Deep learning summer camp
6 Frontiers in deep learning

Ole Winther

Dept for Applied Mathematics and Computer Science
Technical University of Denmark (DTU)

July 6, 2016

Objectives of talk

- Continuing frontiers in deep learning
- Reinforcement learning - AlphaGo
- A few final words

Reinforcement learning

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

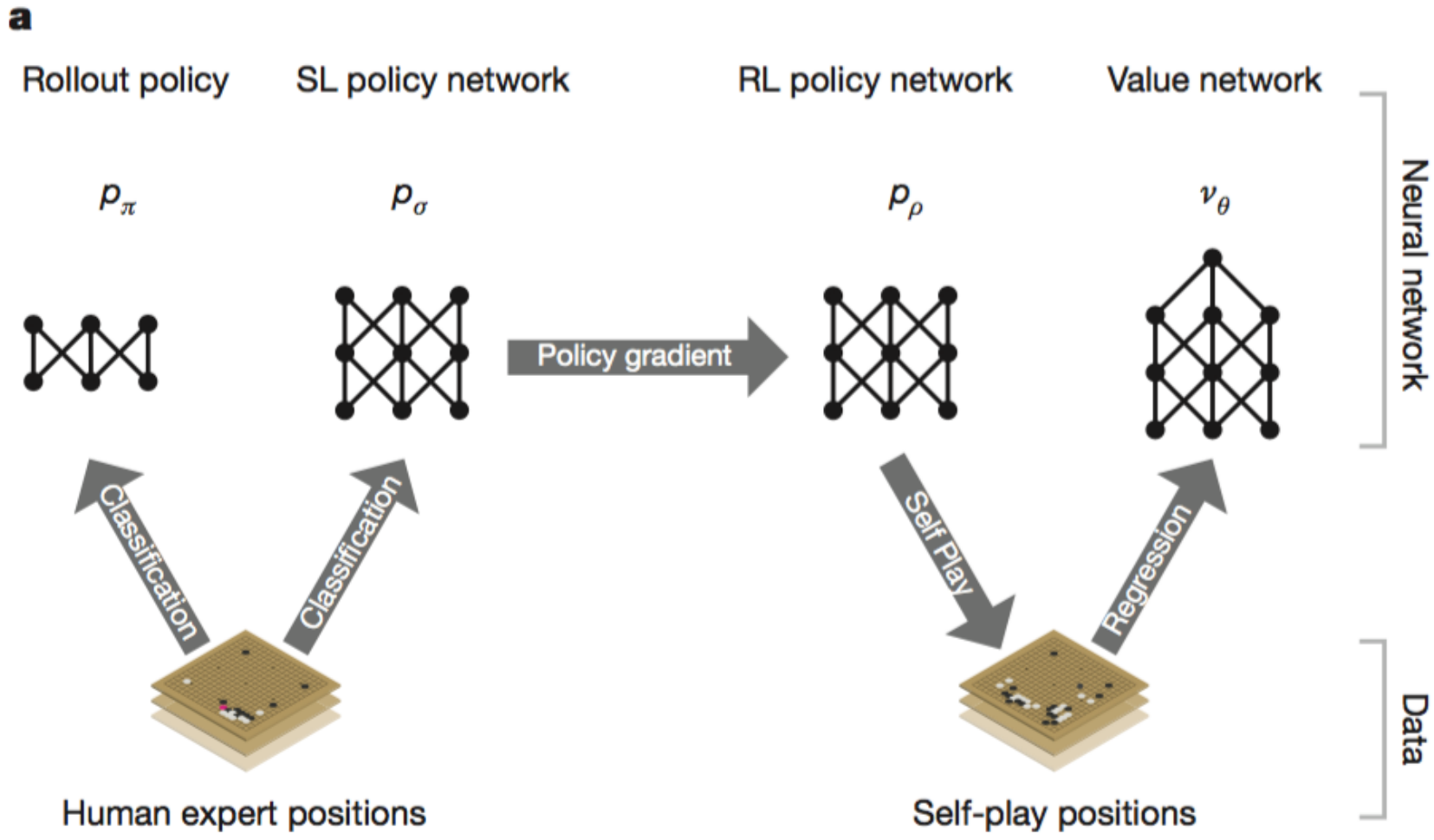
David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

