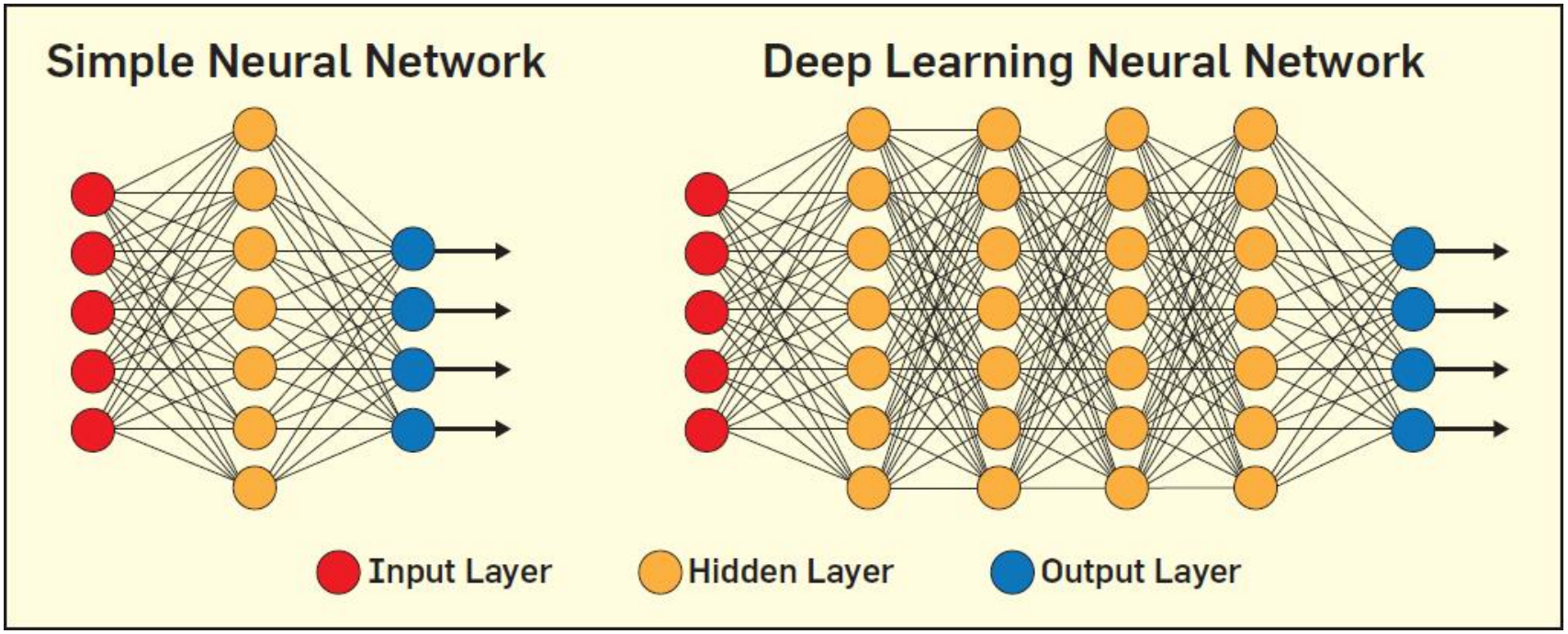
How AI is Screwed up by (using only) Deep Learning

From an academic perspective

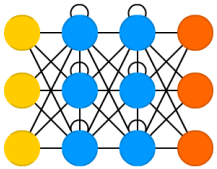
Piyawat Lertvittayakumjorn

2nd January 2019

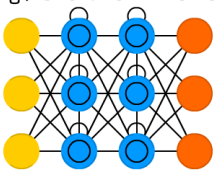
What is Deep Learning (DL) ?



