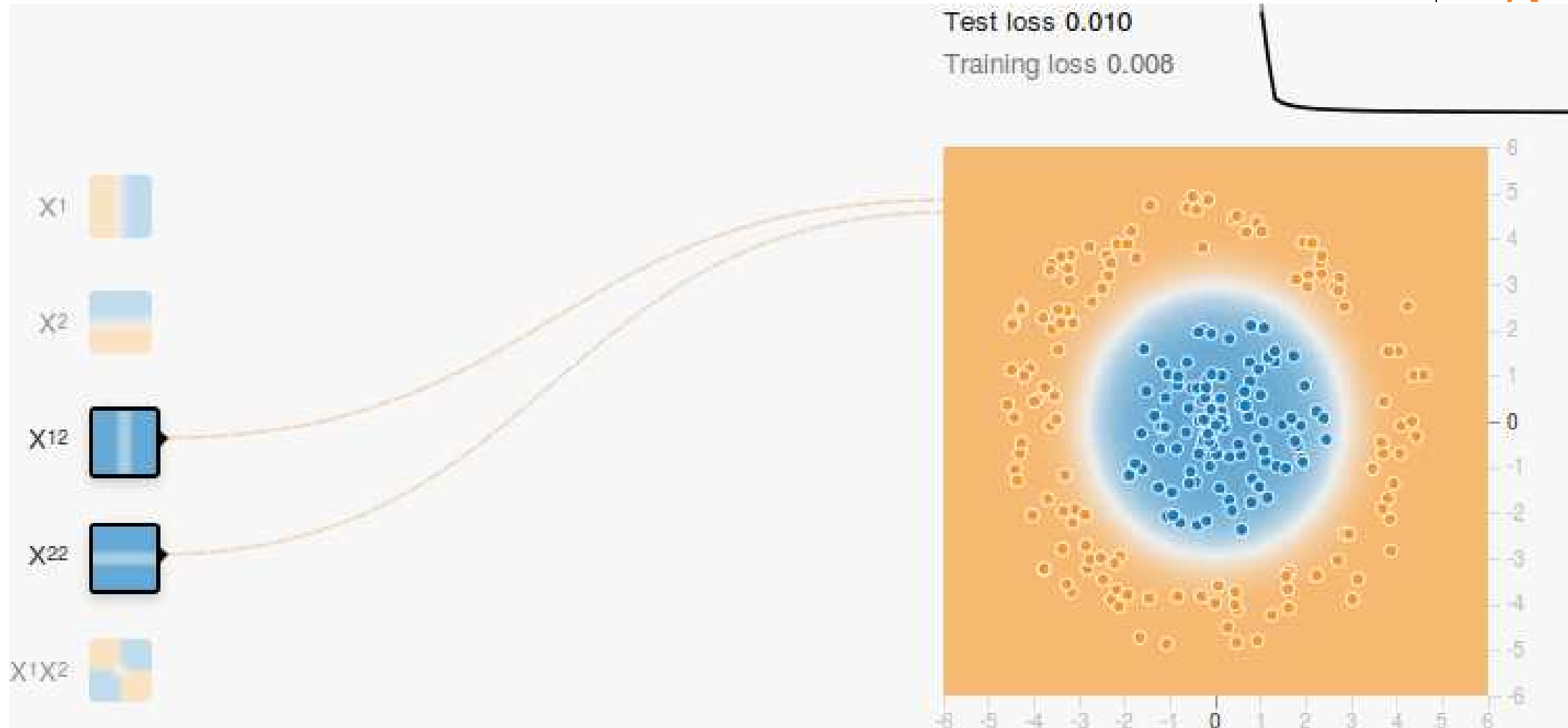
- Neuronové sítě a hluboké učení.
- Neuronový ~~neuronální~~ strojový překlad.
- Učení se reprezentaci, strojové porozumění.

