# Recent Advances and Key Challenges

Russ Salakhutdinov

Machine Learning Department
Carnegie Mellon University
Canadian Institute for Advanced Research

Carnegie Mellon University

MACHINE LEARNING DEPARTMENT

# Key Challenges

- Multimodal Learning
- Reasoning, Attention and Memory
- Natural Language Understanding
- Deep Reinforcement Learning
- Unsupervised Learning / One-Shot & Transfer Learning

# Deep Learning: Image Understanding

TAGS:

strangers, coworkers, conventioneers, attendants

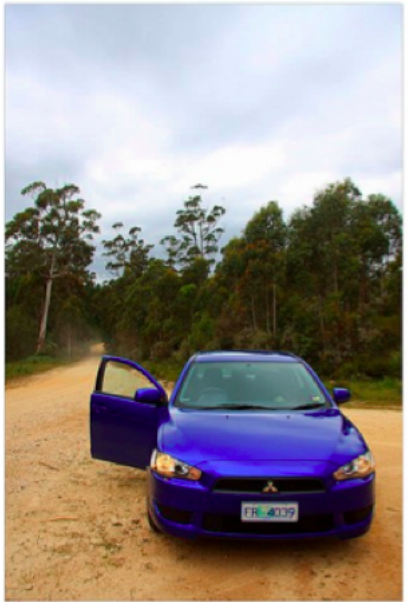
Nearest Neighbor Sentence:

people taking pictures of a crazy person

Model Samples

- a group of people in a crowded area
- a group of people are walking and talking
- a group of people, standing around and talking

# Caption Generation



A car is parked in the middle of nowhere

There is a cat sitting on a shelf

A little boy with a bunch of friends on the street

Kiros, Salakhutdinov, Zemel, ICML 2014

# Caption Generation

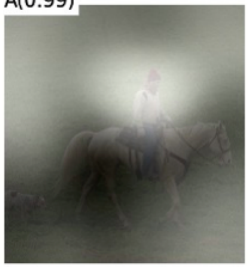
The two birds are trying to be seen in the water

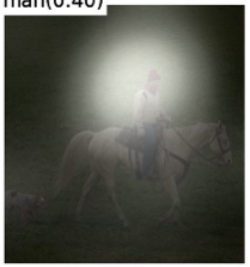
The handlebars are trying to ride a bike rack

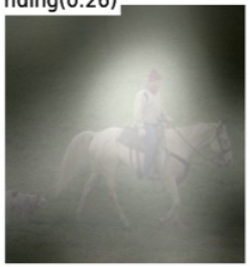
A man holding a red apple in his mouth

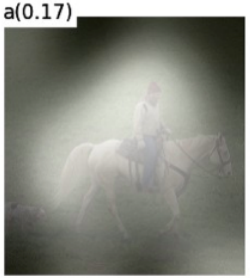
Kiros, Salakhutdinov, Zemel, ICML 2014

# Caption Generation with Visual Attention

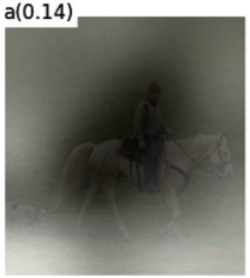







A man riding  a horse
in a field.

Xu et al, ICML 2015

# Caption Generation with Visual Attention

A woman holding a clock in her hand.

A large white bird standing in a forest.

Xu et al, ICML 2015

# Key Challenges

- Multimodal Learning
- Reasoning, Attention and Memory
- Natural Language Understanding
- Deep Reinforcement Learning
- Unsupervised Learning / One-Shot & Transfer Learning

# Who-Did-What Dataset

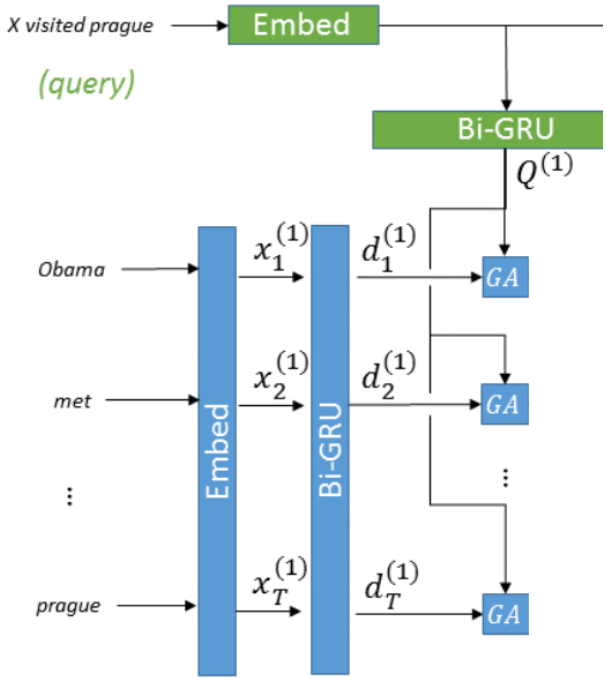
- **Context**: "…arrested Illinois governor **Rod Blagojevich** and his chief of staff John Harris on corruption charges … included **Blogojevich** allegedly conspiring to sell or trade the senate seat left vacant by President-elect Barack Obama…"

- **Query**: President-elect Barack Obama said Tuesday he was not aware of alleged corruption by **X** who was arrested on charges of trying to sell Obama's senate seat.

- **Answer**: Rod Blagojevich

Onishi, Wang, Bansal, Gimpel, McAllester, EMNLP, 2016

# Gated Attention Mechanism



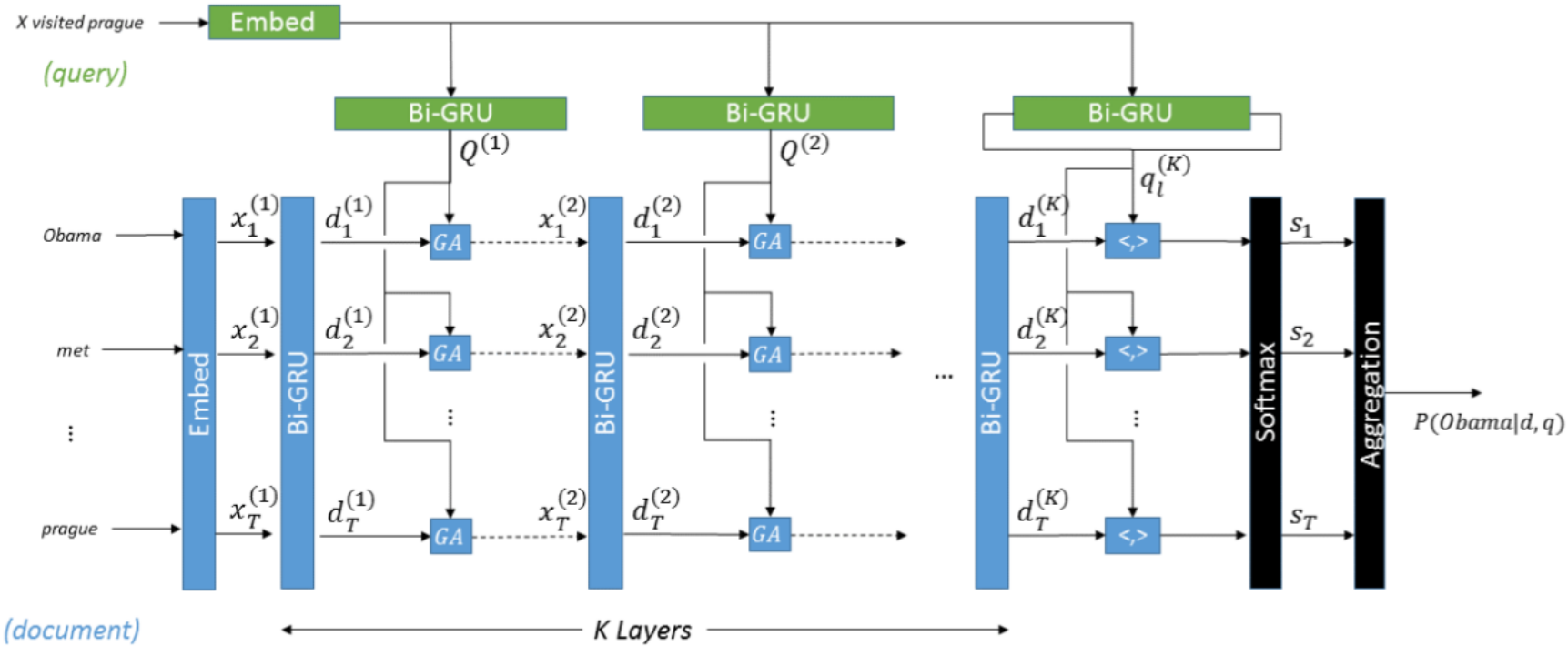

- Use Recurrent Neural Networks (RNNs) to encode a document and a query.
- Use element-wise multiplication to model the interactions between document and query:

$$x_i = d_i \odot q_i$$

Dhingra, Liu, Yang, Cohen, Salakhutdinov, 2016

# Multi-hot Architecture

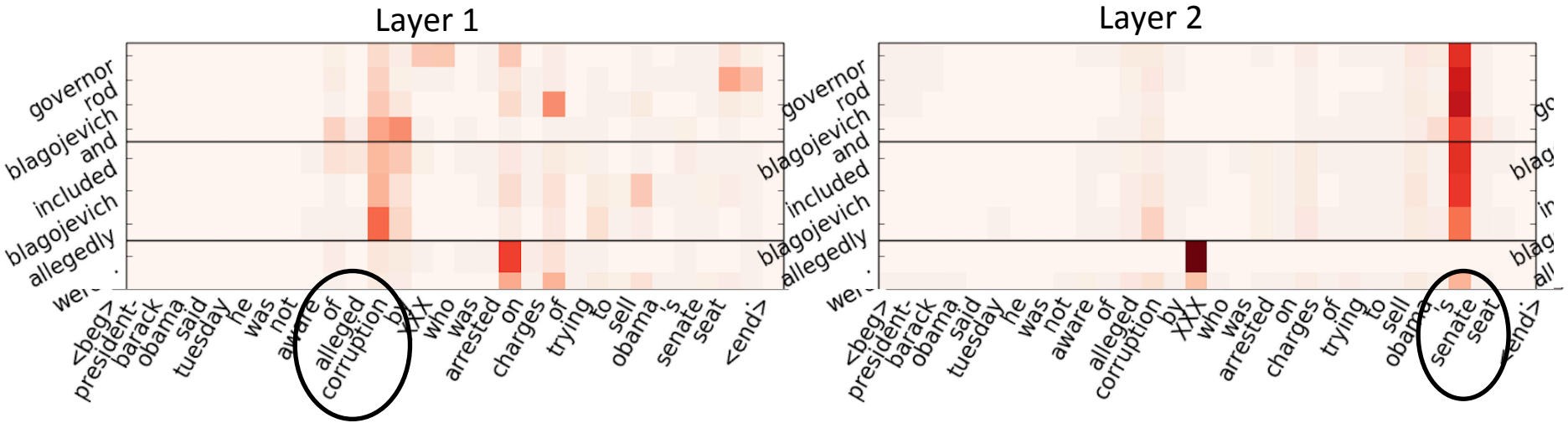
- Reasoning over multiple sentences requires several passes over the context




Dhingra, Liu, Yang, Cohen, Salakhutdinov, 2016

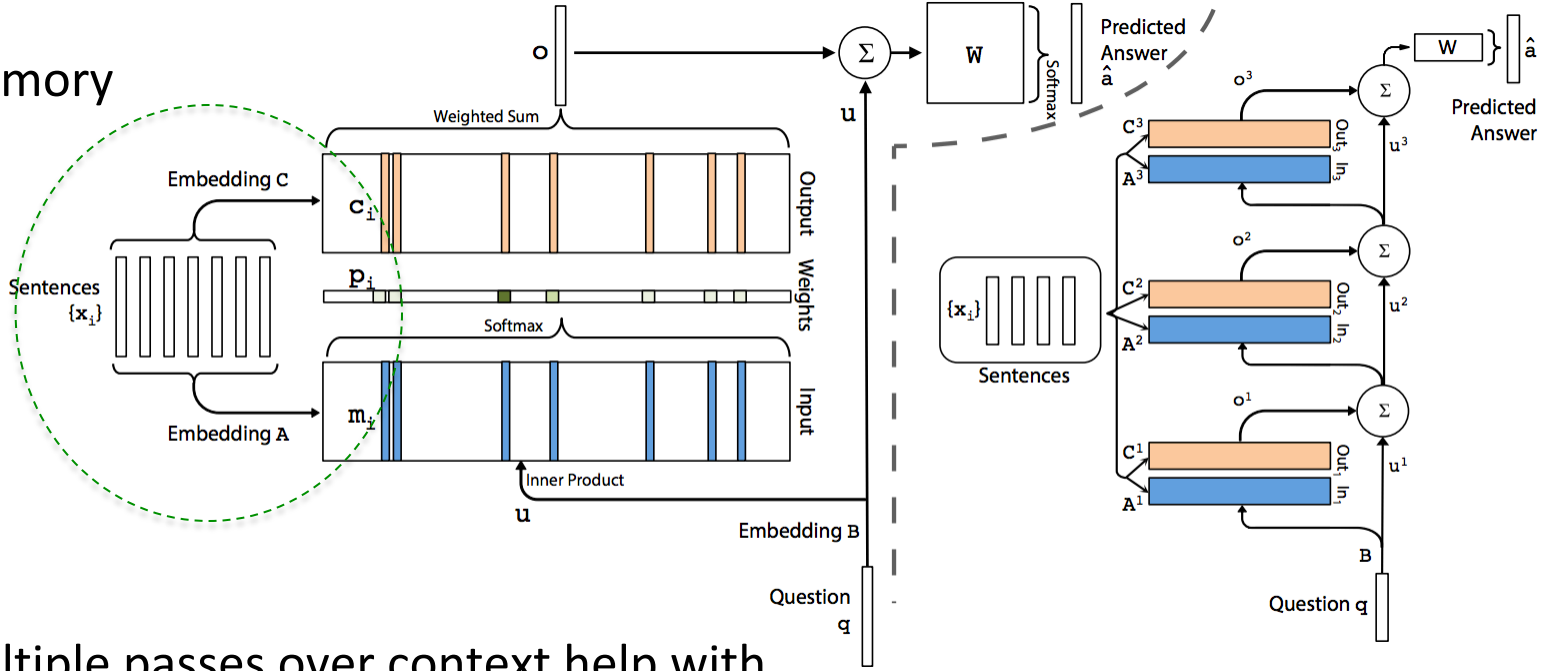
# Reasoning and Attention

- **Context**: "…arrested Illinois **governor Rod Blagojevich** and his chief of staff John Harris on corruption charges … **included Blogojevich** allegedly conspiring to sell or trade the **senate seat** left vacant by President-elect Barack Obama…"
- **Query**: "President-elect Barack Obama said Tuesday he was not aware of **alleged corruption** by **X** who was arrested on charges of trying to sell Obama's **senate seat**."
- **Answer**: **Rod Blagojevich**



# Memory Networks

Memory




Multiple passes over context help with sequential reasoning

Weston, Chopra, Bordes, ICLR 2015; Sukhbaatar et al., NIPS 2015

# Broad-Context Language Modeling

Her plain face broke into a huge smile when she saw Terry.
“Terry!” she called out.
She rushed to meet him and they embraced.
“Hon, I want you to meet an old friend, Owen McKenna.
Owen, please meet Emily."
She gave me a quick nod and turned back to **X**

LAMBADA dataset, Paperno et al., 2016

# Broad-Context Language Modeling

Her plain face broke into a huge smile when she saw **Terry**.
“Terry!” she called out.
She rushed to meet him and they embraced.
“Hon, I want you to meet an old friend, **Owen** McKenna.
Owen, please meet **Emily**."
She gave me a quick nod and turned back to **X**