Value function $v^*(s)$

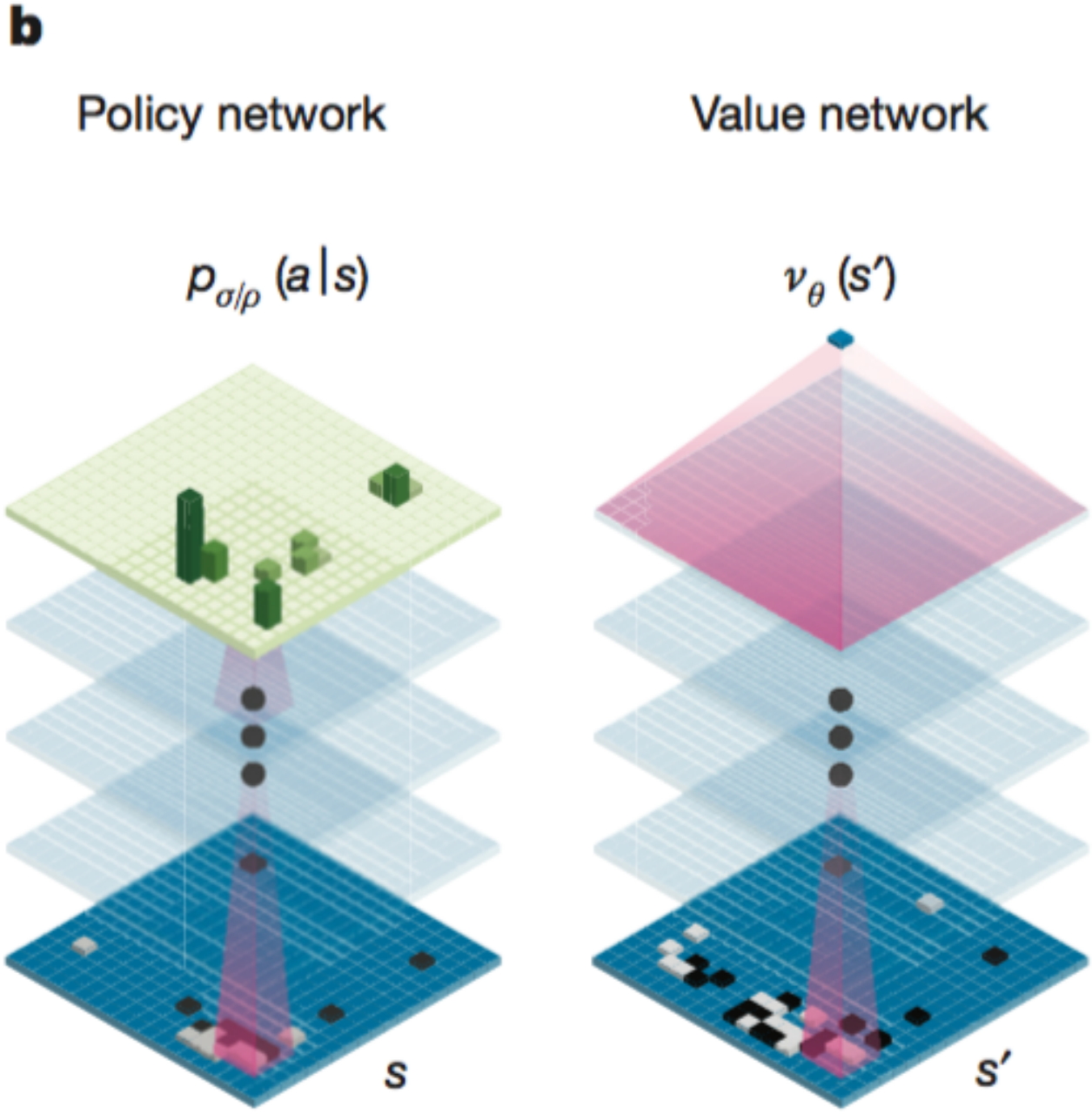
Value function $v^*(s)$

- s = state = game position
- Game breadth: b
- Game depth. d
- Complexity: b^d for calculating $v^*(s)$
- Chess: $b \approx 35$, $d \approx 80$
- Go: $b \approx 250$, $d \approx 150$

Rollout policy, SL and RL policy and value networks



Policy and value networks



Step 1: supervised learning (SL) policy $p_\sigma(a|s)$

- a = action
- s = state
- We have large database of expert games.
- Train classifier to imitate expert moves (s, a):

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial\sigma}$$

- Supervised learning (SL) policy

Step 2: reinforcement learning (RL) policy $p_\rho(a|s)$

- The policy network $p_\rho(a|s)$ plays against (a younger version of itself).
- Record whether it wins/losses: $z_t = r(T) = +1/-1$
- Train classifier to imitate expert moves (s, a):

$$\Delta\rho \propto \frac{\partial \log p_\rho(a_t|s_t)}{\partial \rho} z_t$$

- Better than SL policy.

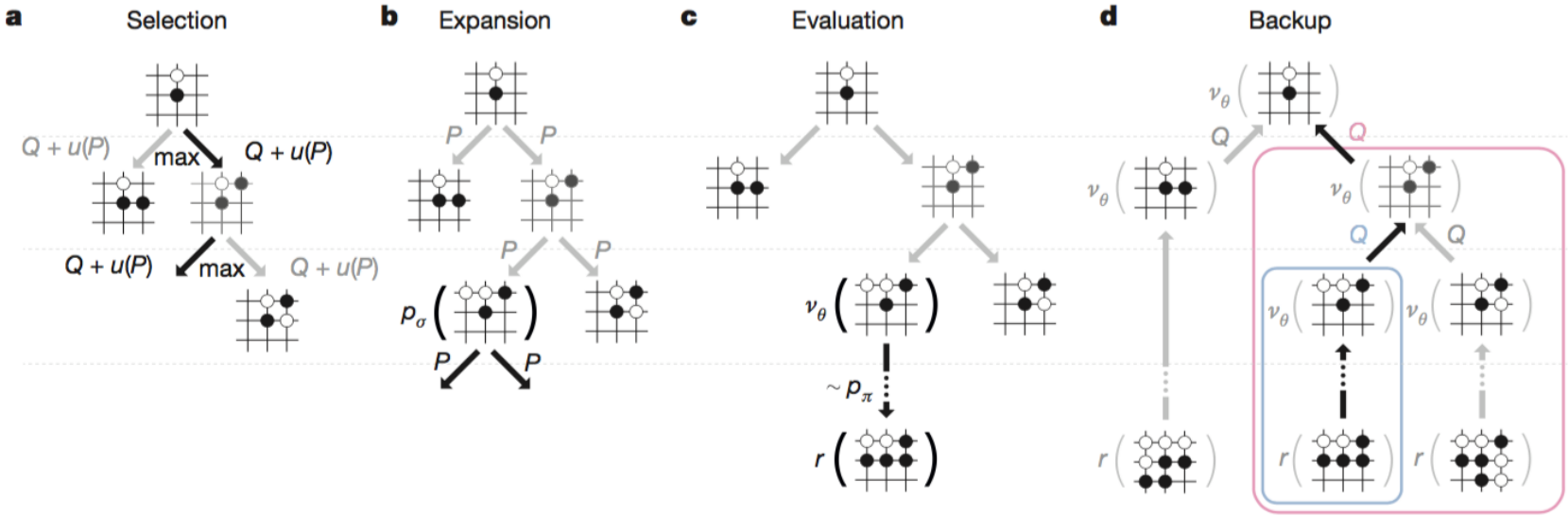
Step 3: RL value function $\nu_\theta(s)$

- Remember that value function $v^*(s)$ is unknown.
- We can try to learn the value function of our RL policy network by a network: $\nu_\theta(s)$:

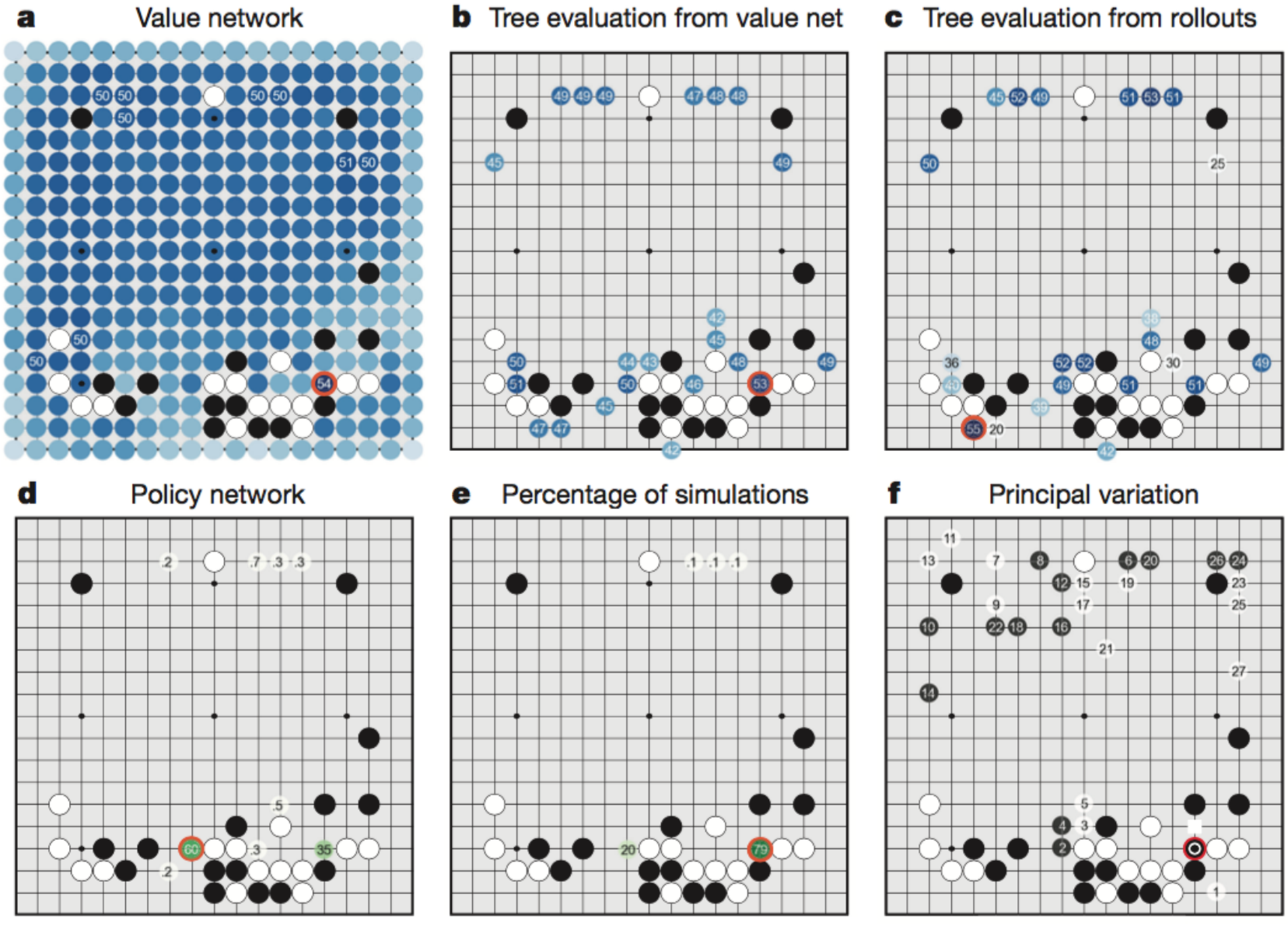
$$\Delta\theta \propto \frac{\partial \log \nu_\theta(s)}{\partial\theta}(z - \nu_\theta(s))$$

