Project Alexandria
Constructing a KB of Common Sense Knowledge

Oren Etzioni

CEO, Allen Institute for AI

(Professor, Allen School, UW)

“..position papers discussing forward-looking work.”

“ ..novel **shared tasks** to help develop a better community.”

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

Outline

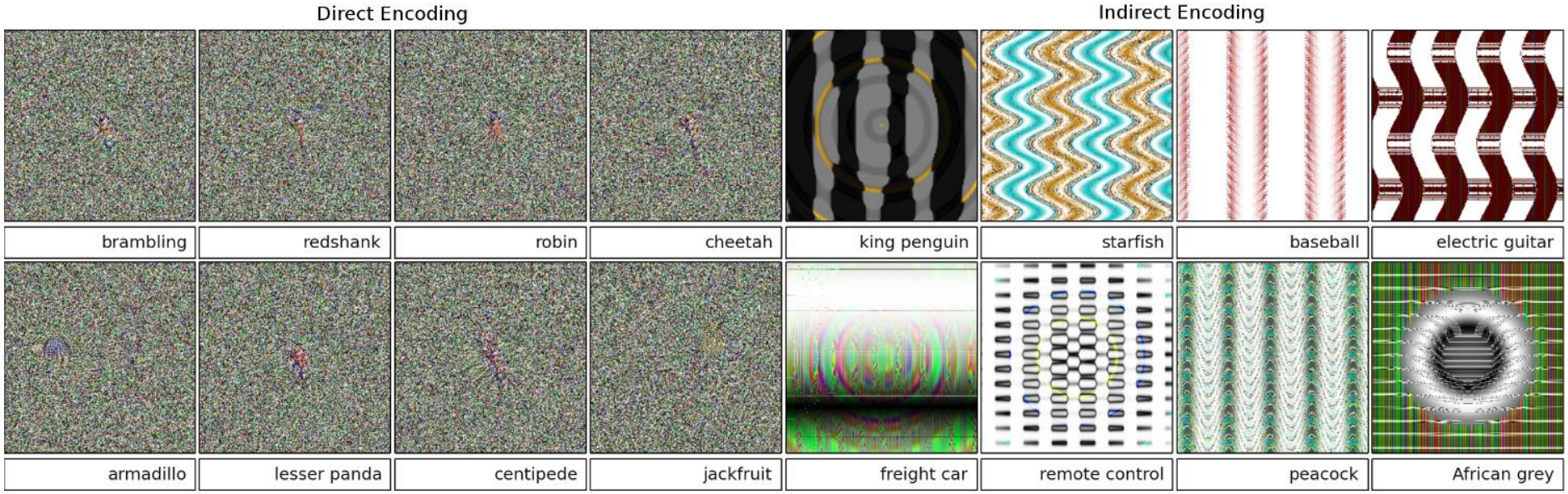
I. Motivation & Formulation
II. Lessons from Cyc
III. Common-sense Benchmark
IV. Acquisition of Common Sense Knowledge

I. Why Bother?

".. performance in their specialized domains is often very impressive.

Nevertheless, hardly any of them have certain commonsense knowledge…This lack makes them **"brittle."** …

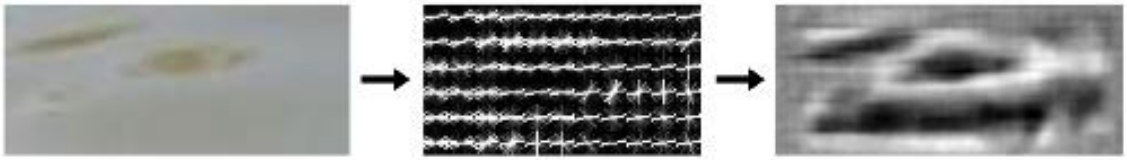
difficult to expand beyond the scope originally contemplated by their designers.."



