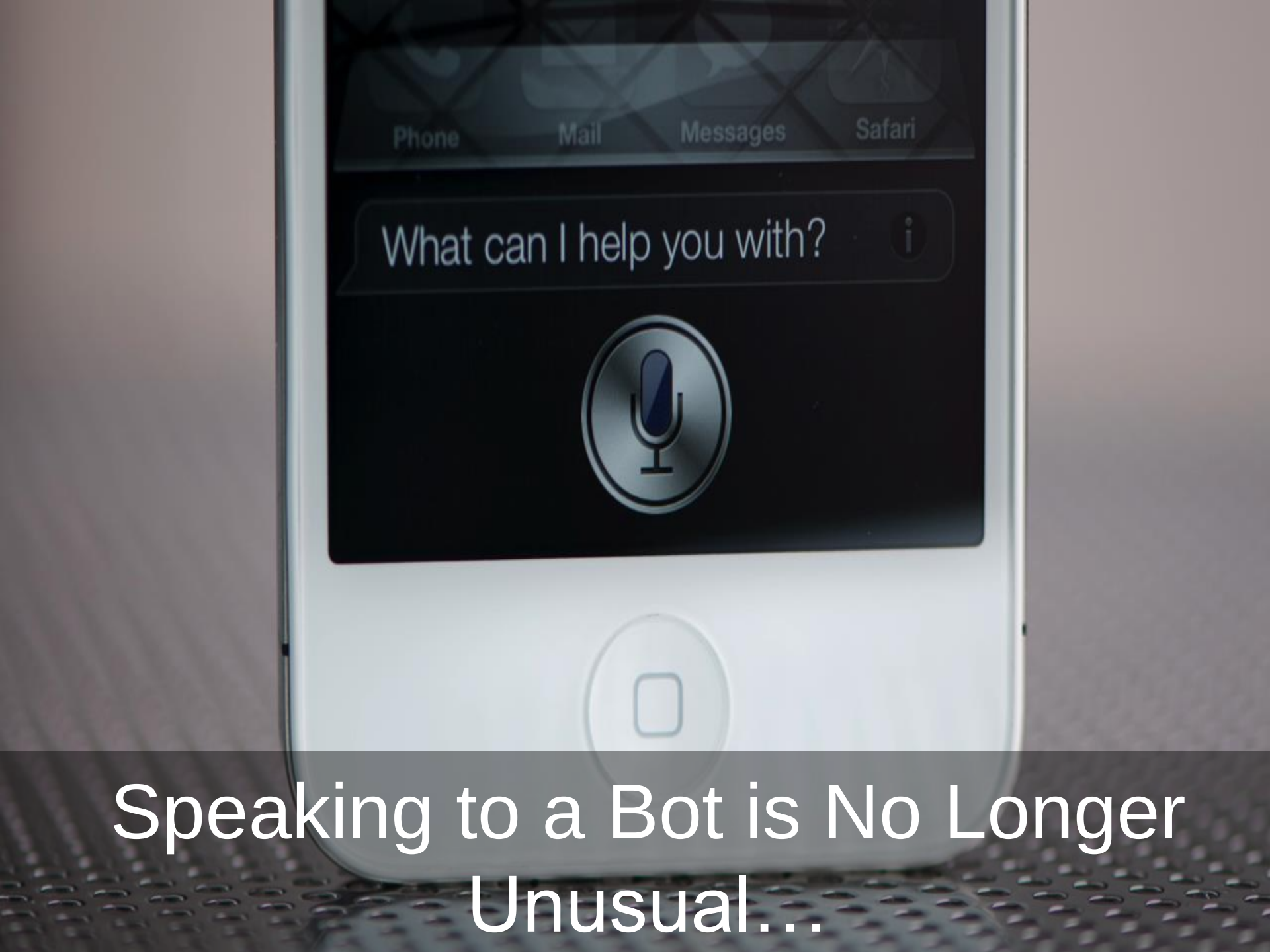
Artificial Intelligence and Deep Learning

MILA

Université de Montréal

Cars are now driving themselves…

(far from perfectly, though)