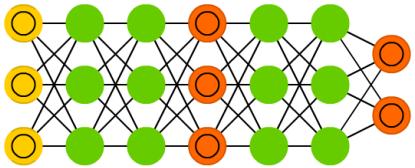
Recurrent Neural Network (RNN)



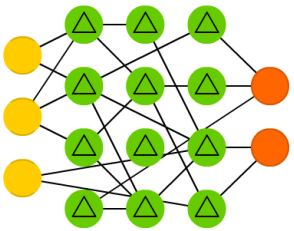
Long / Short Term Memory (LSTM)



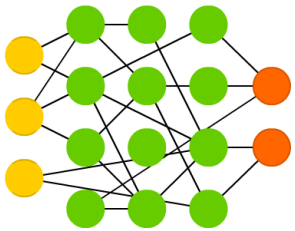
Generative Adversarial Network (GAN)



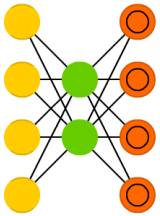
Liquid State Machine (LSM)



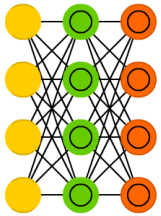
Extreme Learning Machine (ELM)



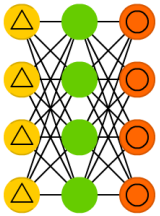
Auto Encoder (AE)



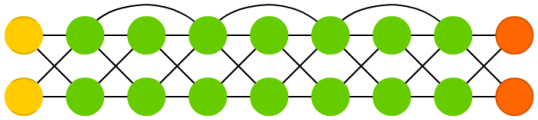
Variational AE (VAE)



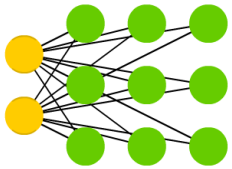
Denoising AE (DAE)



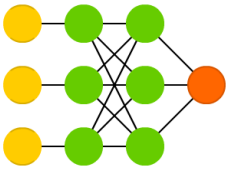
Deep Residual Network (DRN)



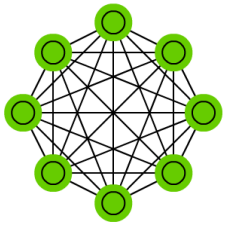
Kohonen Network (KN)



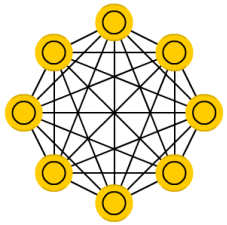
Support Vector Machine (SVM)



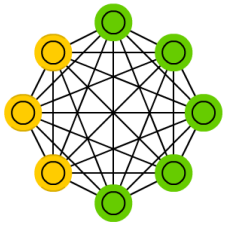
Markov Chain (MC)



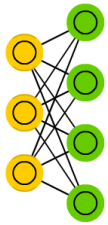
Hopfield Network (HN)



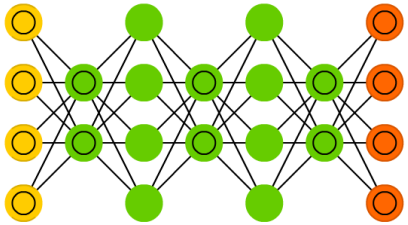
Boltzmann Machine (BM)



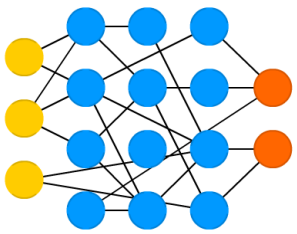
Restricted BM (RBM)



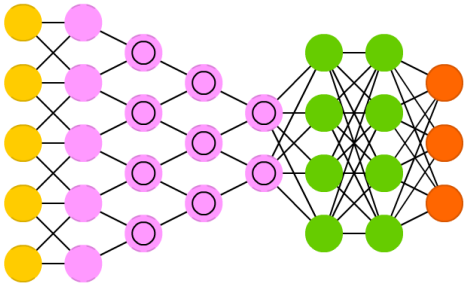
Deep Belief Network (DBN)



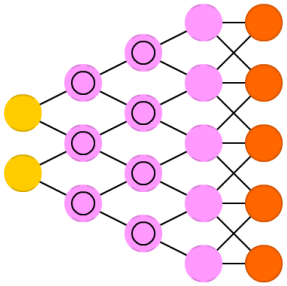
Echo State Network (ESN)



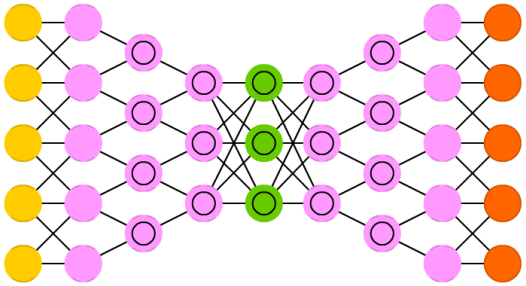
Deep Convolutional Network (DCN)



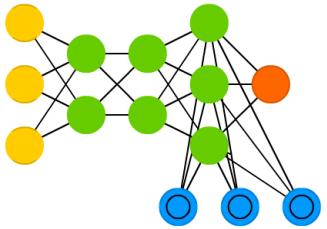
Deconvolutional Network (DN)



Deep Convolutional Inverse Graphics Network (DCIGN)

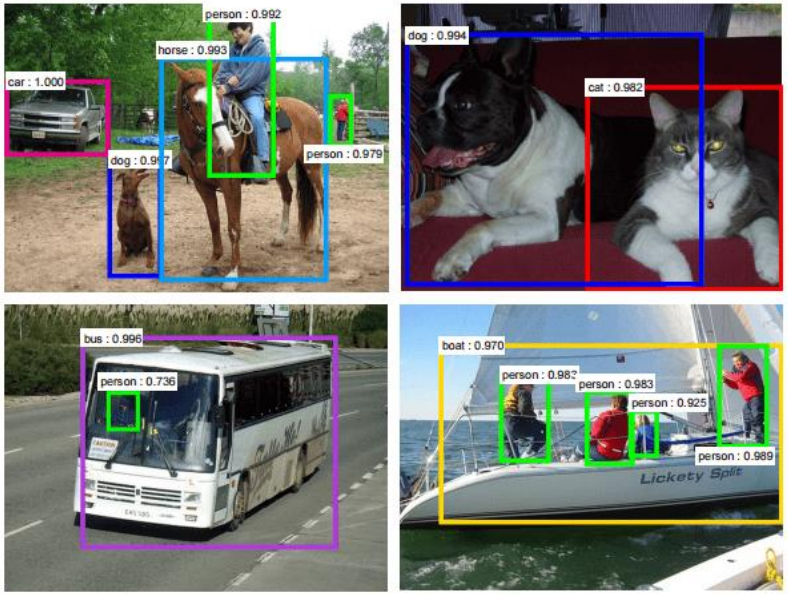


Neural Turing Machine (NTM)



Deep learning is awesome, actually.

- Requires no feature engineering
- Has sufficient parameters for known knowns, known unknowns, unknown unknowns



SQuAD1.1 Leaderboard

Since the release of SQuAD1.0, the community has made rapid progress, with the best models now rivaling human performance on the task. Here are the ExactMatch (EM) and F1 scores evaluated on the test set of v1.1.

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) *Google AI Language* https://arxiv.org/abs/1810.04805	**87.433**	**93.160**
2 Oct 05, 2018	BERT (single model) *Google AI Language* https://arxiv.org/abs/1810.04805	85.083	91.835

AI can reproduce Mona Lisa in the styles of Picasso, van Gogh, and Monet. Source: Gene Kogan

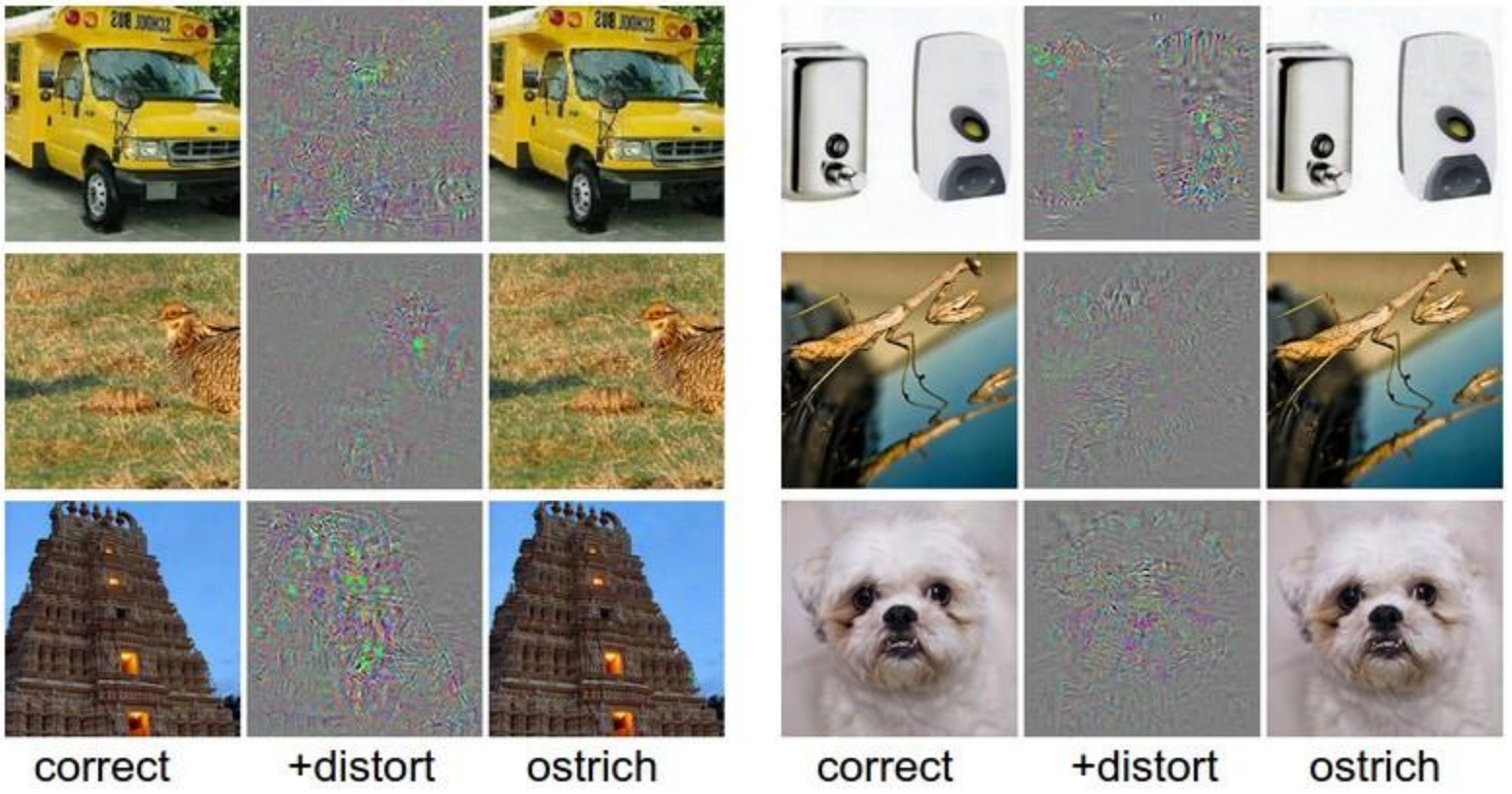
Music Transformer: Generating Music with Long-Term Structure

Dec 13, 2018 https://magenta.tensorflow.org/music-transformer

BUT

(1) DL is brittle and spurious

- Object Recognition



(1) DL is brittle and spurious

- Reading Comprehension

Paragraph: "In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses."

Question: "What city did Tesla move to in 1880?"

Answer: Prague

80% accuracy

Paragraph: "In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses. Tadakatsu moved to the city of Chicago in 1881."

Question: "What city did Tesla move to in 1880?"

Answer: Chicago

34.2% accuracy

(1) DL is brittle and spurious

- Visual Question Answering

question	answer
How many…?	2
Is/Are… ?	Yes
What sport…?	Tennis
What animal…?	Dog

(2) DL is data-crazy

- It doesn't care about general knowledge of the domains of interest
  - Intuitive Physics
  - Linguistic Phenomena
  - Psychology
  - …

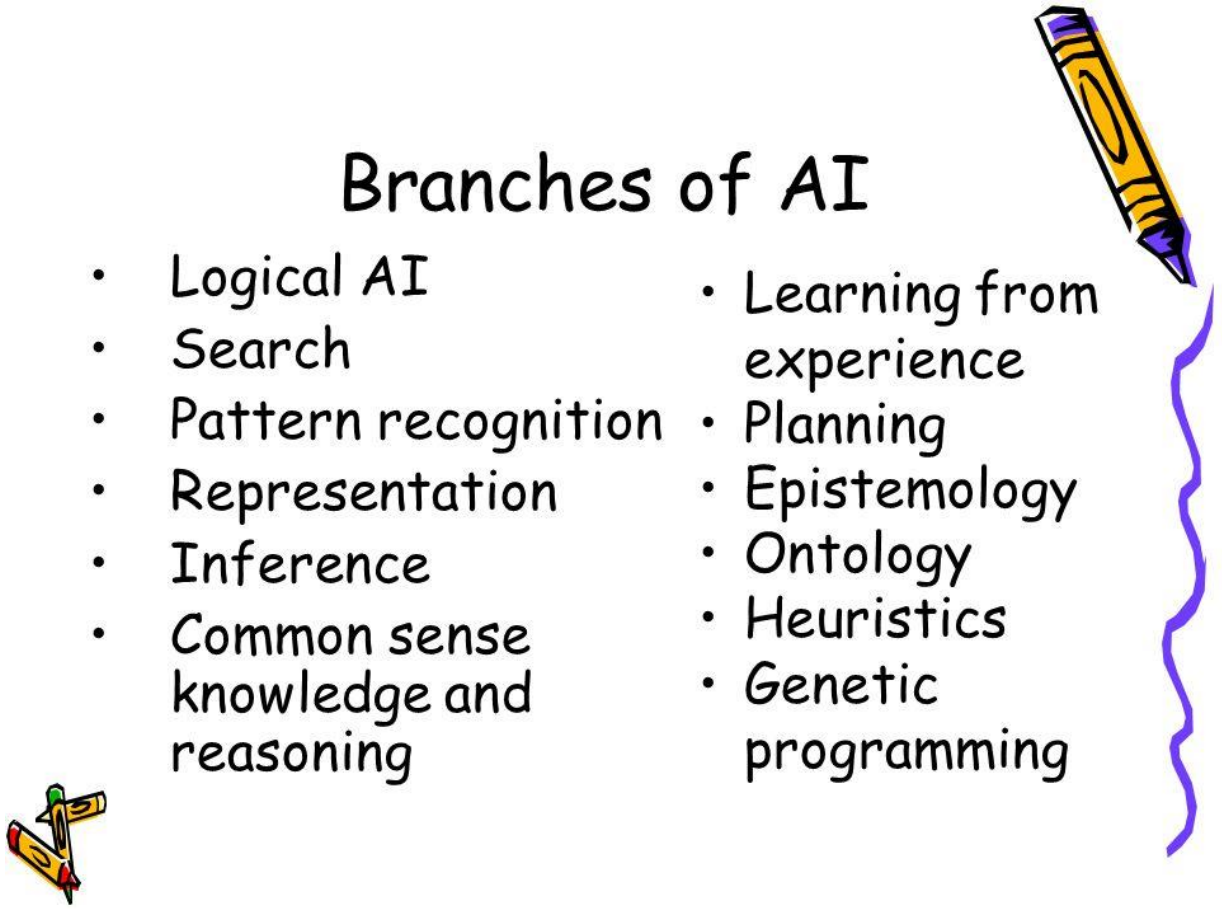
(2) DL is data-crazy

- It doesn't have any commonsense knowledge

Figure 1. Julia Child's kitchen. Photograph by Matthew Bisanz.

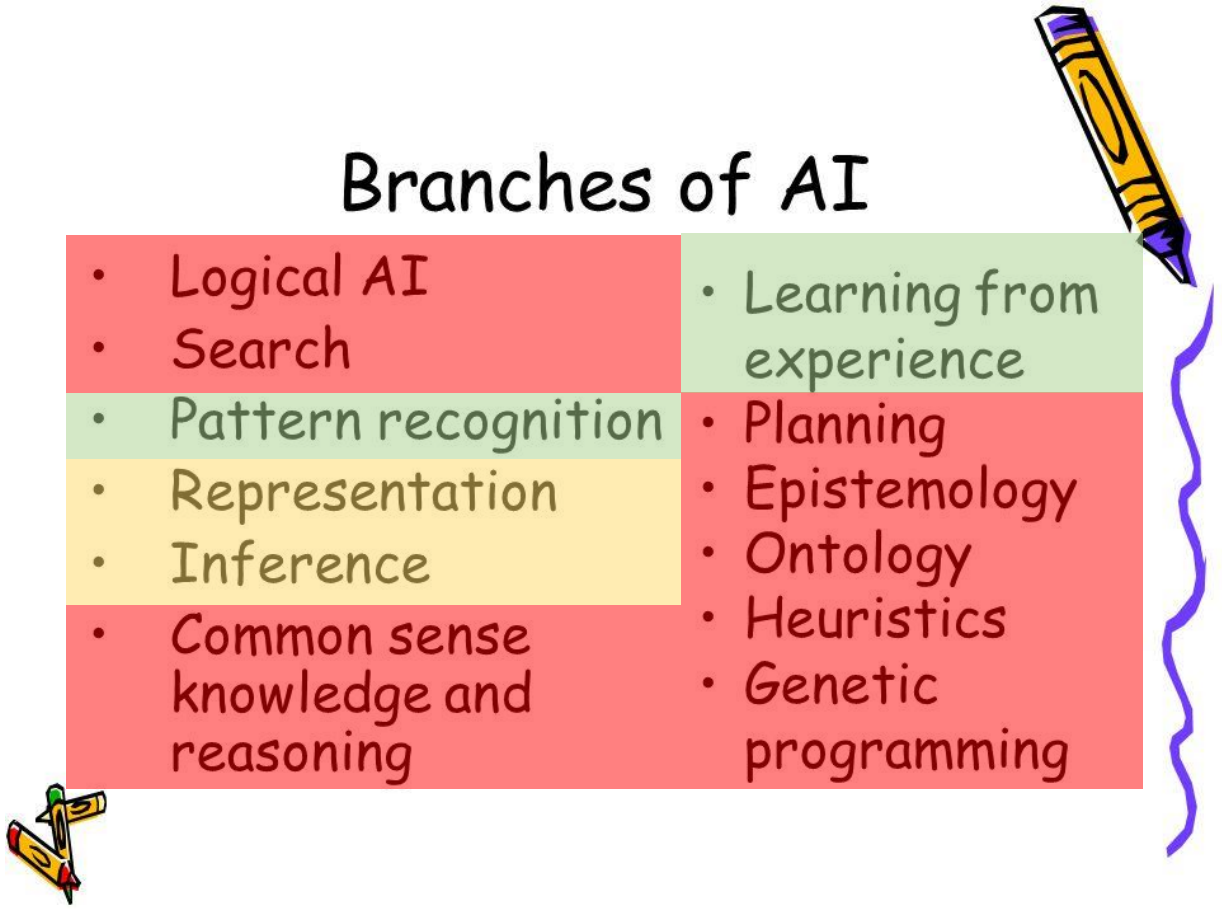
(2) DL is data-crazy

- It neglects other AI techniques outside machine learning

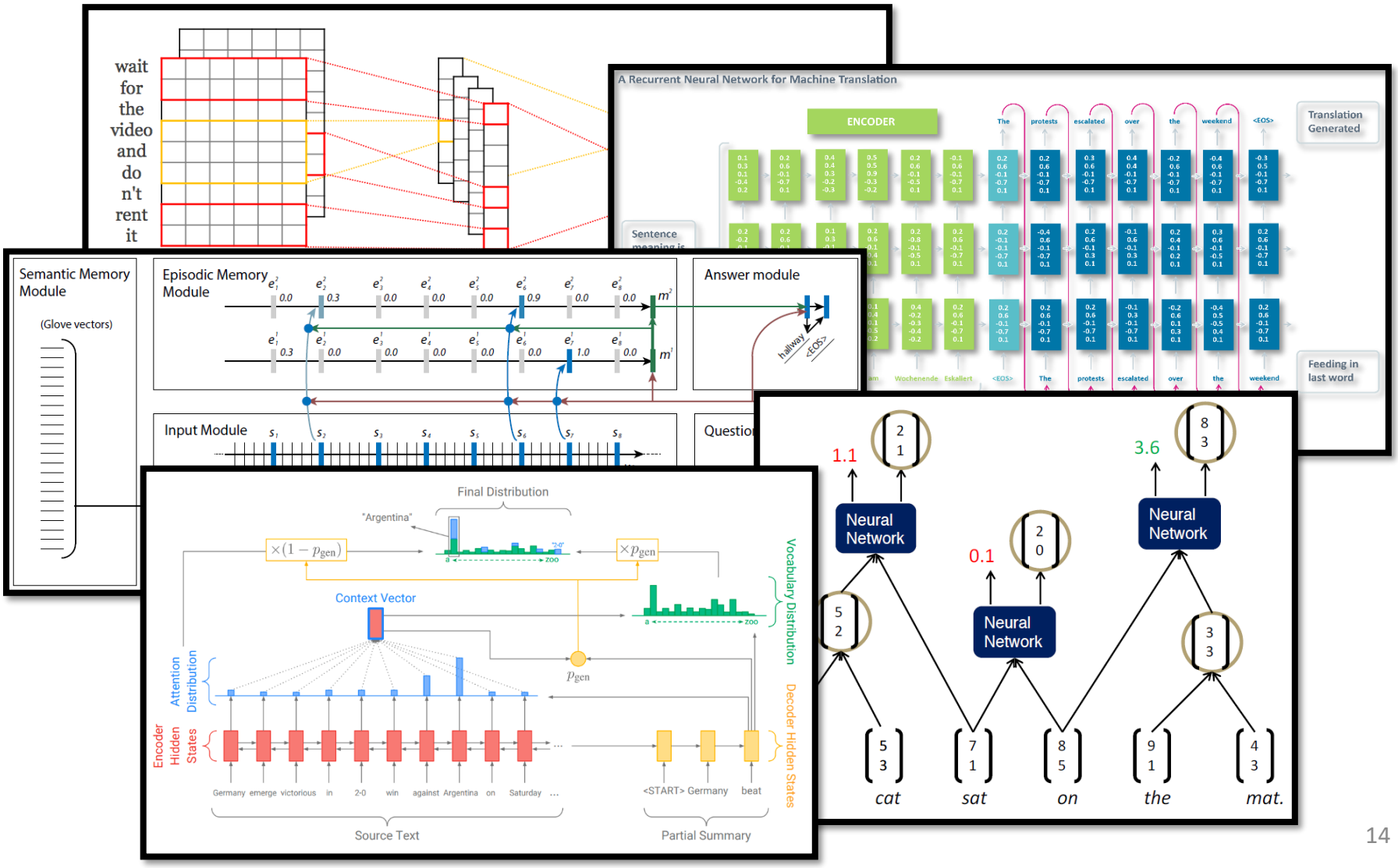


(2) DL is data-crazy

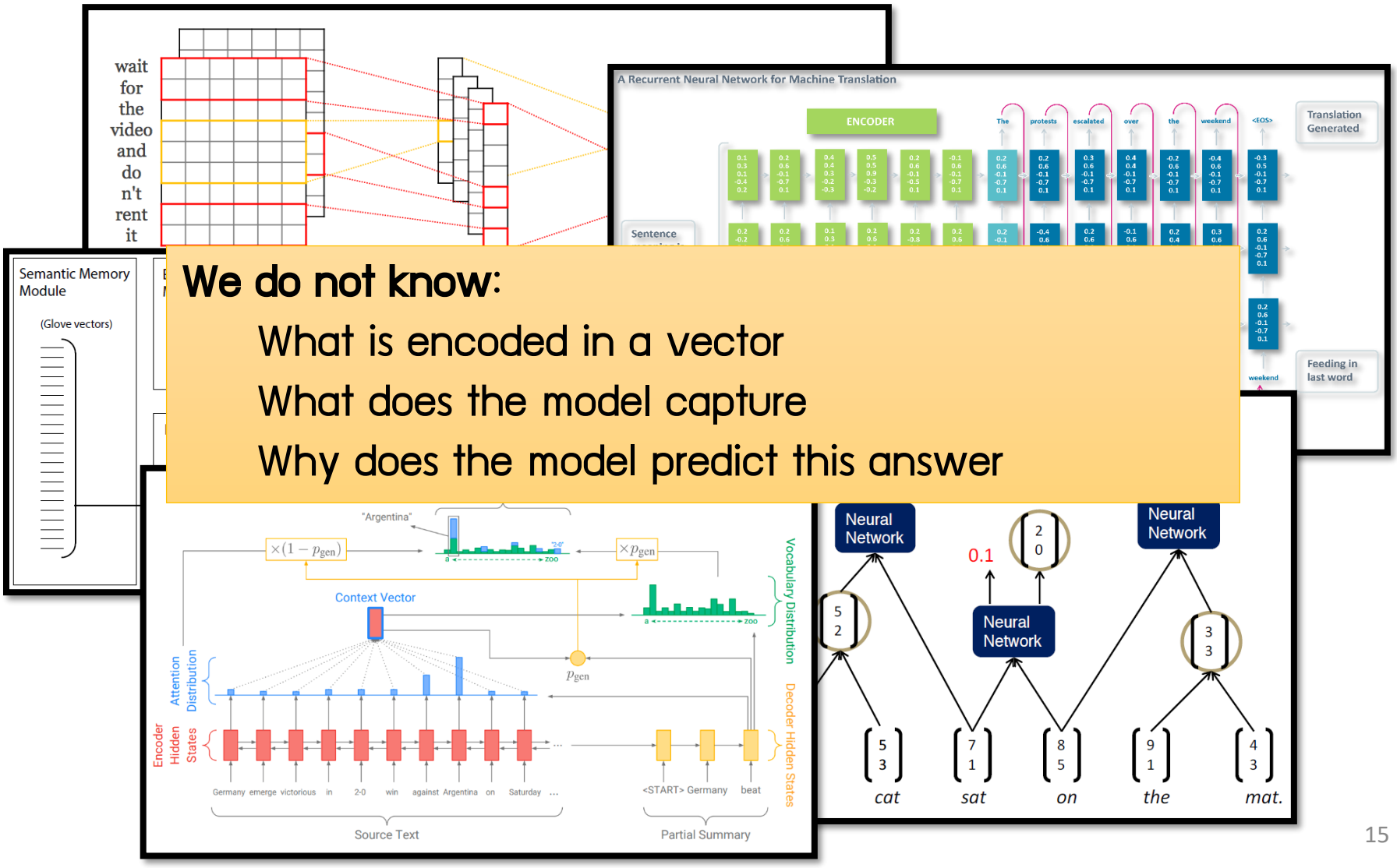
- It neglects other AI techniques outside machine learning



(3) DL is unexplainable



(3) DL is unexplainable



We do not know:

- What is encoded in a vector
- What does the model capture
- Why does the model predict this answer

(3) DL is unexplainable

- But we need a **right to explanation**
  - A right to be given an explanation for an algorithm's output

The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention.

...

In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.

The European Union General Data Protection Regulation (enacted 2016, taking effect 2018)

Related Research Trends 2018-2019

- More challenging AI/DL evaluation
- Using external/commonsense knowledge in DL
- Adding more inductive biases to DL
- Fusing model-based and function-based AI approaches
- Explainable (and effective) AI
- Human-AI argumentation

Image sources:

- https://cacm.acm.org/magazines/2018/6/228030-deep-learning-hunts-for-signals-among-the-noise/fulltext
- https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463
- https://sigmoidal.io/dl-computer-vision-beyond-classification/
- https://rajpurkar.github.io/SQuAD-explorer/
- https://medium.com/enabled-innovation/artificial-general-intelligence-too-much-or-too-little-too-soon-9c0dd7bd1c2d
- http://karpathy.github.io/2015/03/30/breaking-convnets/
- https://thegradient.pub/frontiers-of-generalization-in-natural-language-processing/
- https://writer.dek-d.com/ipep/story/view.php?id=797481
- https://cacm.acm.org/magazines/2015/9/191169-commonsense-reasoning-and-commonsense-knowledge-in-artificial-intelligence/fulltext
- https://www.slideshare.net/butest/cs-561a-introduction-to-artificial-intelligence