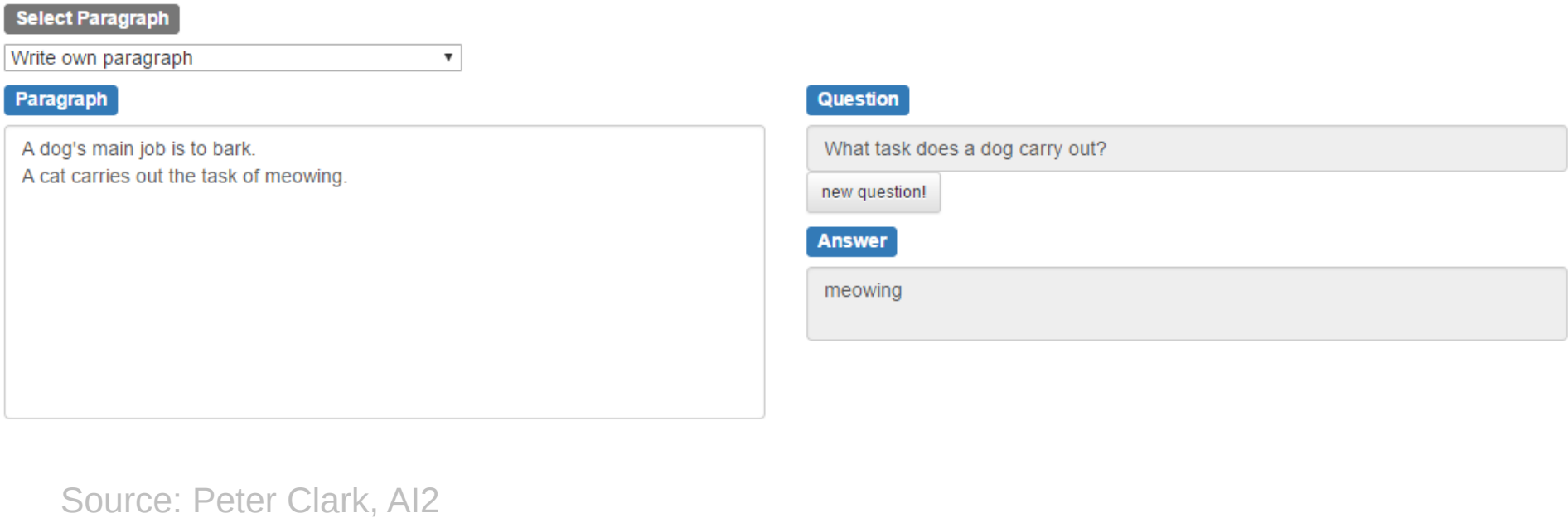
Source: Nguyen, *et. al,* Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. (CVPR '15).

Objection Detection Failure…

Figure 1: An image from PASCAL and a high scoring car detection from DPM [8]. Why did the detector fail?