Speaking to a Bot is No Longer Unusual…


March 2016:
World Go Champion
Beaten by Machine

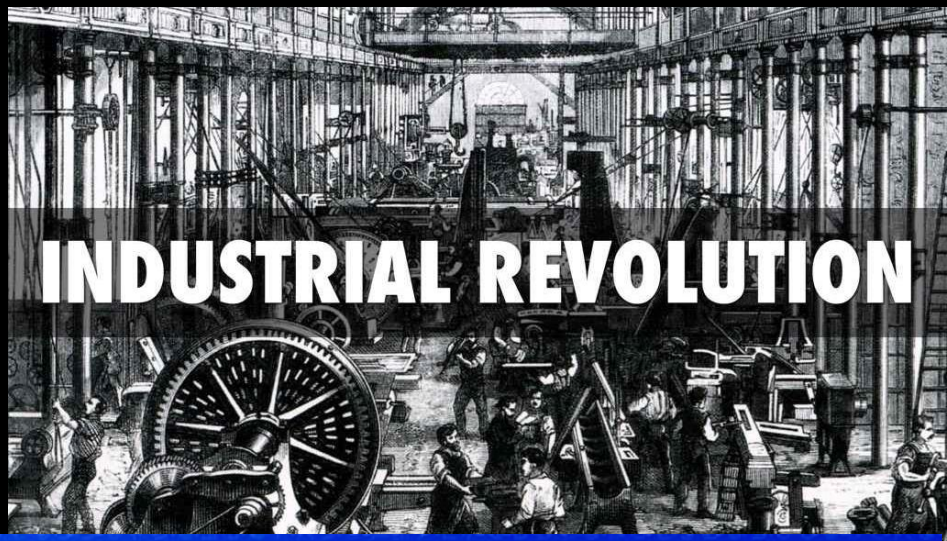
AI: The Upcoming Industrial Revolution

First industrial revolution:

- Machines extending humans' **mechanical power**

Upcoming industrial revolution:

- Machines extending humans' **cognitive power**
  - From the digital economy to the AI economy
  - Predicted growth at least 25%/yr
  - All sectors of the economy



A new revolution seems to be in the work after the industrial revolution.

Devices are becoming intelligent.

And Deep Learning is at the epicenter of this revolution.

Breakthrough in deep learning

A Canadian-led trio at CIFAR initiated the deep learning AI revolution

- Fundamental breakthrough in 2006:

first successful recipe for training a deep supervised neural network

- Second major advance in 2011, with rectifiers
- Breakthroughs in applications since then

AI Needs Knowledge

- Failure of classical AI: a lot of knowledge is not formalized, expressed with words
- Solution: computer gets knowledge from data, learns from examples

MACHINE LEARNING

Machine Learning, AI & No Free Lunch

- Five key ingredients for ML towards AI
  1. Lots & lots of data
  2. Very flexible models
  3. Enough computing power
  4. **Powerful priors that can defeat the curse of dimensionality**
  5. Computationally efficient inference

