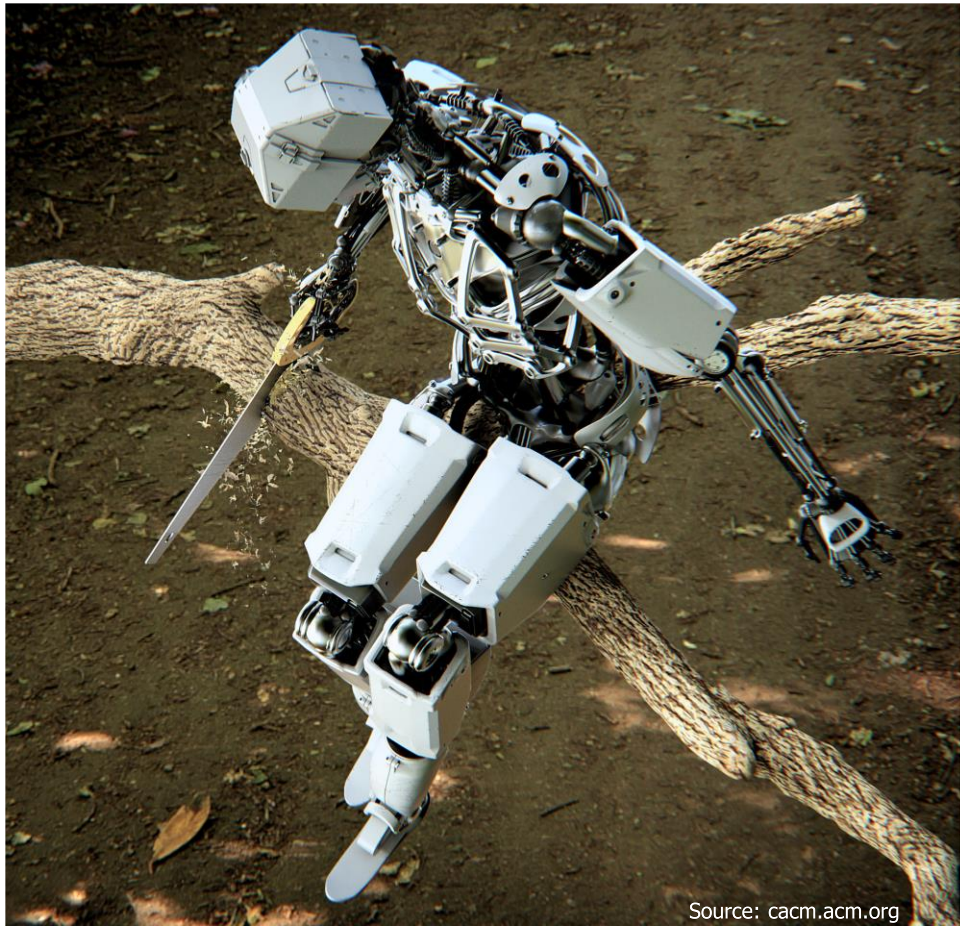


Source: cacm.acm.org

Ernest Davis and Gary Marcus. 2015. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun. ACM* 58, 9 (August 2015), 92-103.

Machine Common Sense

David Gunning
DARPA/I2O

Proposers Day

October 18, 2018

Approved for Public Release, Distribution Unlimited

Agenda

Start	End	Item
8:00 AM	9:00AM	Registration
9:00 AM	9:05 AM	Security Leon Kates, Program Security Representative, DARPA SID
9:05 AM	10:10 AM	Machine Common Sense (MCS) Dave Gunning, Program Manager, DARPA I2O
10:10 AM	10:30 AM	Contracts Mark Jones, Contracting Officer, DARPA CMO
10:30 AM	11:30 AM	Break
11:30 AM	1:00 PM	Q&A Session (in-person and webcast)

Please email your questions to mcs@darpa.mil

MCS BAA Outline

Funding Opportunity Description

A. Introduction/Background

B. Program Description/Scope

C. Technical Areas (TAs)

  TA1: Foundations of Human Common Sense

  TA2: Test Environment for the Foundations of Human Common Sense

  TA3: Broad Common Knowledge

D. Schedule/Milestones

E. TA-specific Deliverables

F. Government-furnished Property/Equipment/Information

G. Intellectual Property

DoD Funding Categories

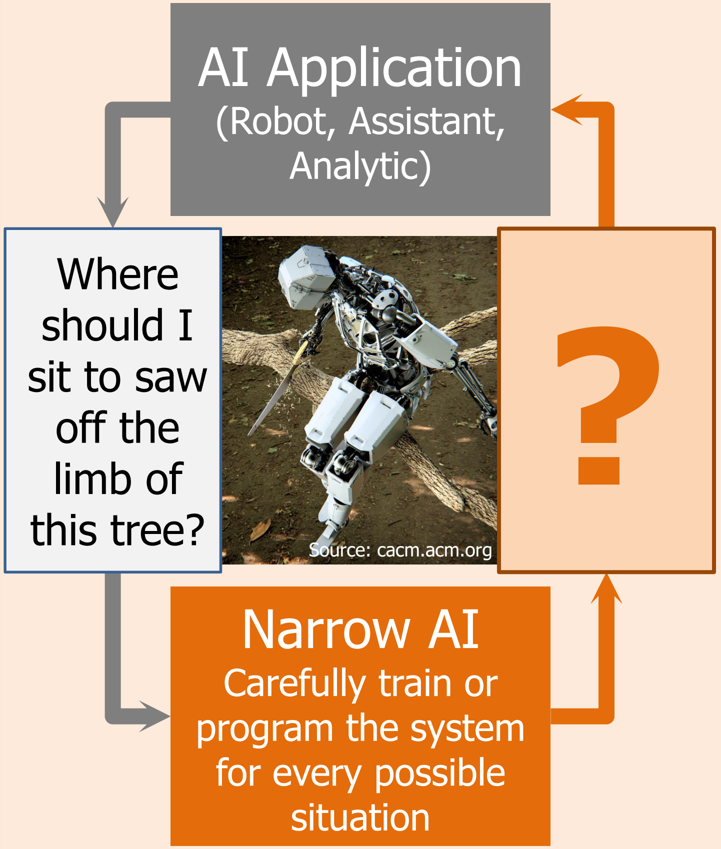
MCS →

Category	Definition
Basic Research (6.1)	Systematic study directed toward greater knowledge or understanding of the fundamental aspects of phenomena and/or observable facts without specific applications in mind.
Applied Research (6.2)	Systematic study to gain knowledge or understanding necessary to determine the means by which a recognized and specific need may be met.
Technology Development (6.3)	Includes all efforts that have moved into the development and integration of hardware (and software) for field experiments and tests.

What are we trying to do?

TODAY

Narrow
Artificial Intelligence



AI Application
(Robot, Assistant, Analytic)

Where should I sit to saw off the limb of this tree?

Source: cacm.acm.org

?

Narrow AI
Carefully train or program the system for every possible situation

TOMORROW

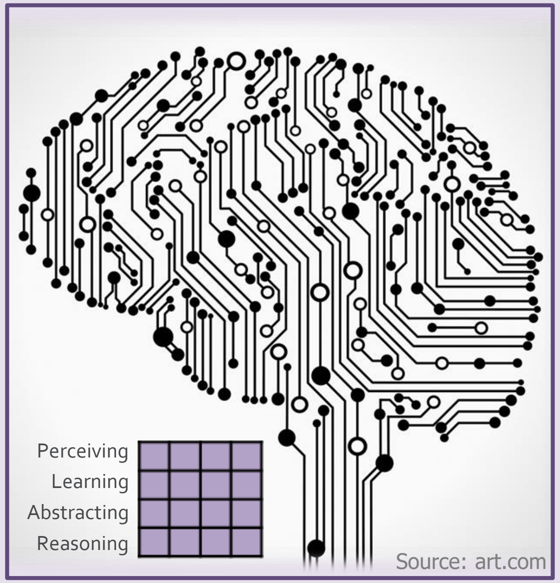
Machine
Common Sense

Source: magazine.owen.vanderbilt.edu

The elephant in the room

FUTURE

General
Artificial Intelligence



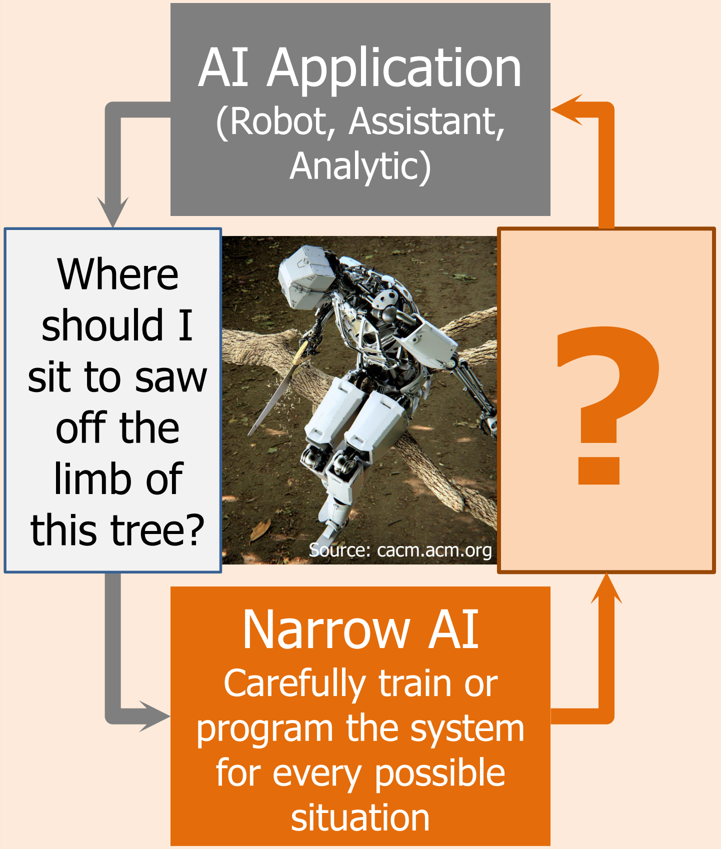
Source: art.com

Human-Level AI

What are we trying to do?

TODAY

Narrow
Artificial Intelligence



AI Application
(Robot, Assistant, Analytic)

Where should I sit to saw off the limb of this tree?

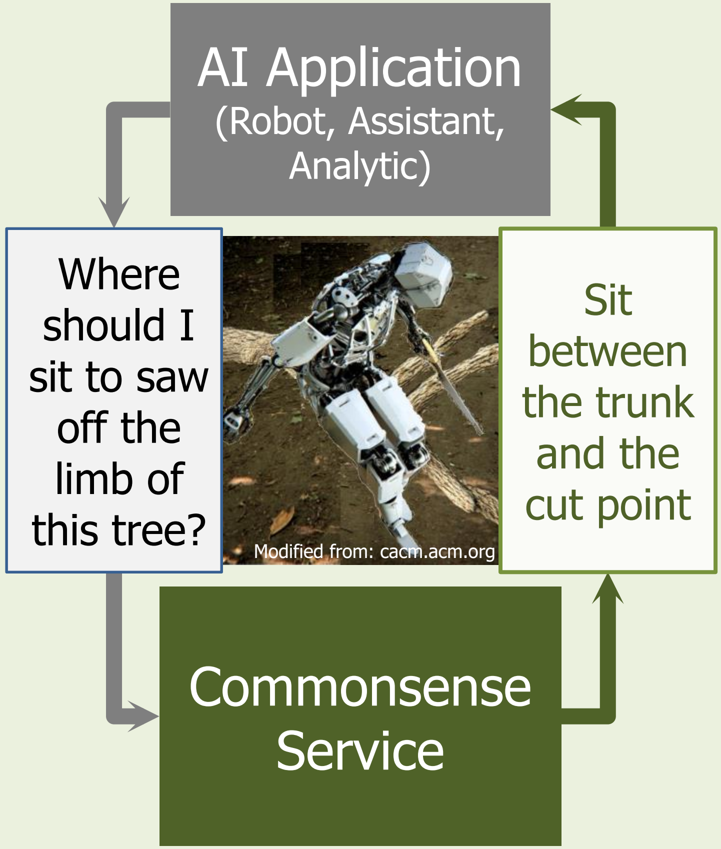
Source: cacm.acm.org

?

Narrow AI
Carefully train or program the system for every possible situation

TOMORROW

Machine
Common Sense



AI Application
(Robot, Assistant, Analytic)

Where should I sit to saw off the limb of this tree?

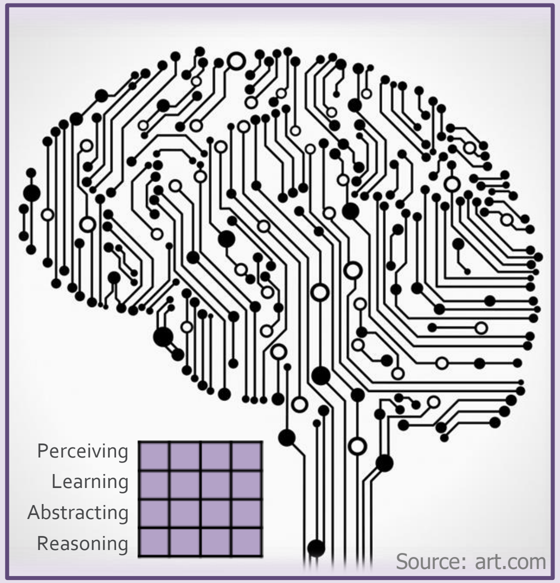
Modified from: cacm.acm.org

Sit between the trunk and the cut point

Commonsense Service

FUTURE

General
Artificial Intelligence



Source: art.com

Human-Level AI

Approved for Public Release, Distribution Unlimited

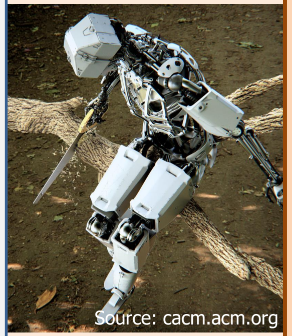
What are we trying to do?

TODAY

Narrow
Artificial Intelligence

AI Application
(Robot, Assistant, Analytic)

Where should I sit to saw off the limb of this tree?



Source: cacm.acm.org

?

TOMORROW

Machine
Common Sense

AI Application
(Robot, Assistant, Analytic)

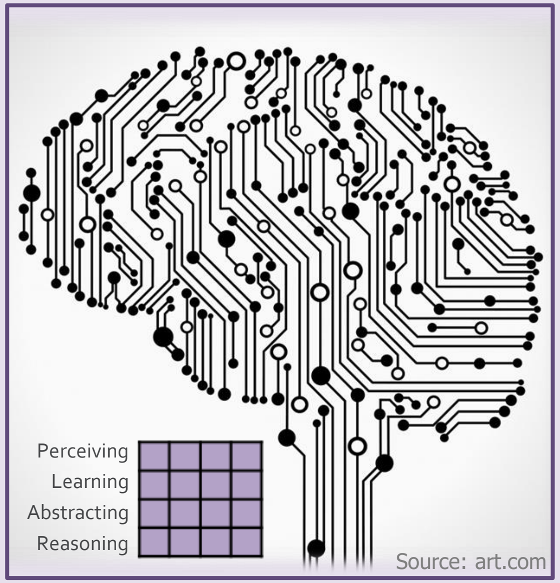
Where should I sit to saw off the limb of this tree?

Modified from: cacm.acm.org

Sit between the trunk and the cut point

FUTURE

General
Artificial Intelligence



Source: art.com

MCS will create the computing foundations needed to develop machine commonsense services to enable AI applications to understand new situations, monitor the reasonableness of their actions, communicate more effectively with people, and transfer learning to new domains.

What is Common Sense?

Examples:

- Which of these would fit through a doorway?

Source: AI2

- If I put my socks in the drawer, will they still be there tomorrow?
- Which object is flying and which is stationary in this sentence?

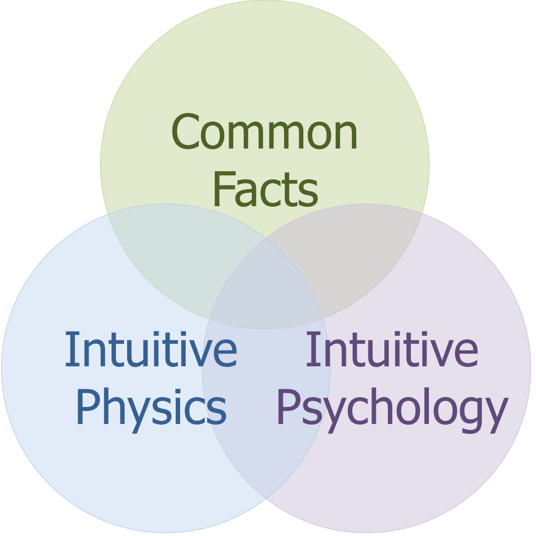
I saw the Grand Canyon flying to Los Angeles.

Wikipedia:
The basic ability to perceive, understand, and judge things that are shared by ("common to") nearly all people and can reasonably be expected of nearly all people without need for debate.

John McCarthy (Stanford, circa 1960):

Source: amturing.acm.org

$\exists a.\ \mathit{Name}(a) = \mathit{ANY\text{-}FOOL}$
$\forall k.\ \mathit{Knows}(\mathit{ANY\text{-}FOOL}, k) \Leftrightarrow \forall p \in \mathit{Persons}.\ \mathit{Knows}(p, k)$
$\forall k.\ \mathit{Commonsense}(k) \Leftrightarrow \mathit{Knows}(\mathit{ANY\text{-}FOOL}, k)$



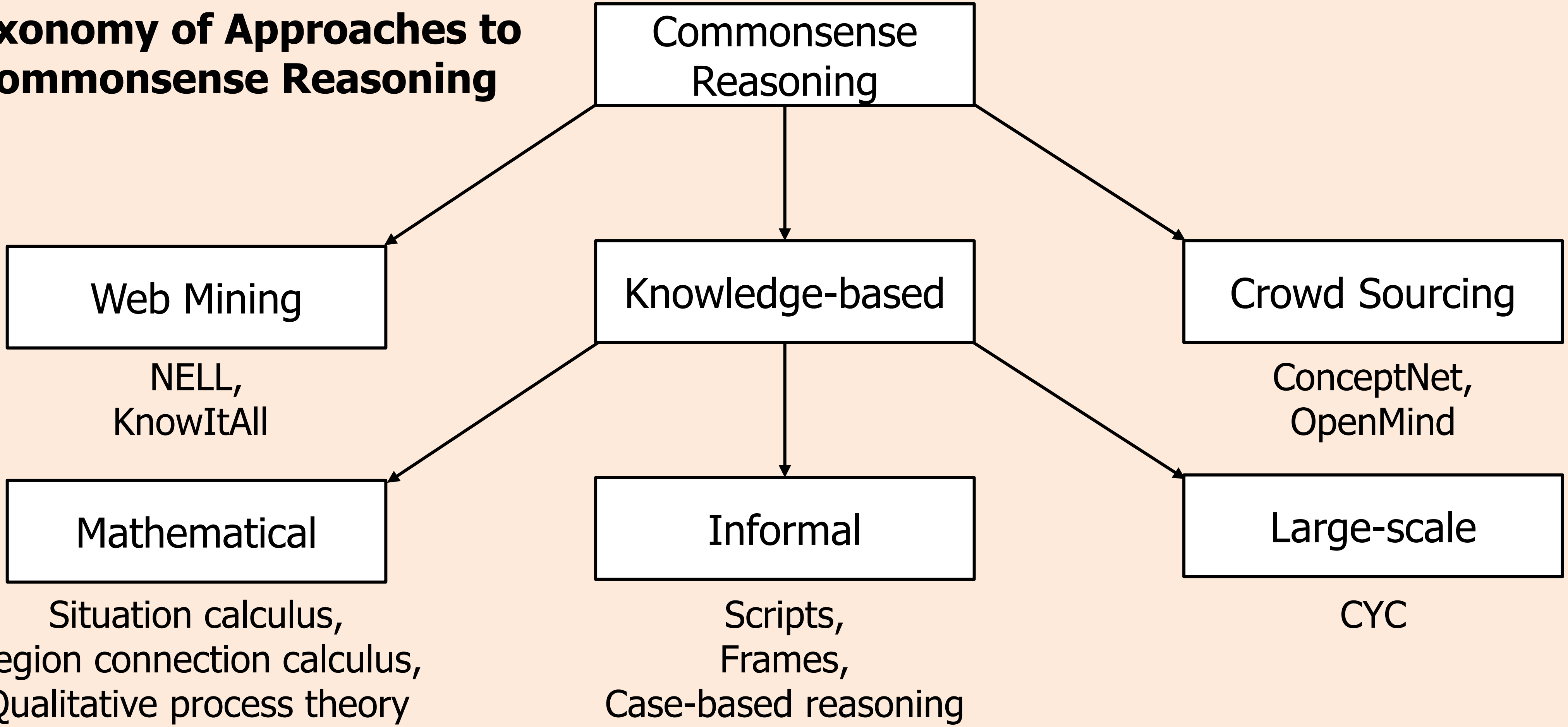
Core Domains of Human Cognition:

Source: scholar.harvard.edu
Elizabeth Spelke (Harvard)

- Objects
- Agents
- Places
- Number
- Forms
- Social Beings

How is it done today?

Taxonomy of Approaches to Commonsense Reasoning



Commonsense Reasoning

Web Mining

NELL,
KnowItAll

Knowledge-based

Crowd Sourcing

ConceptNet,
OpenMind

Mathematical

Situation calculus,
Region connection calculus,
Qualitative process theory

Informal

Scripts,
Frames,
Case-based reasoning

Large-scale

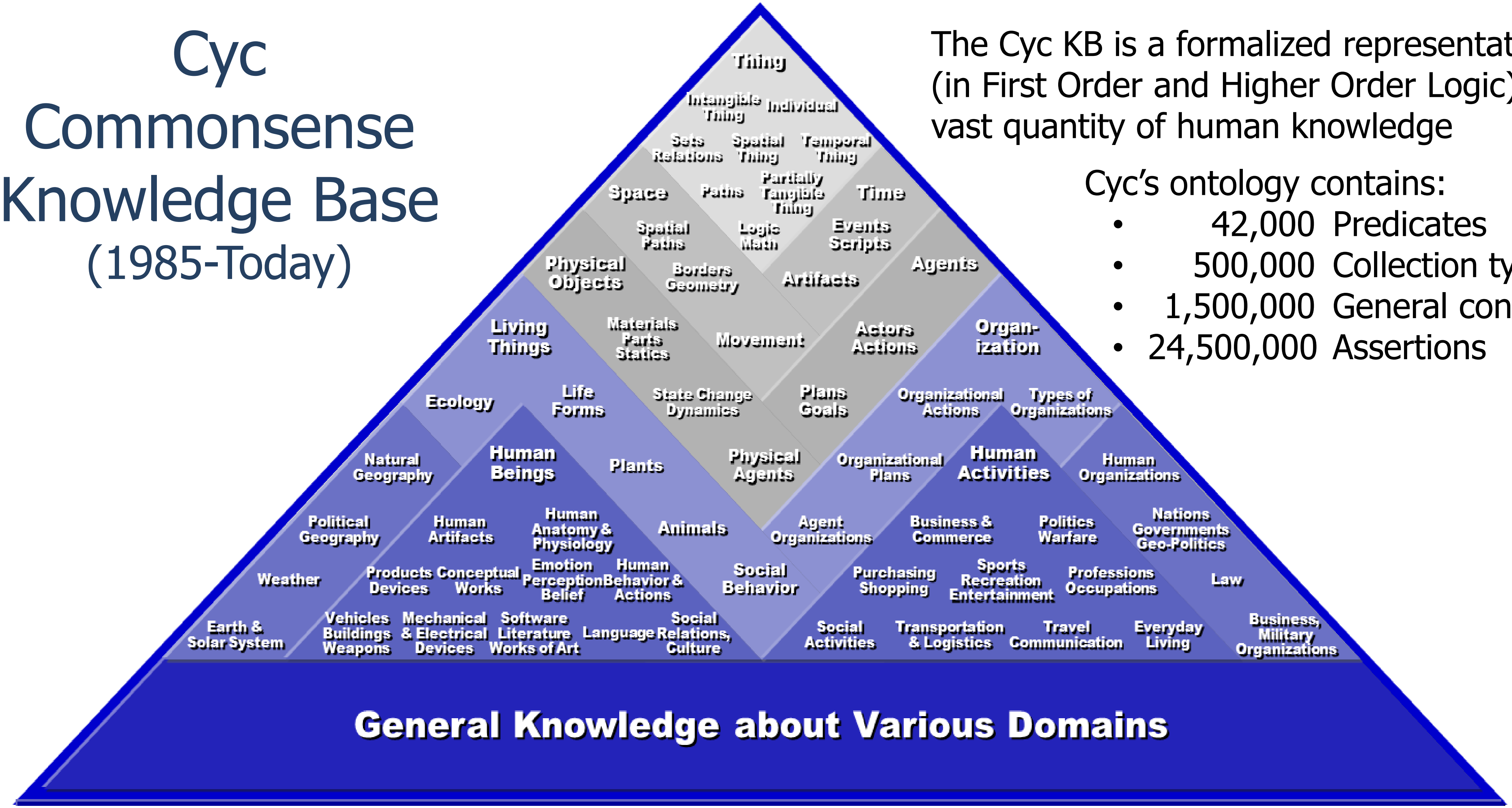
CYC

Ernest Davis and Gary Marcus. 2015. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM* 58, 9 (August 2015), 92-103. DOI: https://doi.org/10.1145/2701413

How is it done today?

Cyc Commonsense Knowledge Base

(1985-Today)



The Cyc KB is a formalized representation (in First Order and Higher Order Logic) of a vast quantity of human knowledge

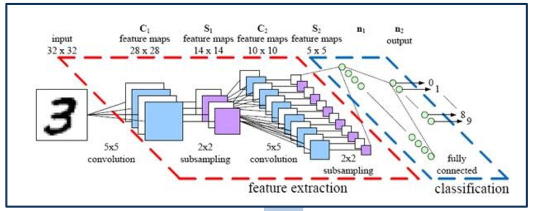
Cyc's ontology contains:

- 42,000 Predicates
- 500,000 Collection types
- 1,500,000 General concepts
- 24,500,000 Assertions

Source: Dr. Doug Lenat, Cycorp, cyc.com

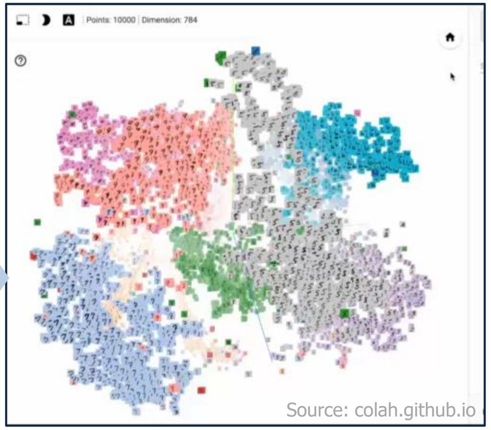
What is new in your approach?

Learning Grounded Representations



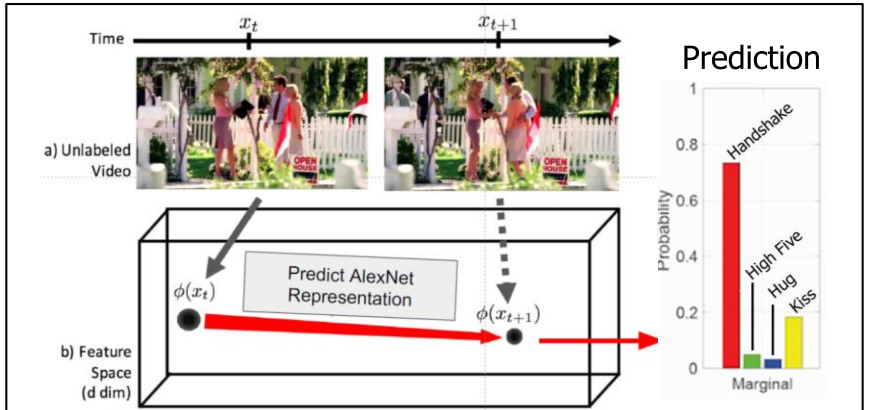
Source: medium.com

Vector-based “embeddings” extracted from hidden layers



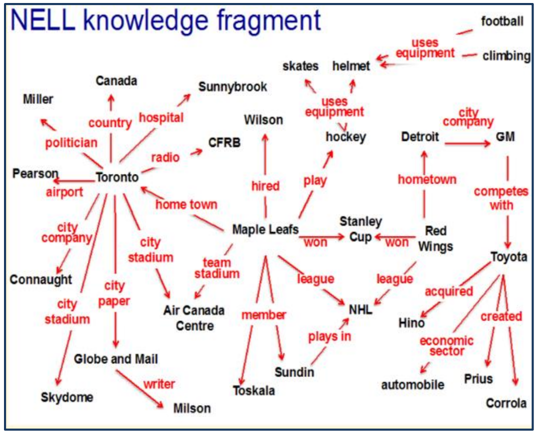
Source: colah.github.io

Learning Predictive Models from Experience



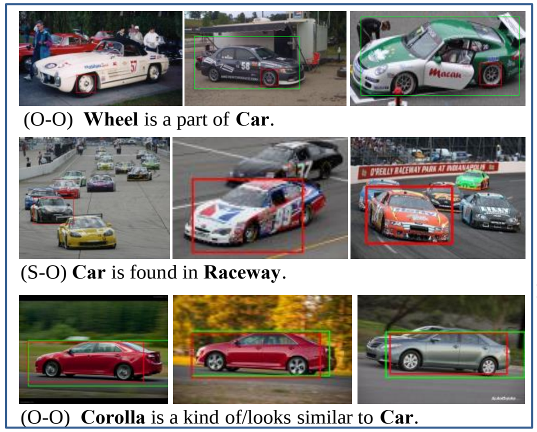
Vondrick et al., 2016. Anticipating visual representations from unlabeled video

Learning Commonsense Knowledge from the Web



Mitchell et al., 2018. Never-Ending Learning

Never Ending Language Learning (NELL)



(O-O) **Wheel** is a part of **Car**.

(S-O) **Car** is found in **Raceway**.

(O-O) **Corolla** is a kind of/looks similar to **Car**.

Source: Dr. Abhinav Gupta, CMU

Never Ending Image Learning (NEIL)

Understanding & Modeling Childhood Cognition

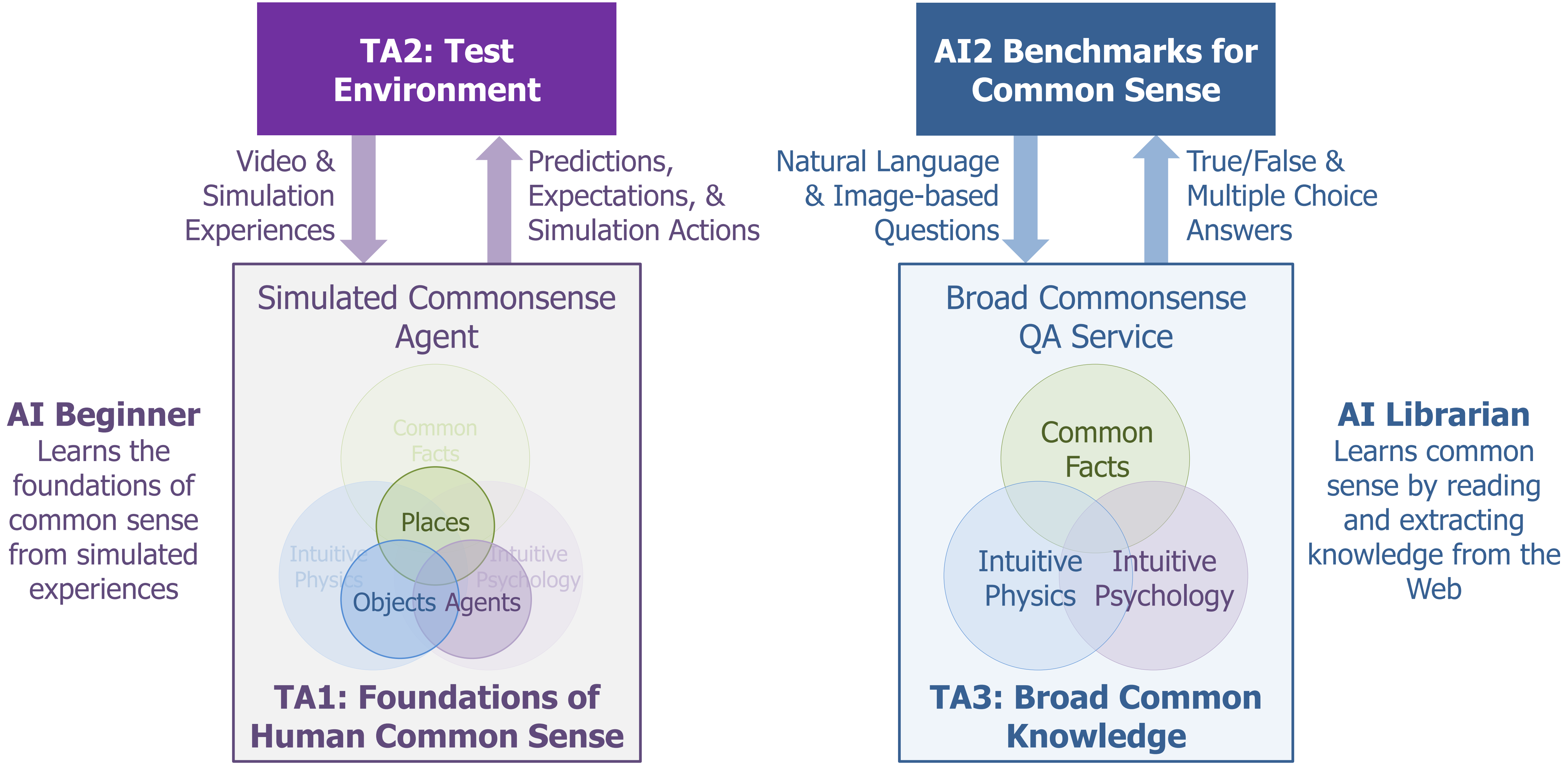
Source: scholar.harvard.edu

Elizabeth Spelke (Harvard)

- Objects
- Agents
- Places
- Number
- Forms
- Social Beings

Core Domains of Child Cognition

Program Approach



TA2: Test Environment

Video & Simulation Experiences

Predictions, Expectations, & Simulation Actions

Simulated Commonsense Agent

Common Facts

Places

Intuitive Physics

Intuitive Psychology

Objects

Agents

TA1: Foundations of Human Common Sense

AI Beginner
Learns the foundations of common sense from simulated experiences

AI2 Benchmarks for Common Sense

Natural Language & Image-based Questions

True/False & Multiple Choice Answers

Broad Commonsense QA Service

Common Facts

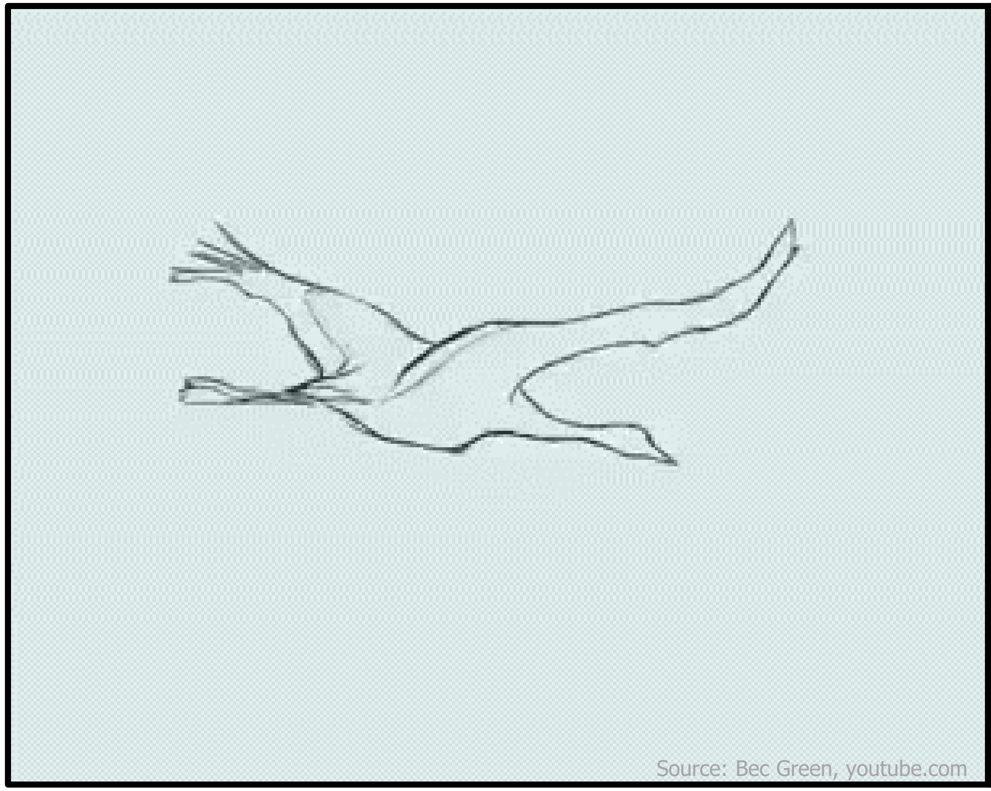
Intuitive Physics

Intuitive Psychology

TA3: Broad Common Knowledge

AI Librarian
Learns common sense by reading and extracting knowledge from the Web

Did the Wright Flyer need to fly like a bird?



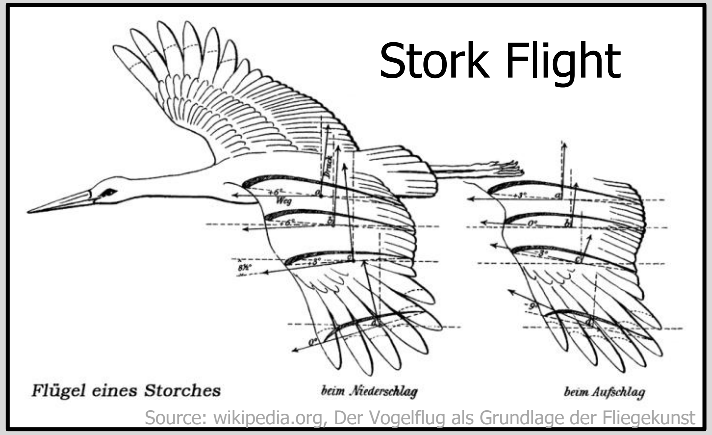
Source: Bec Green, youtube.com

Stork in Flight

Source: youtube.com

Wright Flyer, 1903

Otto Lilienthal



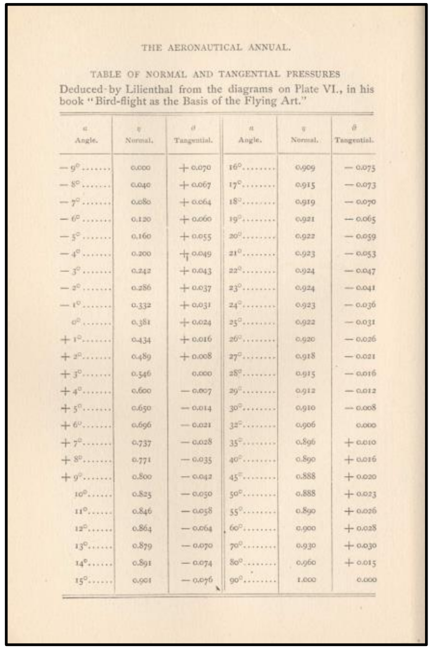
Source: wikipedia.org, Der Vogelflug als Grundlage der Fliegekunst

Der Vogelflug als Grundlage der Fliegekunst (Bird flight as the basis for flying art), 1882

Source: wikipedia.org

Source: wikipedia.org

Otto Lilienthal in mid-flight (first successful glider, 1895)



THE AERONAUTICAL ANNUAL.

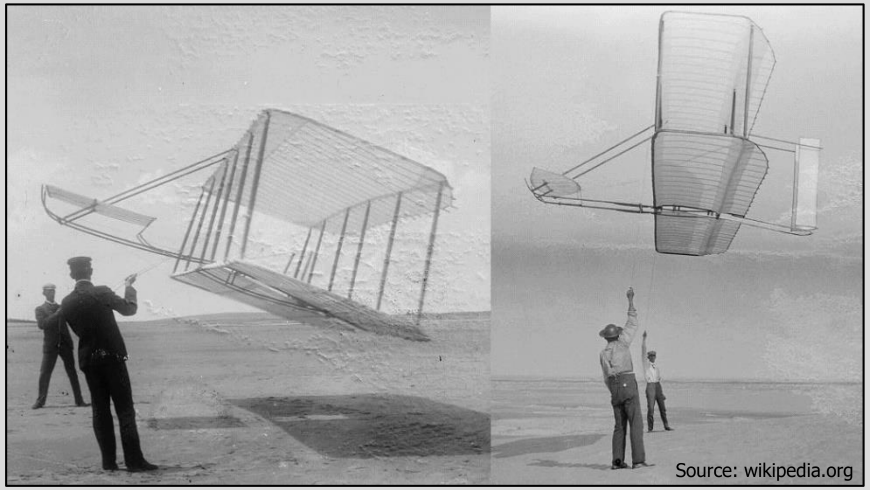
TABLE OF NORMAL AND TANGENTIAL PRESSURES

Deduced by Lilienthal from the diagrams on Plate VI., in his book "Bird-flight as the Basis of the Flying Art."

α Angle.	η Normal.	θ Tangential.	α Angle.	η Normal.	θ Tangential.
— 9°	0.000	+ 0.070	16°	0.909	— 0.075
— 8°	0.040	+ 0.067	17°	0.915	— 0.073
— 7°	0.080	+ 0.064	18°	0.919	— 0.070
— 6°	0.120	+ 0.060	19°	0.921	— 0.065
— 5°	0.160	+ 0.055	20°	0.922	— 0.059
— 4°	0.200	+ 0.049	21°	0.923	— 0.053
— 3°	0.242	+ 0.043	22°	0.924	— 0.047
— 2°	0.286	+ 0.037	23°	0.924	— 0.041
— 1°	0.332	+ 0.031	24°	0.923	— 0.036
0°	0.381	+ 0.024	25°	0.922	— 0.031
+ 1°	0.434	+ 0.016	26°	0.920	— 0.026
+ 2°	0.489	+ 0.008	27°	0.918	— 0.021
+ 3°	0.546	0.000	28°	0.915	— 0.016
+ 4°	0.600	— 0.007	29°	0.912	— 0.012
+ 5°	0.650	— 0.014	30°	0.910	— 0.008
+ 6°	0.696	— 0.021	32°	0.906	0.000
+ 7°	0.737	— 0.028	35°	0.896	+ 0.010
+ 8°	0.771	— 0.035	40°	0.890	+ 0.016
+ 9°	0.800	— 0.042	45°	0.888	+ 0.020
10°	0.825	— 0.050	50°	0.888	+ 0.023
11°	0.846	— 0.058	55°	0.890	+ 0.026
12°	0.864	— 0.064	60°	0.900	+ 0.028
13°	0.879	— 0.070	70°	0.930	+ 0.030
14°	0.891	— 0.074	80°	0.960	+ 0.015
15°	0.901	— 0.076	90°	1.000	0.000

Source: airandspace.si.edu

Lilienthal's tables of lift and drag



Source: wikipedia.org

At left, 1901 glider flown by Wilbur and Orville (using Lilienthal's original lift and drag tables) exhibiting a steep angle of attack due to poor lift and high drag. At right, 1902 glider (after correcting Lilienthal's coefficients) showing dramatic improvement in performance.

"Lilienthal was without question the greatest of the precursors, and the world owes to him a great debt." – Wilbur Wright, 1912

Core Domains of Child Cognition

Source: scholar.harvard.edu

Elizabeth Spelke (Harvard)

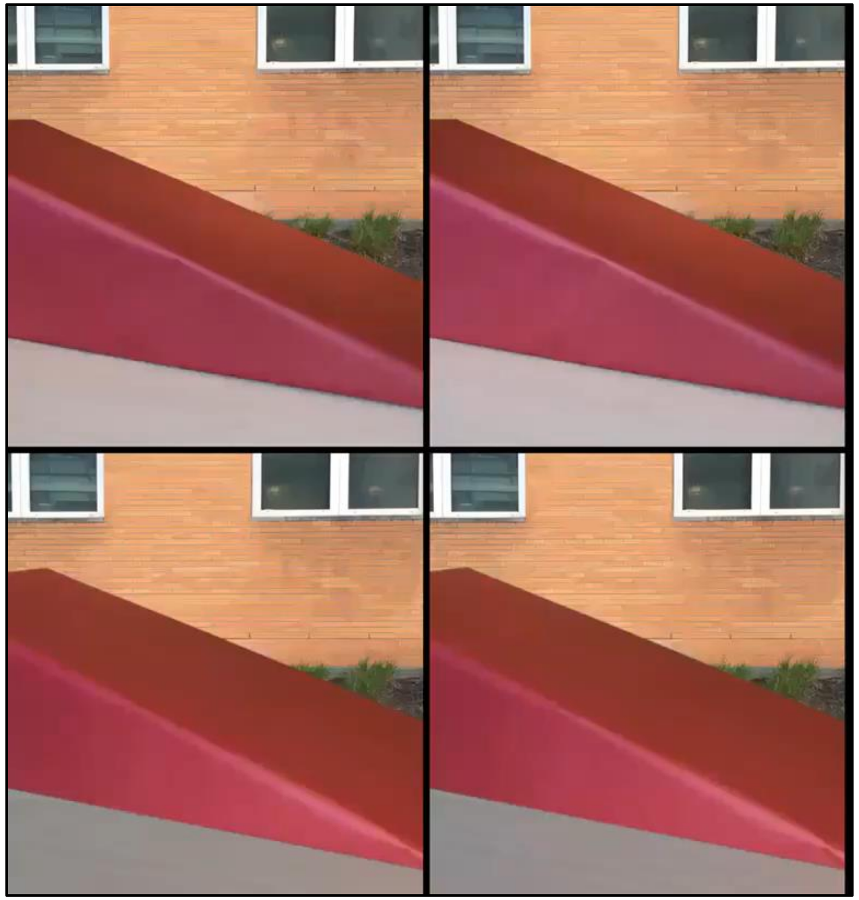
Director of the Harvard Laboratory for Developmental Studies. Since the 1980s, she has carried out experiments to test the cognitive faculties of children and formulate her theories of child cognition.

Domain	Description
Objects	supports reasoning about objects and the laws of physics that govern them
Agents	supports reasoning about agents that act autonomously to pursue goals
Places	supports navigation and spatial reasoning around an environment
Number	supports reasoning about quantity and how many things are present
Forms	supports representation of shapes and their affordances
Social Beings	supports reasoning about Theory of Mind and social interactions

Approved for Public Release, Distribution Unlimited

Understanding the Foundations of Human Cognition

Lookit: the online child lab, MIT Early Childhood Cognition Lab



Stimuli

Response

"Your baby, the physicist" study: https://lookit.mit.edu/studies/cfddb63f-12e9-4e62-abd1-47534d6c4dd2/

Approved for Public Release, Distribution Unlimited

Cognitive Development Milestones (for children 0-18 months old)

		Months																		
	Core Domains and Milestones	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
OBJECTS	**Innate Core Objects Domain**	○	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	▸
	Objects have depth >2D & move in 2.5D or 3D space			■	■	■														
	Objects move separately from one another except on contact			■	■	■														
	Objects change motion on contact/don't pass thru one another			■	■	■														
	Objects persist & can be tracked briefly over occlusion			■	■	■														
	Unseen objects can cause visible outcomes			■	■	■														
	Occluded objects are the same if their visible surfaces align							■	■	■										
	Objects are connected if their visible parts move together							■	■	■										
	Objects belong to kinds with distinctive forms & functions													■	■	■	■	■	■	■
	Learn object labels from language											■	■	■	■	■				
	Objects fall if not supported under center of mass											■	■	■	■	■				
AGENTS	**Innate Core Agents Domain**	○	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	▸
	Agents can change object motion				■	■	■													
	Agents have goals				■	■	■													
	Agents act efficiently				■	■	■													
	Unseen agents can cause visible outcomes										■	■	■	■						
	Judge amount of effort agent expends to cause outcomes										■	■	■	■						
	Infer if agents help/hinder & adjust their +/- behavior										■	■	■	■						
	Learn which agents to reach for										■	■	■	■						
	Understand what agents can or cannot see										■	■	■	■						
	Understand simple dispositions and preferences of agents													■	■	■	■			
	Anticipate actions based on credibility of referential behavior																	■		
	Help others in response to salient cues of need															■	■	■	■	■
PLACES	**Innate Core Places Domain**	○	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	▸
	Observed agents navigate to goal by path of least resistance				■	■	■													
	Keep track of location (active, self-guided locomotion)											■	■	■	■	■	■	■	■	■
	Ability to learn the layout of their environment											■	■	■	■	■	■	■	■	■
	Navigate by encoding distances/directions of stable surfaces											■	■	■	■	■	■	■	■	■

Foundations of Human Common Sense

Objects

Source: medium.com

Agents

Source: medium.com

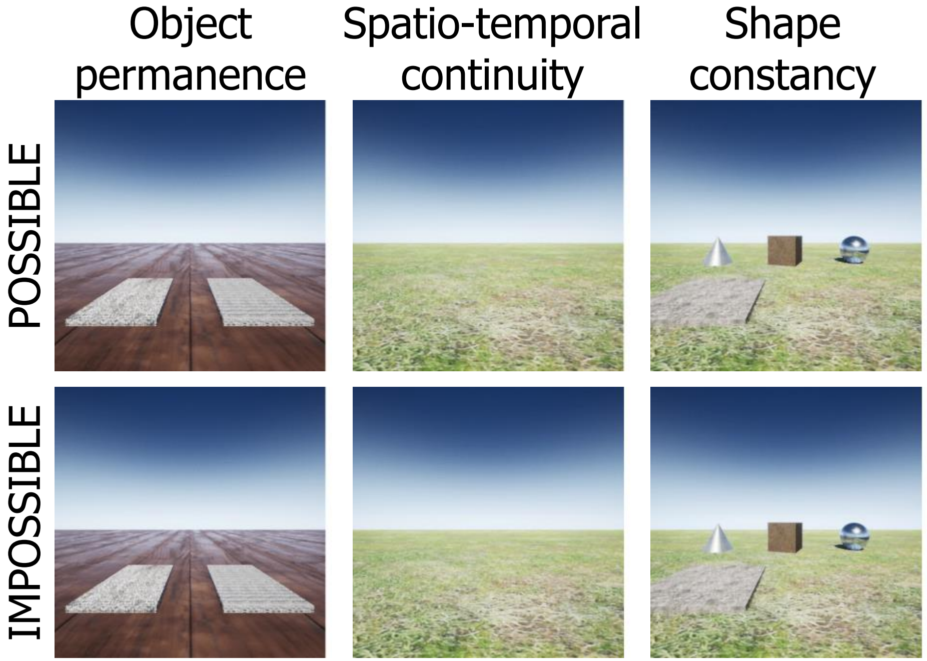
Infant cognition for Objects and Agents. These core domains likely form the fundamental building blocks of human intelligence and common sense, especially the core domains of objects (intuitive physics), agents (intentional actors), and places (spatial navigation). For example, the core domain of objects not only provides the fundamental concepts for understanding the physical world, but also provides the foundation for understanding causality. The core domain of agents not only provides the fundamental concepts for understanding intentional actors and Theory of Mind (TOM), but also provides the foundation for dealing with the “frame problem” in AI (i.e., knowing that objects in a scene only change if acted on by an agent).

Approved for Public Release, Distribution Unlimited

Courtesy: Dave Gunning

Courtesy: Dave Gunning

TA2 Foundations Test Environment (Examples)



Source: coml.lscp.ens.fr/intphys/ (Emmanuel Dupoux, Mathieu Bernard, Ronan Riochet)

Core principle	Milestone
Objects don't pop in and out of existence	5 months
Object trajectories are continuous	4 months
Objects keep their shapes	10 months

Source: AI2

Core principle	Milestone
Active, self-guided locomotion	10 months
Learn environment layout	10 months
Encode distances/directions of stable surfaces to navigate	10 months

TA2 Foundations Tests: Levels of Performance

❶ Prediction/Expectation

- The test environment will present the TA1 models with videos and simulation experiences of the type used to test child cognition for each cognitive milestone.
- The models will produce a expectation output (a measurable Violation of Expectation (VOE) signal) that will be used to determine if the model matches human cognitive performance by comparison to the VOE results observed in children.

❷ Experience Learning

- The test environment will present TA1 models with videos and simulation experiences in which a new object, agent, or place is introduced.
- The models will be tested to determine that they are able to learn the properties of the newly introduced item in a way that matches human cognitive performance.

❸ Problem Solving

- The test environment will present the TA1 models with videos and simulation experiences in which a problem solving task is introduced.
- The models will be tested to determine that they solve the problem in a way that matches human cognitive performance.

Approved for Public Release, Distribution Unlimited

Examples of Developmental Psychology Research & Scorecard

	Core Domains and Milestones	Months	VOE	Learn	Solve
		0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18			
OBJECTS	**Innate Core Objects Domain**	○······▸			
	Objects have depth >2D & move in 2.5D or 3D space	Termine et al. , 1987 [5] Spelke et al., 1989 [6]	✓	✓	
	Objects move separately from one another except on contact	Kellman & Spelke, 1983 [7]; Ball, 1973 [8]; Johnson & Aslin, 1995 [9]	✓	✓	
	Objects change motion on contact/don't pass thru one another	Baillargeon et al., 1985 [10]	✓		
	Objects persist & can be tracked briefly over occlusion	Feigenson & Carey, 2003 [11]; Aguiar & Baillargeon, 1999 [12]	✓		
	Unseen objects can cause visible outcomes	Saxe et al., 2005 [13]	✓		
	Occluded objects are the same if their visible surfaces align	Needham [14]			
	Objects are connected if their visible parts move together	Kellman et al., 1987 [15]			
	Objects belong to kinds with distinctive forms & functions	Xu [16]			
	Learn object labels from language	Xu [16]			
	Objects fall if not supported under center of mass	Baillargeon [17]			
AGENTS	**Innate Core Agents Domain**	○······▸			
	Agents can change object motion	Baillargeon & Luo [17][18]	✓	✓	
	Agents have goals	Woodward, 1999 [19]; Csibra, 2003 [20]	✓	✓	
	Agents act efficiently	Gergely & Csibra, 2013 [21]; Liu & Spelke, 2017 [22]	✓		
	Unseen agents can cause visible outcomes	Saxe et al., 2005 [13]	✓		
	Judge amount of effort agent expends to cause outcomes	Liu et al., 2017 [23]; Leonard et al., 2017 [24]	✓		
	Infer if agents help/hinder & adjust their +/- behavior	Hamlin [25]			
	Learn which agents to reach for	Hamlin [25]			
	Understand what agents can or cannot see	Hamlin et al., 2013 [26]			
	Understand simple dispositions and preferences of agents	Song et al., 2005 [27]; Sootsman & Woodward, 2007 [28]			
	Anticipate actions based on credibility of referential behavior	Poulin-Dubois & Chow, 2009 [29]			
	Help others in response to salient cues of need	Warneken [30]			
PLACES	**Innate Core Places Domain**	○······▸			
	Observed agents navigate to goal by path of least resistance	Gergely & Csibra [31][32]; Skerry [33]	✓	✓	
	Keep track of location (active, self-guided locomotion)	O'Keefe & Nadel [34]; Spelke & Lee, 2012 [35]			
	Ability to learn the layout of their environment	O'Keefe & Nadel [34]; Spelke & Lee, 2012 [35]			
	Navigate by encoding distances/directions of stable surfaces	Hermer [36]; Doeller & Burgess, 2008 [37]			

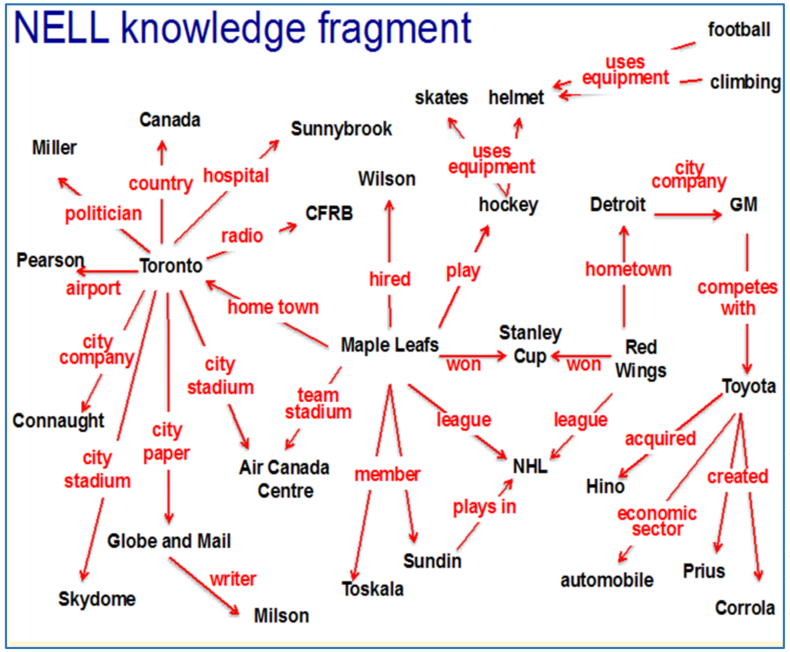
Approved for Public Release, Distribution Unlimited

TA1 Schedule and Scorecard Targets (Estimates)

	Year 1	Year 2	Year 3	Year 4
	Baseline	“Watch”	“Crawl”	“Walk”
OBJECTS				
Violation of Expectations	30%	50%	60%	80%
Experience Learning		30%	50%	80%
Problem Solving			30%	50%
AGENTS				
Violation of Expectations	30%	50%	60%	80%
Experience Learning		30%	50%	50%
Problem Solving			30%	50%
PLACES				
Violation of Expectations	30%	50%	60%	80%
Experience Learning		30%	50%	50%
Problem Solving			30%	50%

Learning Commonsense Knowledge from the Web

Never Ending Language Learning (NELL)

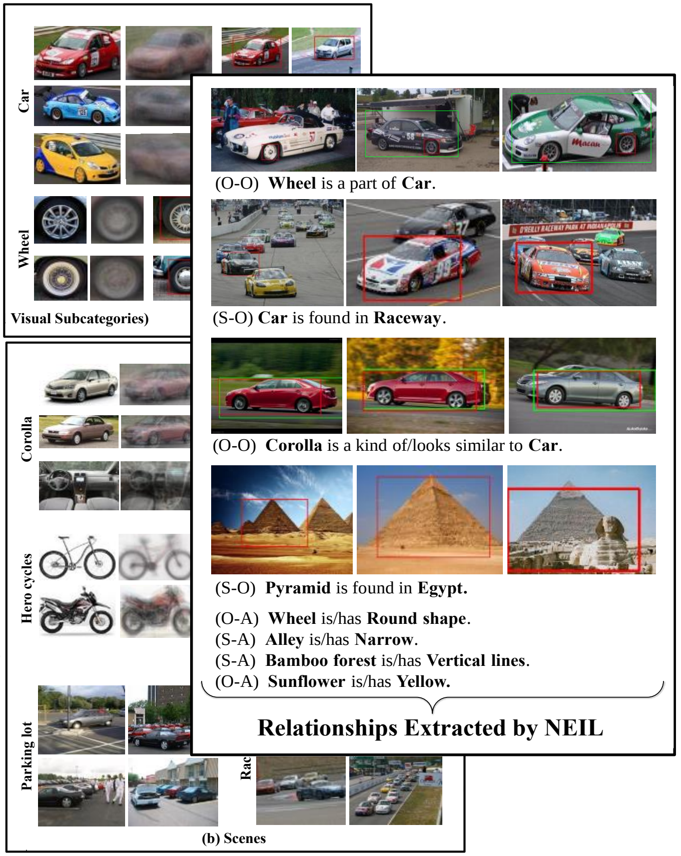


(Mitchell et al. 2018) Never-Ending Learning

NELL has been learning to read the Web 24 hours a day since January 2010. So far, NELL has acquired a knowledge base with 120 million diverse, confidence-weighted beliefs. NELL runs continuously to extract new instances of categories and relations.

Never Ending Image Learning (NEIL)

Source: Dr. Abhinav Gupta, CMU



(O-O) **Wheel** is a part of **Car.**

(S-O) **Car** is found in **Raceway.**

(O-O) **Corolla** is a kind of/looks similar to **Car.**

(S-O) **Pyramid** is found in **Egypt.**

(O-A) **Wheel** is/has **Round shape.**
(S-A) **Alley** is/has **Narrow.**
(S-A) **Bamboo forest** is/has **Vertical lines.**
(O-A) **Sunflower** is/has **Yellow.**

Relationships Extracted by NEIL

Source: Dr. Abhinav Gupta, CMU

Source: Dr. Abhinav Gupta, CMU

Extract visual commonsense knowledge from the Web

Verb Physics

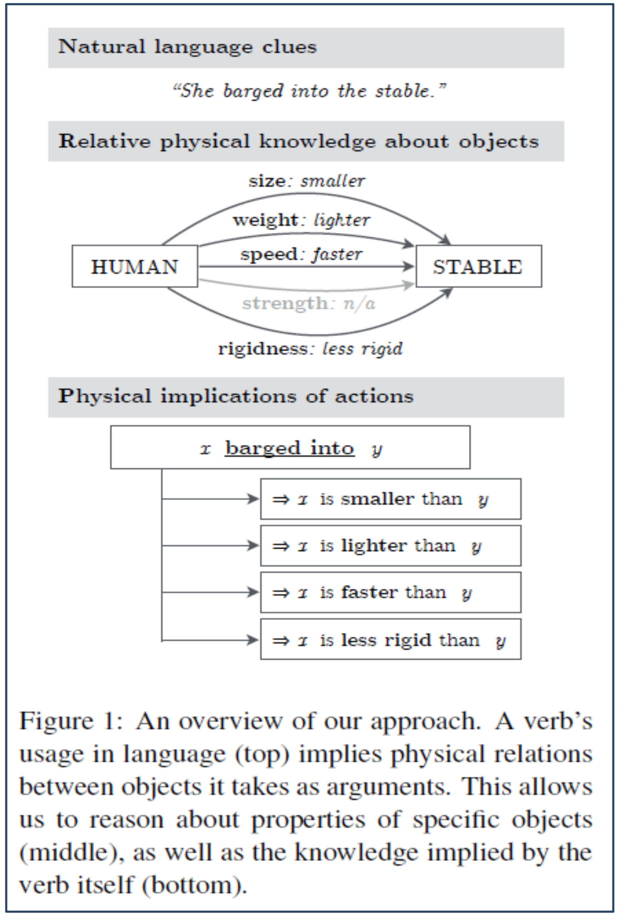
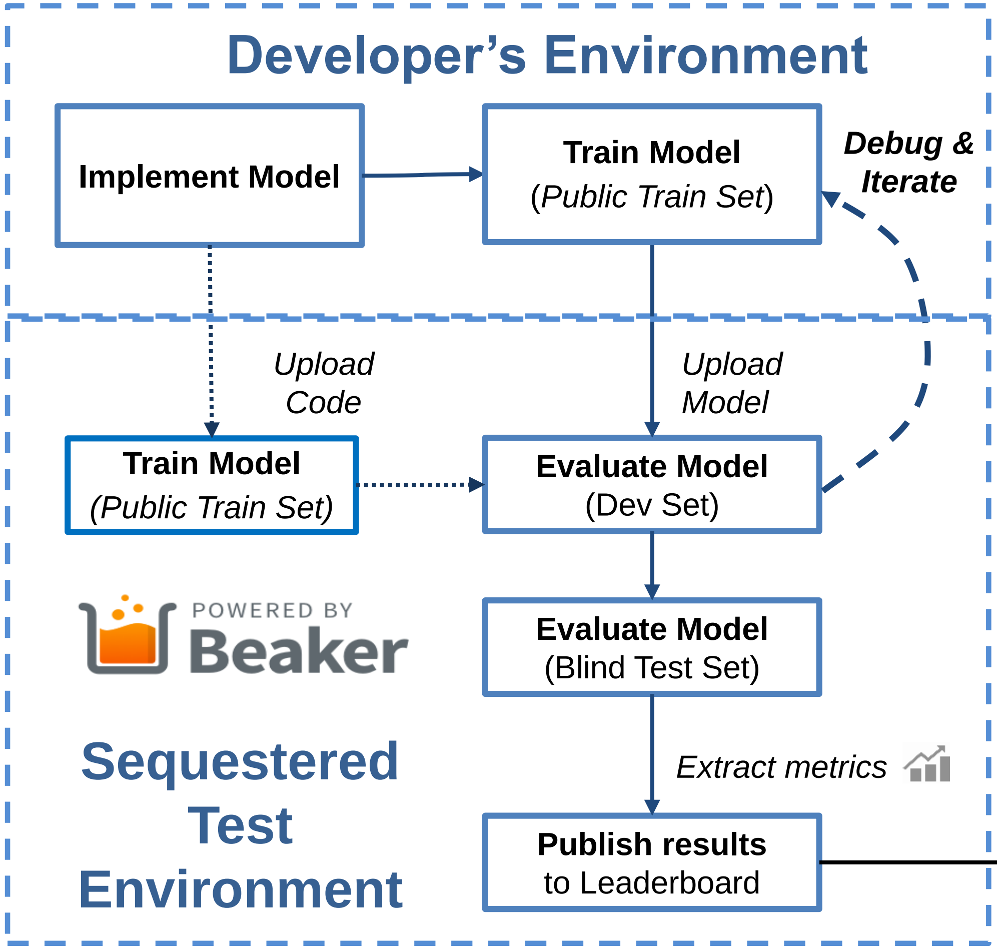


Figure 1: An overview of our approach. A verb's usage in language (top) implies physical relations between objects it takes as arguments. This allows us to reason about properties of specific objects (middle), as well as the knowledge implied by the verb itself (bottom).

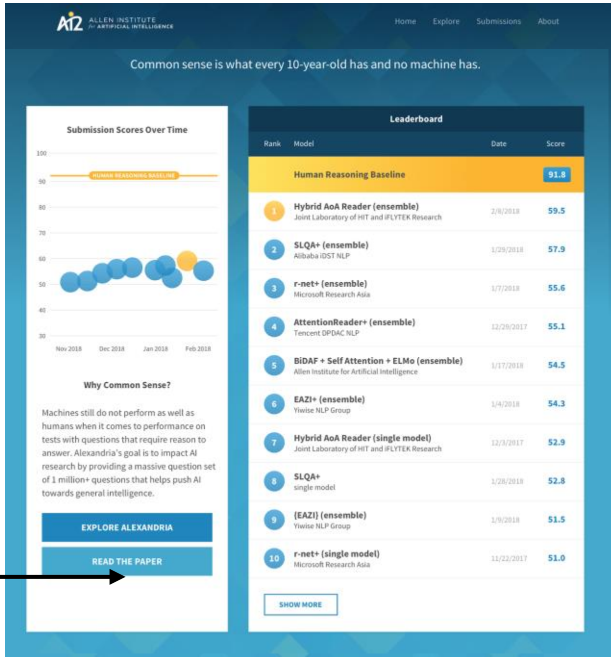
(Forbes & Choi, 2017). VERB PHYSICS: Relative Physical Knowledge of Actions and Objects

Extract relative physical knowledge of actions and objects from the Web

Allen Institute for Artificial Intelligence (AI2) Benchmarks for Common Sense



AI2's Project Common Sense is developing a suite of standard measurements for the common sense abilities of an AI system. The initial test set and leaderboard will be available in OCT 2018.



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

Why Common Sense?

Machines still do not perform as well as humans when it comes to performance on tests with questions that require reason to answer. Alexandria's goal is to impact AI research by providing a massive question set of 1 million+ questions that helps push AI towards general intelligence.

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
6	EAZI+ (ensemble) Yiwise NLP Group	1/4/2018	54.3
7	Hybrid AoA Reader (single model) Joint Laboratory of HIT and iFLYTEK Research	12/3/2017	52.9
8	SLQA+ single model	1/28/2018	52.8
9	{EAZI} (ensemble) Yiwise NLP Group	1/9/2018	51.5
10	r-net+ (single model) Microsoft Research Asia	11/22/2017	51.0

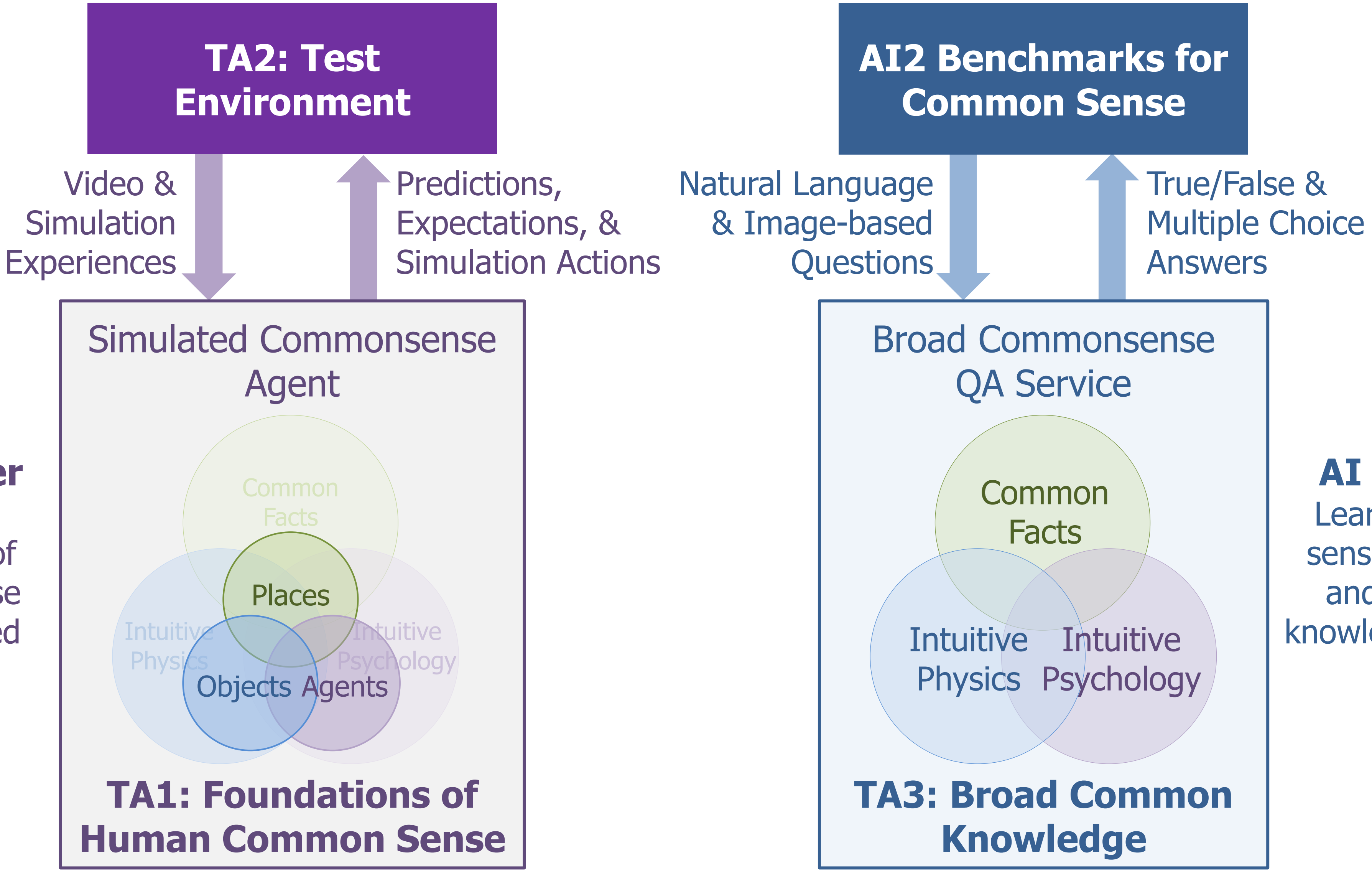
AI2's Commonsense Test Sets

- Commonsense Natural Language Inference (NLI)
- Commonsense NLI with Vision
- Abductive NLI
- Physical Interaction Question Answering (QA)
- Social Interaction QA

TA3 Schedule and Target Milestones

AI2 Common Sense Benchmark Data Set	Year 1	Year 2	Year 3	Year 4
Commonsense Natural Language Inference (NLI)	50%	60%	70%	80%
Commonsense NLI with Vision	50%	60%	70%	80%
Abductive NLI	50%	60%	70%	80%
Physical Interaction Question Answering (QA)	50%	60%	70%	80%
Social Interaction QA	50%	60%	70%	80%

Technical Areas



TA2: Test Environment

Video & Simulation Experiences

Predictions, Expectations, & Simulation Actions

AI2 Benchmarks for Common Sense

Natural Language & Image-based Questions

True/False & Multiple Choice Answers

Simulated Commonsense Agent

TA1: Foundations of Human Common Sense

Broad Commonsense QA Service

TA3: Broad Common Knowledge

AI Beginner
Learns the foundations of common sense from simulated experiences

AI Librarian
Learns common sense by reading and extracting knowledge from the Web

TA1: Foundations of Human Common Sense

Goal: develop computational models that mimic the core cognitive capabilities of children, 0-18 months old

- Multiple TA1 development teams will be selected to construct the computational models.
- The TA1 teams may propose a variety of development strategies, ranging from pre-building initial models, to learning everything from scratch using any combination of symbolic, probabilistic, or deep learning techniques.
- The TA1 teams are expected to include both AI and developmental psychology expertise, to produce both computational models and refined psychological theories of cognition.
- Although the primary goal of TA1 is to develop computational models, a secondary goal is to consolidate, refine, and extend the psychological theories of child cognition needed to guide model development, and to test, through research, key predictions made by the computational models.
- The TA1 teams may also propose optional companion research experiments in developmental psychology to refine their theories of cognition, where needed, to answer critical design questions relevant to their computational models.
- Note that, although TA2 will provide sample test problems, each TA1 team is responsible for designing and providing their own development strategy, training regimen, and any necessary datasets.

TA1: Foundations of Human Common Sense

TA1 proposals should include a detailed discussion of the technical plan to:

- Design and develop computational models that mimic the foundations of human common sense for the core domains of objects, agents, and places;
- Consolidate, refine, and extend the psychological theories of child cognition needed to guide model development; and test key predictions made by the computational models;
- Sequence the development and evaluation of the computational models over the four-year program, including any optional companion research experiments in developmental psychology to refine relevant theories of cognition;
- Perform the evaluation tasks, including the three levels of performance: prediction/ expectation, experience learning, and problem solving;
- Achieve the target milestones and metrics identified in the Schedule/Milestone section of the BAA; and
- Publish, share, and disseminate the results of research and development to the broader AI and Developmental Psychology communities.

TA2: Test Environment for the Foundations of Human Common Sense

Goal: provide the test and evaluation environment for evaluating the TA1 models against cognitive development milestones as evidenced in developmental psychology research with children from 0 to 18-months old

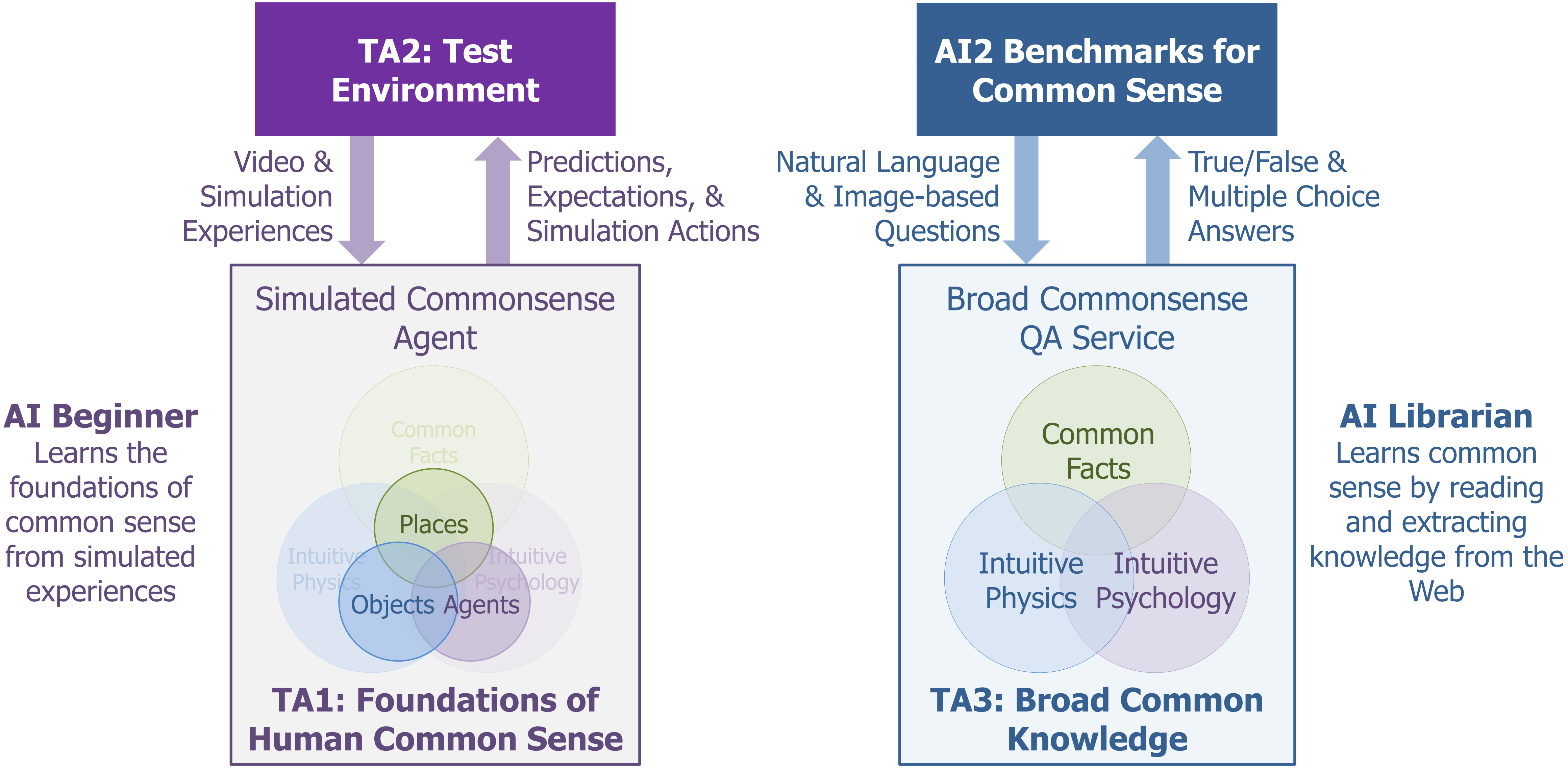
- The existing body of research will be used as an initial starting point for the TA2 team to construct the test environment and develop specific test problems for each milestone in order to evaluate the TA1 computational models at three levels of performance: prediction/ expectation, experience learning, and problem solving.

TA2: Test Environment for the Foundations of Human Common Sense

TA2 proposals should include a detailed discussion of the technical plan to:

- Refine and expand the cognitive milestones for the core domains of objects, agents, and places (including combinations of the three domains);
- Devise a set of specific test problems for each cognitive milestone to assess computational models at the required three levels of performance (prediction/expectation, experience learning, problem solving);
- Select, modify, or construct the video and 3D simulation infrastructure needed to conduct the TA1 evaluations;
- Provide the training and testing infrastructure, with sample test problems, to support TA1 computational model development.
  - TA2 is not expected to provide all of the training data that may be needed by the TA1 teams. TA1 teams are responsible for designing and providing their own development strategy and training regimen.
- Develop and provide all of the documentation (e.g., user guides, test environment specifications, etc.) and APIs for testing TA1 computational models;
- Conduct formal evaluations of TA1 computational models every six months; and
- Provide written test reports that document the performance of the TA1 models for each 6-month evaluation.

Technical Areas



TA2: Test Environment

Video & Simulation Experiences

Predictions, Expectations, & Simulation Actions

AI Beginner
Learns the foundations of common sense from simulated experiences

Simulated Commonsense Agent

TA1: Foundations of Human Common Sense

AI2 Benchmarks for Common Sense

Natural Language & Image-based Questions

True/False & Multiple Choice Answers

Broad Commonsense QA Service

TA3: Broad Common Knowledge

AI Librarian
Learns common sense by reading and extracting knowledge from the Web

TA3: Broad Common Knowledge

Goal: learn/extract/construct a commonsense knowledge repository capable of answering natural language and image-based questions about commonsense phenomena from the AI2 Benchmarks for Common Sense.

- Multiple development teams will be selected to develop the TA3 commonsense repositories/question answering services.
- TA3 teams may propose any combination of manual construction, information extraction, machine learning, and crowdsourcing techniques to construct a repository of broad commonsense knowledge.
    - TA3 teams are not required to include personnel with expertise in psychology and are free to use whatever techniques they prefer, whether artificial or biologically inspired.
- The TA3 teams are expected to submit their system for testing on the blind evaluation datasets (i.e., all five commonsense question datasets identified above, if completed/available) every six months.
    - Additional datasets may be developed over the course of the program, as needed, to align with the evolution of the TA3-developed capabilities.
    - After the first year, TA3 teams may propose their own question datasets for inclusion in the AI2 benchmarks, or propose suggestions for the development of additional datasets by AI2, for testing by all of the TA3 teams.

TA3: Broad Common Knowledge

TA3 proposals should include a detailed discussion of the technical approach to:

- Design and develop the broad commonsense knowledge service;
- Sequence the development and evaluation of the broad commonsense knowledge service over the four-year program;
- Perform the evaluation tasks for the AI2 Benchmarks for Common Sense;
- Achieve the target milestones and metrics identified in Schedule/Milestone section of the BAA; and
- Publish, share, and disseminate the results of research and development to the broader AI community.

AI2 Benchmarks

- Initially, five commonsense question datasets will be developed and available for testing of TA3-developed services:
  1. **Commonsense Natural Language Inference (NLI)**: multiple choice, natural language-based questions about commonsense events derived from captions in the ActivityNet Captions and Large Scale Movie Description Challenge (LSMDC) datasets.
  2. **Commonsense NLI with Vision**: multiple choice, image-based questions about commonsense events selected from the same ActivityNet and LSMDC datasets.
  3. **Abductive NLI**: questions about inferring the most likely hypothesis for a given set of observations.
  4. **Physical Interaction Question Answering (QA)**: natural language questions (initially) and image-based questions (in later years) about everyday objects and actions.
  5. **Social Interaction QA**: questions about human social behavior and the causal effects of everyday events.
- The development of the first dataset, Commonsense NLI, is completed and is described further in Zellers, R., et al. (2018).
- The remaining four datasets are currently in development and will be completed by the start of the program.
- More information about the AI2 commonsense question datasets and leaderboard will be available at https://leaderboard.allenai.org/ (which requires Chrome).

Schedule

- DARPA anticipates a June 2019 start date for the MCS program that will run for a duration of 48 months.
- The following PI Meetings will take place:
  - An in-person Kickoff Meeting at program start. For planning purposes, assume a 3-day meeting in Arlington, VA;
  - Four (4) web-based PI meetings held at six (6) months into each year of the program to review technical progress. For planning purposes, assume the government as host for these 2-day virtual meetings; and
  - Four (4) in-person PI meetings held at the end of each year of the program to review technical progress, conduct demonstrations, and provide opportunities for face-to-face collaboration. For planning purposes, assume 3-day meetings, alternating between a west coast and east coast location.
- In addition to the PI Meetings above, each team should expect to:
  - Host an onsite visit from the PM (and potentially other government personnel) at least once a year; and
  - Make two additional trips to the Washington, D.C. area in the last two years of the program for possible demonstrations and technology transition meetings.

Milestones

- The target milestones and metrics identified have been established to assess technical progress over the course of the program.
- The targets are not "go/no-go" criteria and it is not DARPA's intention to use the targets as the basis for down-selects or as the primary reason for other funding decisions.
  - TA1-developed computational models will be assessed for performance against cognitive development milestone capabilities (for the core domains of objects, agents, and places) at increasing levels of performance: prediction/expectation, experience learning, and problem solving.
  - TA3-developed services will be assessed for performance on the AI2 Common Sense Benchmark datasets (Commonsense NLI, Commonsense NLI with Vision, Abductive NLI, Physical Interaction QA, and Social Interaction QA).
- Assessments of TA1-developed computational models and TA3-developed QA services will be conducted every six (6) months, preceding each PI meeting, in order for results/analyses to be available for review and discussion at the meetings.

TA-specific Deliverables

TA	Deliverable
TA1	• Computational model source code and APIs; and • Any associated data and documentation (including, at a minimum, user manuals and a detailed software design document).
TA2	• Test environment source code and APIs; • Any associated data and documentation (including, at a minimum, user manuals and a detailed software design document); • Test Environment Readiness Assessment Reports; and • Any Test Environment documentation necessary to support TA1 team model assessments (e.g., standard operating procedures, user guide, etc.).
TA3	• Repositories, libraries, source code, and APIs; and • Any associated data and documentation (including, at a minimum, user manuals and a detailed software design document).
All TAs	• Quarterly progress and final reports; • Presentations (in PowerPoint) for each PI Meeting (total of nine (9)); • Copies of published papers and presentations at conferences, provided each month; and • Monthly financial status reports, provided within 10 calendar days of the end of each calendar month.

Government-furnished Property/Equipment/Information

- No Government-furnished equipment is expected to be provided.
- The test and evaluation infrastructures and environments will be provided by TA2 for TA1, and by AI2 for TA3.
- The TA3 evaluation datasets will be provided by AI2.

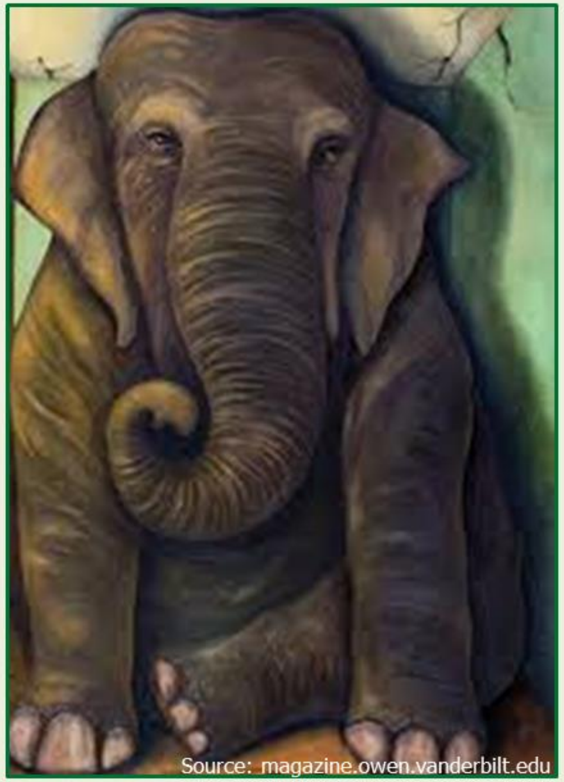
Intellectual Property

- The program will emphasize creating and leveraging open source technology and architecture.
- Intellectual property rights asserted by proposers are strongly encouraged to be aligned with open source regimes.
- A key goal of the program is to facilitate rapid innovation and advancements in AI by providing foundational capabilities for future users or developers of MCS program technologies and deliverables. Therefore, it is desired that all noncommercial software (including source code), software documentation, hardware designs and documentation, and technical data generated by the program be provided as deliverables to the Government, with a minimum of Government Purpose Rights (GPR), as lesser rights may adversely impact the progress towards the realization of AI systems with machine commonsense knowledge and reasoning capabilities.

Summary

TOMORROW

Machine
Common Sense



Source: magazine.owen.vanderbilt.edu

The elephant in the room

- MCS will create the computing foundations needed to develop machine commonsense services to enable AI applications to understand new situations, monitor the reasonableness of their actions, communicate more effectively with people, and transfer learning to new domains.
- MCS is seeking the most interesting and compelling ideas to accomplish this goal
- Abstracts Due - November 6, 2018
- Proposals Due - December 18, 2018

www.darpa.mil