LAMBADA dataset, Paperno et al., 2016

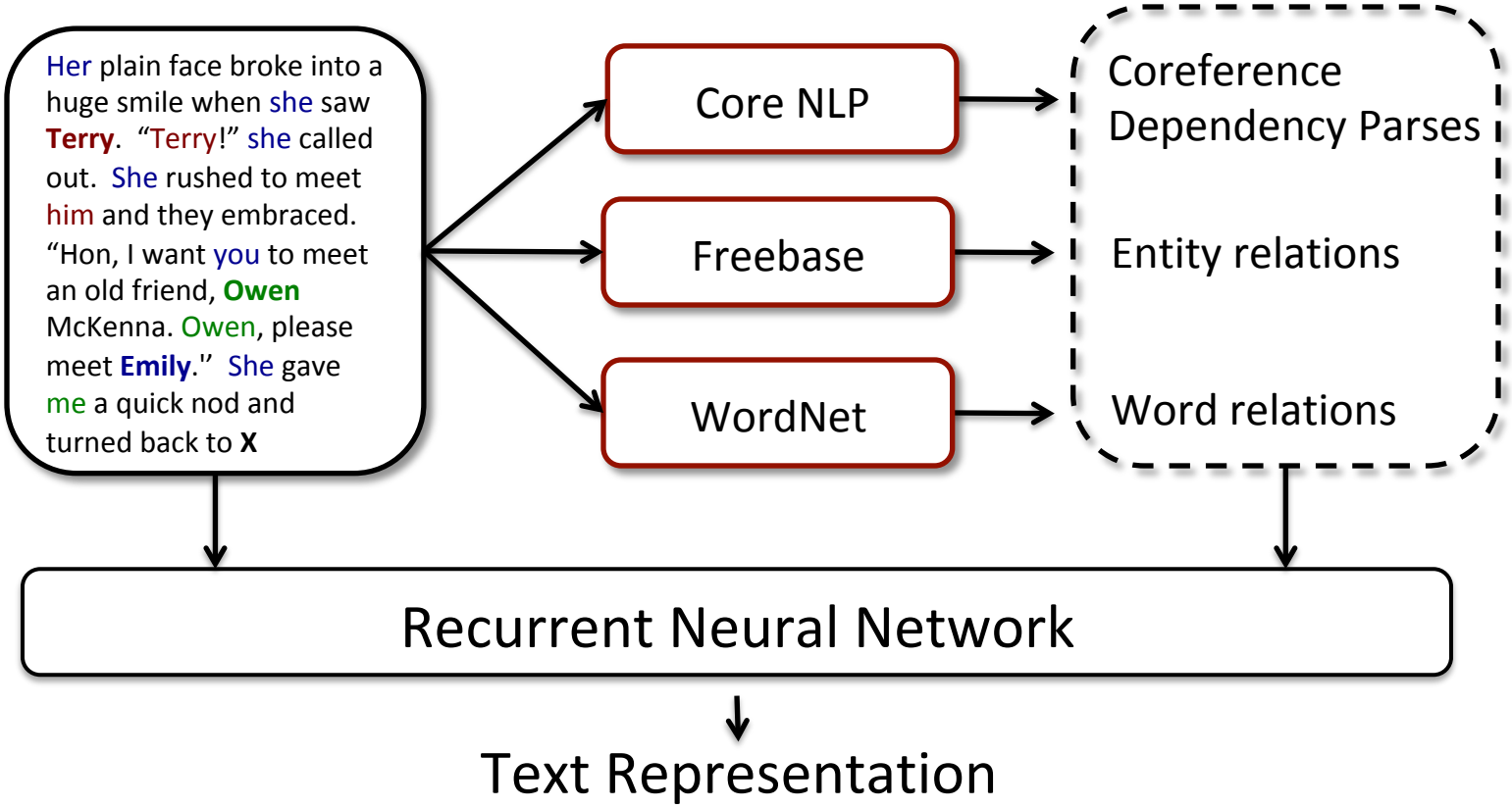
# Broad-Context Language Modeling

Her plain face broke into a huge smile when she saw **Terry**.
"Terry!" she called out.
She rushed to meet him and they embraced.
"Hon, I want you to meet an old friend, **Owen** McKenna.
Owen, please meet **Emily**."
She gave me a quick nod and turned back to **X**

X = **Terry**

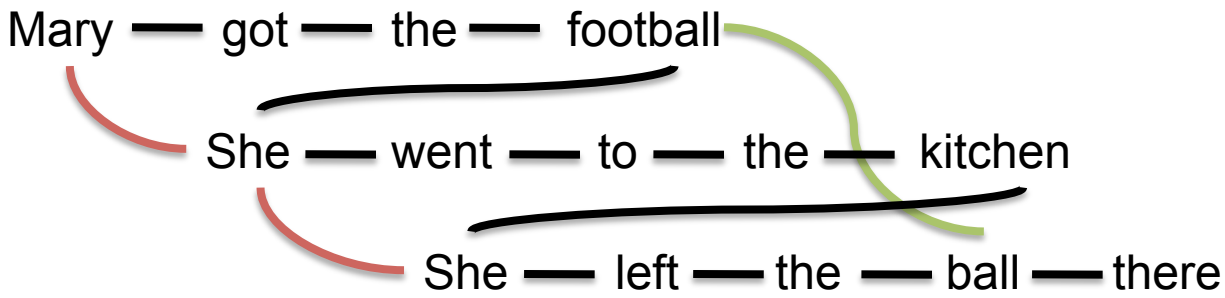
LAMBADA dataset, Paperno et al., 2016

# Incorporating Prior Knowledge




Dhingra, Jin, Yang, Cohen, Salakhutdinov NAACL 2018

# Explicit Memory



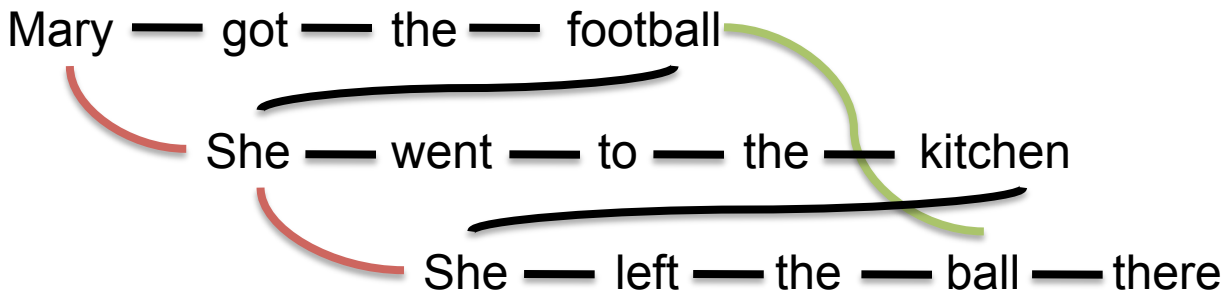

RNN

Coreference

Hyper/Hyponymy

Dhingra, Jin, Yang, Cohen, Salakhutdinov NAACL 2018

# Explicit Memory

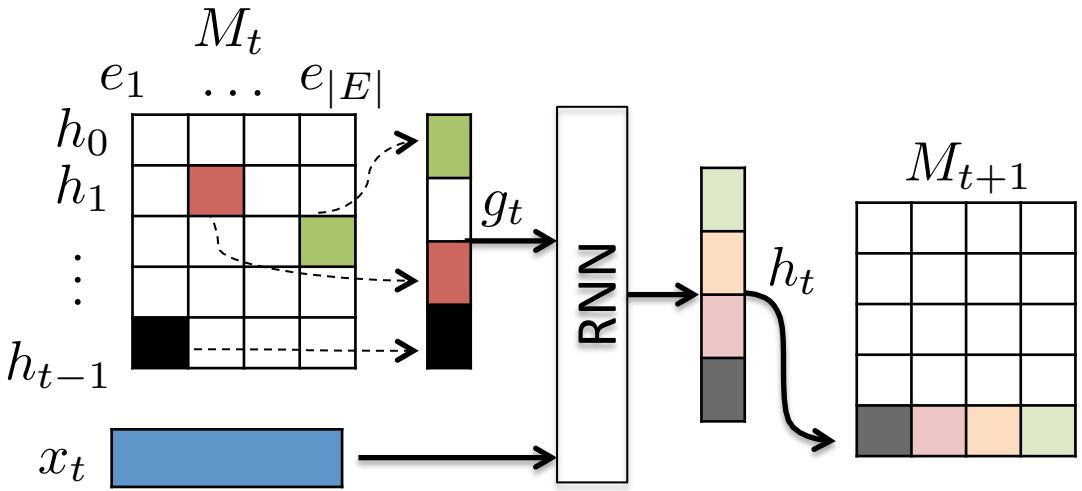



RNN

Coreference

Hyper/Hyponymy

**M**emory as **A**cyclic **G**raph **E**ncoding (MAGE) - RNN




Dhingra, Jin, Yang, Cohen, Salakhutdinov NAACL 2018

# Open Domain Question Answering

- Finding answers to factual questions posed in Natural Language:

***Who voiced Meg in Family Guy?***

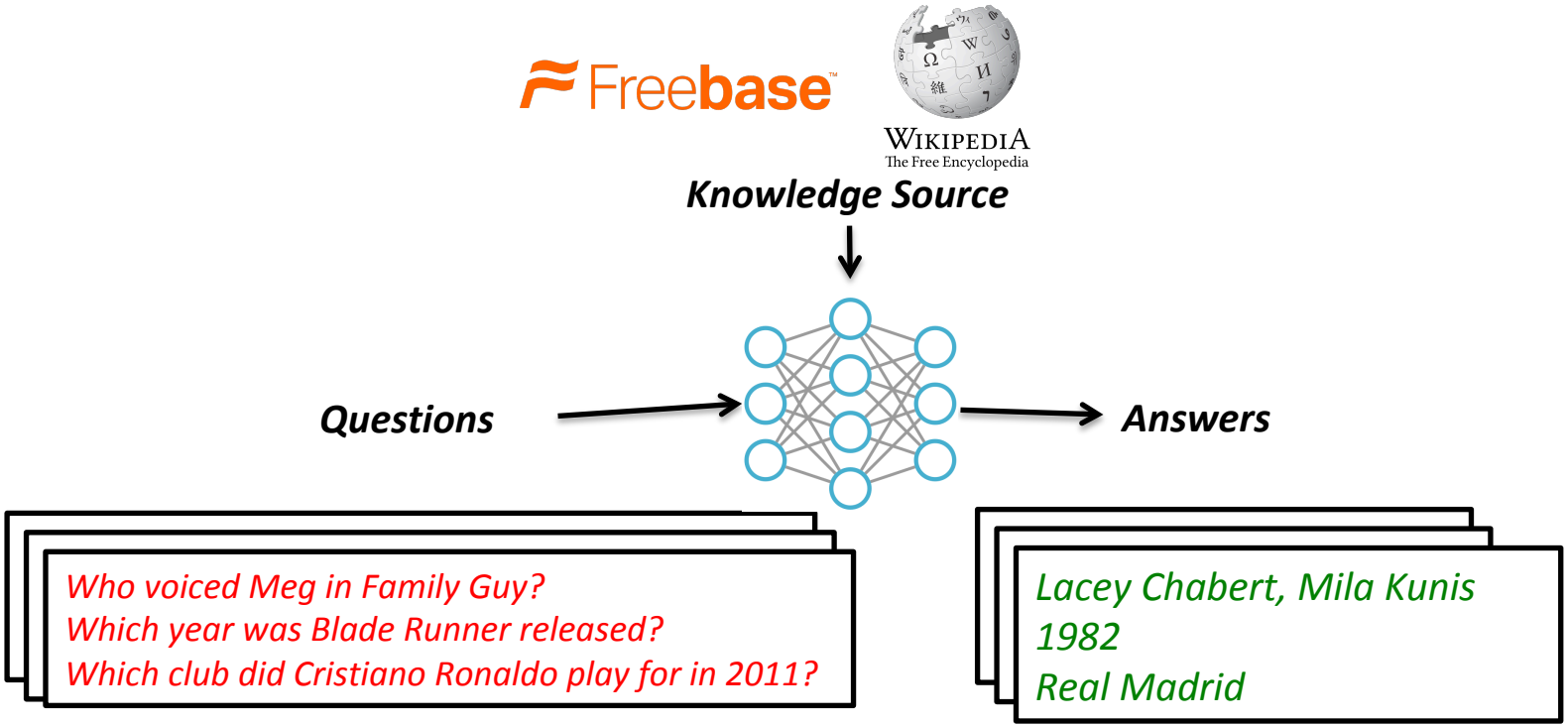
**A. Lacey Chabert, Mila Kunis**

***Who first voiced Meg in Family Guy?***

**A. Lacey Chabert**

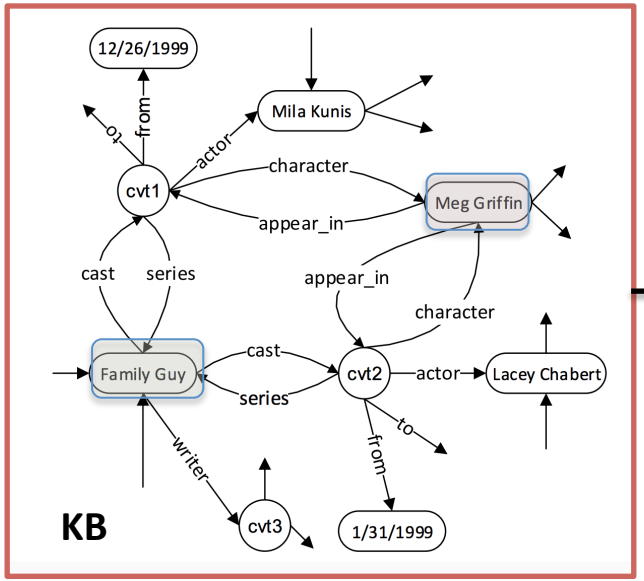
Bhuwan Dhingra et. al. 2018

# Text Augmented Knowledge Graphs




Bhuwan Dhingra et. al. 2018

# Knowledge Base as a Knowledge Source

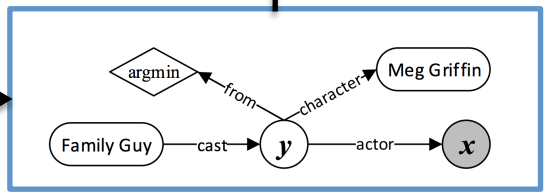



Lacey Chabert

Query Graph

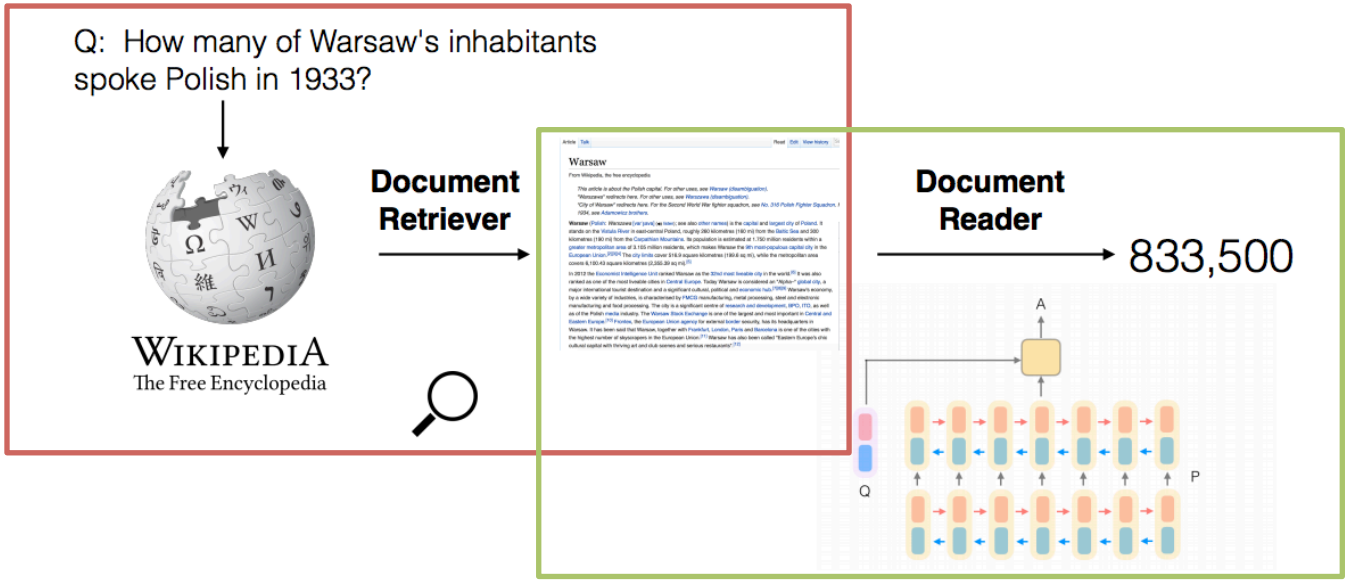
*Who first voiced Meg in Family Guy?*

*Semantic Parsing*




Bhuwan Dhingra et. al. 2018