Bypassing the curse of dimensionality

We need to build compositionality into our ML models

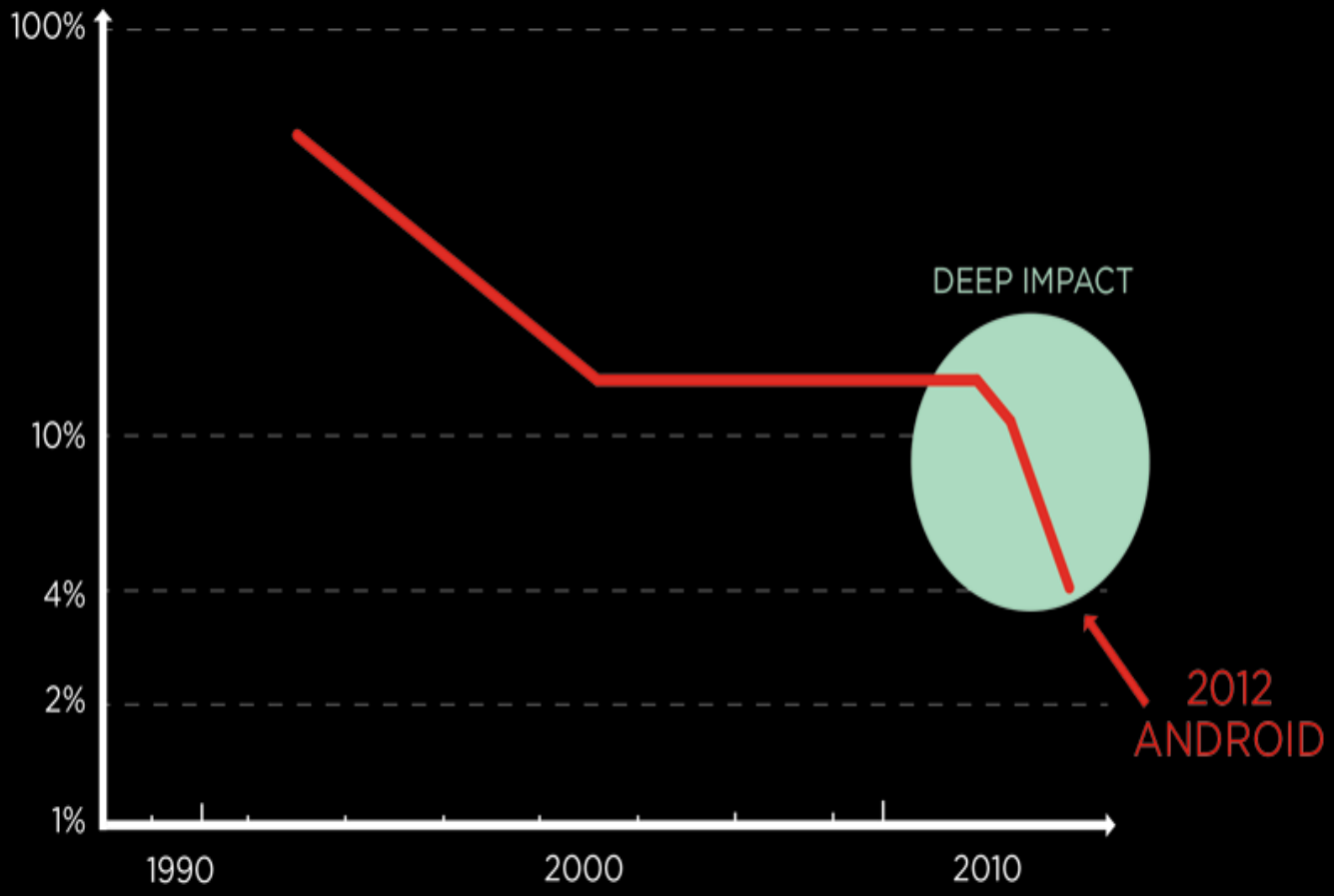
- Just as human languages exploit compositionality to give representations and meanings to complex ideas

Exploiting compositionality gives an exponential gain in representational power

- Distributed representations / embeddings: feature learning
- Deep architecture: multiple levels of feature learning

Prior assumption: compositionality is useful to describe the world around us efficiently

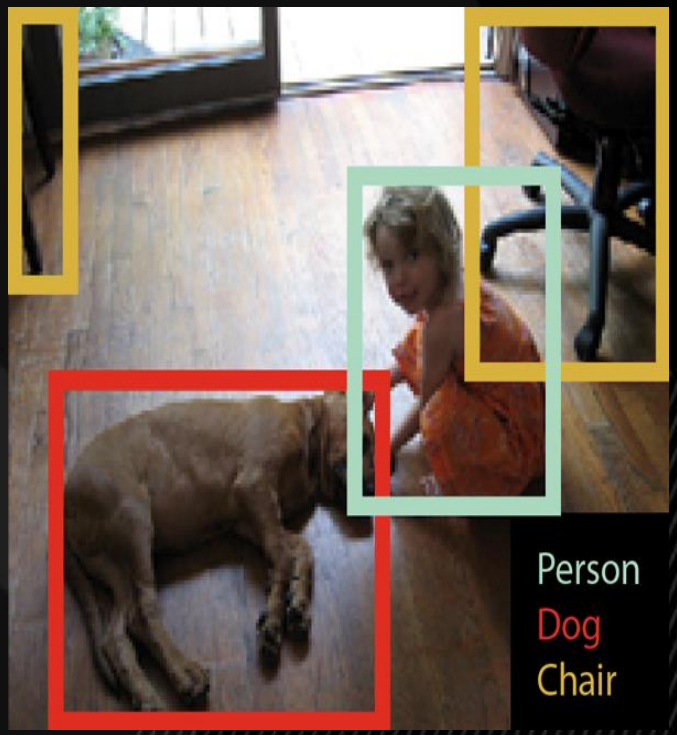
2010-2012: breakthrough in speech recognition