- We can use this as an ingredient in a Monte Carlo tree search

Step 4: Monte Carlo tree search



An example


a Value network
b Tree evaluation from value net
c Tree evaluation from rollouts
d Policy network
e Percentage of simulations
f Principal variation

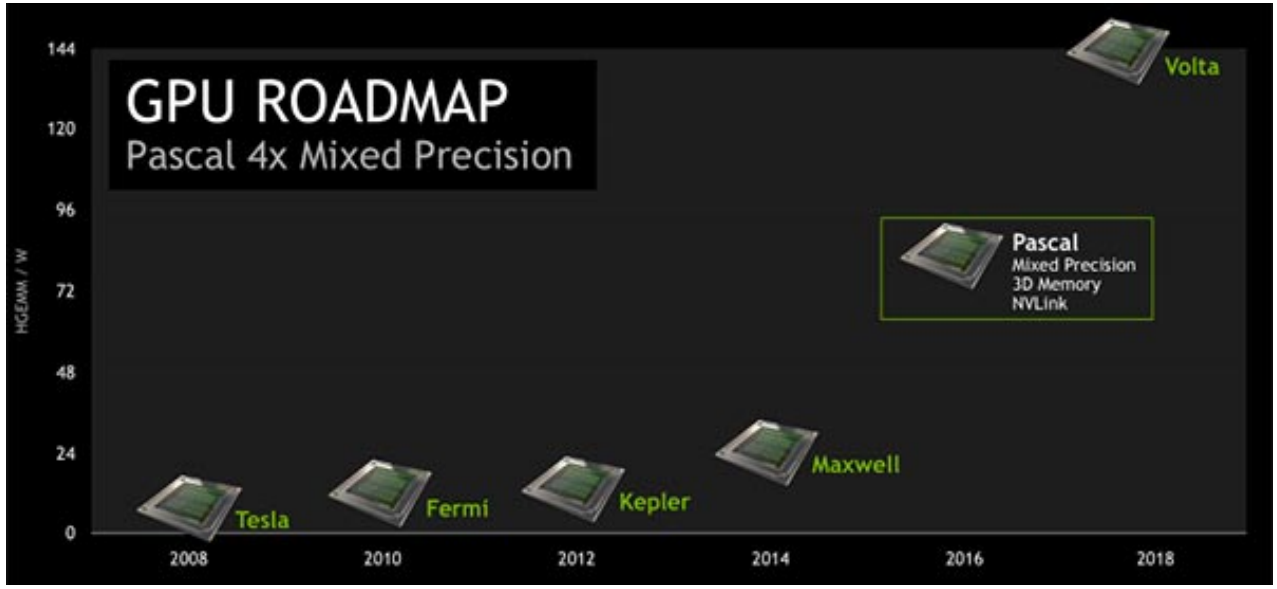
Does it scale?

Joaquin Quiñonero Candela, Director of Applied Machine Learning at Facebook on Quora:

- In computer vision we have a system that processes every single image and video. . . well over 1B items per day.
- . . . predict the content. . . . . to generate captions for the blind,. . . detect and take down offensive content, improve media search results, automate visual captcha. . . .
- In language technology,. . . translate over 2B posts every single day, with over 1800 language directions representing more than 40 unique languages.
- . . . these models are used to rank news feed stories (1B users every day, 1.5K stories per user per day on average), ads, search results (1B+ queries a day), trending news, friend recommendations and even rank notifications that a user receives, or rank the comments on a post.

Deep learning prediction about the future

- Expect rapid progress in coming years!
- Still hard to make it work!
- You are here to learn the trade!



Thanks!
Ole Winther