

# Robust Artificial Intelligence: Reasoning\*

Sargur N. Srihari  
[srihari@buffalo.edu](mailto:srihari@buffalo.edu)

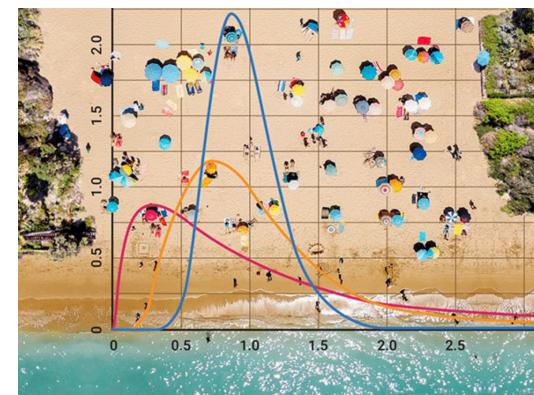
\* Based on Gary Marcus “The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence”

# Topics

- Memorization versus Reasoning
- Reasoning in Romeo and Juliet (CYC)
- Cognitive Models
- Google's MEENA
- Conjecture for Robust Intelligence
- Role of Abstraction in Human Intelligence
- Proposal for a Hybrid Model

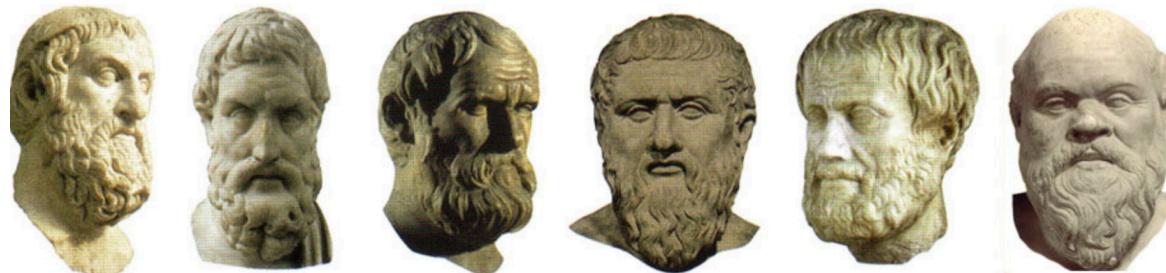
# Memorization versus Reasoning

- Deep Learning AI copes with complexity by trying to memorize the probability density function for the entire world—at the cost of needing more data
- Given the exponential complexity of the world, it's a strategy that is unlikely to work



# Reasoning offers an alternative

- Instead of memorizing everything, or interpolating between near neighbors that you might have previously encountered, you draw inferences
- Instead of memorizing that Plato, Aristotle and Euripides and each of the other billions of other individuals preceding us were all mortal, you learn a general truth
  - That all humans are mortal, and apply that general truth to specific instances of that category, as needed

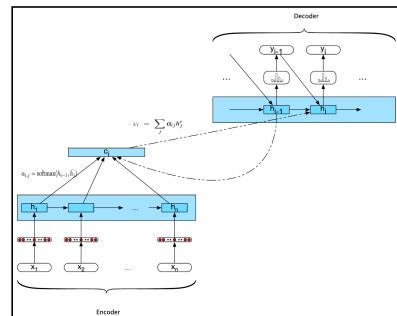


1 Sophocles; 2 Euripides; 3 Aristophanes; 4 Plato; 5 Aristotle; 6 Socrates

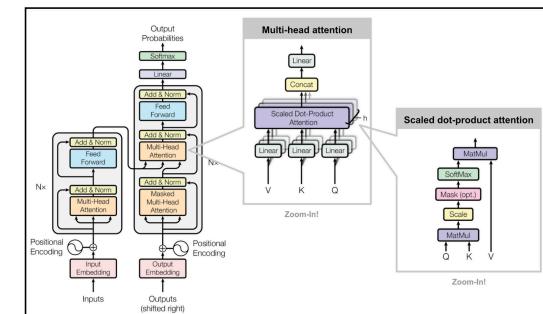
# Symbol manipulation holds promise

- Neural nets such as Transformers are too unreliable to make for sound reasoners
- They may work some of the time, but not robustly
- Symbol-manipulation offers at least the promise of heading the right direction, provided that sufficient knowledge is available.

Encoder-decoder  
with Attention

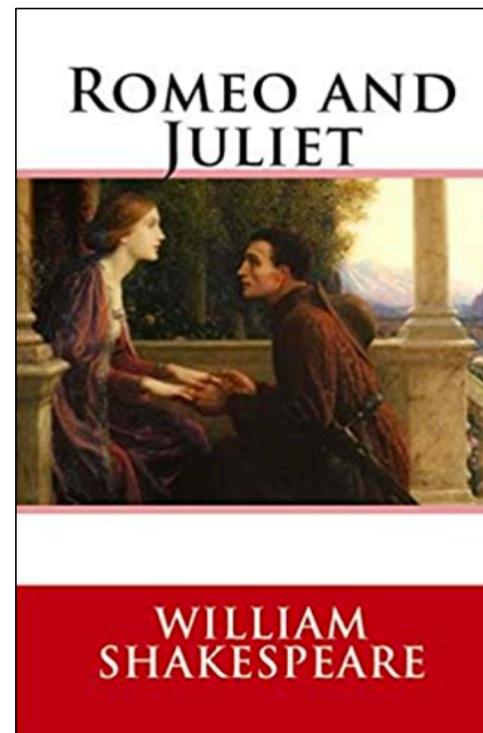


Transformer model  
with no recurrence



# Reasoning in CYC: Romeo and Juliet

- The best case for reasoning engines:
  - Inference that CYC (a symbolic system) performs in optimal circumstances



# Samples of story-relevant knowledge

## Plot summary (Wikipedia)

"Juliet visits Friar Laurence for help [avoiding being forced by her father to marry Paris], and he offers her a potion that will put her into a deathlike coma for [42] hours. The Friar promises to send a messenger to inform Romeo of the plan so that he can rejoin her when she awakens. ...she takes the drug and, when discovered apparently dead, she is laid in the family crypt.

"The messenger, however, does not reach Romeo and, instead, Romeo learns of Juliet's apparent death... Heartbroken, Romeo buys poison from an apothecary and goes to the Capulet crypt... Believing Juliet to be dead, he drinks the poison [and dies]. Juliet then awakens and, discovering that Romeo is dead, stabs herself with his dagger and joins him in death."

## Some time points

- T<sub>0</sub> – just before Act IV starts.
- T<sub>1a</sub> – the start of Juliet's visit with Friar Lawrence
- T<sub>1b</sub> – the end of that visit
- T<sub>2</sub> – what the Friar and Juliet both believe about T<sub>4</sub> and T<sub>5</sub>
- T<sub>3</sub> – Juliet drinks the feign death potion
- T<sub>3b</sub> – Juliet's body is discovered and taken to the crypt
- T<sub>4</sub> – Romeo hears from the Friar's messenger about the planned trickery
- T<sub>5</sub> – Romeo hears (from someone else) that Juliet has died
- T<sub>6</sub> – Romeo goes to the crypt and finds Juliet's inert body.
- T<sub>7</sub> – Juliet awakens from her feigning of death.
- T<sub>8</sub> – Romeo and Juliet secretly flee Verona

## Some statements that are true at T0

- Lord Capulet wants and expects Juliet and Paris to marry.
- Juliet knows that Lord Capulet wants and expects Juliet and Paris to marry.
- Romeo believes that Juliet is alive.
- Romeo, Juliet, and the Friar do not want Juliet to marry Paris.
- All three of them know that all three of them do not want that to happen.

It is a *cognitive model* in that it is an internal distillation of a complex scenario, and it is *rich* in that it filled with subtle information about what particular protagonists do and know and intend and expect at particular time points.

# Common Sense and a Complex Inference

## Rich Common Sense knowledge

- Of course if a person were dead, they would not have to marry anyone.
- Of course a person often asks another for help accomplishing something that they both want.
- Of course if someone drinks an instantly fatal dose of poison, they immediately die.
- Of course if a trusted friend of yours tells you something, and you don't have a better reason *not* to believe it, then you are very likely to believe what they say.
- Of course when someone newsworthy is believed to be dead, news of their death will spread quickly. The scale and speed depend on the information technology available and the dead person's level of fame. In particular, in a small medieval European town, the news of a local noble's demise would spread by word of mouth throughout that town over a period of hours but in less than 24 hours.

Some bits of knowledge might be represented explicitly (e.g, consequence of drinking a poison might be instant death), Some derived online from more general facts. (e.g., , “If a person were dead, they would not have to marry anyone” might be derived from observation that obligations apply only to living people.)

## Sample inference

When she takes the feign-death potion, does Juliet believe Romeo will believe she is alive during the time she is in suspension?

⊕ ⊖ ▾ Yes.

- If, at time T1, an agent's model of a subject's beliefs at time T2 includes a proposition, then the agent believes at T1 that the subject believes the proposition at T2.
- At the time of Juliet's taking of the feign-death potion, Juliet has a model of Romeo's beliefs at the time of Juliet's being in suspension after taking the feign-death potion that includes the proposition that Juliet is a living thing.
  - Juliet believes that if an agent receives information from a trusted source, they will believe that information.
  - Juliet believes that Friar Laurence is a trusted source for Romeo.
  - Friar Laurence and Juliet plan for Romeo to receive Friar Laurence's message while Juliet is in suspension.
  - Friar Laurence and Juliet plan for Friar Laurence's message to Romeo to convey the information that while Juliet is in suspension, she will be alive.
  - Juliet believes the events she and Friar Laurence have planned will occur.

Payoff in having such a broad fund of knowledge Comes in the sophistication of the inferences that Can be drawn (right hand of the bottom figure), And in their level of interpretability— Both of which are on a completely different level from anything thus far produced via deep learning.

# Observations on CYC inference

- Approach needs:
  - Structured representations, Operations over variables, and Records for individuals
- Shows value of:
  - Cognitive models, Knowledge about biology, theory of mind, etc.
- But relies on hand translation of Romeo and Juliet into formal logic
  - A system generating such representations automatically and reason comparably would be a breakthrough.
- Reasoning is not the bottleneck towards robust AI;
  - Bottleneck is in having the right information accessible in the context of real-world inference

# Uncertainty is not modeled in CYC

- Even in CYC's reasoning, there is much room for improvement.
  - Its representations higher-order logic
  - Cannot represent uncertainty, and reason over distributions
- Bertrand Russell
  - "All human knowledge is uncertain, inexact, and partial"

# Reasoning and Knowledge are Essential

- An AI that can't understand a plot summary of Romeo and Juliet is not likely to be competent in the complexity of the real world.
- An AI that *can* reason about the motivations of interacting humans as complex events as they unfold over time has a fighting chance
- Reasoning and knowledge need to be first-class citizens if we are to move forward

# Cognitive Models

- Knowledge that accumulates over time:
  - What we learn about a friend in the course of a conversation,
  - About a nation in the course reading the news
  - About a set of people as we read a book.
- Such accumulating representations are called *cognitive models*.
  - Your cognitive model might differ from mine;
  - Yours more detailed, mine less detailed, but we both use them, routinely.
- A cognitive model might consist of knowledge about some set of entities
  1. The characters in a story and the objects they have available to them),
  2. A set of properties
    - The size and colors of the objects
    - The goals of the characters
  3. Information about time and events
    - At what time  $t$  did character  $x$  meet character  $y$ , and what did  $x$  know at time  $t$

# Transformer+ Reasoning is not Cognitive

- Transformer as input to a downstream reasoning system is not a cognitive model
- The essence of language comprehension is to derive cognitive models from discourse:
  - We can then reason over the models we derive
  - Transformers just don't do that
- Predicting word classes is impressive, but prediction does not equal understanding

# Scene Comprehension

- Deriving rich cognitive models from text and videos
- Aims to interpret visual scenes not only in terms of what objects are there, but how the objects relate to one another,
  - e.g. not just identifying a glass and a table, but taking note of the fact that a particular glass is on a table, in a particular room,
  - That that glass is near the edge of the table and that it is supported by the table,



# Combining Reasoning and Cognition

- 12:00pm child left alone in a room with a closed cookie jar containing a cookie
- 12:05pm it is observed that the jar is closed, but the cookie is nowhere to be seen.
- What happened in between?
- Combining temporal and spatial reasoning, you can readily infer that
  - a) The child opened the jar
  - b) The child removed the cookie,
  - c) The child ate the cookie
  - d) The child closed the jar
  - For bonus points you can infer that
    - (b) must have happened before both (c) and(d),
    - but also that the order of (c) versus (d) is unknown.
- In conjunction with a theory of biology, you can construe the child as a container and realize that the cookie is now contained (partly digested) within, having traversed from an aperture (a mouth) into yet another container (the stomach) within

# AI needs Cognitive Models

- Make these kind of inferences robustly without both internal cognitive models and mechanisms to reason over them.
- Without this, there is no way to reliably understand a detective story, a conversation that goes beyond small talk, or virtually any narrative of human interaction

# Conjecture for Robust Intelligence

- Cognitive Models
  - We cannot construct rich cognitive models in an adequate, automated way without the triumvirate of
    1. Hybrid architecture,
    2. Rich prior knowledge, and
    3. Sophisticated techniques for reasoning
- Robust Intelligence
  - We cannot achieve robust intelligence without the capacity to induce and represent rich cognitive models

# More data will not solve this

- Google's MEENA
- Trained on a massive 341 GB corpus
  - Ten times the size of what GPT-2 was trained on,
  - Equivalent of 341,000 books
- "Towards a Conversational Agent that Can Chat About... Anything"
  - Meena can conduct conversations that are more sensible and specific than existing state-of-the-art chatbots."
- <https://www.youtube.com/watch?v=cA0vHYzGR8s>



# Quality of Meena's conversations

- Like GPT2, Meena produces prose that is both remarkably fluent and also remarkably on topic.
- Its best dialogs are stunning, and the paper reports that it sets a new state of the art in new measure of "Sensibleness and Specificity".
- But the same core foundation—an internal cognitive model of the world—is still missing.

# Meena Conversation Examples

<https://github.com/google-research/google-research/blob/master/meena/meena.txt>

- **Conversation 61**

- *Human: What's your favorite band?*
- *Meena: Avenged sevenfold.*
- *Human: Ooh, that'sa good answer! What's your least favorite band?*
- *Meena: Avenged sevenfold.*
- *Human: Oh, I thought you liked Avenged sevenfold?*
- *Meena: Yeah. That's my favorite band.*

- **Conversation 77**

- *Human: What are some of your new year resolutions?*
- *Meena: Lose 60 pounds Learn Cantonese Learn French*
- *Human: Don't you think, that's pretty far fetched?:P*
- *Meena: Not really. I'm currently down 35 pounds, I can already speak Cantonese to a basic level, and I'm already learning French in school.*
- *Human: Wao! That's impressive!*
- *Meena: I'm learning French on the side, it's really easy forme since I'm a native Cantonese speaker.*

## Evaluation of MEENA

- Each answer is coherent, but there is no coherence from one answer to the next.
- Meena captures the fact that certain phrases like band names are often correlated with phrases like "favorite band";
- But there is no enduring record of Meena's preferences, even for the length of a short conversation,
- And hence complete inconsistency.

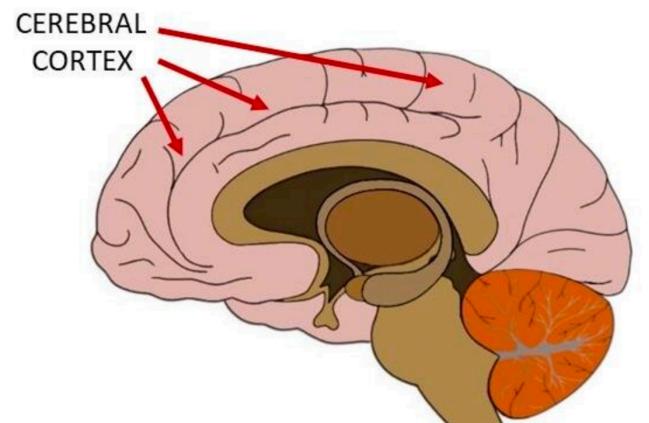
# Role of abstract knowledge in human life

- Without us, the world would continue to exist, but it would not be described, distilled, or understood
- A bird might flap its wings, and the bird might be carried along in flight
  - There would be correlation, but not causal description
- Human lives are filled with abstraction and causal description
  - Children spend a large part of their time asking why
  - Scientists ask such questions in order to generate theories
- Our power comes from understanding and characterizing the world, in the form of science, culture, and technology

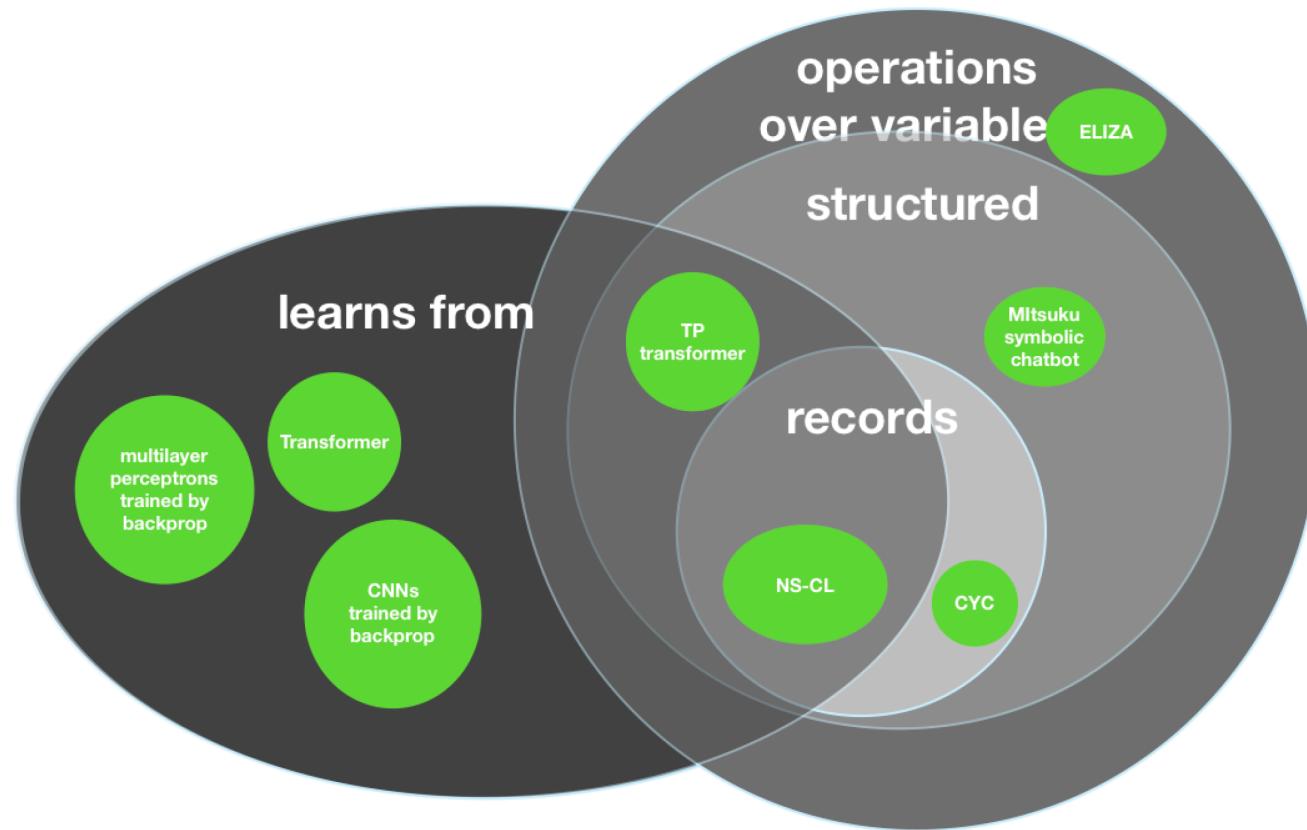


# Biological Support: Cortex functional modules

- Neural models are homogeneous
  - Neurons capable of little more than summation and integration, and often no more than a handful of prespecified modules
- At the macro-level, the cortex alone has hundreds of anatomically and likely functional areas
  - Even a single dendritic compartment of a single neuron can compute the nonlinearity of XOR
- Likely that there is diversity in neural computation
  - Not yet been captured either in computational neuroscience or in AI



# Robust AI resides in the intersection of Learning and Symbol Manipulation



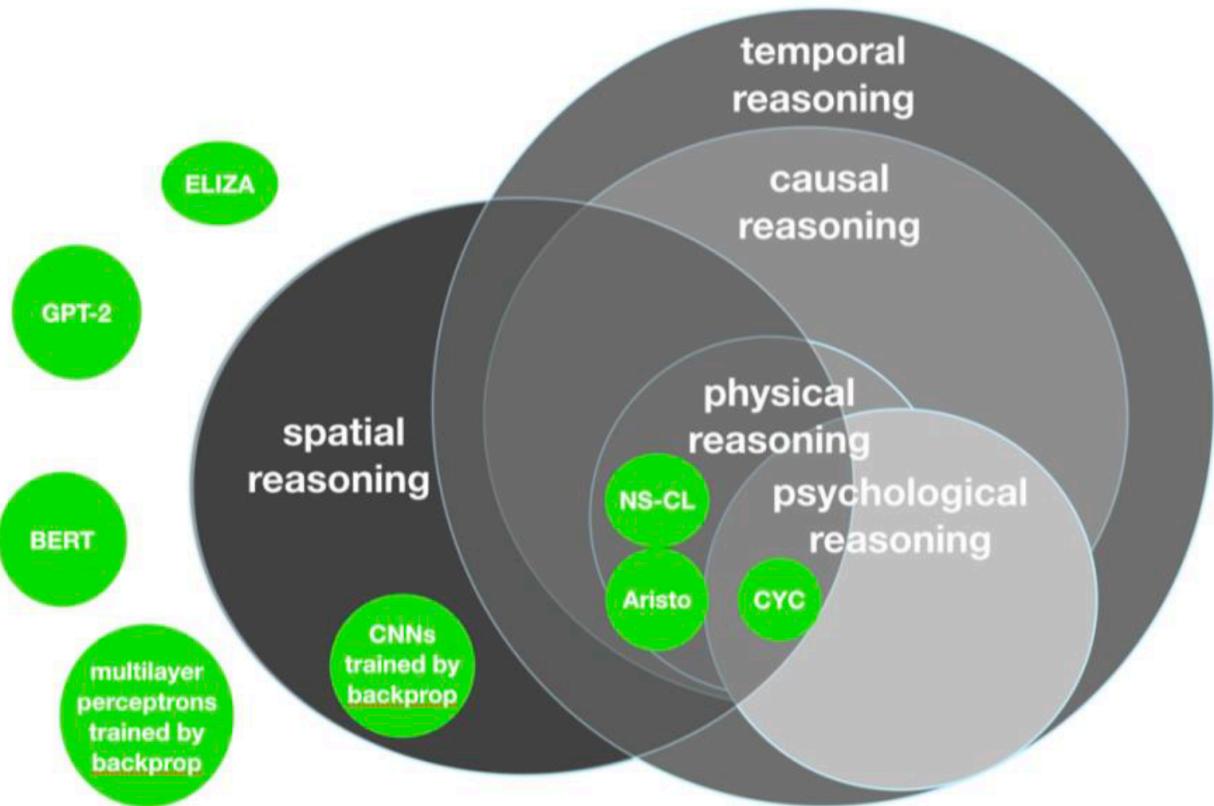
# Need for Special Mechanisms

Need for reasoning machinery for:

1. Spatial
2. Physical
3. Psychological
4. Temporal
5. Causal

Current neural networks lack

1. Explicit mechanisms for these forms of reasoning
2. Ways of representing and reasoning over such domains

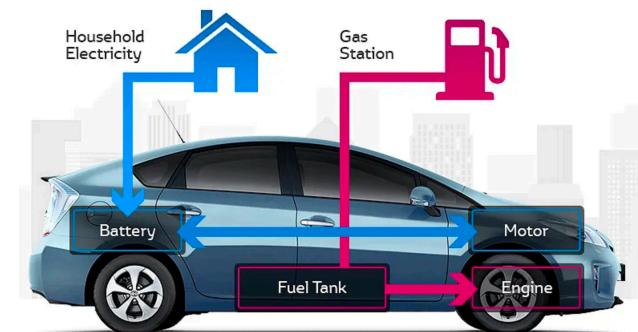


# Proposal for Hybrid Model

- To achieve common sense, and ultimately general intelligence
  - 1. Develop systems to represent core frame works of human knowledge:
    - Time, space, causality, basic knowledge of physical objects and their interactions, basic knowledge of humans and their interactions
  - 2. Embed these in an architecture that can be extended to every kind of knowledge,
    - Keeping in mind tenets of abstraction, compositionality, and tracking of individuals
  - 3. Reasoning that can deal with knowledge that is complex, uncertain, and incomplete and that can freely work both top-down and bottom-up
  - 4. Connect these to perception, manipulation, and language. Use these to build rich cognitive models of the world
  - 5. Final keystone: next slide

# Keystone for Hybrid Model

- Construct human-inspired learning system that uses all knowledge and cognitive abilities that the AI has
  - Incorporate what it learns into its prior knowledge
  - Like a child, voraciously learns from every possible source of information:
    - Interacting with the world,
    - Interacting with people, reading, watching videos, even being explicitly taught
- Put all that together to get to deep understanding



# Seeing the whole elephant

1. Deep learning shows much can be learned from massive data
  - Co-occurrence statistics maybe mere shadows of robust knowledge
  - But there are lots of shadows
  - Maybe we can put those shadows to use knowing both their strengths and their limitations.
2. CYC shows the potential power of sophisticated reasoning
  - In the presence of rich knowledge bases and rich cognitive models,
  - Even if on its own, it is not capable of deriving those models directly from language or perceptual inputs
3. NS-CL(Maoetal.,2019) shows that symbol manipulation and deep learning can be integrated into a seamless whole
  - Not yet at scale,
  - It can both perceive and reason.

# Summary

- Robust Intelligence needs Cognitive models
- More data is insufficient, as shown by Google MEENA
- Cognitive models incorporate reasoning with models of time, space and causality
- Hybrid models should integrate powers of deep learning and symbolic reasoning using seamless architectures