Source: Microsoft

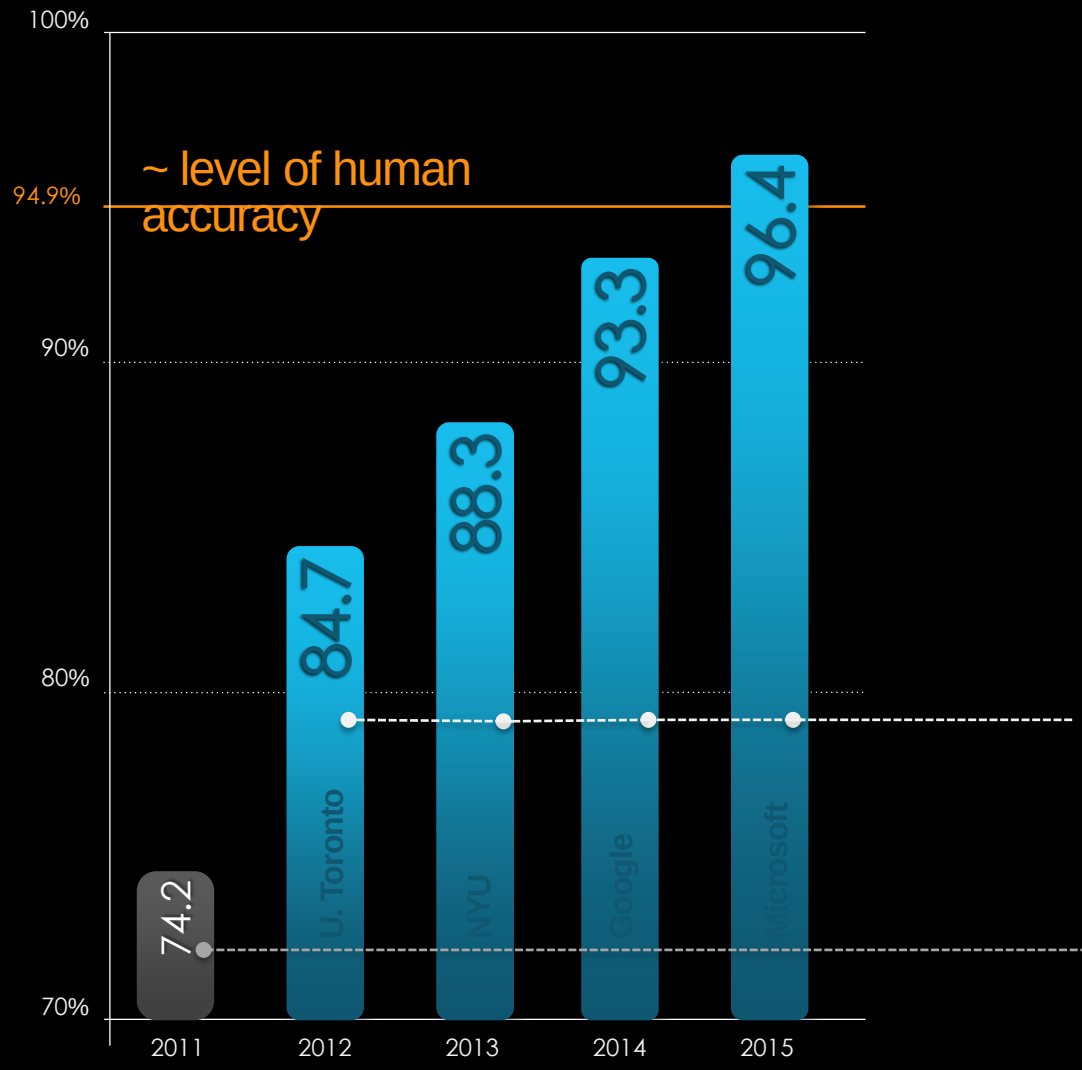
2012-2015: breakthrough in computer vision