Hluboké učení na 4 slidy



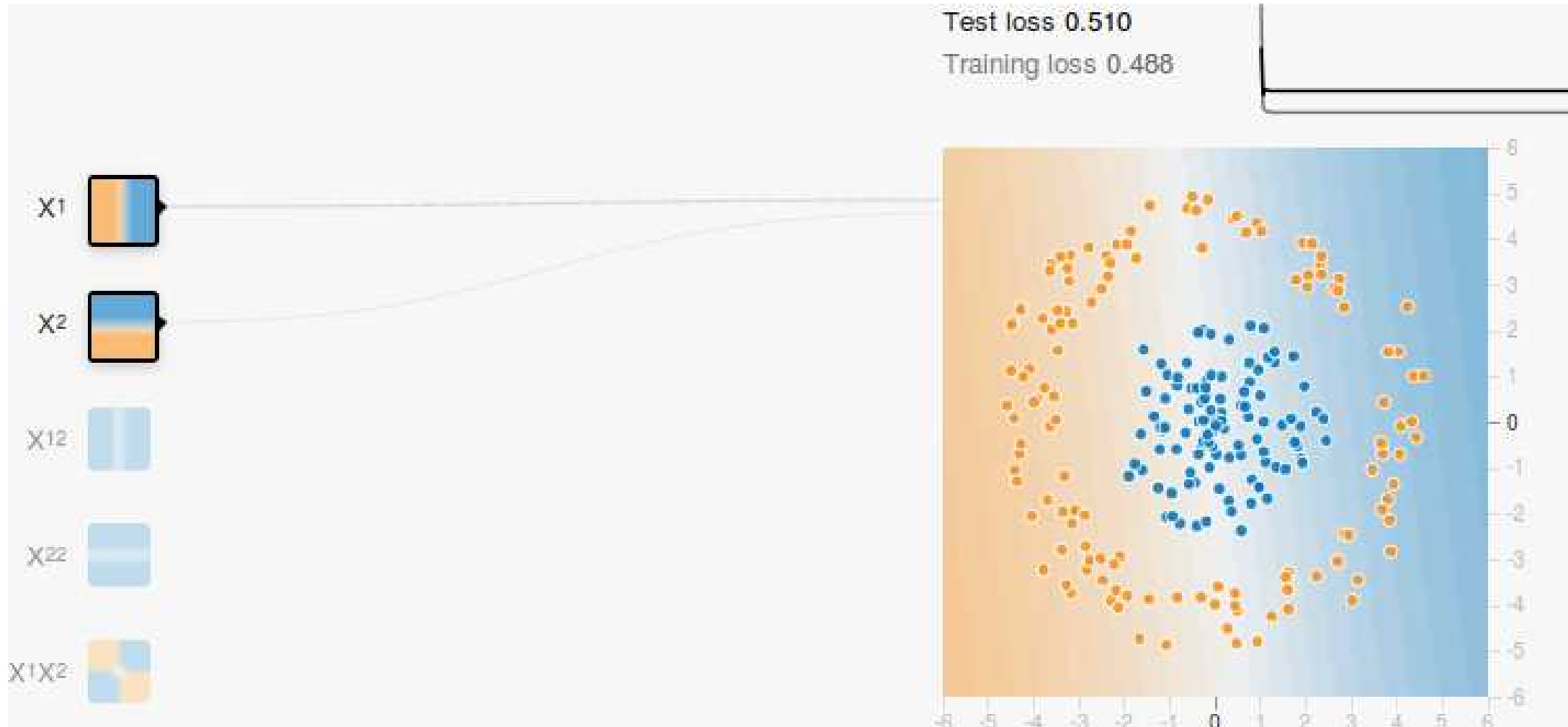
Parametry se natrénují: $1 \tanh(-0.43x_1 - 0.89x_2 + 2.0)$
 $+1 \tanh(-0.67x_1 + 0.89x_2 + 2.1)$
 $+1 \tanh(1.4x_1 - 0.067x_2 + 2.3) - \pi/2 > 0$