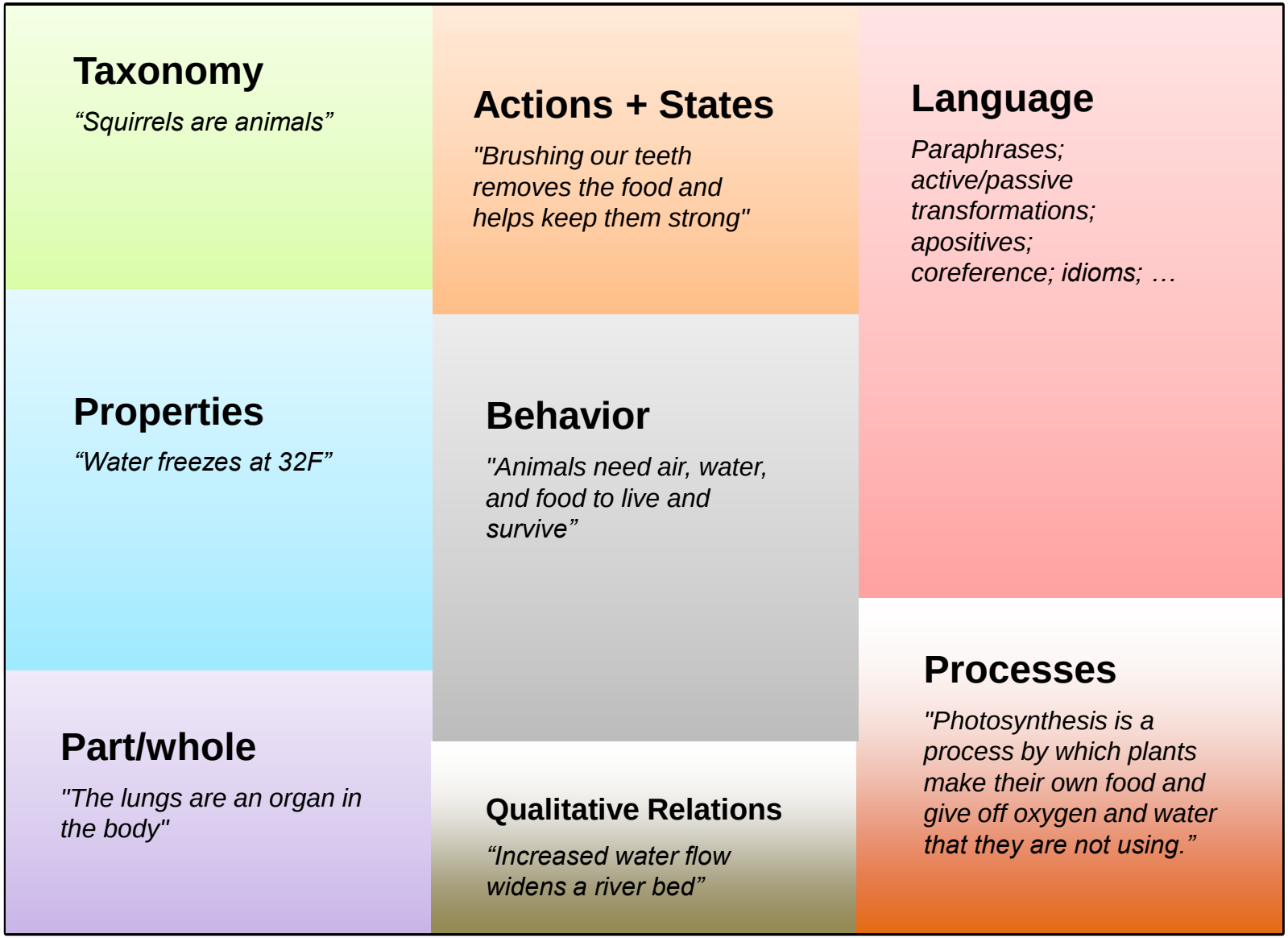
Tricking SqUAD “Readers”



Source: Peter Clark, AI2

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

Knowledge for Middle School Science (Aristo)



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

What is Common Sense?

Knowledge about the world that most 10 year olds have, but most AI systems do not

"…Common-sense facts and methods are only very partially understood today, and extending this understanding is the key problem facing artificial intelligence."

John McCarthy, (1983)

2018: No AI system can answer simple questions such as:

- What's bigger the sun or a Giraffe?
- If I put my socks in the drawer, will they still be there tomorrow?
- What would you typically find in a trash can?

Why do AI systems need Common Sense?

- **Robustness:** adversarial examples; zero-short learning
- **Data Efficiency:** learn with fewer training examples
- **Generality:** transfer learning, *etc.*
- **Performance:** NLP, robotics, medical diagnosis, *etc.*
- **Safety:** how can an AI system avoid harm if doesn't know what is harmful?

II. Lessons from Cyc (Lenat *et al.*, 1984)

"If I have seen further then others, it's because I have stood on the faces of giants…"

- Implicit knowledge is critical
- Size matters--scalability
- But you need to know how to use it! (reasoning..)
- Consistency is not realistic (micro-theories)

Crowd sourcing, machine vision, and modern NLP are an opportunity to re-visit this grand challenge!

Benchmark/performance metric is essential

III. Creating a Benchmark for Common Sense

Questions

- **Breath:** what topics are covered?
- **Depth:** what is the sophistication of knowledge?
- **Language:** should benchmark factor out linguistic challenges (e.g., parapharses)?
- **Vision**: is visual/robotic common included?

How to create it?