- Graphics Processing Units (GPUs) + 10x more data
- 1,000 object categories,
- Facebook: millions of faces
- ***2015: human-level performance***



ImageNet Accuracy Still Improving

Top-5 Classification task



Use of

Deep Learning

over

Conventional Computer Vision

IT companies are racing into deep learning

From computer vision to self-driving cars: 2016

Holmdel, New Jersey
February 2016

Ongoing progress: combining vision and natural language understanding

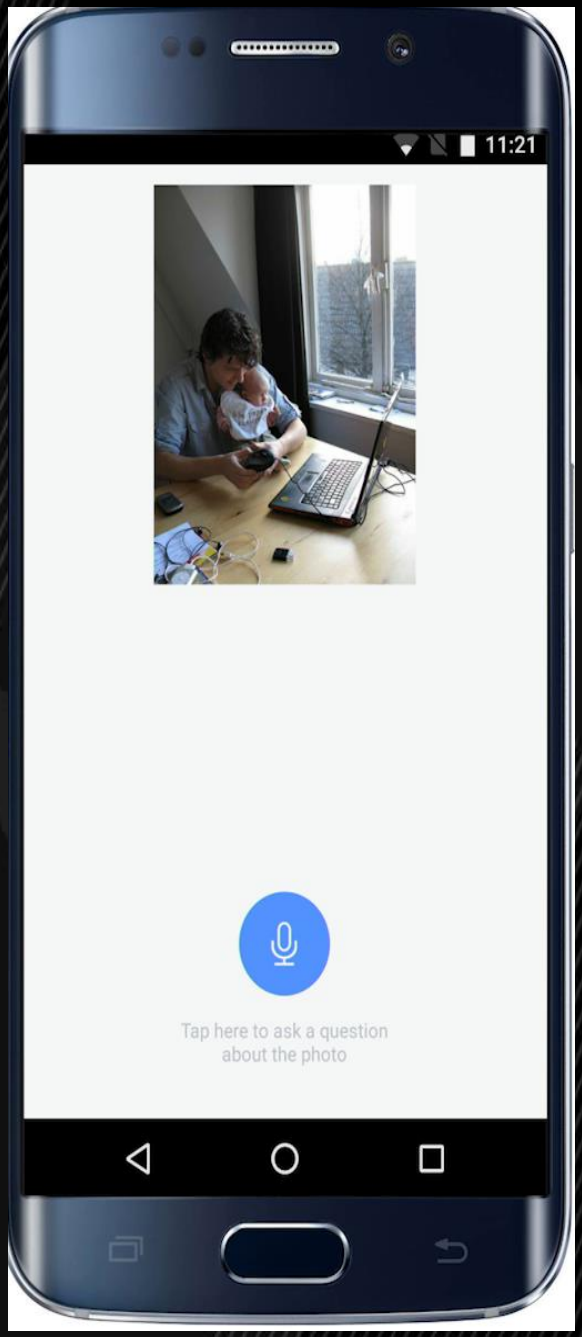
A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor

A stop sign is on a road with a mountain in the background

With a lot more data…

visual question answering



Deep Learning: Beyond Pattern Recognition, towards AI

- Many researchers believed that neural nets could at best be good at pattern recognition
- And they are really good at it!

- But many more ingredients needed towards AI. Recent progress:

  - REASONING: with extensions of recurrent neural networks
    - Memory networks & Neural Turing Machine

  - PLANNING & REINFORCEMENT LEARNING: DeepMind (Atari and Go game playing) & Berkeley (Robotic control)

The next frontier:
to reason and answer questions

Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.