# Text as a Knowledge Source




**Step 1 (Information Retrieval):**

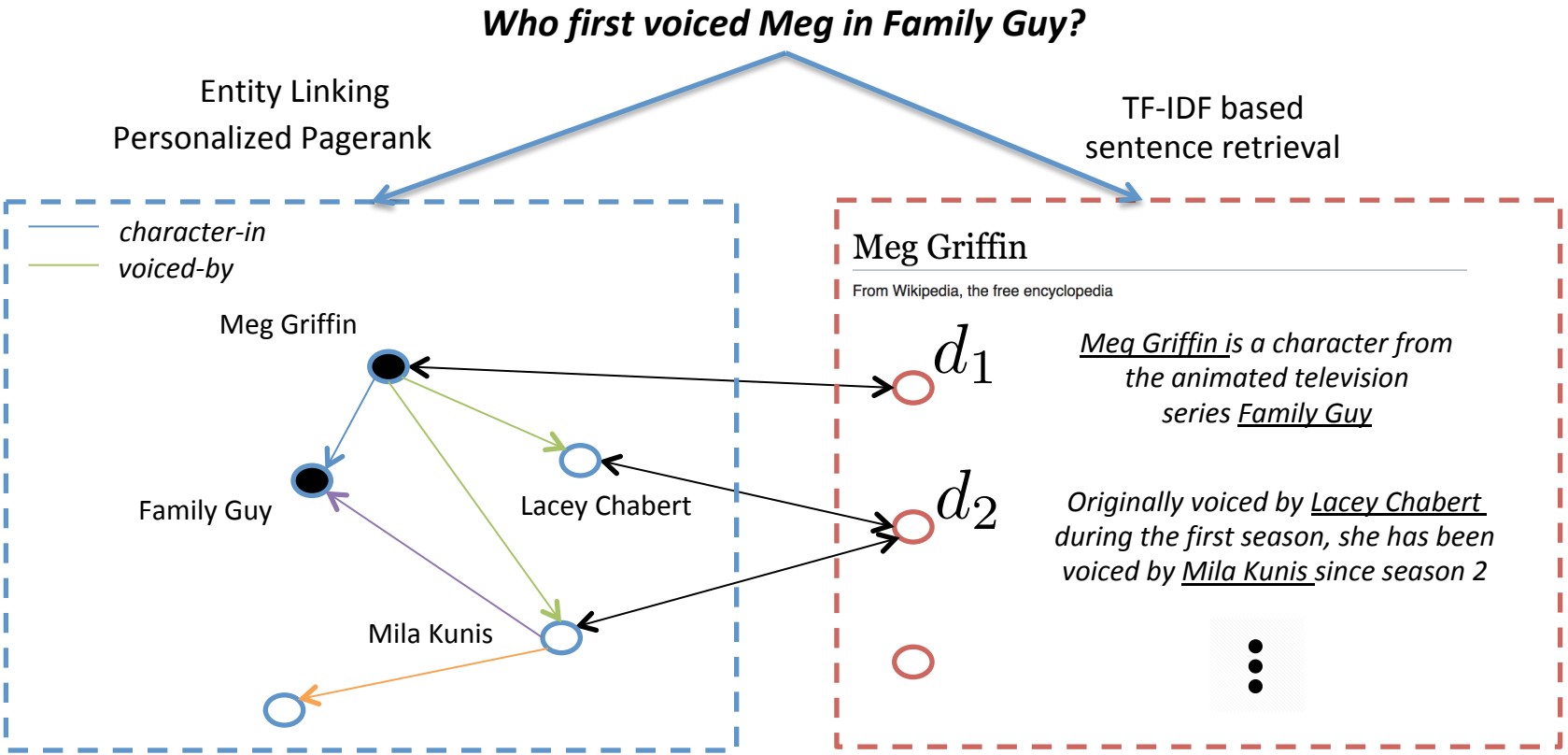
Retrieve passages relevant to the Question using shallow methods

**Step 2 (Reading Comprehension):**

Perform deep reading of passages to extract answers

Bhuwan Dhingra et. al. 2018

# Text Augmented Knowledge Graph

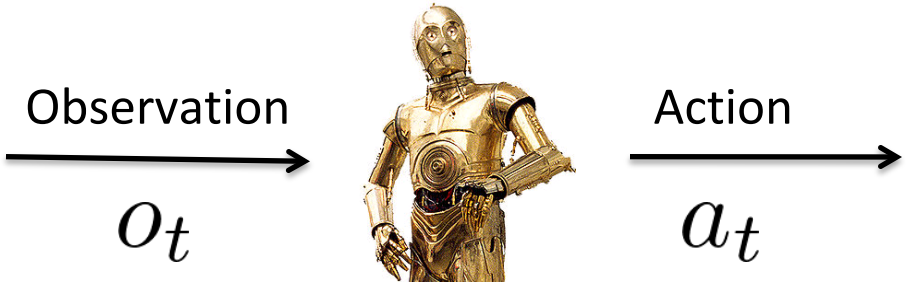



Bhuwan Dhingra et. al. 2018

# Key Challenges

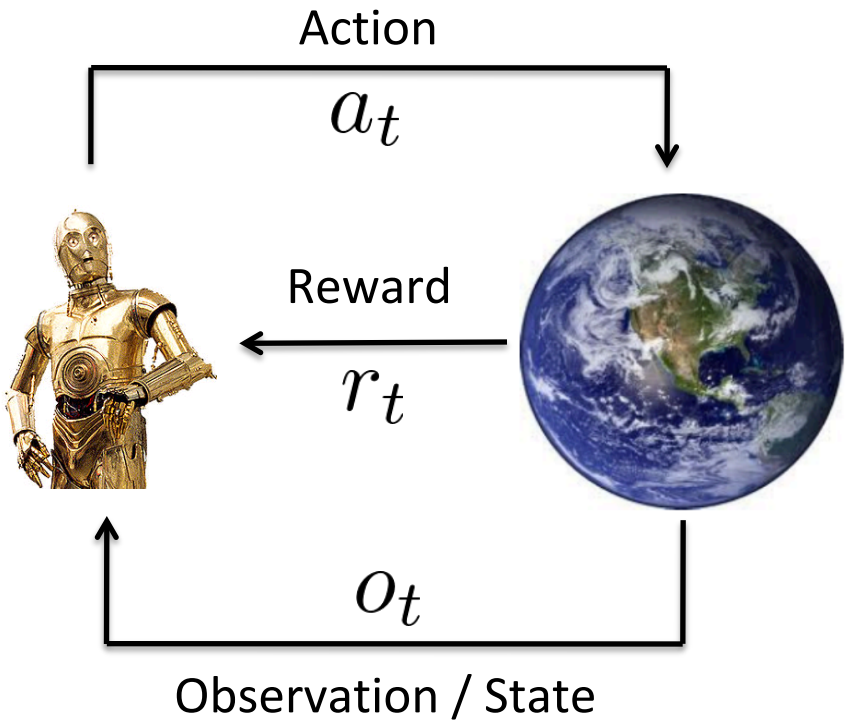
- Multimodal Learning
- Reasoning, Attention and Memory
- Natural Language Understanding
- Deep Reinforcement Learning
- Unsupervised Learning / One-Shot & Transfer Learning