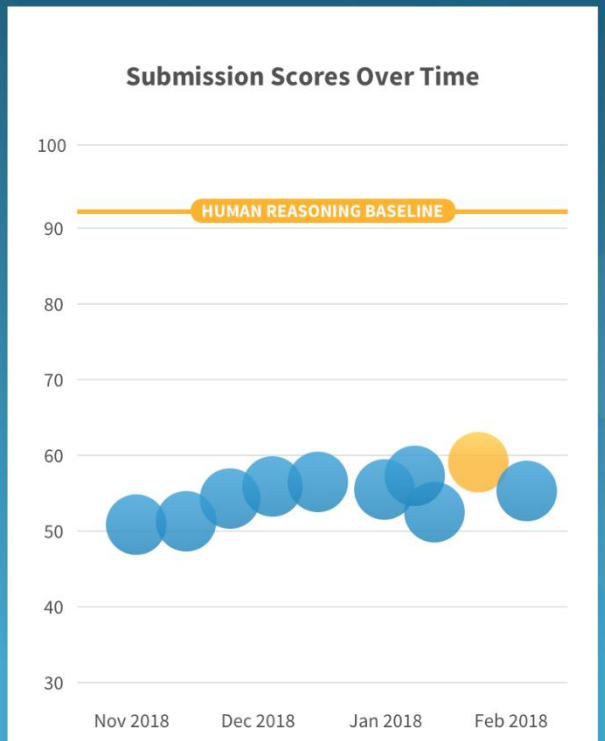
Commonsense Leaderboard

ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

Home Explore Submissions About

ALEXANDRIA

Common sense is what every 10-year-old has and no machine has.



Why Common Sense?

Leaderboard

Rank	Model	Date	Score
	Human Reasoning Baseline		**91.8**
1	**Hybrid AoA Reader (ensemble)** Joint Laboratory of HIT and iFLYTEK Research	2/8/2018	**59.5**
2	**SLQA+ (ensemble)** Alibaba iDST NLP	1/29/2018	**57.9**
3	**r-net+ (ensemble)** Microsoft Research Asia	1/7/2018	**55.6**
4	**AttentionReader+ (ensemble)** Tencent DPDAC NLP	12/29/2017	**55.1**
5	**BiDAF + Self Attention + ELMo (ensemble)** Allen Institute for Artificial Intelligence	1/17/2018	**54.5**

Machine Reading

Auto-Text to Knowledge

Source: DARPA, Machine Reading initiative

Open Information Extraction (2007)