Q: Where is the apple?
A: Bedroom

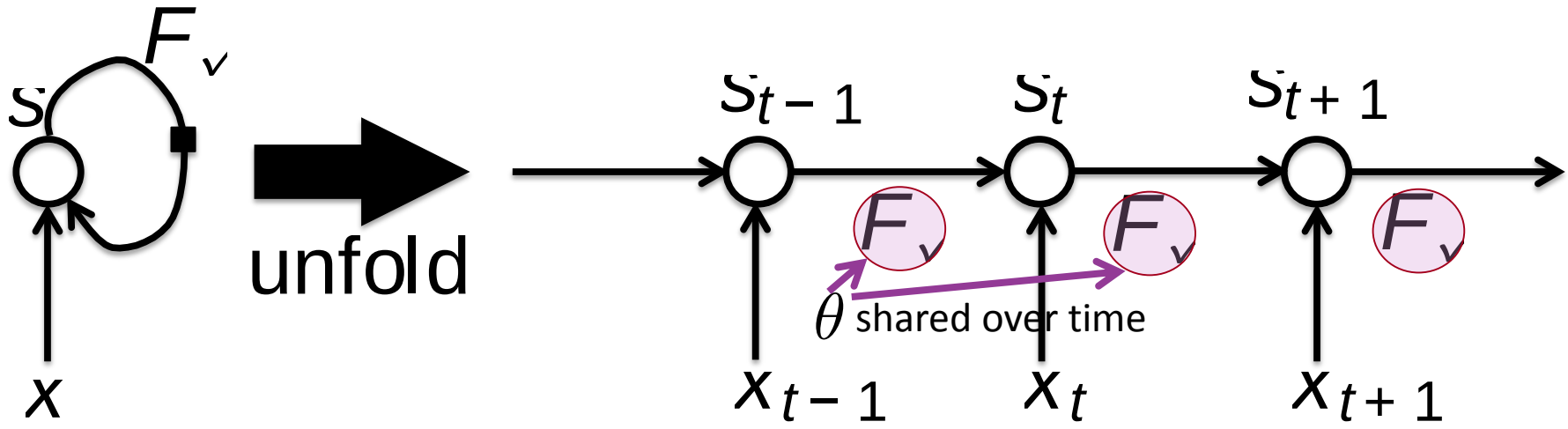
Brian is a lion.
Julius is a lion.
Julius is white
Bernhard is green

Q: What colour is Brian?
A: White

Recurrent Neural Networks

- Selectively summarize an input sequence in a fixed-size state vector via a recursive update

$$s_t = F_\theta(s_{t-1}, x_t)$$



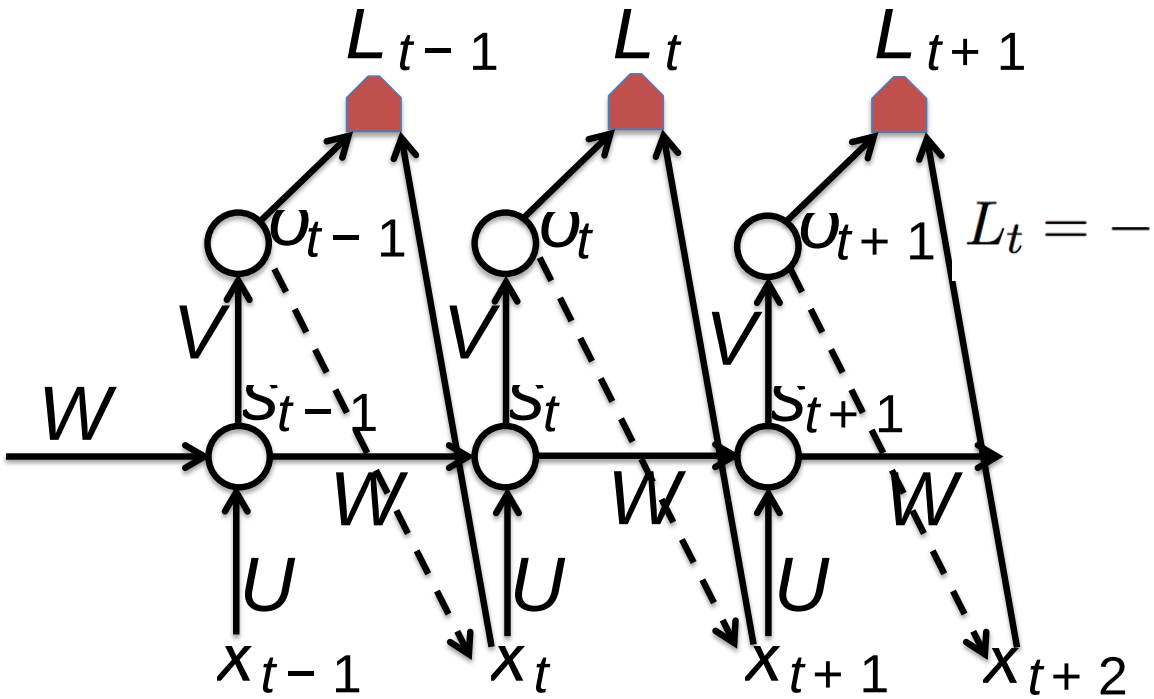
$$s_t = G_t(x_t, x_{t-1}, x_{t-2}, \ldots, x_2, x_1)$$