Ideální vstupy

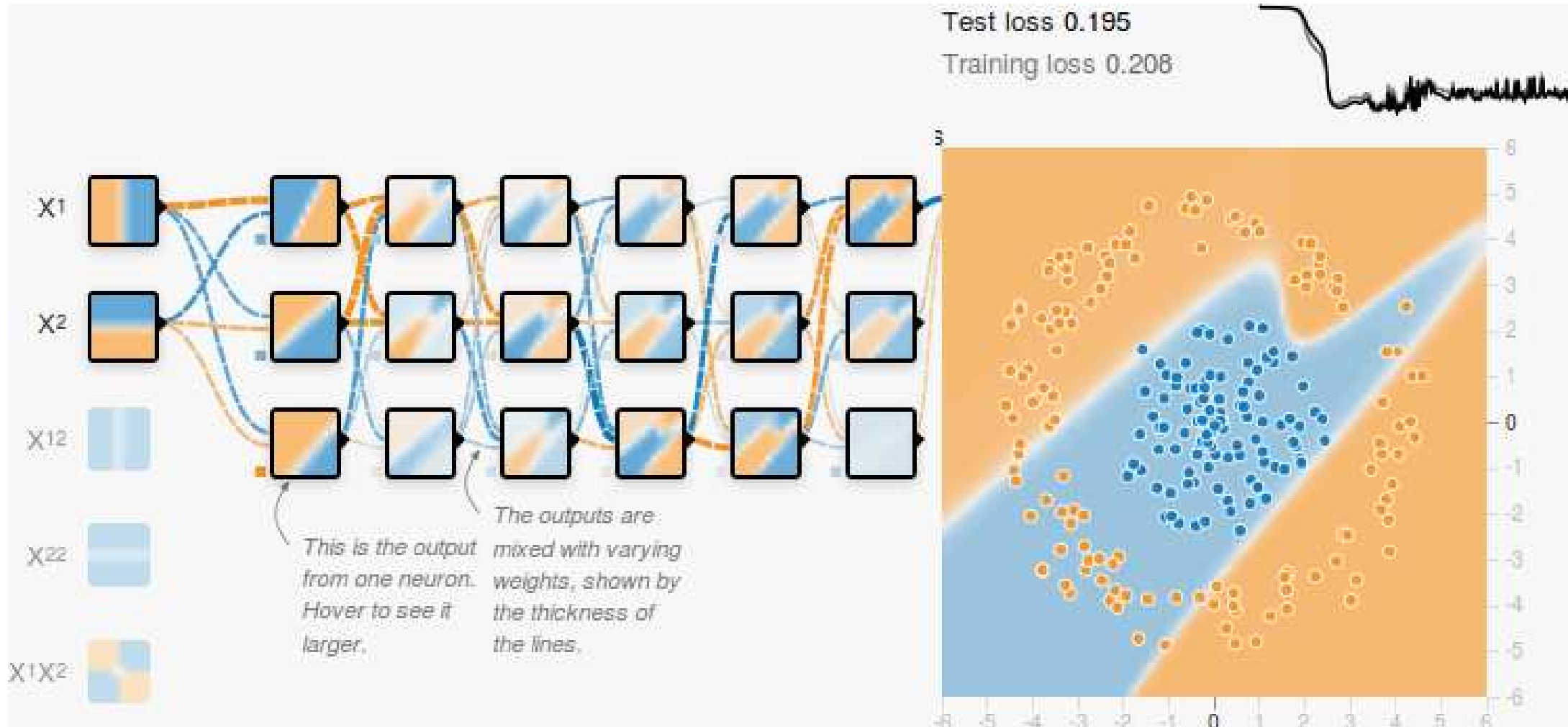


$$1x_1^2 + 1x_2^2 - 1 < 0$$

Nevhodné vstupy a malá hloubka

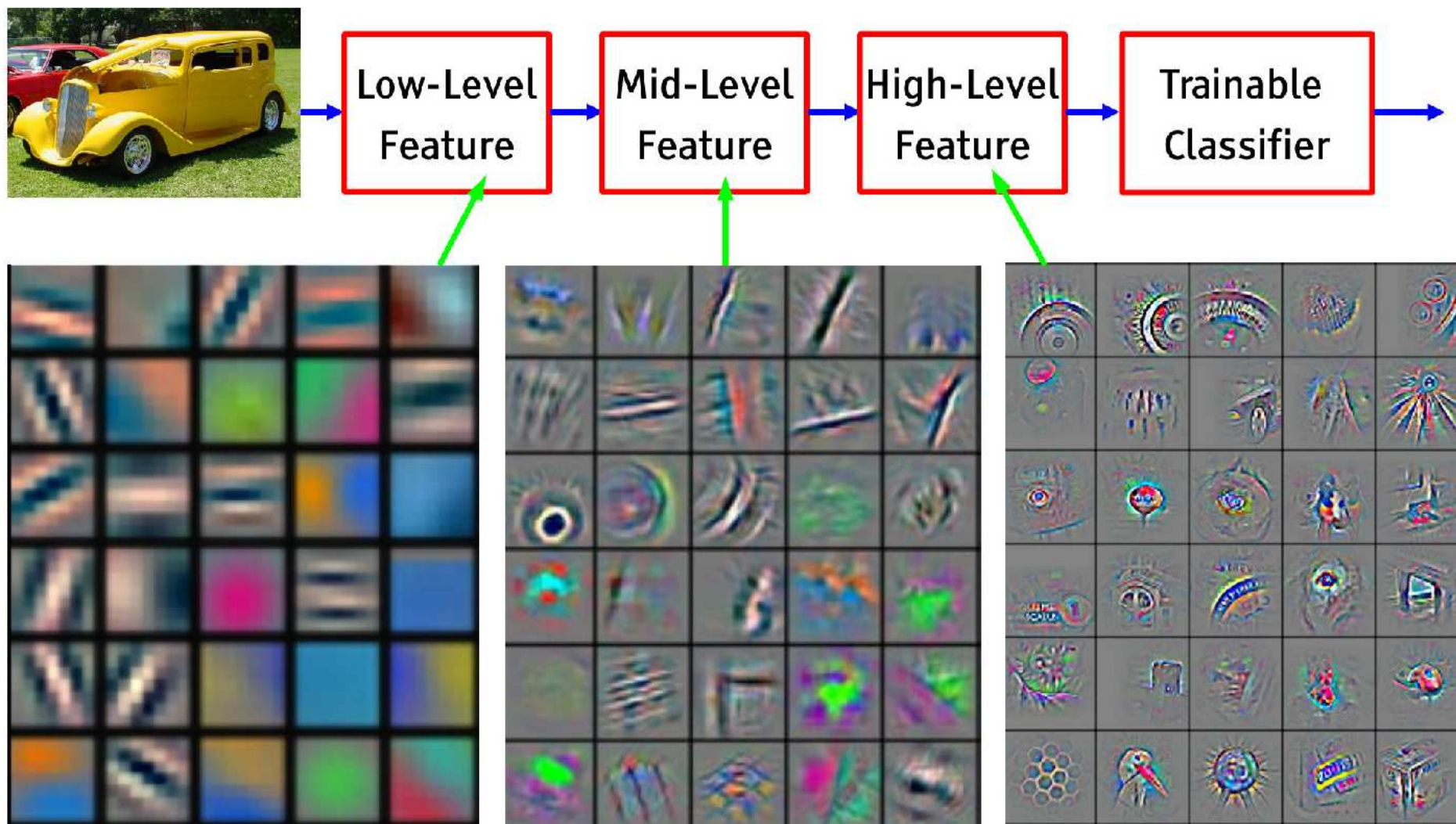


Příliš složitá síť se nenatrénuje



Hluboké NN pro klasifikaci obrázků

■ It's deep if it has more than one stage of non-linear feature transformation



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Hluboké sítě se učí reprezentací