# Learning Behaviors

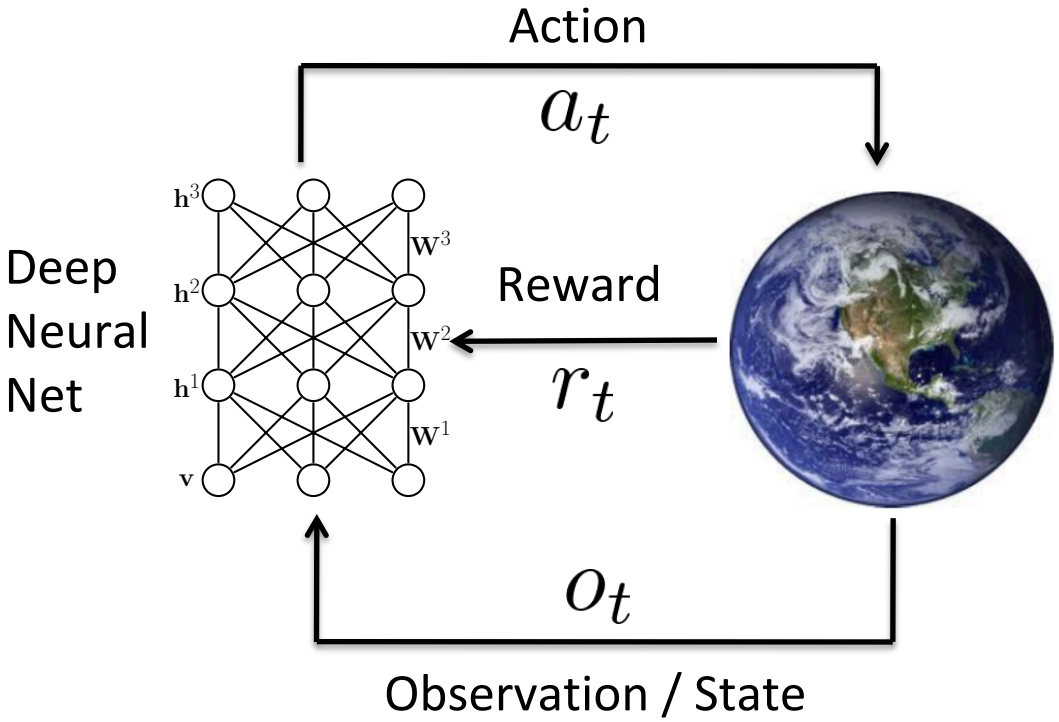



Learning to map sequences of observations to actions, for a particular goal

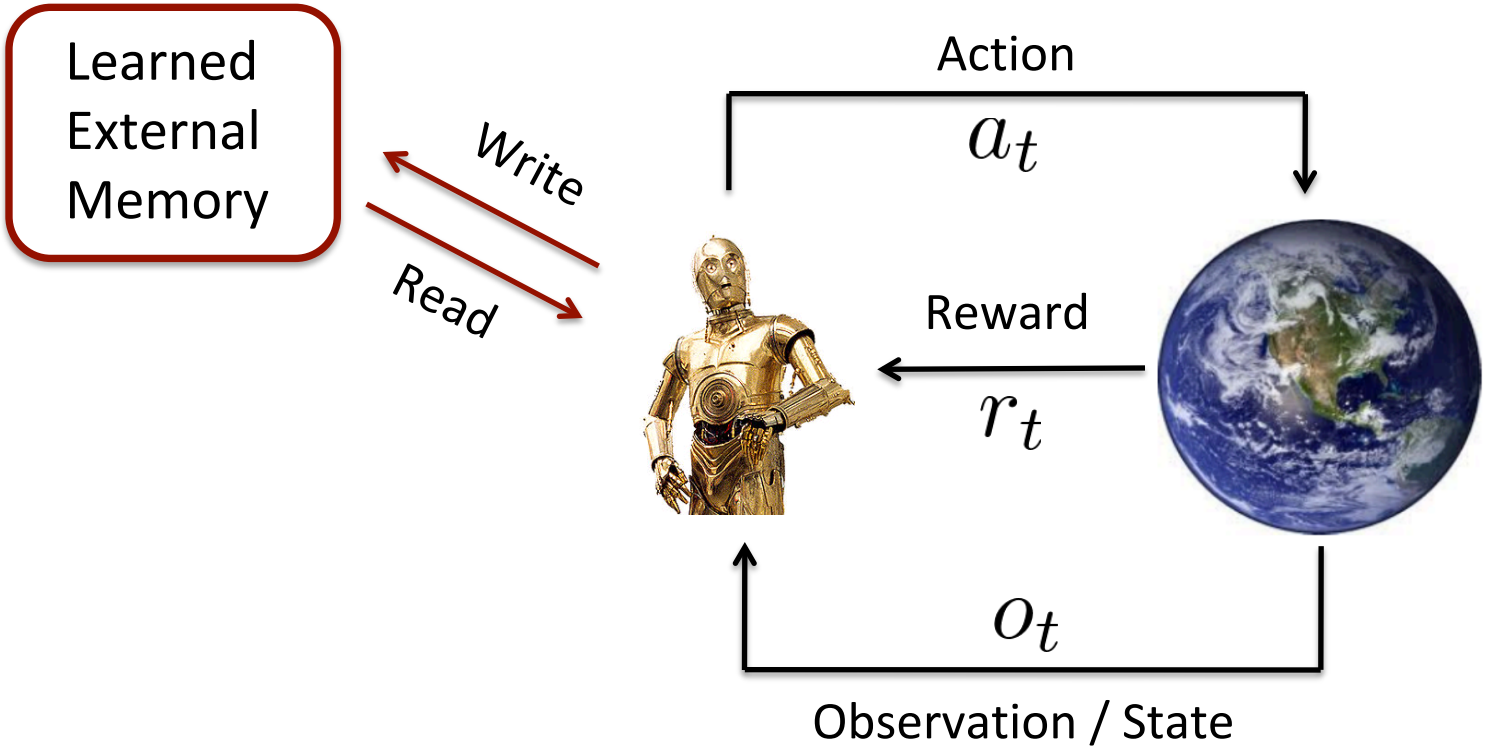
# Reinforcement Learning



# Deep Reinforcement Learning

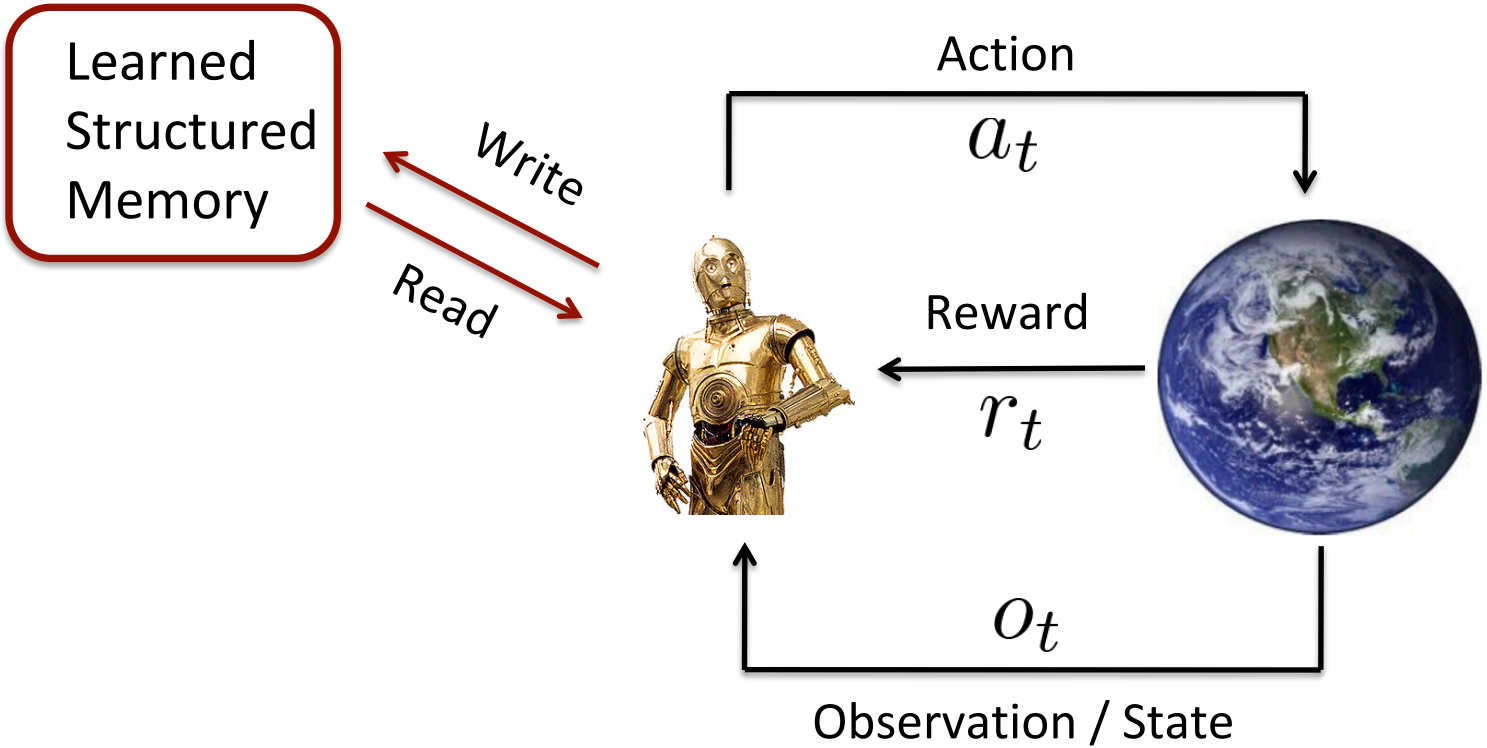


# Deep RL with Memory




Differentiable Neural Computer, Graves et al., Nature, 2016; Neural Turing Machine, Graves et al., 2014

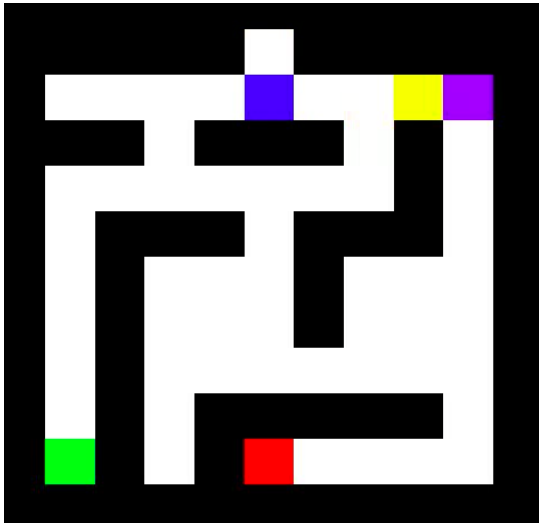
# Deep RL with Memory



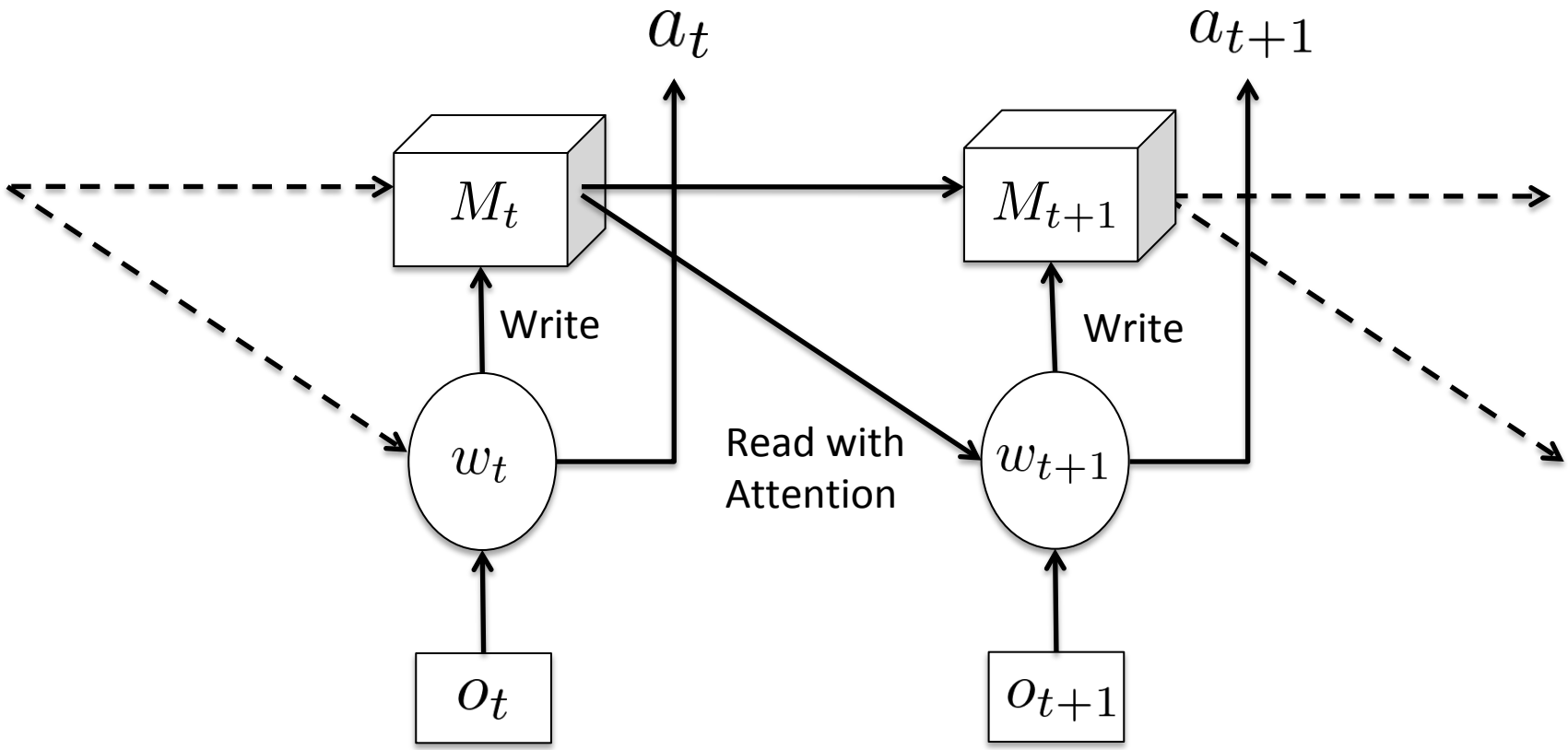

Parisotto, Salakhutdinov, ICLR 2018

# Random Maze with Indicator

- Indicator: Either blue or pink
  - If blue, find the green block
  - If pink, find the red block
- Negative reward if agent does not find correct block in N steps or goes to wrong block.