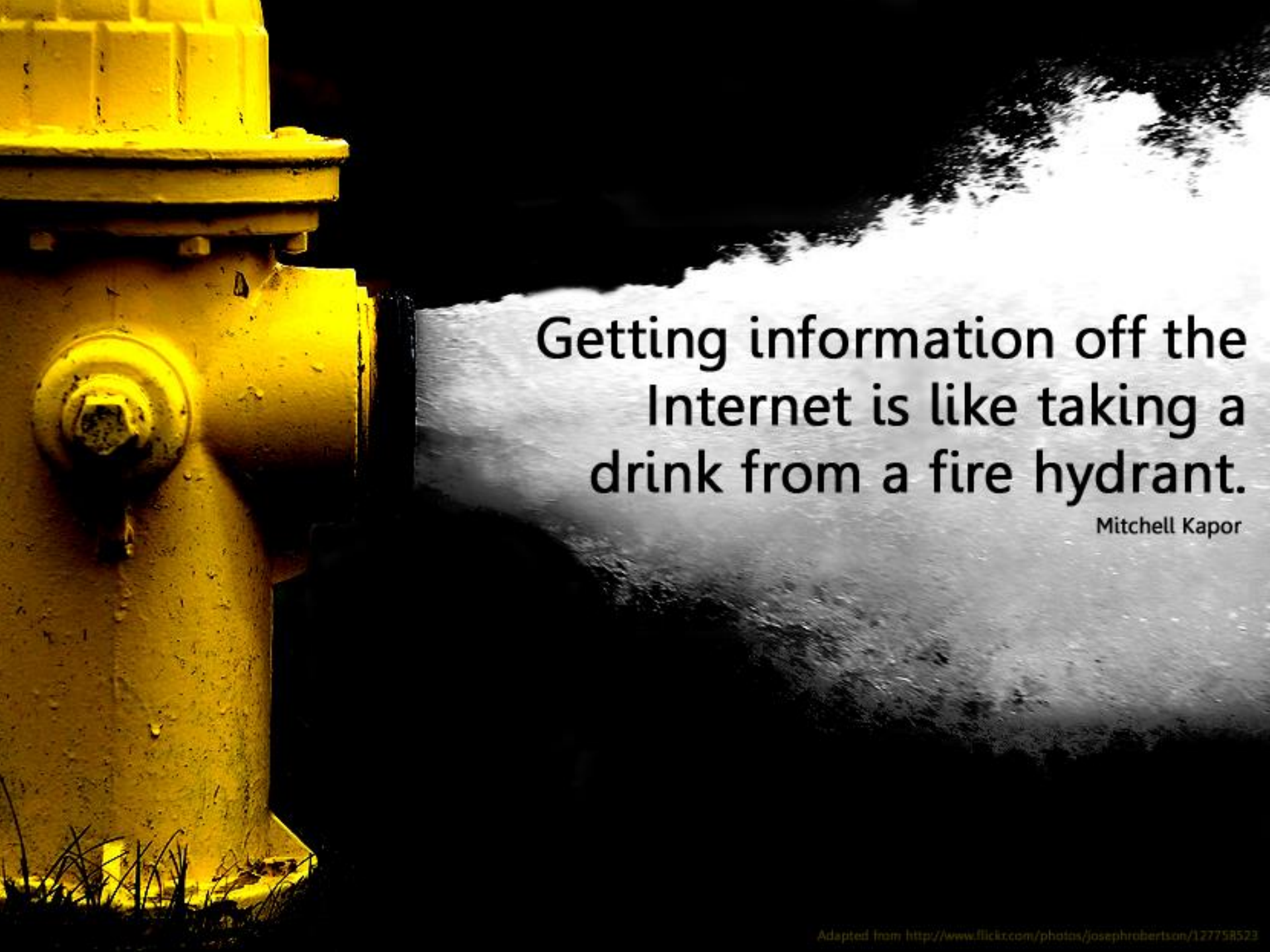
Question: can we leverage regularities in language to extract information in a relation-independent way?

Relations often:

- anchored in verbs.
- exhibit simple syntactic form

Virtues:

- Minimal hand-labeled data
- “No sentence left behind”
- Exploit redundancy & serendipity of Web
- Robust to parser errors


Getting information off the Internet is like taking a drink from a fire hydrant.
Mitchell Kapor
Adapted from http://www.flickr.com/photos/josephroberts/on/127758523

TextRunner

First Web-scale, Open IE system

(Banko, Cafarella, Etzioni *et al* IJCAI '07)

1,000,000,000 distinct extractions

Peak Precision = 0.9 (limited recall)