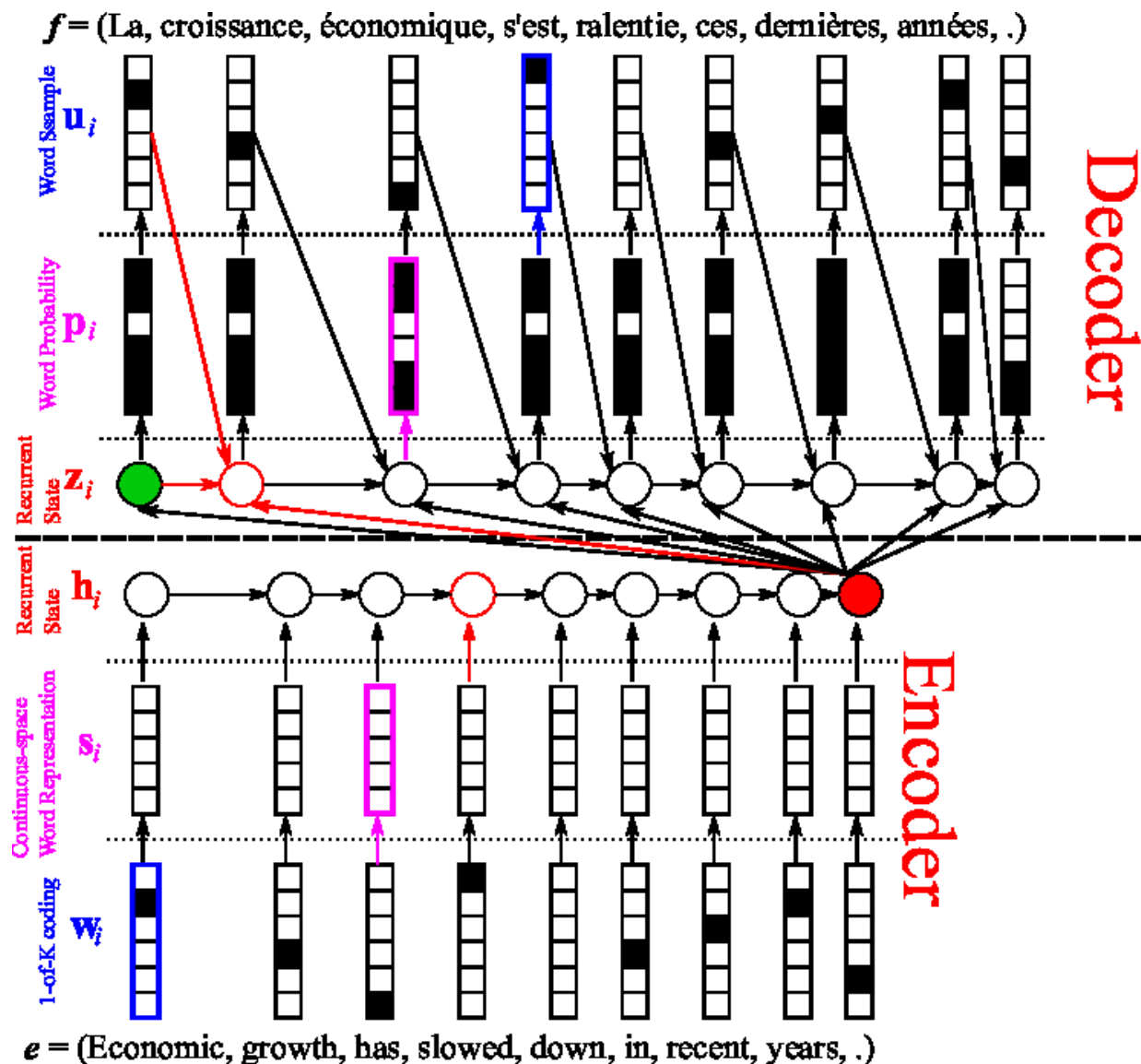


- Na základě trénovacích dat (ukázkové vstupy a očekávané výstupy)
- se neuronová síť (neural network, NN) sama naučí
- čeho si ve vstupech všímat.

“**Reprezentace**” je nový souřadný systém.

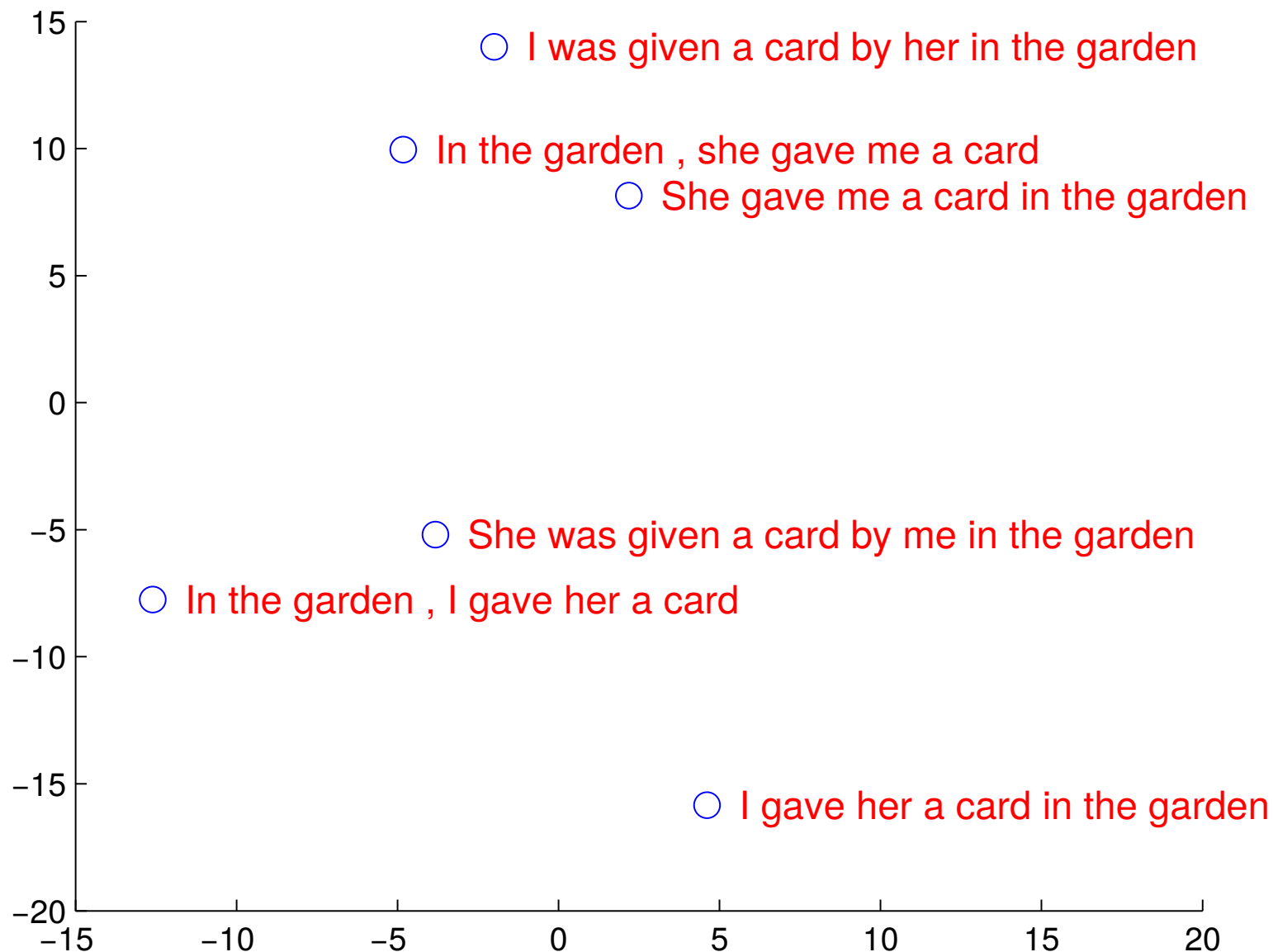
- Místo 3 rozměrů (x, y , barva) dostaneme
- 2000 rozměrů: (slonovitost, počet čápů, modrost, . . .)
- nalezených tak, aby nejvíc pomáhaly uhodnout výstup.

NMT: Enkodér-Dekodér



<https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/>

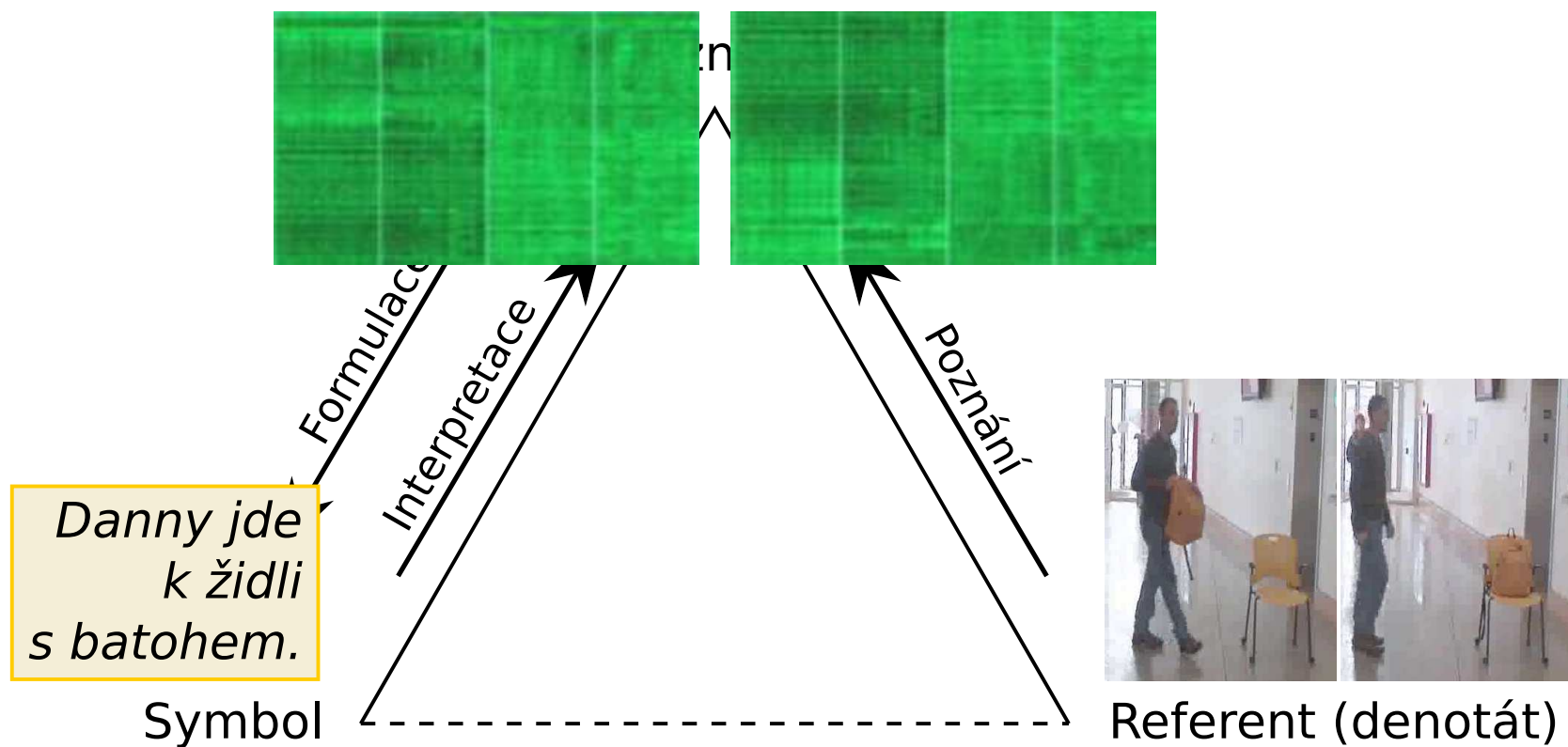
Vektorová reprezentace vět



8000-rozměrný prostor reprezentací vět promítnut do 2D (Sutskever et al., 2014).

Rozumí už počítače?

- Hluboké učení se učí reprezentace.
- Není to vlastně přesně totéž, co dělají lidé?
- Nevede hluboké učení k opravdickému porozumění?



Vstup:

legendární slovenská punkrocková kapela extip se letos vrátila na pódia poté, co vyšla v reedici její debutová deska pekný, škaredý deň, kterou přehraje 1. prosince na sedmičce na strahově. soubor nezanikl, i když bratislavskou punkovou scénu v devadesátých letech rozložily drogy. své zkušenosti s tím má kytarista sveto korbel, který odpovídal na otázky noviněk.

Lidský výstup:

slovenská punková legenda extip se vrátila

Vstup:

legendární slovenská punkrocková kapela extip se letos vrátila na pódia poté, co vyšla v reedici její debutová deska pekný, škaredý deň, kterou přehraje 1. prosince na sedmičce na strahově. soubor nezanikl, i když bratislavskou punkovou scénu v devadesátých letech rozložily drogy. své zkušenosti s tím má kytarista sveto korbel, který odpovídal na otázky noviněk.

Lidský výstup:

slovenská punková legenda extip se vrátila

“Sumarizace překladem:”

slovenská kapela extip se vrací do prahy

Automatická sumarizace chápe?



Vstup:

legendární slovenská punkrocková kapela extip se letos vrátila na pódia poté, co vyšla v reedici její debutová deska pekný, škaredý deň, kterou přehraje 1. prosince na sedmičce na **strahově**. soubor nezanikl, i když bratislavskou punkovou scénu v devadesátých letech rozložily drogy. své zkušenosti s tím má kytarista sveto korbel, který odpovídal na otázky noviněk.

Lidský výstup:

slovenská punková legenda extip se vrátila

“Sumarizace překladem:”

slovenská kapela extip se vrací do **prahy**

Automatická sumarizace nechápe.



na strahově	slovenská kapela extip se vrací do prahy
v o2 aréně	slovenská kapela extip se vrací do prahy
na hradecku	slovenská kapela extip se vrací do čech
u vajgaru	slovenská kapela extip se vrací do prahy

Automatická sumarizace nechápe!



na strahově
v o2 aréně
na hradecku
u vajgaru
ve stromovce

slovenská kapela extip se vrací do **prahy**
slovenská kapela extip se vrací do **prahy**
slovenská kapela extip se vrací do **čech**
slovenská kapela extip se vrací do **prahy**

slovenská kapela extip se vrací na scénu.
tentokrát kvůli drogám v reedici. s. s. m. m.
m. m. m. m. m. m. m. m. m. m. m. m.
m. m. i. m. m. . . . m.
m. m. m. m. m. m. m. m. m. m. m. m.
m. m. m. m. m. m. . m. m. m. m. m. m.
m. m. m. m.

- Co čekat od neuronového překladu?
 - Mimořádný skok v plynulosti výstupu.
 - Velký skok i v adekvátnosti překladu.
 - Často nelze odlišit lidský a strojový překlad.
- Kdy strojový překlad nahradí překladatele?
 - V řadě oblastí se to stalo už dávno.
 - Neuronový překlad vyřešil morfologii.
 - K **opravdickému porozumění** je ještě daleko.
... ale jsme na dobré cestě.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. pages 3104–3112.