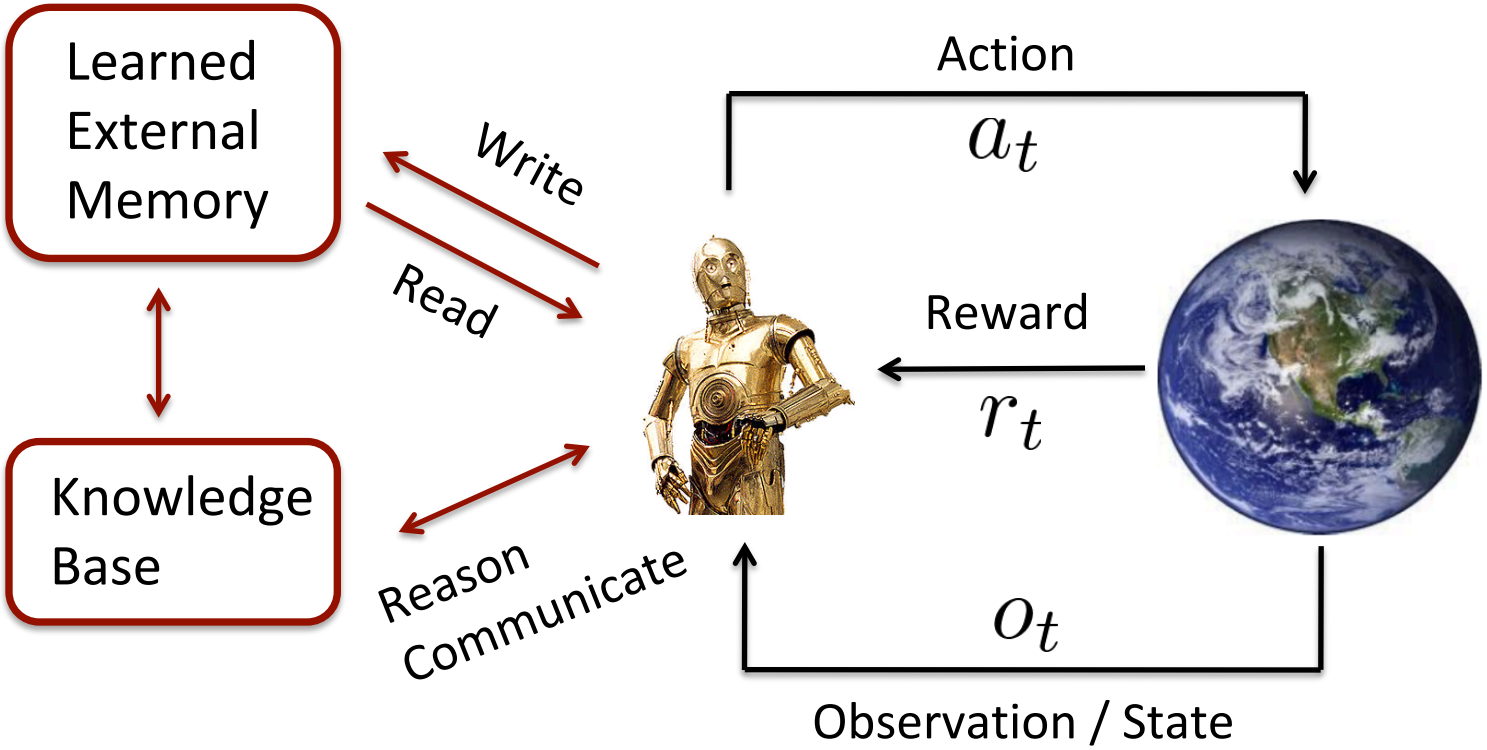


# Deep RL with Structured Memory

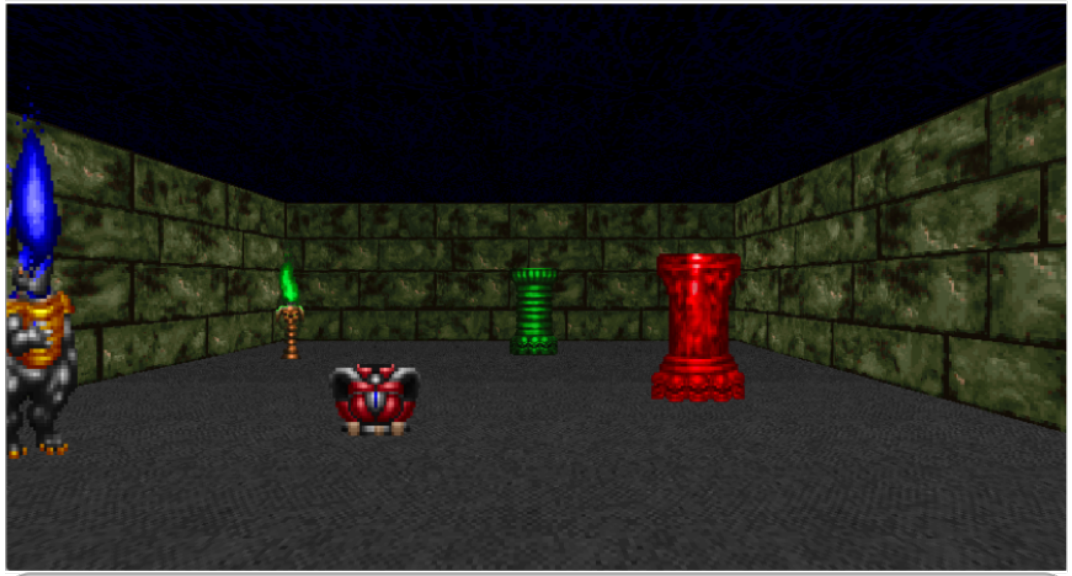



Parisotto, Salakhutdinov, 2017

# Building Intelligent Agents



# Task-oriented Language Grounding



**Go to the green torch**

**Train**

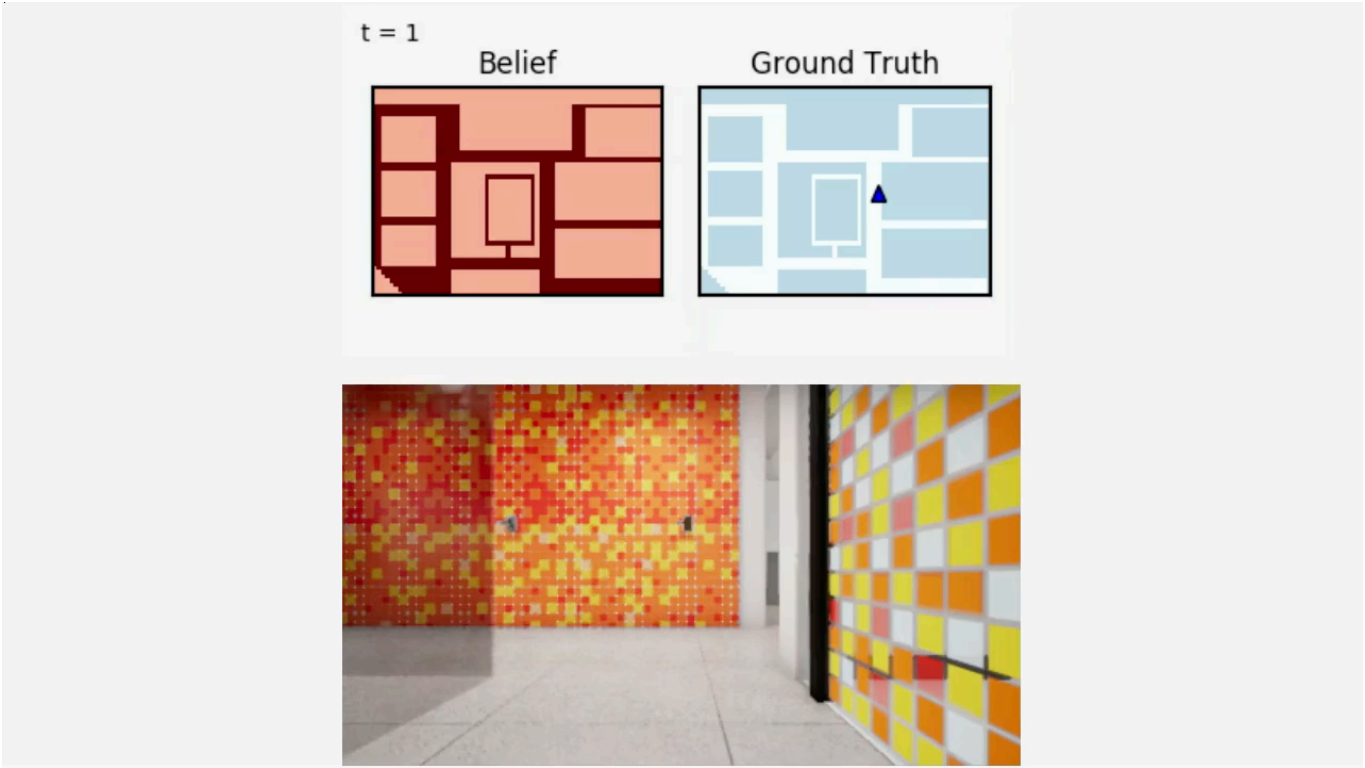
Go to the short red torch
Go to the blue keycard
Go to the largest yellow object
Go to the green object

**Test**

Go to the tall green torch
Go to the red keycard
Go to the smallest blue object

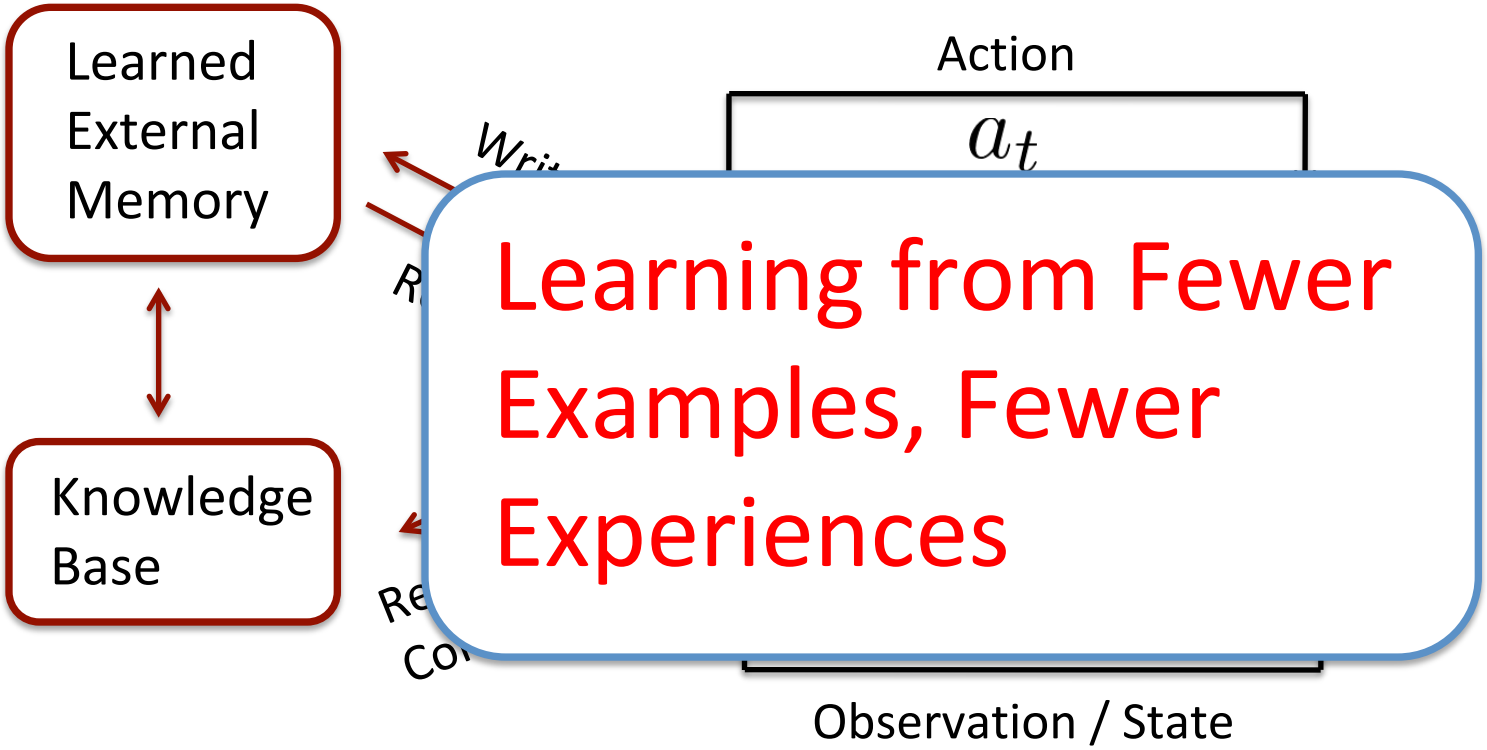
Chaplot et al., AAAi 2019

# Active Neural Localization and SLAM




Chaplot, Parisotto, Salakhutdinov, ICLR2018

# Building Intelligent Agents



Thank you