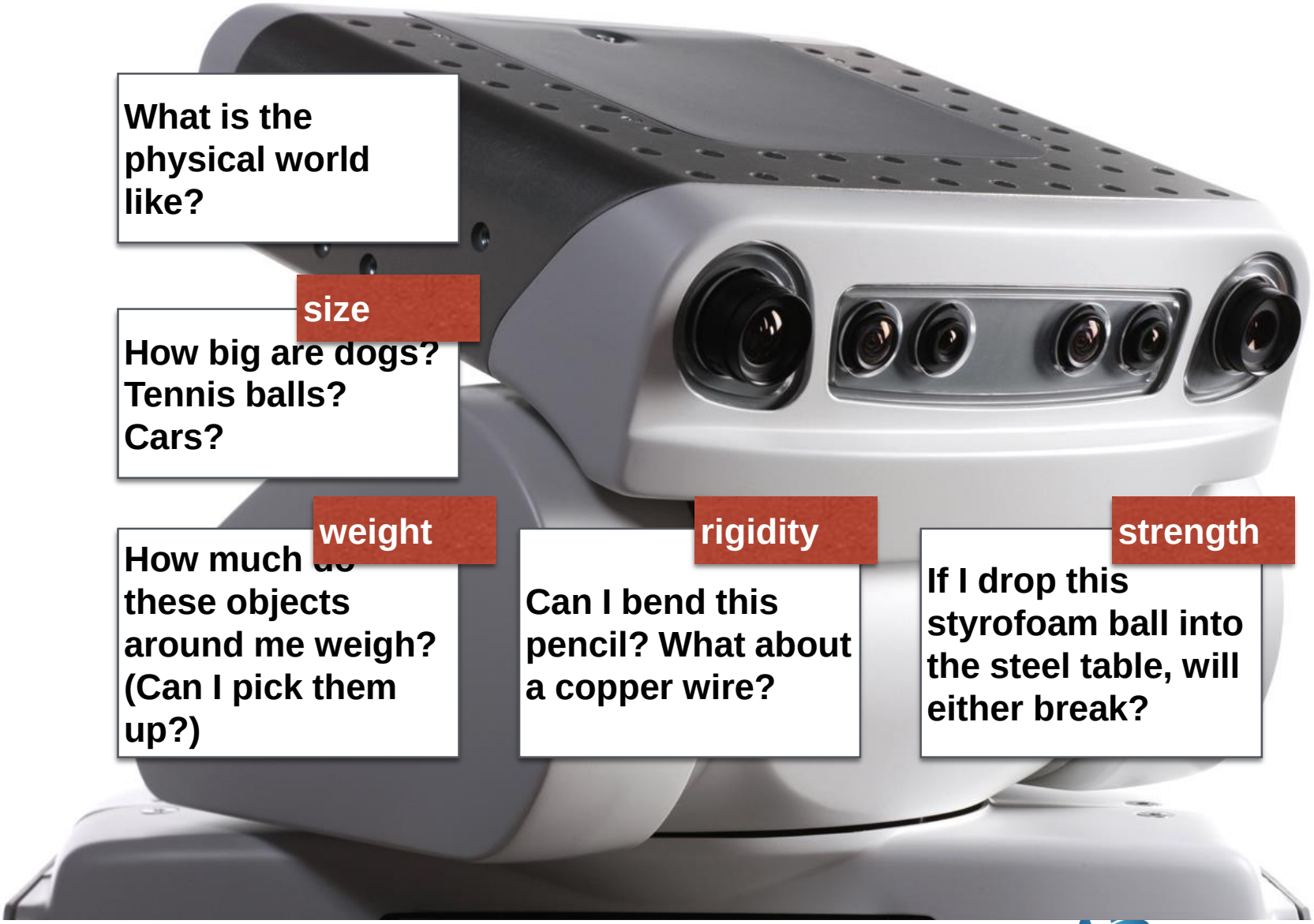
Openie.allenai.org

Verb Physics

relative physical knowledge about actions and objects

Maxwell Forbes et al. (ACL 2017)

Physical properties of objects



What is the physical world like?

size

How big are dogs? Tennis balls? Cars?

weight

How much do these objects around me weigh? (Can I pick them up?)

rigidity

Can I bend this pencil? What about a copper wire?

strength

If I drop this styrofoam ball into the steel table, will either break?

"I am larger than a chair"