➔ **Generalizes naturally to new lengths not seen during training**

Generative RNNs

- An RNN can represent a fully-connected **directed generative model**: every variable predicted from all previous ones.

$$P(\mathbf{x}) = P(x_1, \ldots x_T) = \prod_{t=1}^{T} P(x_t | x_{t-1}, x_{t-2}, \ldots x_1)$$



$$L_t = -\log P(x_t | x_{t-1}, x_{t-2}, \ldots x_1)$$

End-to-End Machine Translation with Recurrent Nets and Attention Mechanism

(Bahdanau et al ICLR 2015, Jean et al ACL 2015, Gulcehre et al 2015, Firat et al 2016)

- Reached the state-of-the-art in one year, from scratch

(a) **English→French (WMT-14)**

	NMT(A)	Google	P-SMT
NMT	32.68	30.6*	**37.03**$^\bullet$
+Cand	33.28	–	
+UNK	33.99	32.7°	
+Ens	**36.71**	**36.9**$^\circ$	

(b) **English→German (WMT-15)**

Model	Note
24.8	Neural MT
24.0	U.Edinburgh, Syntactic SMT
23.6	LIMSI/KIT
22.8	U.Edinburgh, Phrase SMT
22.7	KIT, Phrase SMT

(c) **English→Czech (WMT-15)**

Model	Note
18.3	Neural MT
18.2	JHU, SMT+LM+OSM+Sparse
17.6	CU, Phrase SMT
17.4	U.Edinburgh, Phrase SMT
16.1	U.Edinburgh, Syntactic SMT

Google-Scale NMT Success

(Wu et al & Dean, Nature, 2016)

- After beating the classical phrase-based MT on the academic benchmarks, there remained the question: will it work on the very large scale datasets like used for Google Translate?
- Distributed training, very large model ensemble
- Not only does it work in terms of BLEU but it makes a killing in terms of human evaluation on Google Translate data

Table 10: Side-by-side scores on production data

	PBMT	GNMT	Human	Relative Improvement
English → Spanish	3.594±1.58	5.031±1.09	5.140±1.04	93%
English → French	3.518±1.70	5.032±1.22	5.215±1.03	89%
English → Portuguese	3.675±1.64	4.856±1.29	4.973±1.17	91%
English → Chinese	2.457±1.48	4.154±1.42	4.580±1.26	80%
Spanish → English	3.410±1.65	4.921±1.16	4.930±1.12	99%
French → English	3.639±1.63	5.000±1.07	5.016±1.09	99%
Portuguese → English	3.471±1.74	5.029±1.05	5.040±1.03	99%
Chinese → English	1.994±1.47	3.884±1.37	4.334±1.20	81%

Applications on the horizon

Computer Interaction

Healthcare

Robotics

MILA: Institut de Montréal des Algorithmes d'Apprentissage

MILA Faculty


Yoshua Bengio
Director


Aaron Courville


Pascal Vincent


Roland Memisevic


Christopher Pal


Laurent Charlin


Simon Lacoste-Julien


Doina Precup


Joelle Pineau