~~"I am larger than a pen"~~

~~"I am larger than a stone"~~

~~"I am larger than a chair"~~

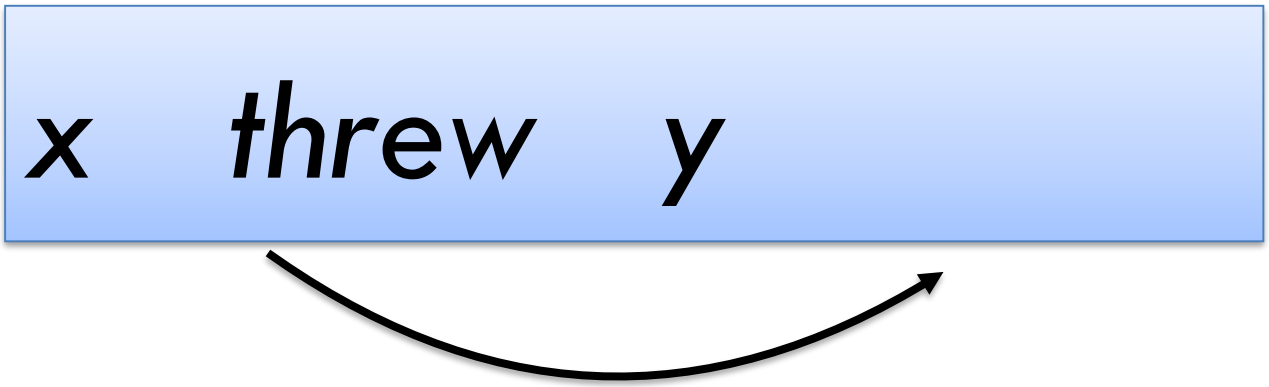
~~"I am larger than a ball"~~

~~"I am larger than a towel"~~

➔ **reporting bias**: people don't state the obvious

(Grice 1975, Van Durme 2010, Sorower et al., 2011, Mistra et al., 2016)

Inference to Overcome Reporting Bias



x is bigger than y

x weighs more than y

as a result, y will be moving faster than x

Vision as Basis for Common Sense (Imsitu.org)

Users/orene/Desktop/situations.pdf

Situation Recognition: Visual Semantic Role Labeling for Image Understanding

Mark Yatskar[1], Luke Zettlemoyer[1], Ali Farhadi[1,2]
[1]Computer Science & Engineering, University of Washington, Seattle, WA
[2]Allen Institute for Artificial Intelligence (AI2), Seattle, WA
[my89, lsz, ali]@cs.washington.edu

CLIPPING

ROLE	VALUE	ROLE	VALUE
AGENT	MAN	AGENT	VET
SOURCE	SHEEP	SOURCE	DOG
TOOL	SHEARS	TOOL	CLIPPER
ITEM	WOOL	ITEM	CLAW
PLACE	FIELD	PLACE	ROOM

JUMPING

ROLE	VALUE	ROLE	VALUE
AGENT	BOY	AGENT	BEAR
SOURCE	CLIFF	SOURCE	ICEBERG
OBSTACLE	-	OBSTACLE	WATER
DESTINATION	WATER	DESTINATION	ICEBERG
PLACE	LAKE	PLACE	OUTDOOR

SPRAYING

ROLE	VALUE	ROLE	VALUE
AGENT	MAN	AGENT	FIREMAN
SOURCE	SPRAY CAN	SOURCE	HOSE
SUBSTANCE	PAINT	SUBSTANCE	WATER
DESTINATION	WALL	DESTINATION	FIRE
PLACE	ALLEYWAY	PLACE	OUTSIDE

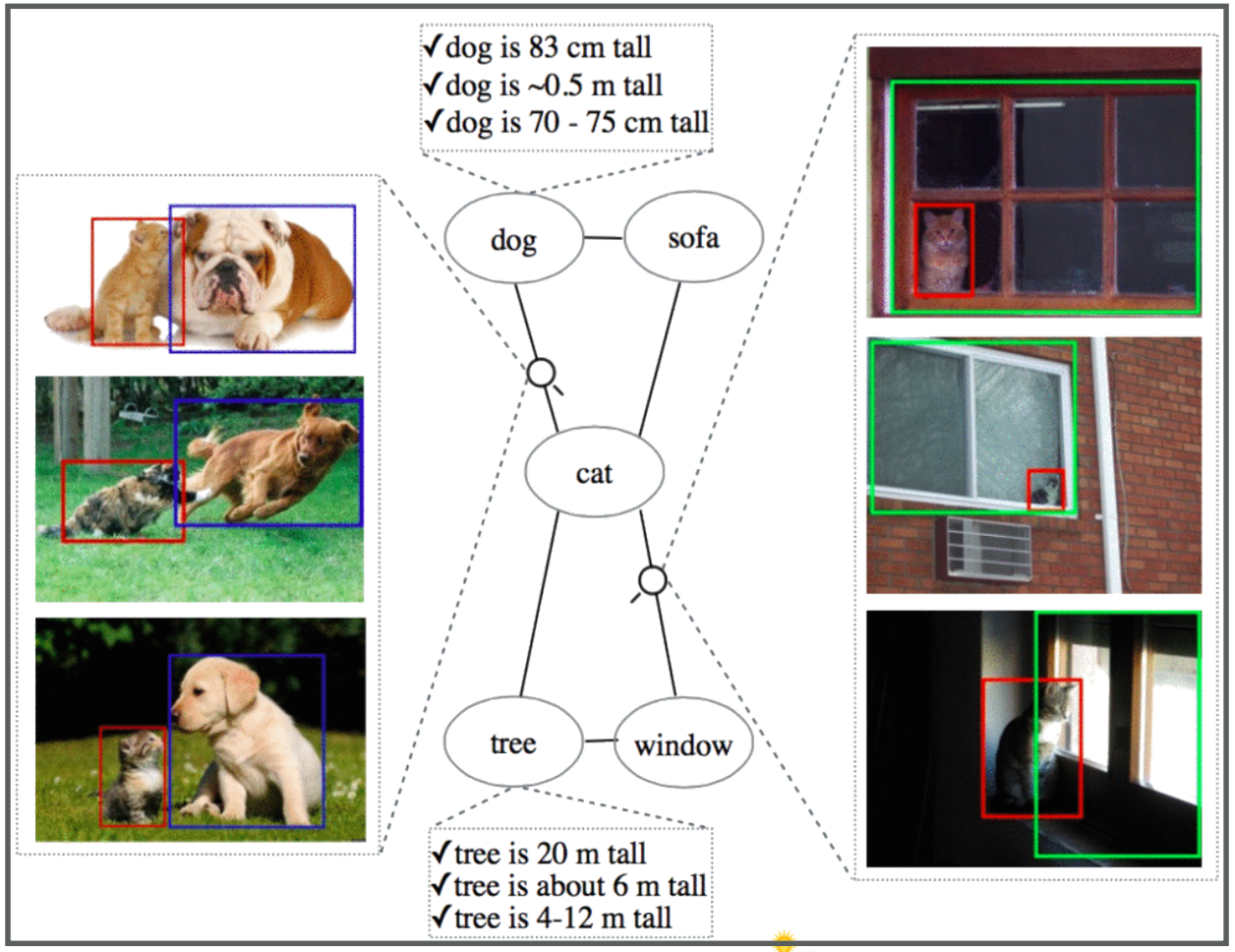
Figure 1. Six images that depict situations where actors, objects, substances, and locations play roles in an activity. Below each image is a *realized frame* that summarizes the situation: the left columns (blue) list activity-specific roles (derived from FrameNet, a broad coverage verb lexicon) while the right columns (green) list values (from ImageNet) for each role. Three different activities are shown, highlighting that visual properties can vary widely between role values (e.g., clipping a sheep's wool looks very different from clipping a dog's nails).

Are Elephants Bigger than Butterflies?

(Bagherinezhad et al. @ AAAI 2016)

- **Language – absolute estimation**
  - **“car is * x * m”**
  - **“person is * m tall”**
- **Vision – relative estimation**

$$\frac{size(O_i)}{size(O_j)} = \frac{area(box_1)}{area(box_2)} \cdot \frac{depth(box_1)^2}{depth(box_2)^2}$$



Commonsense from the Crowd

Desiderata for designing crowdsourcing procedures

- Scalability: Is it economically viable?
- Class distribution: truth fraction of interesting facts?

- Difficulty: Are the collected facts easily derived from simple data-driven methods (e.g., Google, PMI, LM)?

- Coverage: Can the procedure combat the reporting bias and collect facts comprehensively?

Crowdsourcing Commonsense with 20 Questions

Input: Object (or event)

Output: 20 ‘natural’ (Q, A)

- Setup:
  - Person A: I have an object

Does it bark?

  - Person B: Ask yes or no questions or guess what the object is
  - Goal: identify the object in fewer than 20 questions

It’s a dog!

Conclusions

- Common sense is critical for AI
- Progress has been limited, despite DL successes
- We need a benchmark and crisp metrics
- We have some new ideas on acquisition
- We're launching Project Alexandria

ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE