



DeUmbra

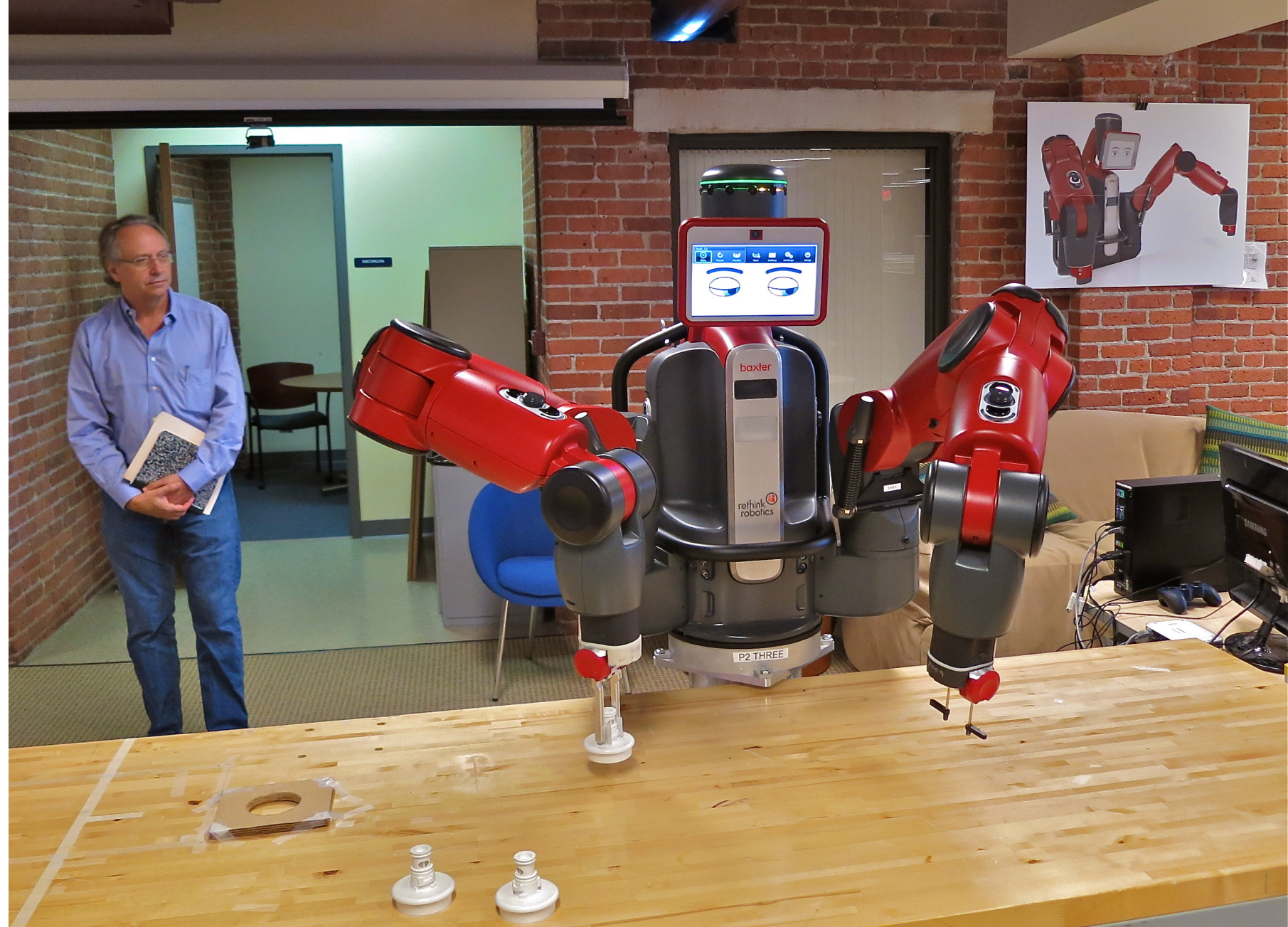
How to build someone we can talk to

Jonathan Mugan
@jmugan

Data Day Texas
January 28th, 2023

So many have left us

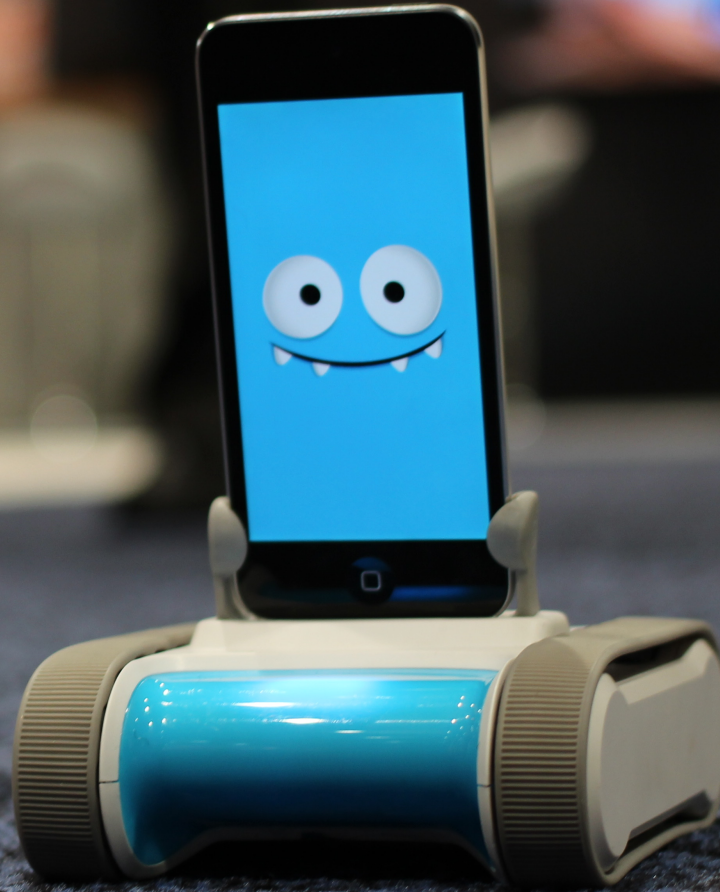
Many intelligent
robots have come
and gone, failing
to become a
commercial
success



https://commons.wikimedia.org/wiki/File:Rethink_Robotics_%E2%80%94_Brooks_and_Baxter_%28000143255%29.jpg

Steve Jurvetson from Los Altos, USA, CC BY 2.0 <<https://creativecommons.org/licenses/by/2.0/>>, via Wikimedia Commons

Romo is gone





Jibo never seemed to get off the ground

Cynthiabreazeal, CC BY-SA 4.0
<<https://creativecommons.org/licenses/by-sa/4.0/>>, via
Wikimedia Commons

SONY



**Mighty
Sony
couldn't
make
Aibo stay**



Sony photographer, CC0, via Wikimedia Commons

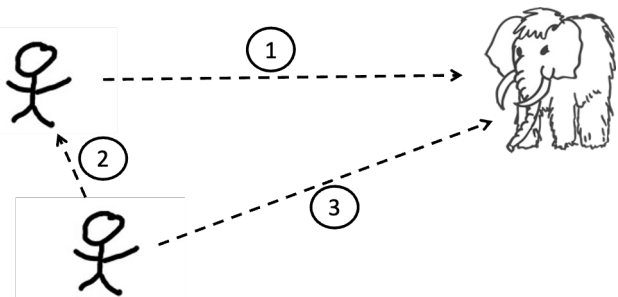
Even Alexa is reducing staff

These robots didn't reach their potential because you can't talk to them, not really.



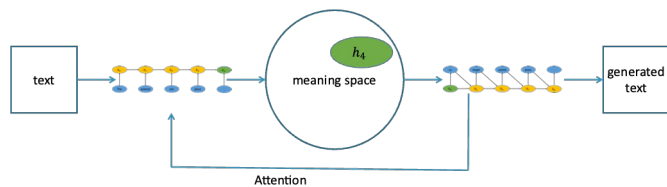
<https://arstechnica.com/gadgets/2022/11/amazon-alexa-is-a-colossal-failure-on-pace-to-lose-10-billion-this-year/>

Data Day 2017



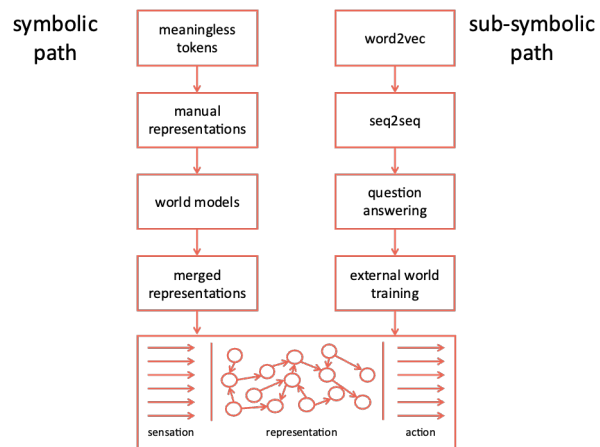
Talked about how language evolved and the current state-of-the-art for chatbots and how they won't work without understanding

Data Day 2018



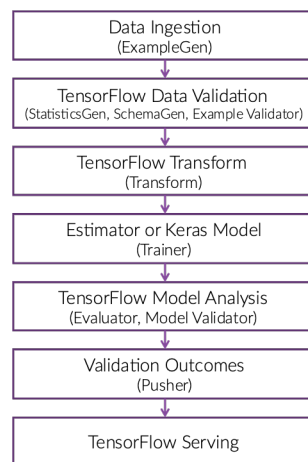
Went into detail about how neural networks can generate text

Data Day 2019



Talked about two paths to getting understanding

Data Day 2020

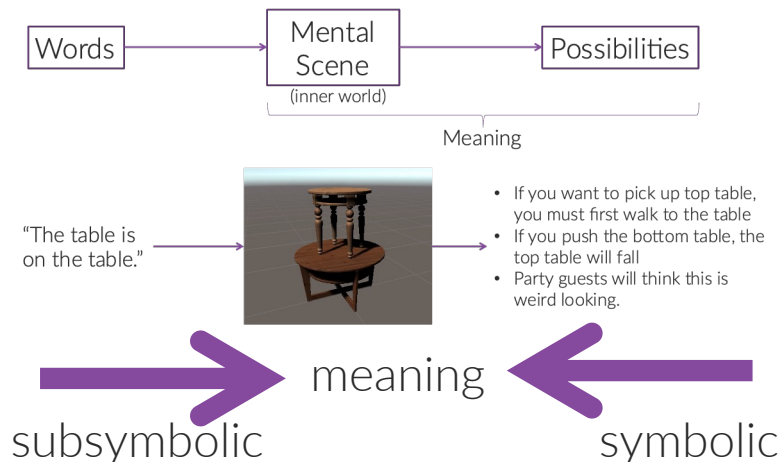


We got distracted and did MLOps with TensorFlow Extended

Data Day 2021

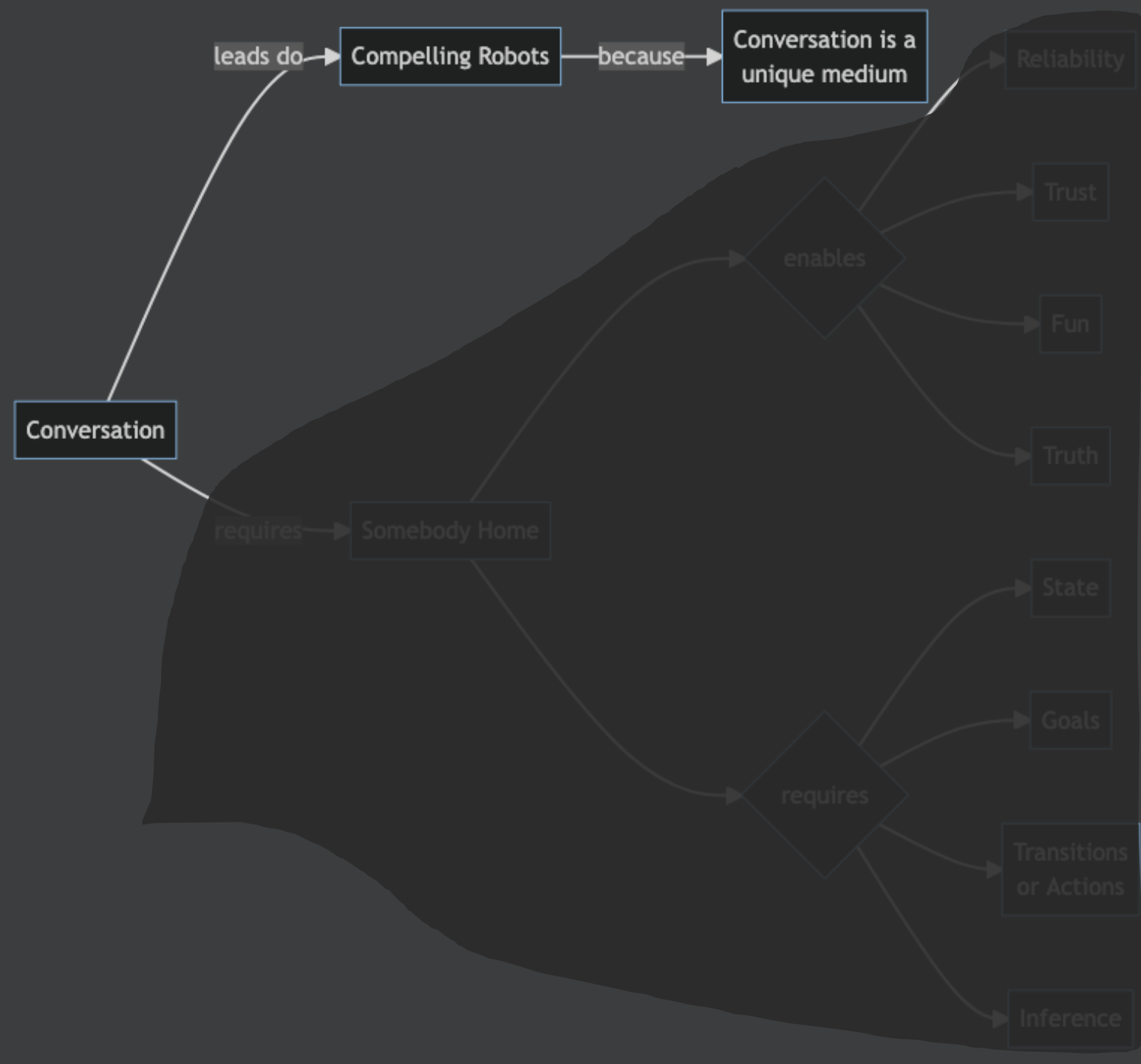
Covid

Data Day 2022



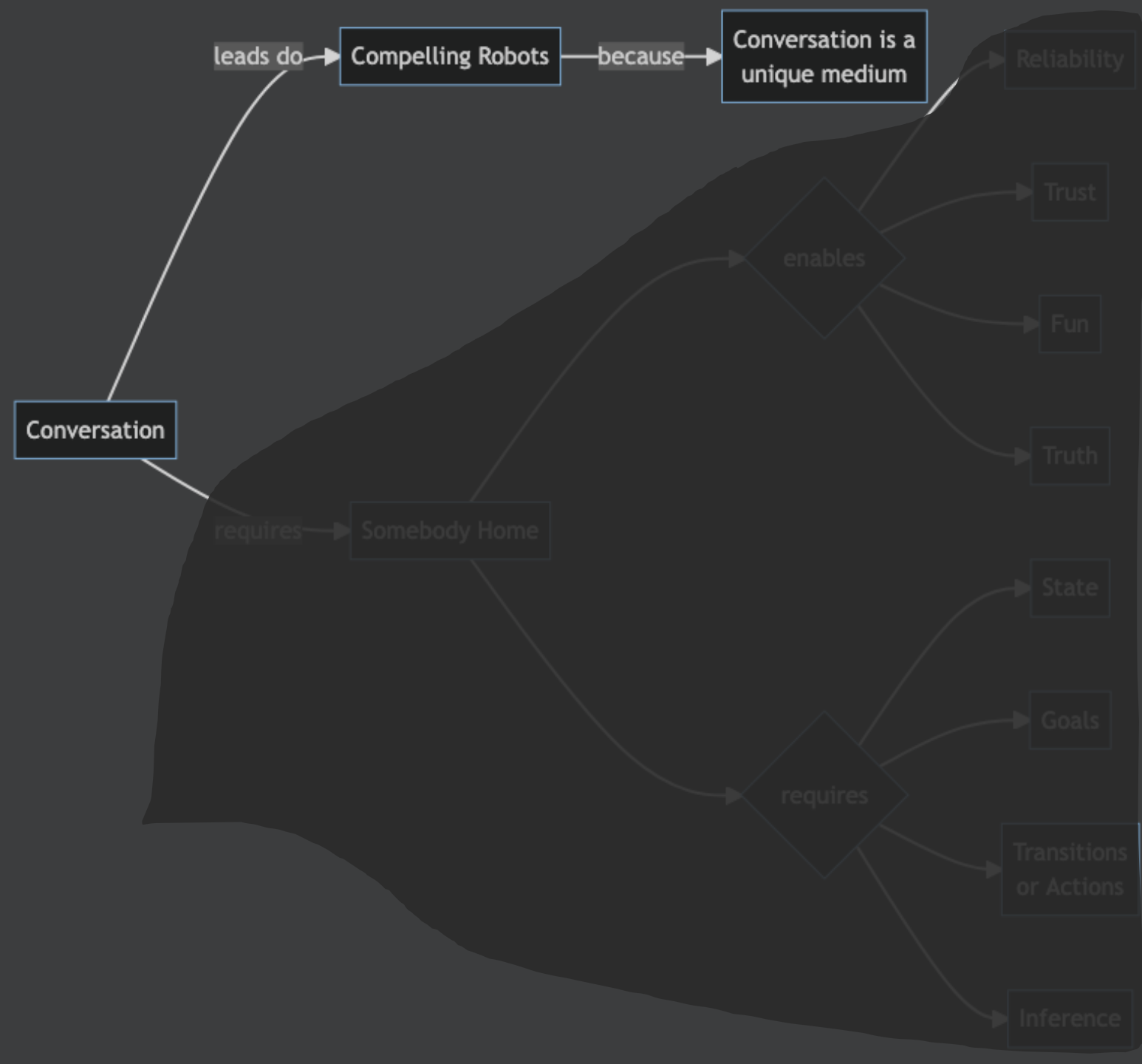
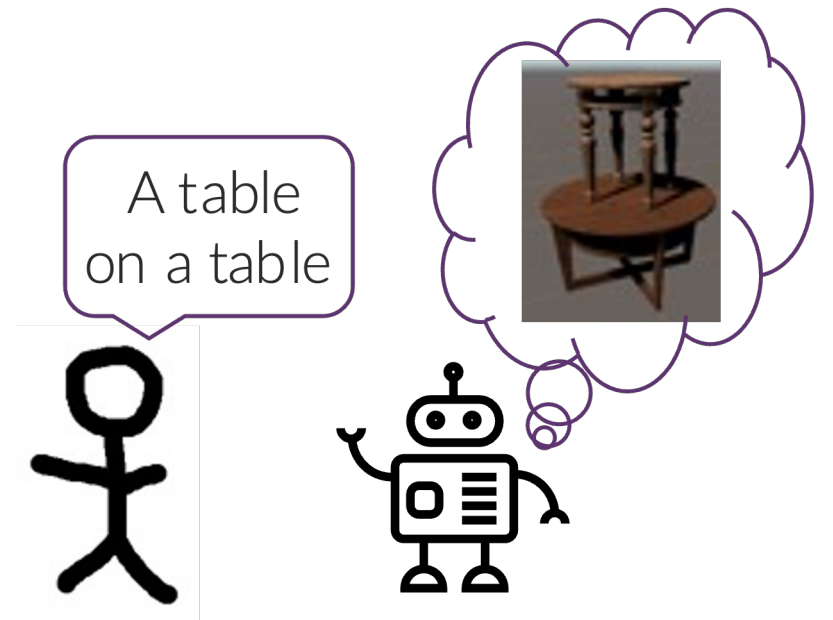
Conversation is a unique medium

- Conversations are interactive
- but they are also hyperlocal and hyper-specific to the current context
 - context entails the current situation and conversation history



Conversation is a unique medium

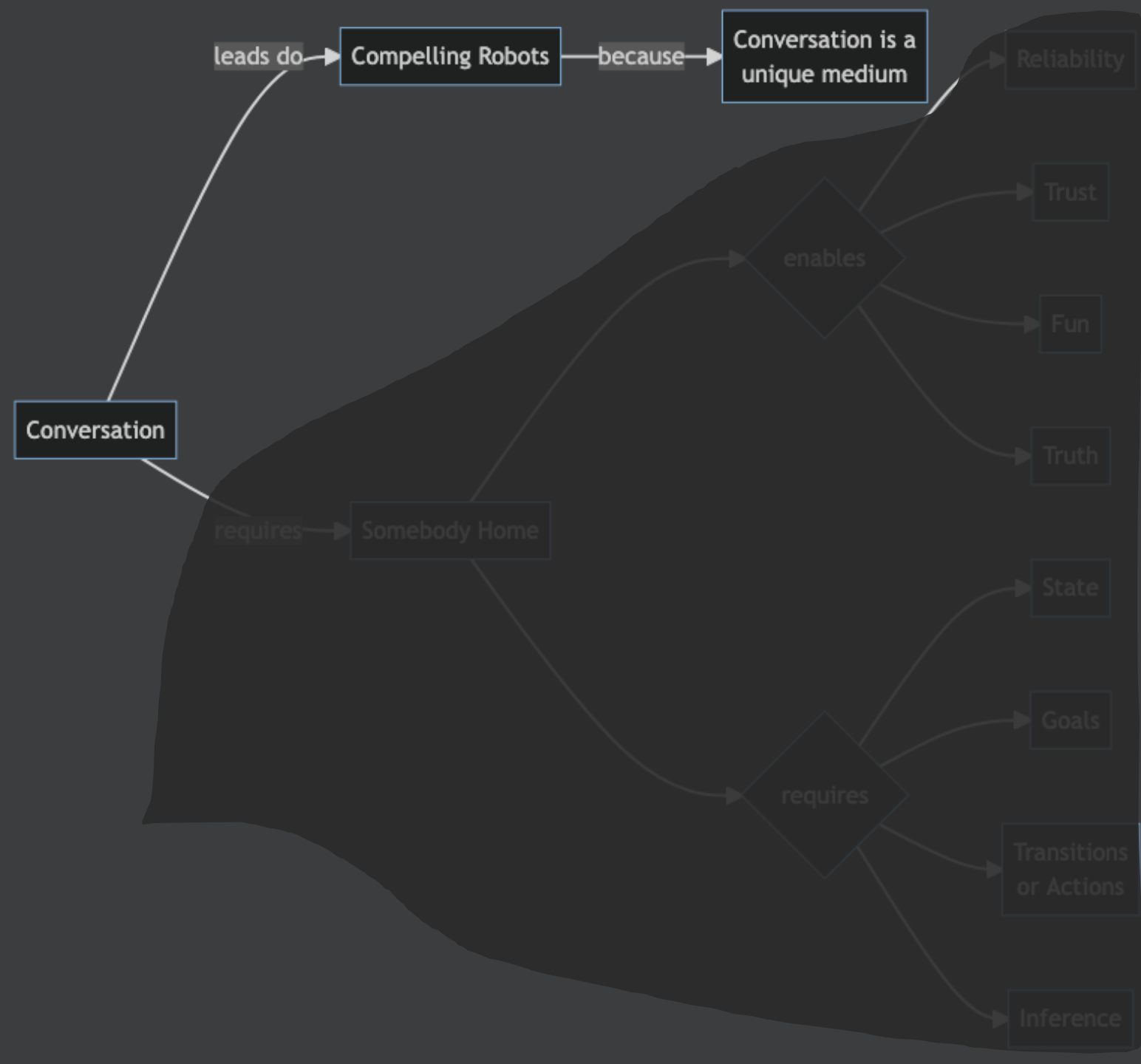
When you have a conversation with someone, you direct their imagination and they yours



Conversation is a unique medium

We don't have any other medium that has those properties

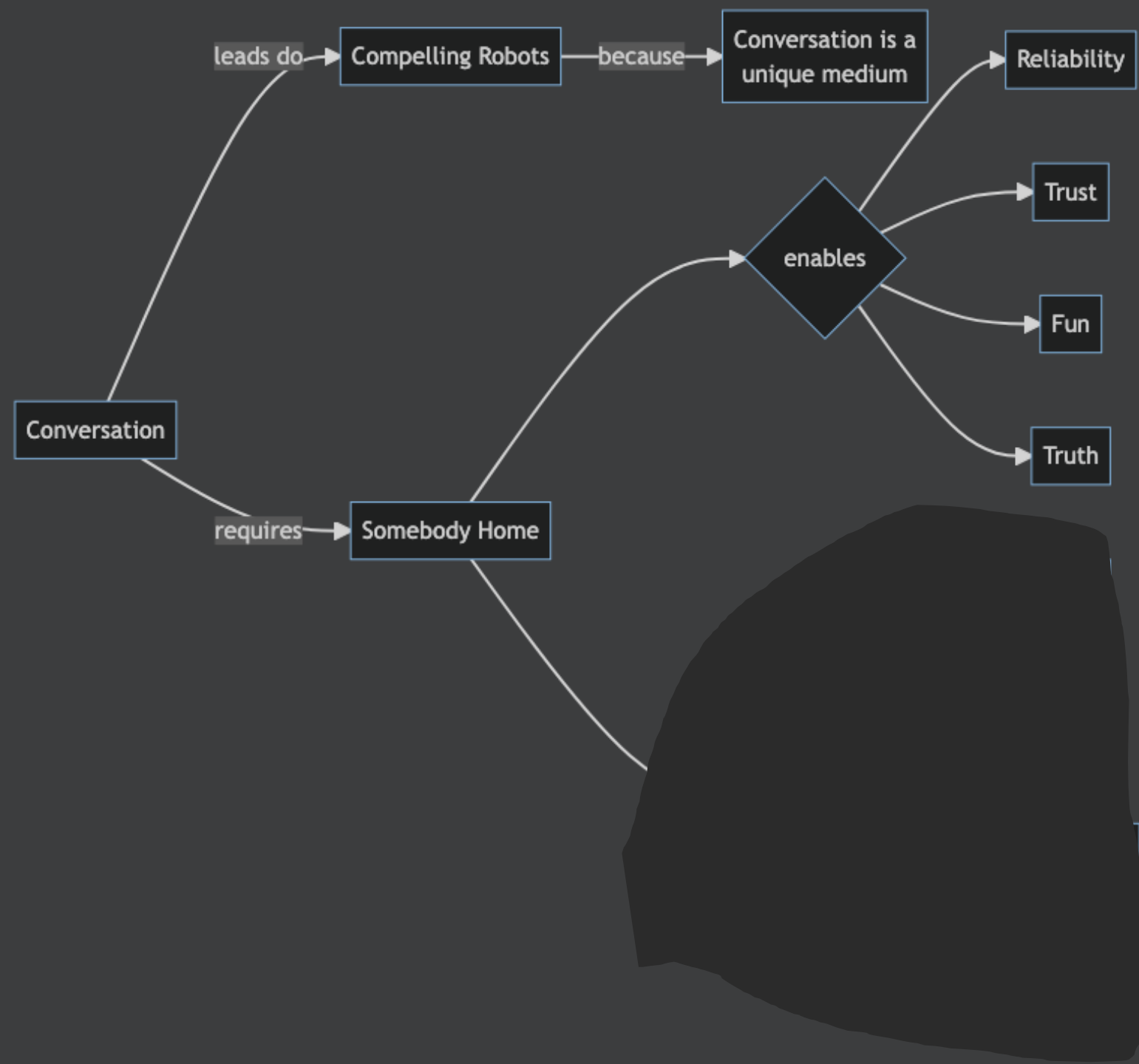
- movies, tv, videos, don't
- The best conversations teach
- Conversation is one of the hardest things we do as humans—why they say people need to socialize to keep their thinking sharp

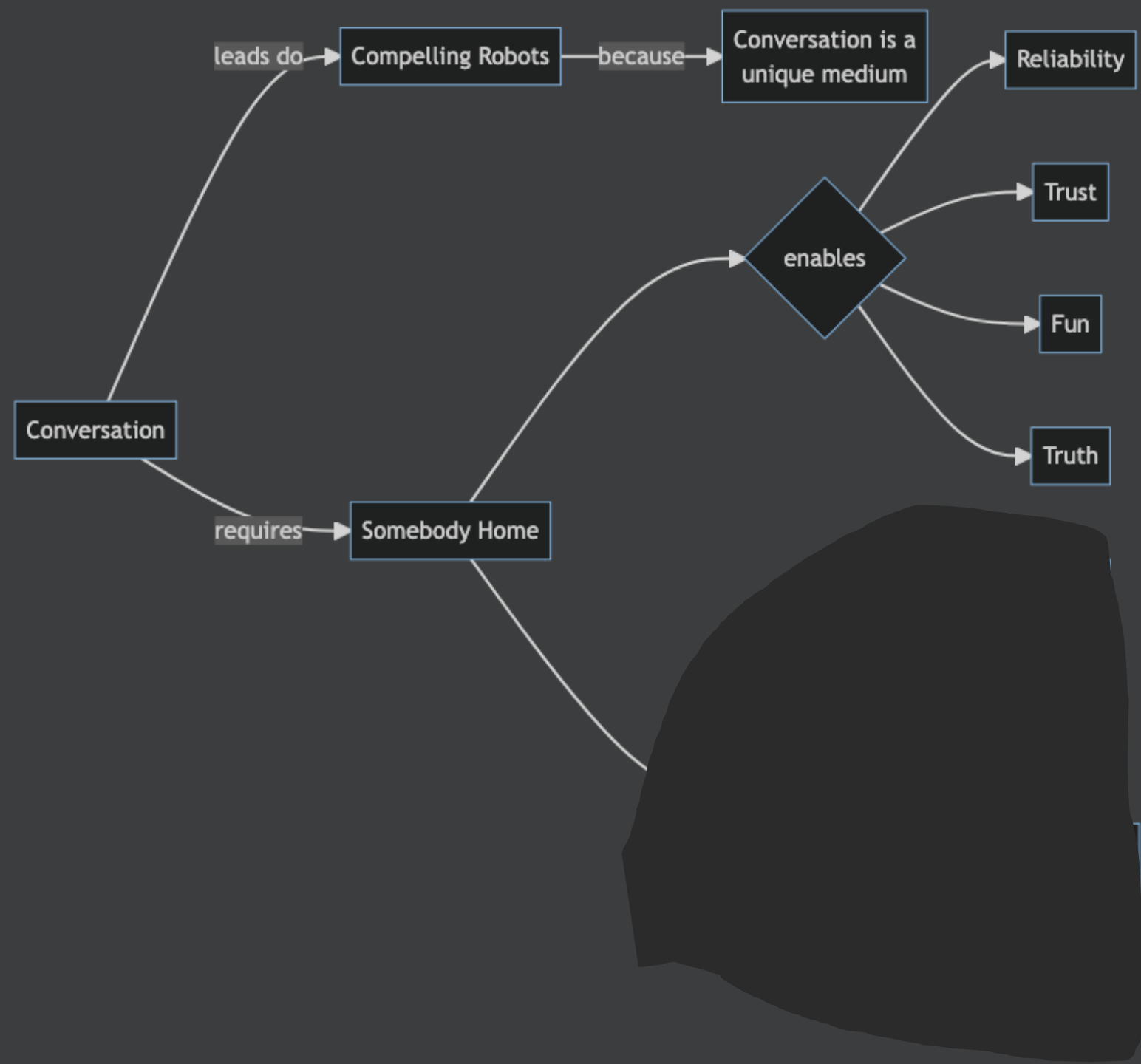


Why a conversation partner needs to be “somebody”

Reliability: If it represents the world the way that we do less likely to make a catastrophic mistake
(since its knowledge structure is built on our representation)

Trust: need the mistakes it does make to make sense to us





Why a conversation partner needs to be “somebody”

Fun: much more interesting to talk to somebody.

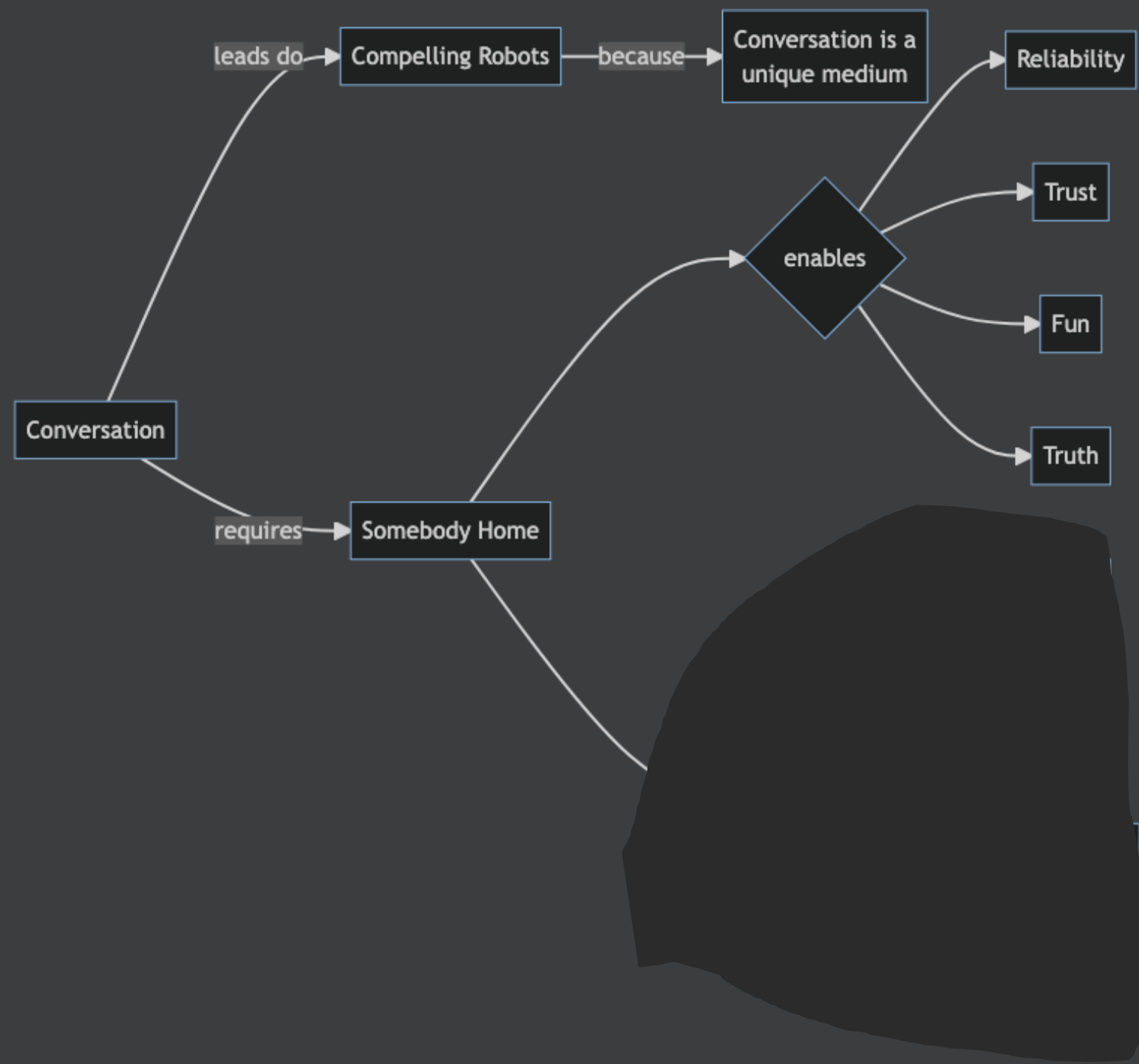
Consider the humor

- Household robot: should have some goal obscure to the owner, like tapping on the walls. “What the hell is it doing?”
- NPC: “You’re playing a video game, right? Can you get me out of here?”

Why a conversation partner needs to be “somebody”

Truth: most important

- Truth comes from interaction with the real world
- A person seeks to bend the world to their goals
 - A person strives for things
- ChatGPT: you can kind-of have a conversation with it, but it isn't there



Somebody home requires

State

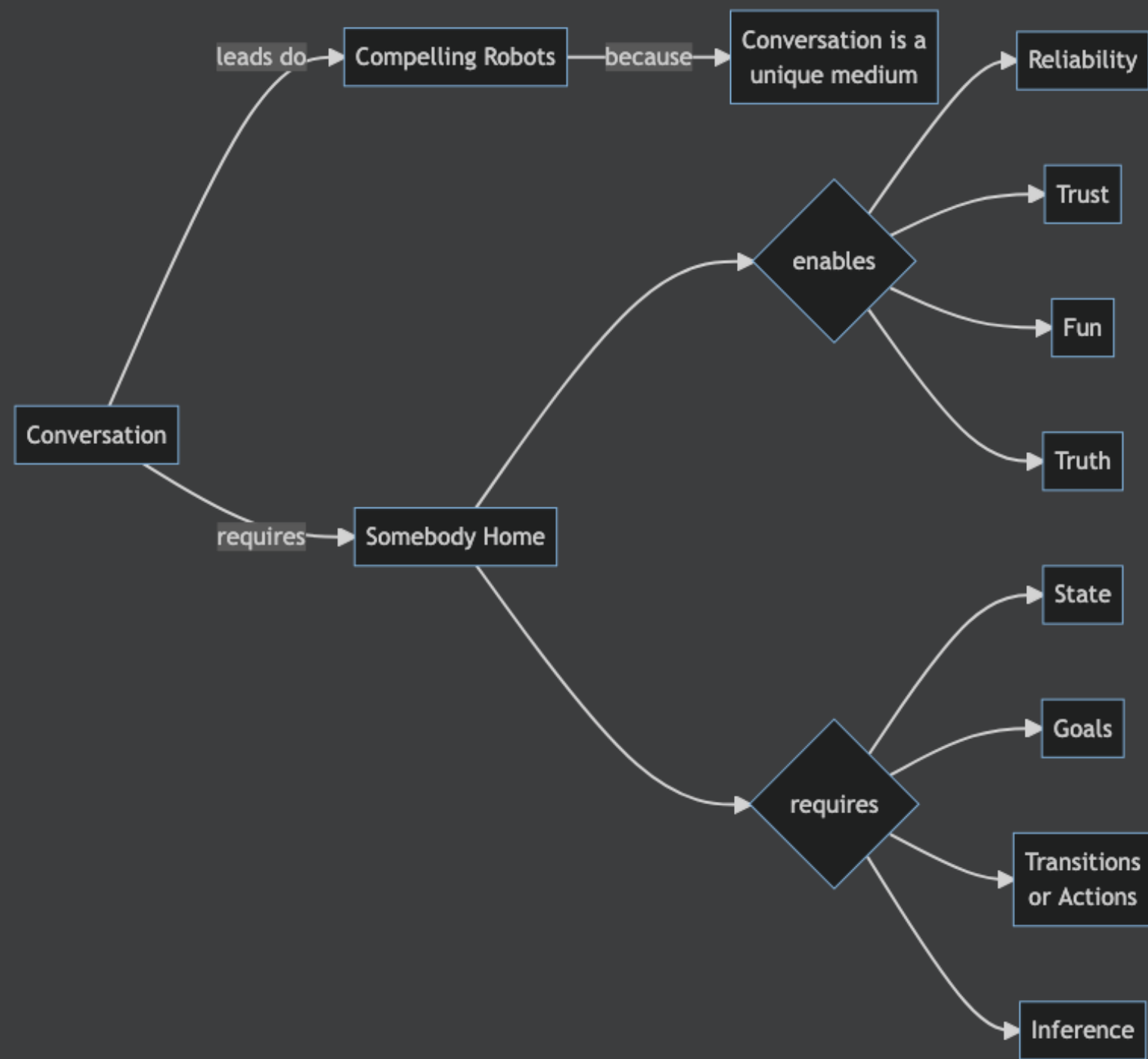
- The world around it

Goals

- Transitions enable one to talk about goals
- Why would it talk at all?
 - Thinking beyond natural-language interface

Transitions (actions)

- How the world changes

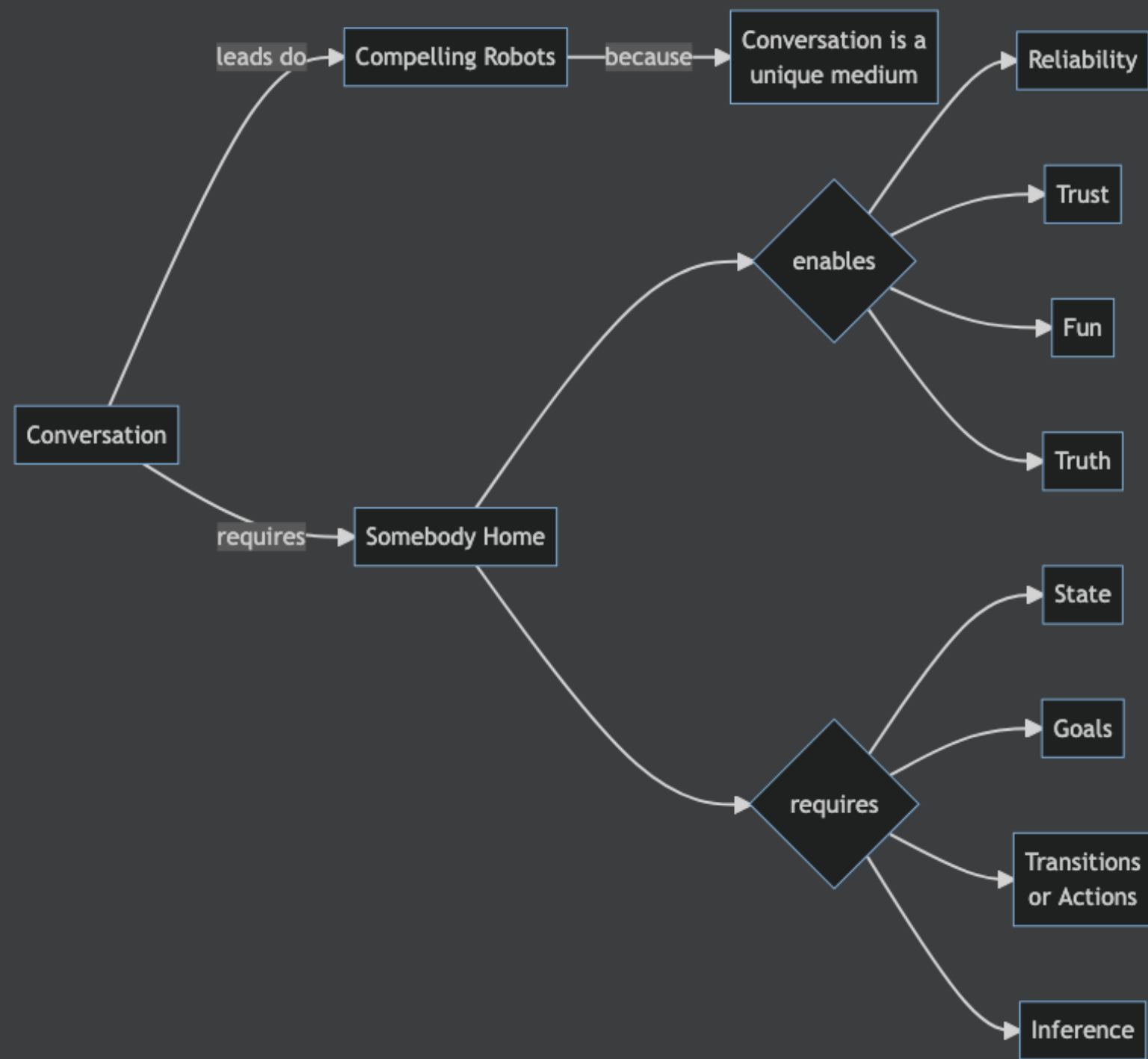


Somebody home requires

Inference

- Inference is mapping a continuous situation to some space of values
- Keeps us from being brittle
- See Eiffel tower, in Paris

- Back to Kant and before
 - He calls it Judgment
 - <https://plato.stanford.edu/entries/kant-judgment/>



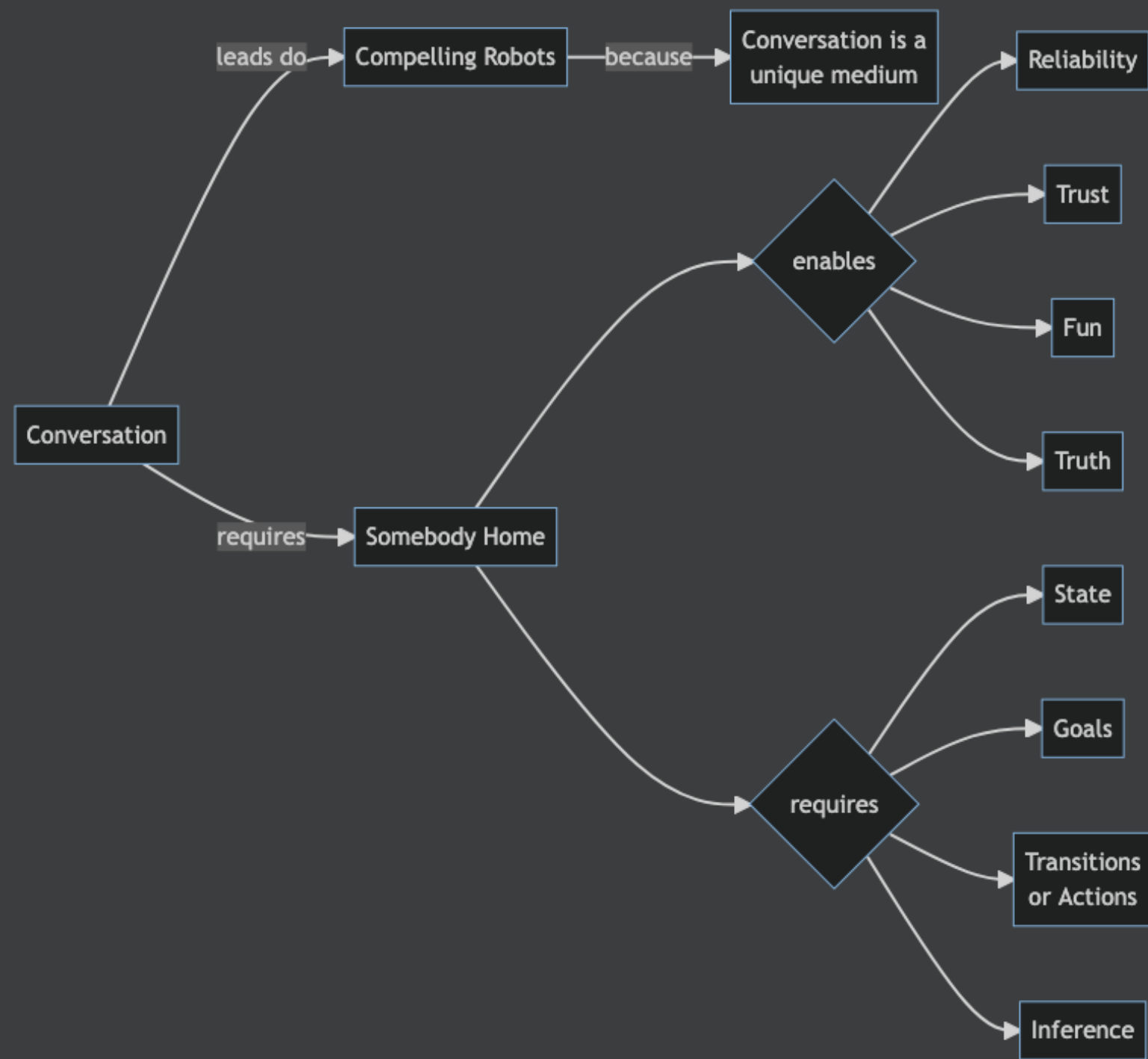
Somebody home requires

Inference

- Inference is mapping a continuous situation to some space of values
- Keeps us from being brittle
- See Eiffel tower, in Paris

“The Beginning of Infinity,” by David Deutsch

- “People” are entities that can create explanations.
- When we see stars, we don’t see what they are, we just see points of light.



Inference has a precise formulation

it's the Bayesian equation

$$P(\text{hypothesis}|\text{observation}) = \frac{P(\text{observation}|\text{hypothesis})P(\text{hypothesis})}{P(\text{observation})}$$

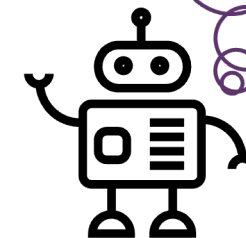
$$P(\text{hypothesis}|\text{observation}) \propto P(\text{observation}|\text{hypothesis})P(\text{hypothesis})$$

$P(\text{observation}|\text{hypothesis})$ is your model of the world
Hypothesis: is an explanation in the David Deutsch sense

“A big burning ball of hydrogen and helium burning millions of miles away might look like that”

Or more simply

A table
on a table



Inference has a precise formulation

it's the Bayesian equation

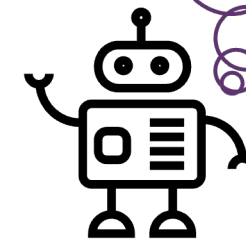
$$P(\text{hypothesis}|\text{observation}) = \frac{P(\text{observation}|\text{hypothesis})P(\text{hypothesis})}{P(\text{observation})}$$

$$P(\text{hypothesis}|\text{observation}) \propto P(\text{observation}|\text{hypothesis})P(\text{hypothesis})$$

$P(\text{observation}|\text{hypothesis})$ must be composable and you have to come up with it on the fly

Or more simply

A table
on a table



Inference leads to “meaning”

Reminiscent of Robert Brandom through Richard Evans

<https://www.doc.ic.ac.uk/~re14/Evan-s-R-2020-PhD-Thesis.pdf>



“The table is on the table.”



- If you want to pick up top table, you must first walk to the table
- If you push the bottom table, the top table will fall
- Party guests will think this is weird looking.

Inference leads to “meaning”



“The table is on the table.” → *What?*

Inference leads to “meaning”



Two ways to not understand in a conversation:

1. Wrong mental scene
2. Not knowing the possibilities

Inference leads to “meaning”



$P(\text{mental scene} | \text{stimuli})$

$\propto P(\text{stimuli} | \text{mental scene})P(\text{mental scene})$

Meaning

$P(\text{next state} | \text{mental scene})$

Transition model

$P(\text{stimuli} | \text{mental scene})$

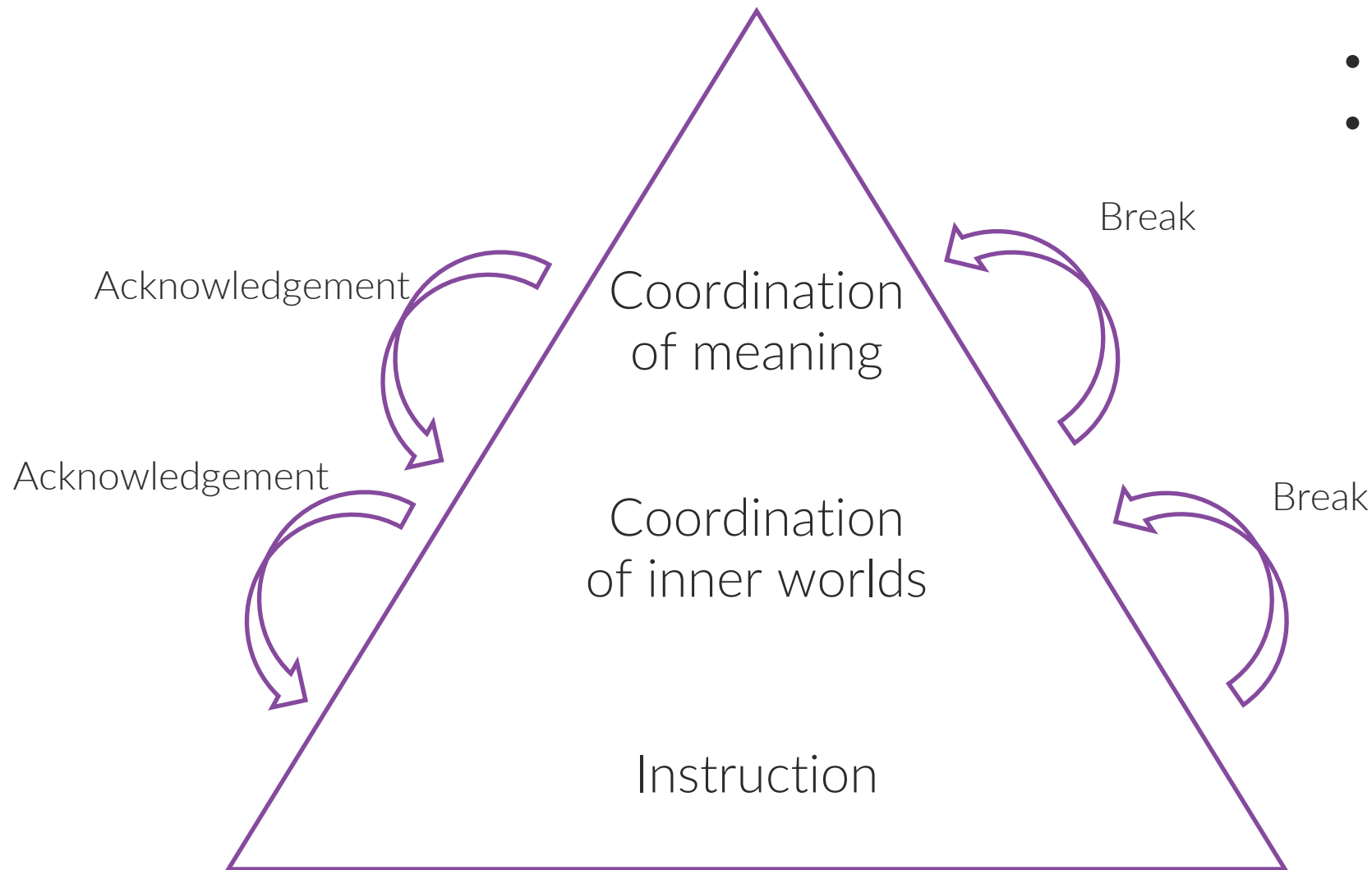
State model

$P(\text{mental scene})$

State prior

- These two models must be composable.
- They must be dynamic and can't be precomputed, as we will see.

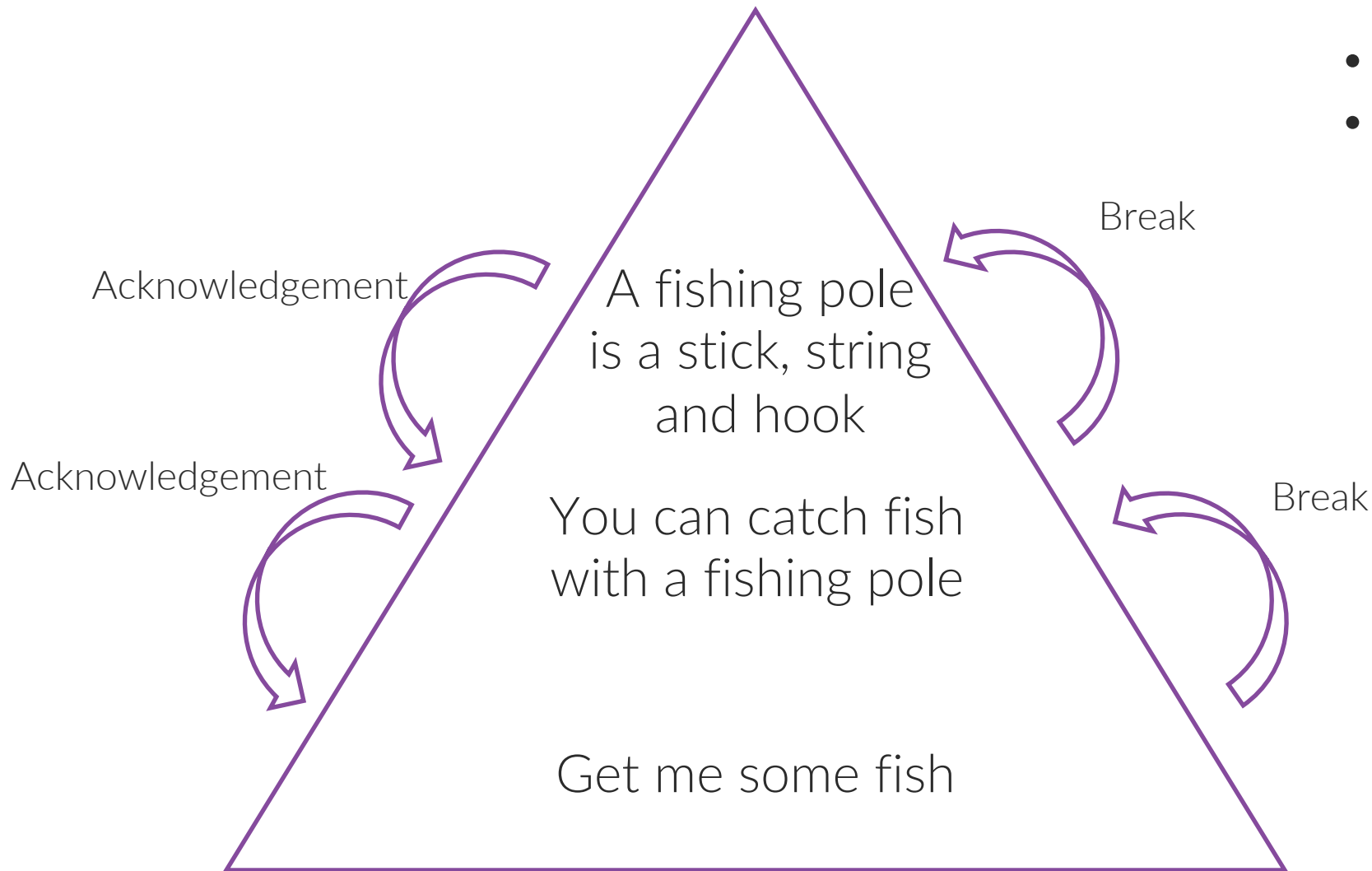
Inference for conversation requires dynamic composition



- Levels of discourse
- Complicated to go up and down the pyramid

Modified from
Gärdenfors (2014),
which was based on
Winter (1998)

Inference for conversation requires dynamic composition



- Levels of discourse
- Complicated to go up and down the pyramid

Modified from
Gärdenfors (2014),
which was based on
Winter (1998)

Inference is over more than words: conversation has its own rules (pragmatics)

- Conversational maxims: Grice (1975, 1978)
- Breaking these rules is a way to communicate more than the meaning of the words.

Maxim of Quantity: Say only what is not implied.

Yes: "Bring me the table."

No: "Bring me the table by transporting it to my location."

What did she mean by that?

Maxim of Quality: Say only things that are true.

Yes: "I hate carrying tables."

No: "I love carrying tables, especially when they are covered in fire ants."

She must be being sarcastic.

Maxim of Relevance: Say only things that matter.

Yes: "Bring me the table."

No: "Bring me the table and birds sing."

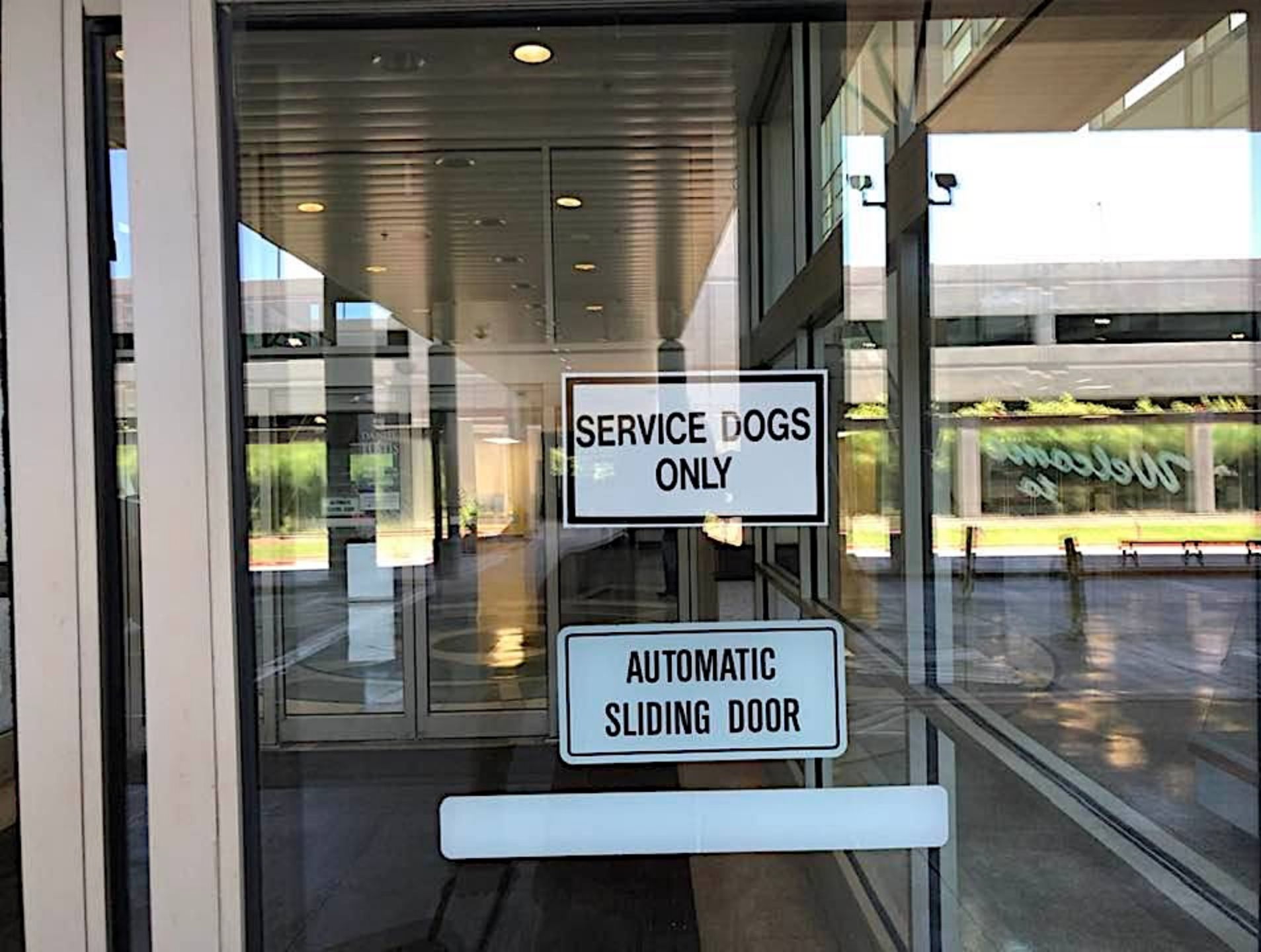
What did she mean by that?

Maxim of Manner: Speak in a way that can be understood.

Yes: "Bring me the table."

No: "Use personal physical force to levitate the table and transport it to me."

What did she mean by that?



Words are only hints at possible meanings

When I saw this, my first thought was, “Where do people enter?”

Our inference models need to compose for maximum coverage

Multiplicative: Most general sense, a bird that can drive a car

Foundational Metaphors (See Mark Johnson and others. Steven Pinker talks about two main ones)

- Force
 - An offer or a person can be *attractive*
 - A broken air conditioner can *force* you to move a meeting
- Location in space
 - AI has come *a long way* in the last 15 years

Analogies (Melanie Mitchell and Douglas Hofstadter)

- We can broadly apply the story of sour grapes (Hofstadter in *Surfaces and Essences*)

Conceptual Blending (*The Way We Think* by Gilles Fauconnier and Mark Turner)

- That running back is a truck
- Dall-e 2 could generate a good picture of this, but it couldn't imagine what it is like to tackle a vehicle

OpenAI Dall-e 2

- Trained on combinations of images and text; [CLIP](#) and diffusion models
- Can create images from whatever you type

<https://openai.com/dall-e-2/>

Thanks to you [@hardmaru](#) for the images!

Prompt:

“Darth Vader on the cover of Vogue magazine”



Outline

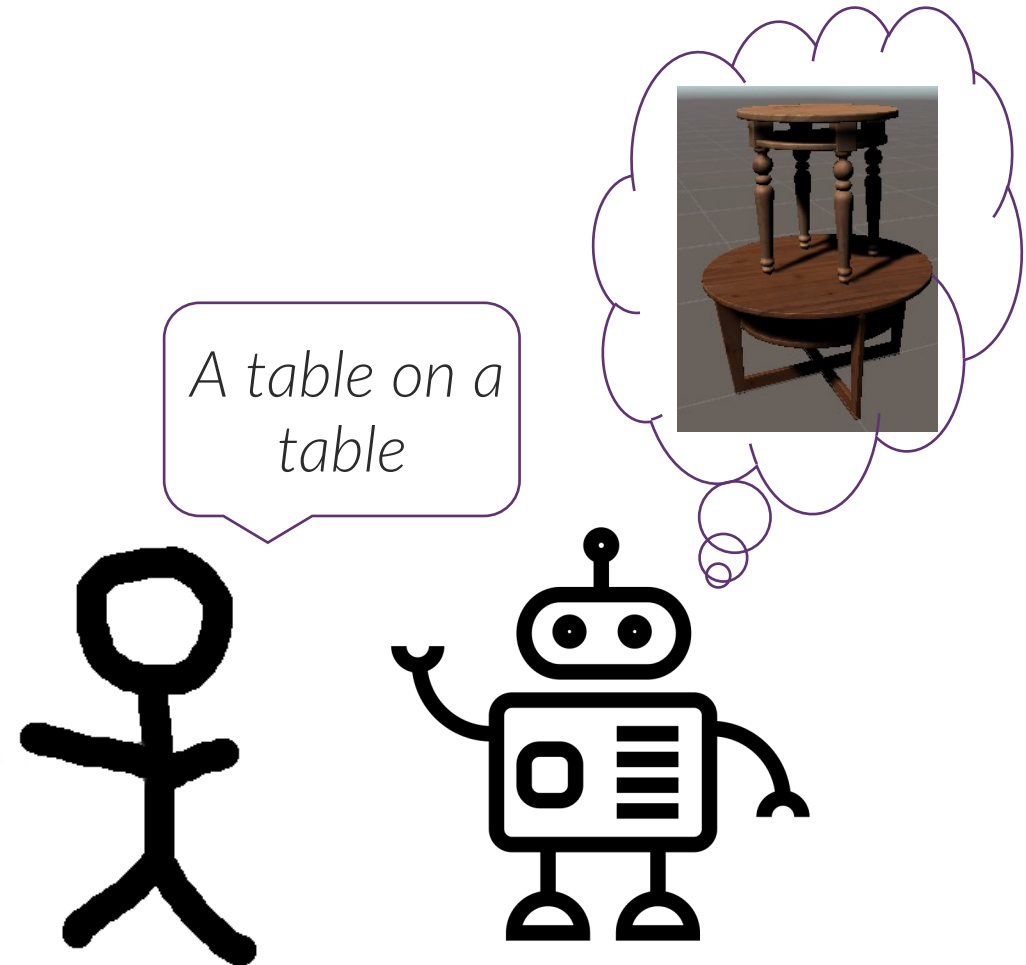
- Conversation makes compelling robots and what conversation requires
- The symbolic path of autonomous simulation
- The sub-symbolic path of coaxing neural networks

Outline

- Conversation makes compelling robots and what conversation requires
- The symbolic path of autonomous simulation
- The sub-symbolic path of coaxing neural networks

Robots can use simulation for inference and meaning

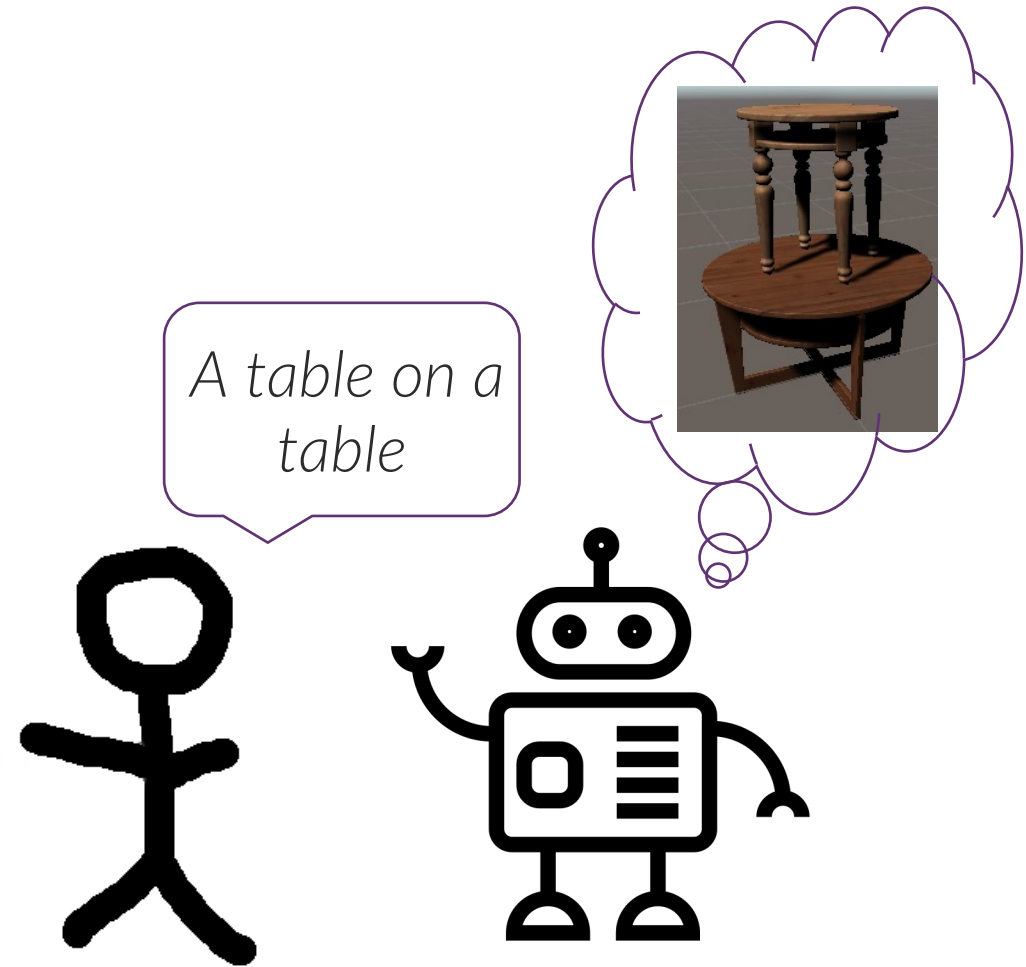
- One way to build supple intelligence is to have a robot build a simulation of its environment.
- The robot simulates what you say in something like Unity.



See "Computers could understand natural language using simulated physics" from 2017 for additional details.
<https://chatbotlife.com/computers-could-understand-natural-language-using-simulated-physics-26e9706013da>

Robots can use simulation for inference and meaning

- Simulation enables robust inference because it provides an unbroken description of the dynamics of the environment, even if it is not complete in full detail.
- This unbroken description is a **grounding**.



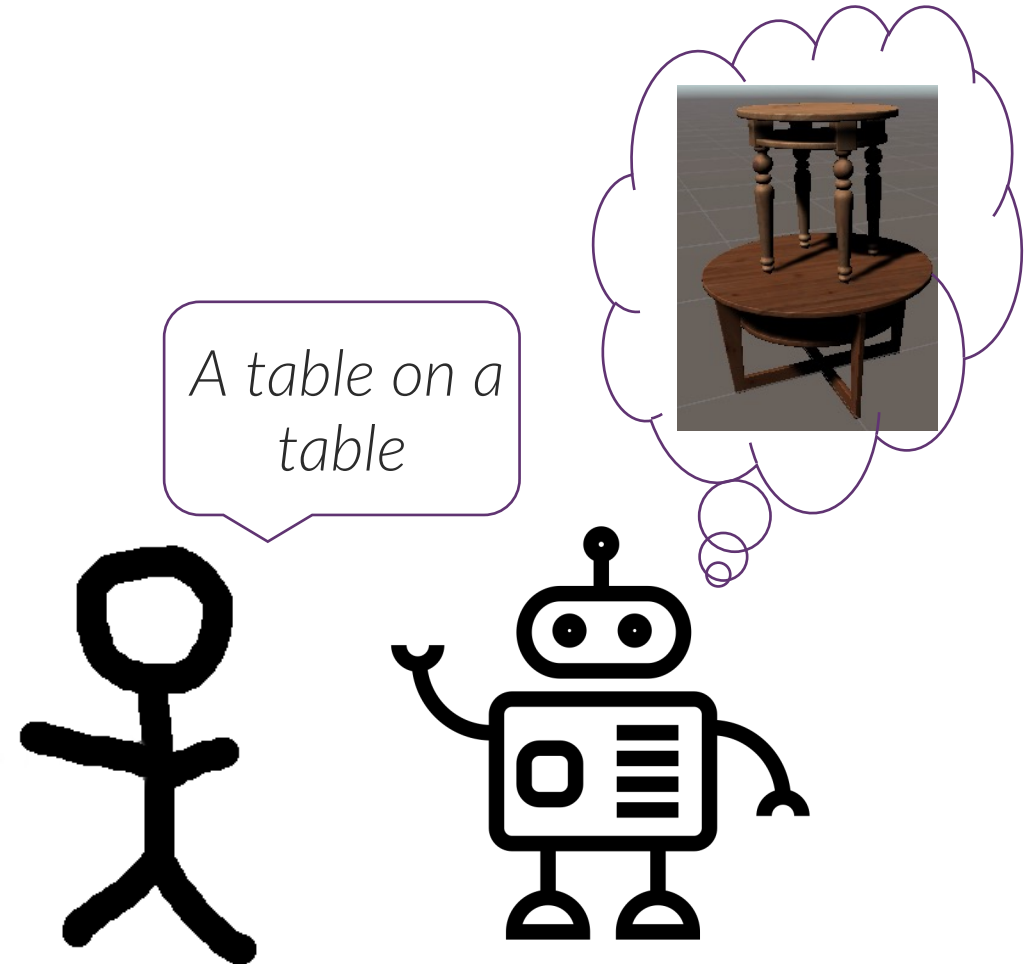
See "Computers could understand natural language using simulated physics" from 2017 for additional details.
<https://chatbotlife.com/computers-could-understand-natural-language-using-simulated-physics-26e9706013da>

Robots can use simulation for inference and meaning

What will happen if I push the bottom table?



Well, mentally push it and see.



See "Computers could understand natural language using simulated physics" from 2017 for additional details.
<https://chatbotlife.com/computers-could-understand-natural-language-using-simulated-physics-26e9706013da>

Robots can use simulation for inference and meaning

Classic AI question, can birds fly?

$$\forall x \text{ bird}(x) \rightarrow \text{can_fly}(x)$$

Okay, okay

$$\forall x \text{ bird}(x) \wedge \text{flying_bird}(x) \rightarrow \text{can_fly}(x)$$

Okay, okay

$$\forall x \text{ bird}(x) \wedge \text{flying_bird}(x) \wedge \text{not broken_wing}(x) \rightarrow \text{can_fly}(x)$$



WKRP "As God as my witness, I thought turkeys could fly"

https://www.youtube.com/watch?v=lf3mgmEdfwg&ab_channel=EpicHouston

Okay, okay, what if it is covered in maple syrup?

The AI must build its own simulations



“The table is on the table.”

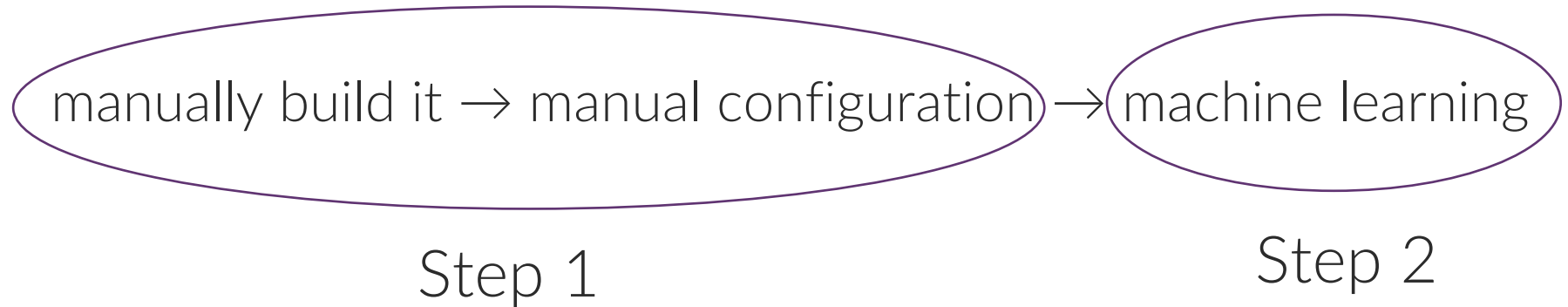


- If you want to pick up top table, you must first walk to the table
- If you push the bottom table, the top table will fall
- Party guests will think this is weird looking.

How to Build an AI that Imagines through Simulation

We don't have to put as much information in as if we were using logic because the simulations will automatically combine things, but we still have a challenging problem ahead of us. How do we do it?

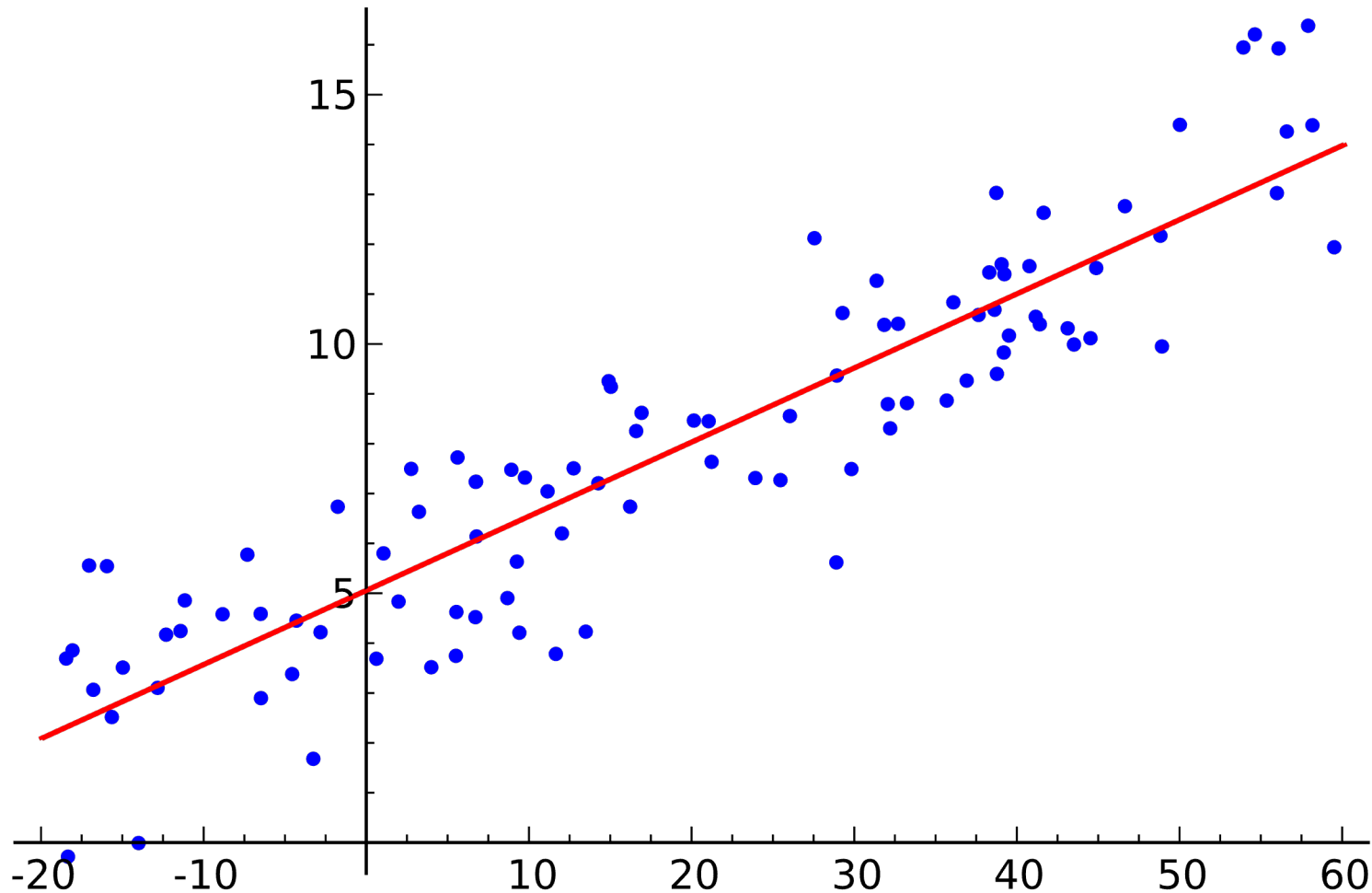
Gradually
working
our way
to real
machine
learning



Build it in two steps

1. Design hypothesis space
2. Human-like machine learning

The hypothesis space is the foundation of machine learning



Simplest hypothesis space, the space of all real values of x and b

$$y = mx + b$$

The hypothesis space is the foundation of machine learning

Recall from 2015

Neural networks are just millions of linear regressions with little nonlinear functions thrown in between the layers to keep it from squashing down.

Deep learning begins with a little function

It all starts with a humble linear function called a perceptron.

$$\begin{array}{r} \text{weight1} \times \text{input1} \\ + \text{weight2} \times \text{input2} \\ + \text{weight3} \times \text{input3} \\ \hline \text{sum} \end{array}$$

Perceptron:

If $\text{sum} > \text{threshold}$: output 1

Else: output 0

In math, with x being an input vector and w being a weight vector.

$$\text{sum}(x) = \sum_{i=1}^n w_i x_i = w^T x$$

Example: The inputs can be your data. Question: Should I buy this car?

$$\begin{array}{r} 0.2 \times \text{gas milage} \\ + 0.3 \times \text{horse power} \\ + 0.5 \times \text{num cup holders} \\ \hline \text{sum} \end{array}$$

If $\text{sum} > \text{threshold}$: buy car

Else: walk

But it doesn't have to be simple like that

Structured hypothesis spaces

- Space of Bayesian networks
- Space of probabilistic programs
- Space of antenna shapes
- Space of relational database schemas
- Space of DNA sequences

Once you set up the space, then you can search in that space

- Genetic algorithms
- Hillclimbing
- Backpropagation (like in neural networks, if the space is continuous)

Interlude: Best book if you are interested in the fundamentals of machine learning

You can read it for free!

Tom Mitchell,
“Machine Learning”

<http://www.cs.cmu.edu/~tom/mlbook.html>

So, we got to set up a particularly complicated hypothesis space

Domain, Classes, Relations <https://web.stanford.edu/~jurafsky/slp3/21.pdf>

Frames: frames, slots, classes, instances, types, and values

- My view: you can't get very far with formal representation and reasoning.
 - Even simple problems are hard.
 - Yale shooting problem https://en.wikipedia.org/wiki/Yale_shooting_problem
- Humans don't do it that way. We have to be taught *How the Mind Works*, Steven Pinker

So, it has to be some kind of embodied (problem specific) representation, but it needs to be structured enough that you can add to it without having to write Python code.

manually build it → manual configuration → machine learning

Step 1

Building the hypothesis space: Manual Configuration

The screenshot displays the ECG Workbench interface. The main window shows the sentence "robot1, dash to box1!" being analyzed. The central pane displays a hierarchical structure of the hypothesis space for the sentence, starting with "ROOT ('robot1, dash to box1!')". The structure includes a "DiscourseElement" with various slots: "speechAct" (5), "addressee" (8), "mood" (12) "Imperative", "speaker" (10), and "attentional_focus" (11). The "DiscourseElement" contains an "RD" (Robot Description) with slots for "number" (27), "amount" (21), "extensions" (18), "gender" (15), "referent" (25), "ontological-category" (14), "scale" (20), "extras" (22), "givenness" (19), "hedge" (17), and "bounding" (16). Below the "RD" is an "EventDescriptor" containing a "MotionPath" with slots for "actionary" (52) and "distance" (42). The left sidebar shows a list of grammar categories, and the right sidebar shows a tree view of the hypothesis space. The bottom of the interface includes a "Text Output" pane showing "SemSpec 1, cost -25.841331" and a "Debug Console" pane.

Protégé is another familiar example
<https://protege.stanford.edu/>

Embodied construction grammar <https://github.com/icsi-berkeley>

Build for composability

Multiplicative

- Build it so code reuses code with “duck” typing

Foundational Metaphors and Analogies

- Build it so code reuses code with “duck” typing
- Evolution progresses by finding new uses for existing structures

Conceptual Blending

- Focus on properties; create new objects using unions of properties.

2. Use the hypothesis space for machine learning

manually build it → manual configuration → machine learning

Step 2

With the right hypothesis space, we can do real machine learning. We don't need special algorithms; basic rule learning and decision-tree type stuff will be enough.

If A and B were there when great thing C happened, A and B are good and $A, B \rightarrow C$

Real (human-like) machine learning means you have all the pieces, you just have to label some subsets.

2. Use the hypothesis space for machine learning

manually build it → manual configuration → machine learning

Step 2

Different from the standard paradigm:

while not converged

run training examples through and update parameters based on errors

Real (human-like) machine learning means you have all the pieces, you just have to label some subsets.

Less of a pushing in continuous space and more of labeling groups.

2. Use the hypothesis space for machine learning

manually build it → manual configuration → machine learning

Step 2

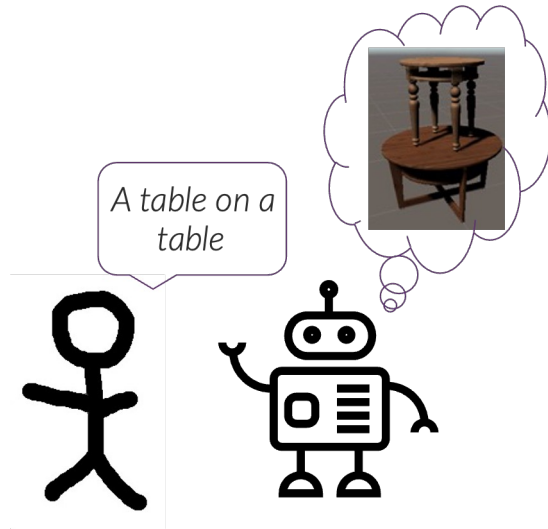
Real (human-like) machine learning also means you have all the pieces those pieces should causally fit together

If A and B were there when great thing C happened, A and B are good and $A, B \rightarrow C$

Based on other associations between A and B and C there should be a causal mechanism that ties them together

“Oh, that makes sense because $A \rightarrow M \rightarrow N$ and $B \rightarrow Q$ and, of course and everyone knows that $N \& Q \rightarrow C$ ”

How to build someone to talk to using the symbolic method



Build autonomous simulation through
manually build it → manual configuration → machine learning

You don't build an AI that can understand stories. You build an AI that can *build* stories. Then, it can "understand" stories by constructing a sequence of events that meets the constraints of the story.

Outline

- Conversation makes compelling robots and what conversation requires
- The symbolic path of autonomous simulation
- The sub-symbolic path of coaxing neural networks



how to get match_id for match pattern spacy



In the context of natural language processing with the spaCy library, a "match_id" is a unique identifier for a pattern match. This identifier can be used to refer to the specific match when working with the results of a matching operation.



To obtain the match_id for a specific pattern match, you can use the `Span.get_id()` method. This method is called on a `Span` object that represents the matched pattern, and it returns the match_id as an integer.

Here is an example of how you might use this method to get the match_id for a pattern match:

```
import spacy
from spacy.matcher import Matcher

# Load a spaCy model and create the matcher
nlp = spacy.load("en_core_web_sm")
matcher = Matcher(nlp.vocab)

# Define a pattern to match
pattern = [{"LOWER": "example"}, {"LOWER": "pattern"}]

# Add the pattern to the matcher and apply it to a text
matcher.add("EXAMPLE_PATTERN", None, pattern)
doc = nlp("This is an example of a pattern match.")
matches = matcher(doc)

# Iterate over the matches and print the match_id for each one
for match_id, start, end in matches:
    matched_span = doc[start:end]
    print(matched_span.get_id())
```

In this example, the output would be the integer `0`, which is the match_id for the first (and in this case, only) pattern match that was found in the text.

ChatGPT is amazing



how do you match the id 0 with "EXAMPLE_PATTERN"



To match a specific pattern by its `match_id`, you can use the `Matcher.get_pattern_id()` method. This method is called on a `Matcher` object, and it takes the `match_id` as an argument. It returns the pattern definition (i.e. the sequence of tokens that the pattern is intended to match) associated with the specified `match_id`.



Here is an example of how you might use this method to get the pattern associated with the `match_id` `0`:

```
import spacy
from spacy.matcher import Matcher

# Load a spaCy model and create the matcher
nlp = spacy.load("en_core_web_sm")
matcher = Matcher(nlp.vocab)

# Define a pattern to match
pattern = [{"LOWER": "example"}, {"LOWER": "pattern"}]

# Add the pattern to the matcher and apply it to a text
matcher.add("EXAMPLE_PATTERN", None, pattern)
doc = nlp("This is an example of a pattern match.")
matches = matcher(doc)

# Iterate over the matches and print the pattern associated with each one
for match_id, start, end in matches:
    pattern_id = matcher.get_pattern_id(match_id)
    print(pattern_id)
```

Copy code

In this example, the output would be the pattern that we defined earlier: `[{"LOWER": "example"}, {"LOWER": "pattern"}]`. This is the pattern that was matched by the `Matcher` when it was applied to the text.

Try again

ChatGPT is amazing



spacy get_pattern_id



All Images Shopping Videos News More Tools

About 1,240 results (0.43 seconds)

https://spacy.io > api > top-level

Top-level Functions · spaCy API Documentation

spaCy is a free open-source library for Natural Language Processing in Python. It features NER, POS tagging, dependency parsing, word vectors and more.

Missing: `get_pattern_id` | Must include: `get_pattern_id`

https://spacy.io > usage > processing-pipelines

Language Processing Pipelines · spaCy Usage Documentation

spaCy is a free open-source library for Natural Language Processing in Python. It features NER, POS tagging, dependency parsing, word vectors and more.

Missing: `get_pattern_id` | Must include: `get_pattern_id`

https://spacy.io > models

Trained Models & Pipelines · spaCy Models Documentation

Downloadable trained pipelines and weights for spaCy.

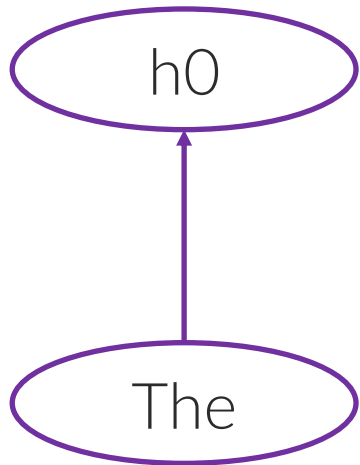
Missing: `get_pattern_id` | Must include: `get_pattern_id`

You've visited this page 2 times. Last visit: 12/3/22

ChatGPT is amazing but it has no sense of truth

Encoding sentence meaning into a vector

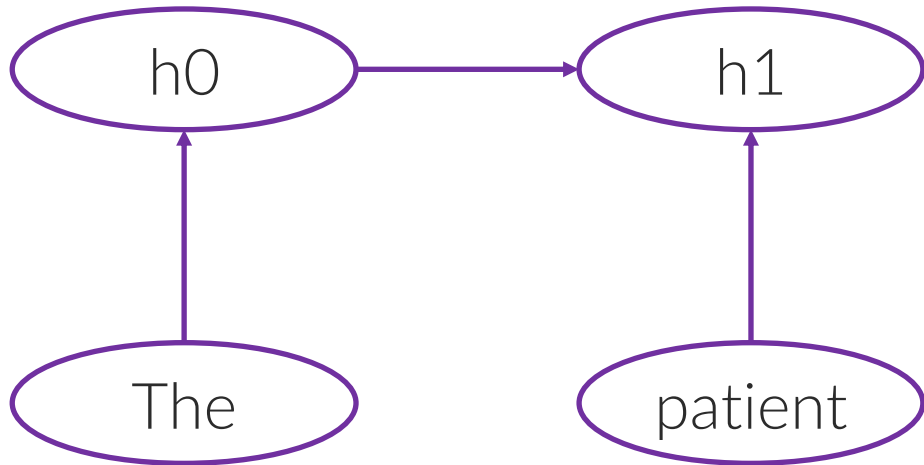
“The patient fell.”



Using a recurrent neural network (RNN).

Encoding sentence meaning into a vector

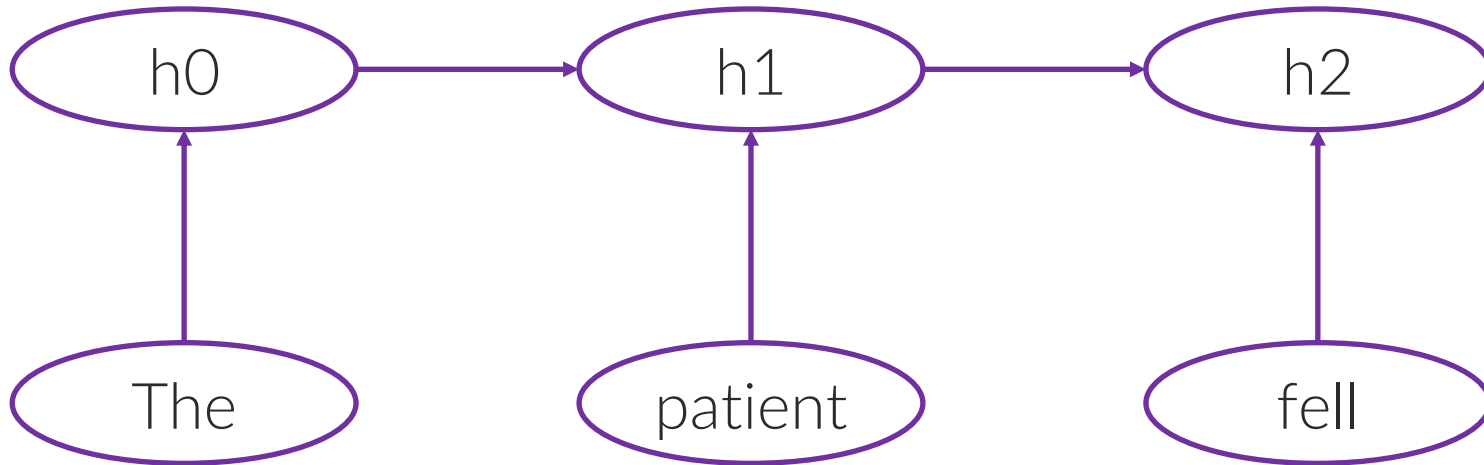
“The patient fell.”



Using a recurrent neural network (RNN).

Encoding sentence meaning into a vector

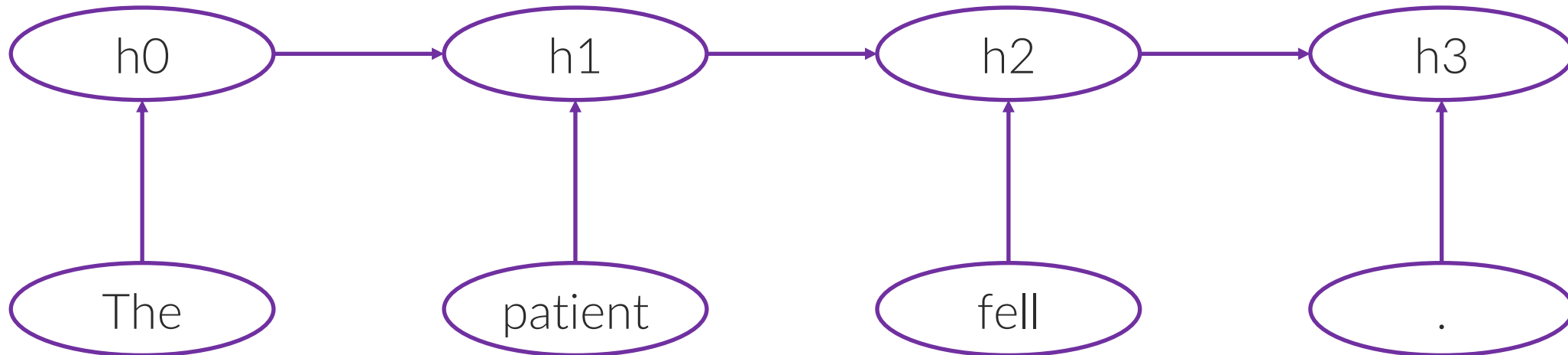
“The patient fell.”



Using a recurrent neural network (RNN).

Encoding sentence meaning into a vector

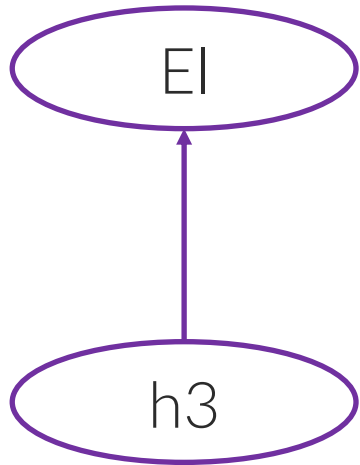
“The patient fell.”



RNN is like a hidden Markov model but doesn't make the Markov assumption and benefits from a vector representation.

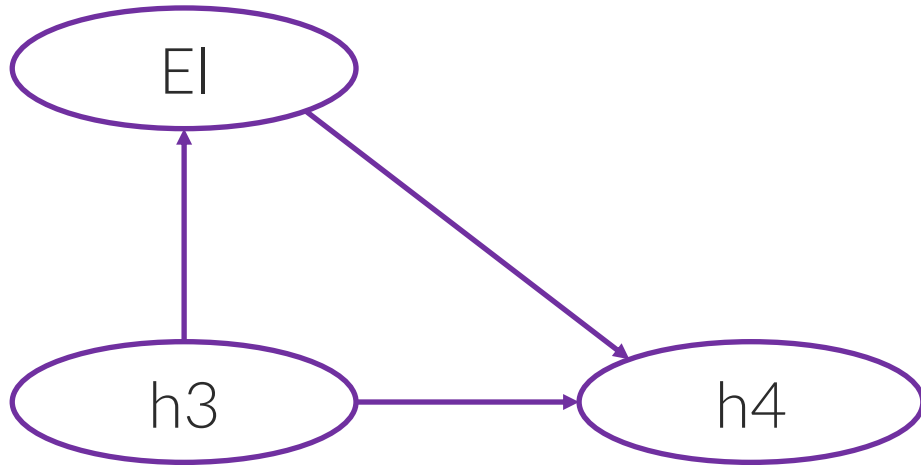
Decoding sentence meaning

Machine translation.



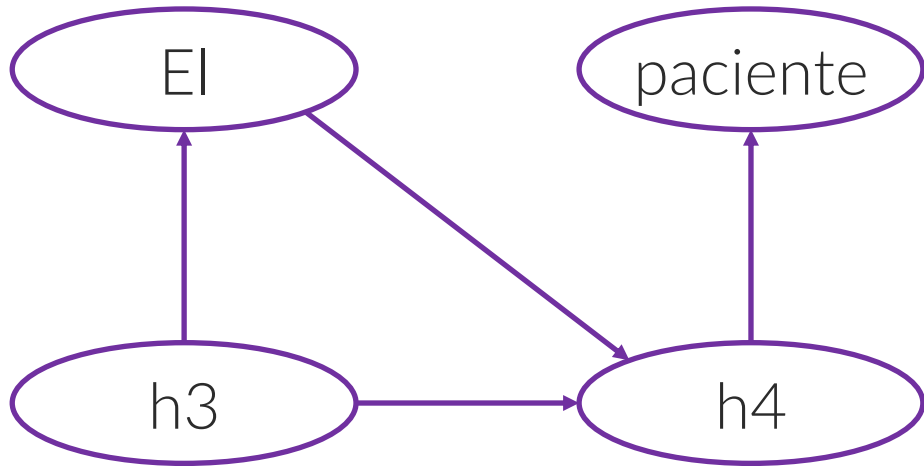
Decoding sentence meaning

Machine translation.



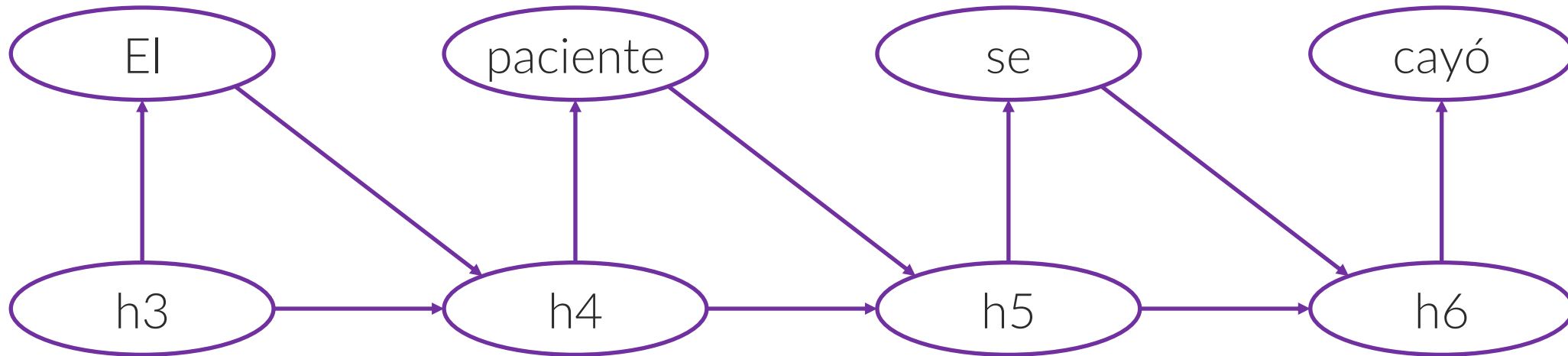
Decoding sentence meaning

Machine translation.



Decoding sentence meaning

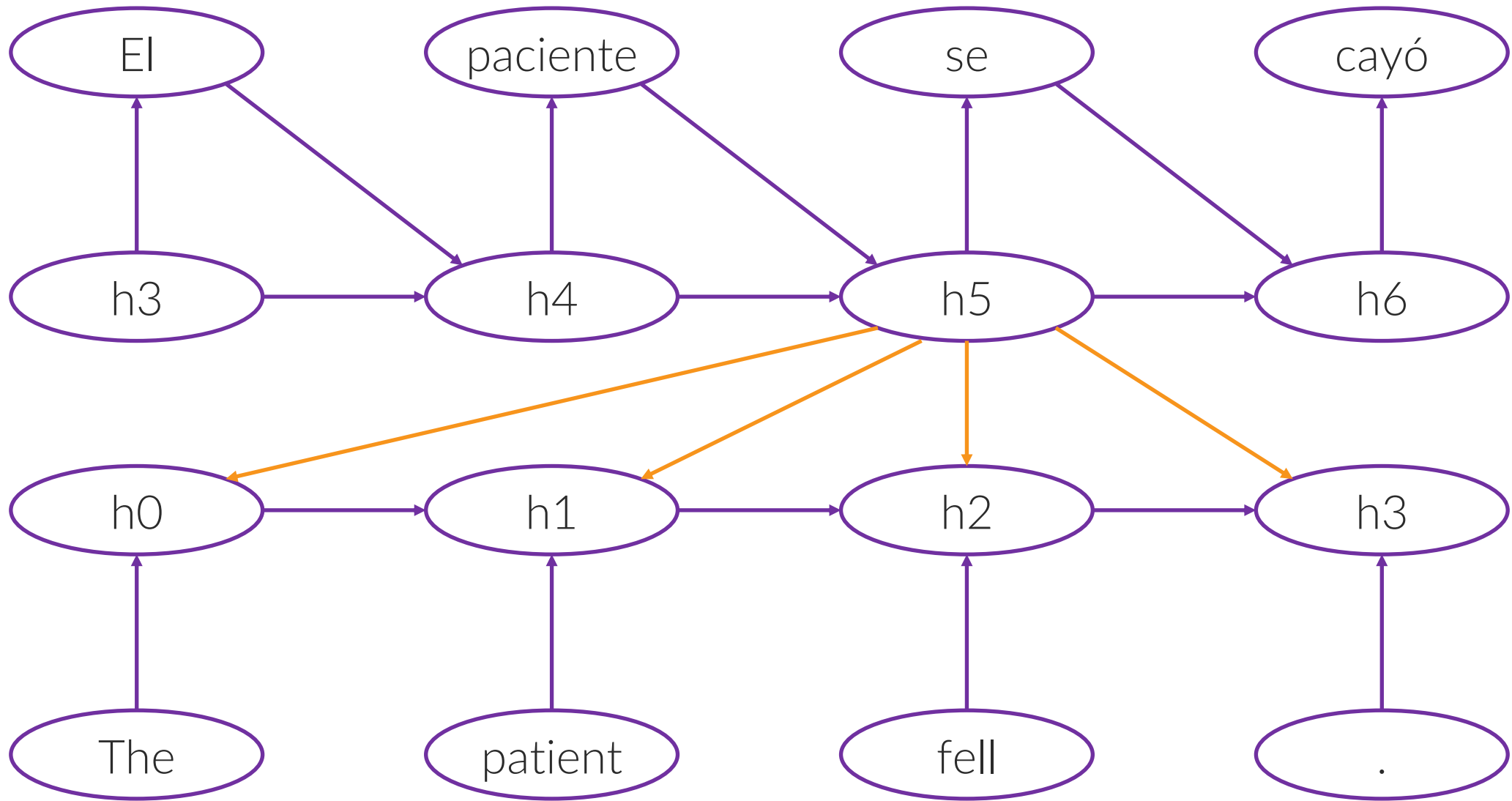
Machine translation.



[Cho et al., 2014]

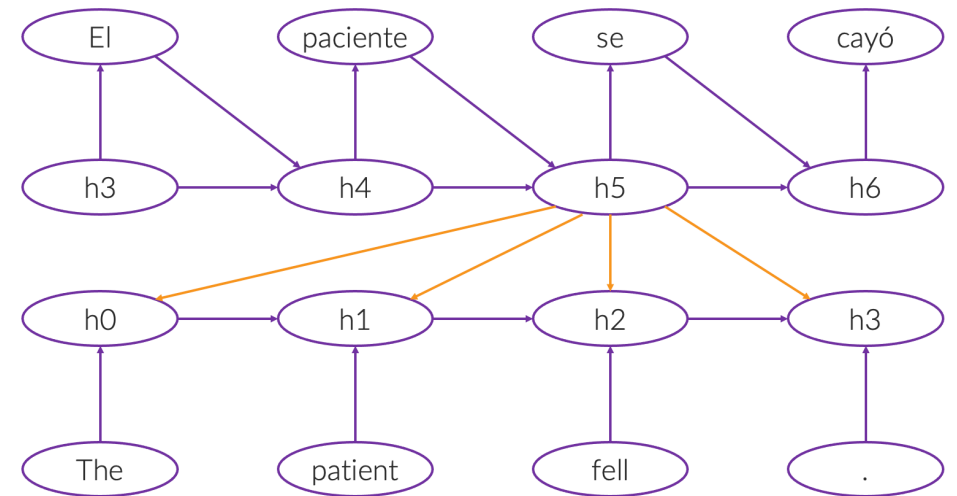
- It keeps generating until it generates a stop symbol.
- It is using a kind of interpolation from a huge set of training data.

Attention [Bahdanau et al., 2014]

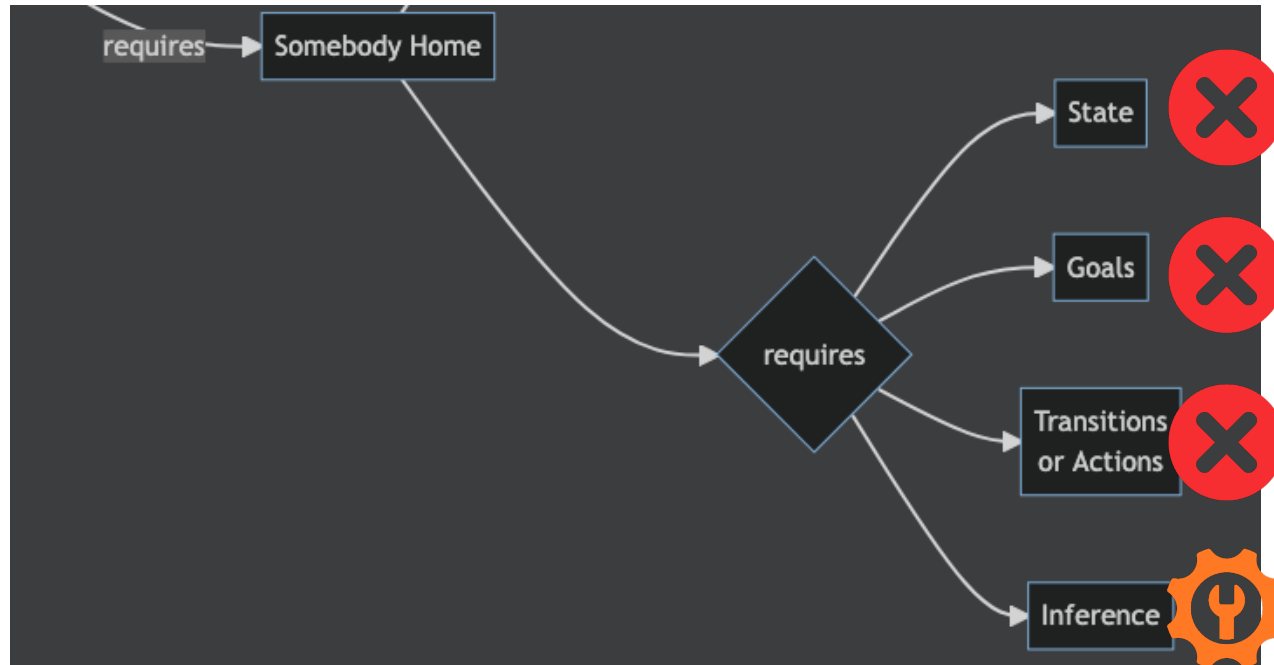


Transformers: Attention is all you need

<https://arxiv.org/abs/1706.03762>



You can't have truth without world interaction



We need to build an AI that knows what a toddler knows before we can build one that understands Wikipedia.

Although ChatGPT could interact with a virtual machine
<https://www.engraved.blog/building-a-virtual-machine-inside/>

Language is a representation medium for the world—it isn't the world itself.

- When we talk, we only say what can't be inferred because we assume the listener has a basic understanding of the dynamics of the world (e.g., if I push a table the things on it will also move)

The inferences are not about the world

Is this stove hot? Ouch!

Imitation Learning from Video with Transformers

Robots can watch YouTube and learn to imitate, analogous to ChatGPT

Multimodal, language and object manipulation

The trick is the *tokenization* of events in video, but Google has made some good progress

Robotics Transformer 1 (RT-1)

- Transformer model trained by copying demonstrations
- Predict the next most likely action based on what it has learned from the demonstrations

<https://blog.google/technology/ai/helping-robots-learn-from-each-other/>

<https://ai.googleblog.com/2022/12/rt-1-robotics-transformer-for-real.html>

<https://robotics-transformer.github.io/assets/rt1.pdf>



noah mugan



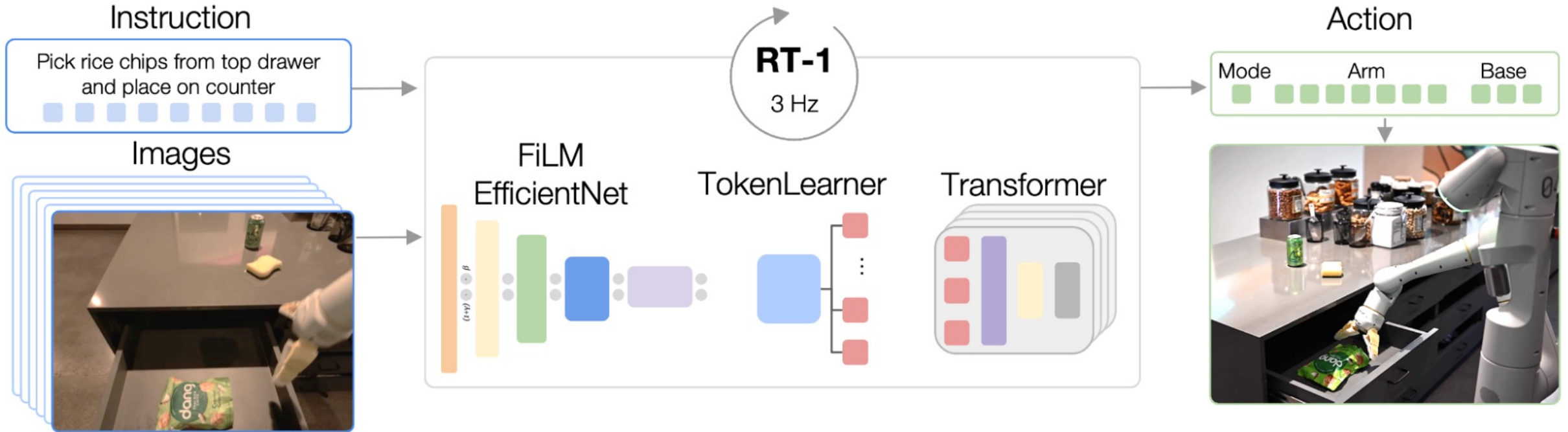
Hero Factory 4.0 Review-Surge



Noah Mugan
16 subscribers



Imitation Learning from Video with Transformers



The trick is the tokenization of events in video, but Google has made some good progress

Image used with permission.
Thanks Keerthana Gopalakrishnan!
@keerthanpg

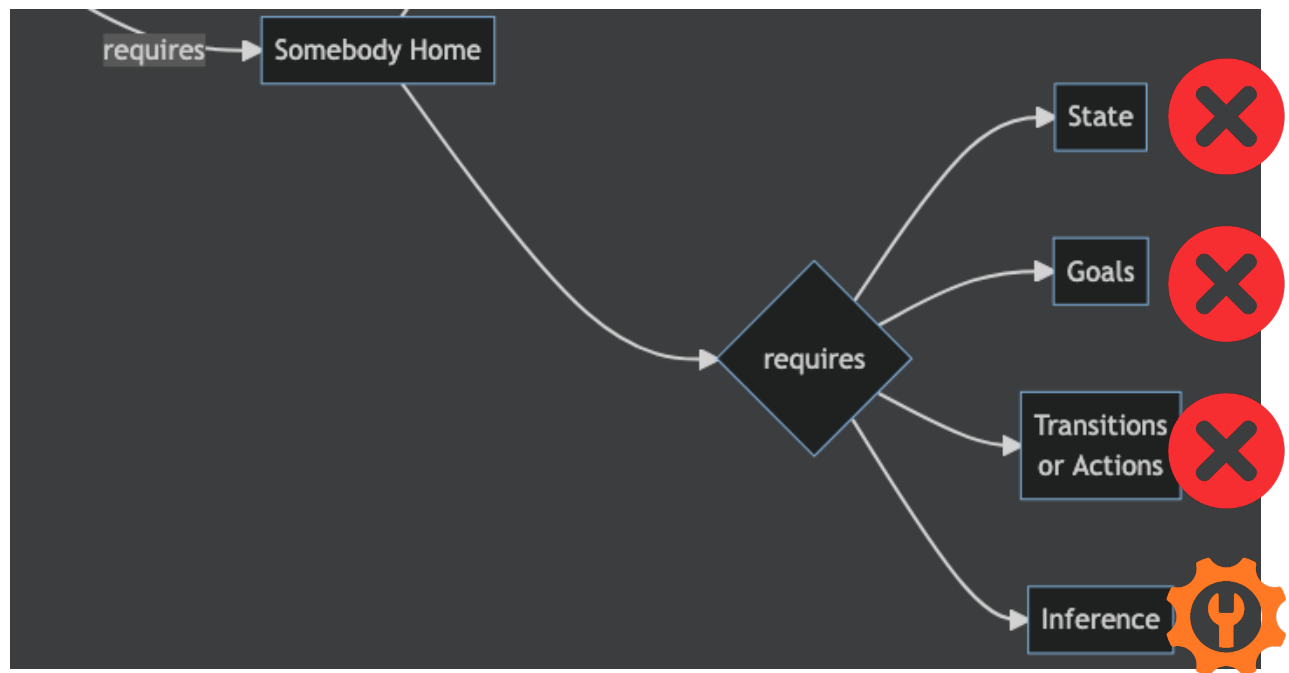
Robotics Transformer 1 (RT-1)

- Transformer model trained by copying demonstrations
- Predict the next most likely action based on what it has learned from the demonstrations

<https://blog.google/technology/ai/helping-robots-learn-from-each-other/>

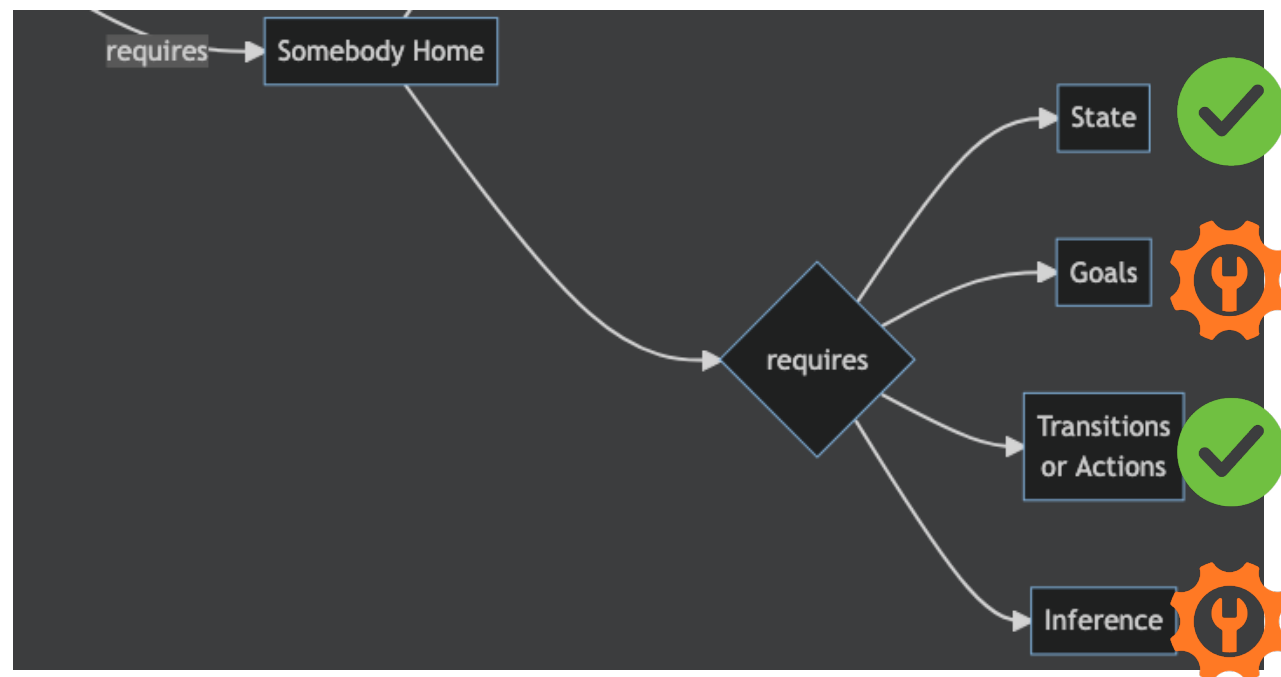
<https://ai.googleblog.com/2022/12/rt-1-robotics-transformer-for-real.html>

<https://robotics-transformer.github.io/assets/rt1.pdf>



Recall the
report card for
transformer
trained on text

Imitation Learning from Video with Transformers



Using robots in the real or simulated world provides State.

Goals are weak since it is imitation learning.

We have transitions/actions that happen in the real world.

Also limited due to imitation learning

Reinforcement learning trained in simulation

Instead of building a system that can learn to build its own simulations, we train robots in simulated worlds that we build



AI2Thor by Allen AI
<https://ai2thor.allenai.org/>

Microsoft Flight Simulator
<https://www.flightsimulator.com/>

Reinforcement learning trained in simulation

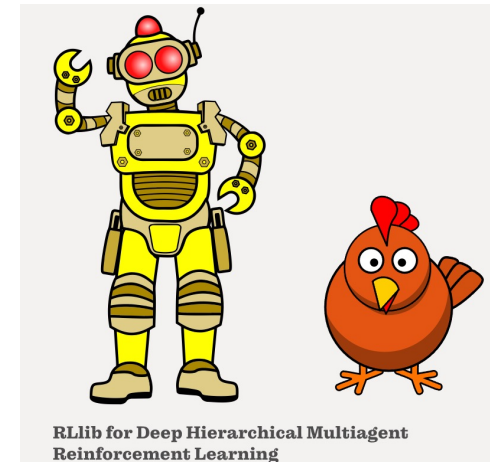
Instead of building a system that can learn to build its own simulations, we train robots in simulated worlds that we build

There has been some exciting progress in the area from DeepMind and others
<https://www.deepmind.com/blog/from-motor-control-to-embodied-intelligence>

Soccer!

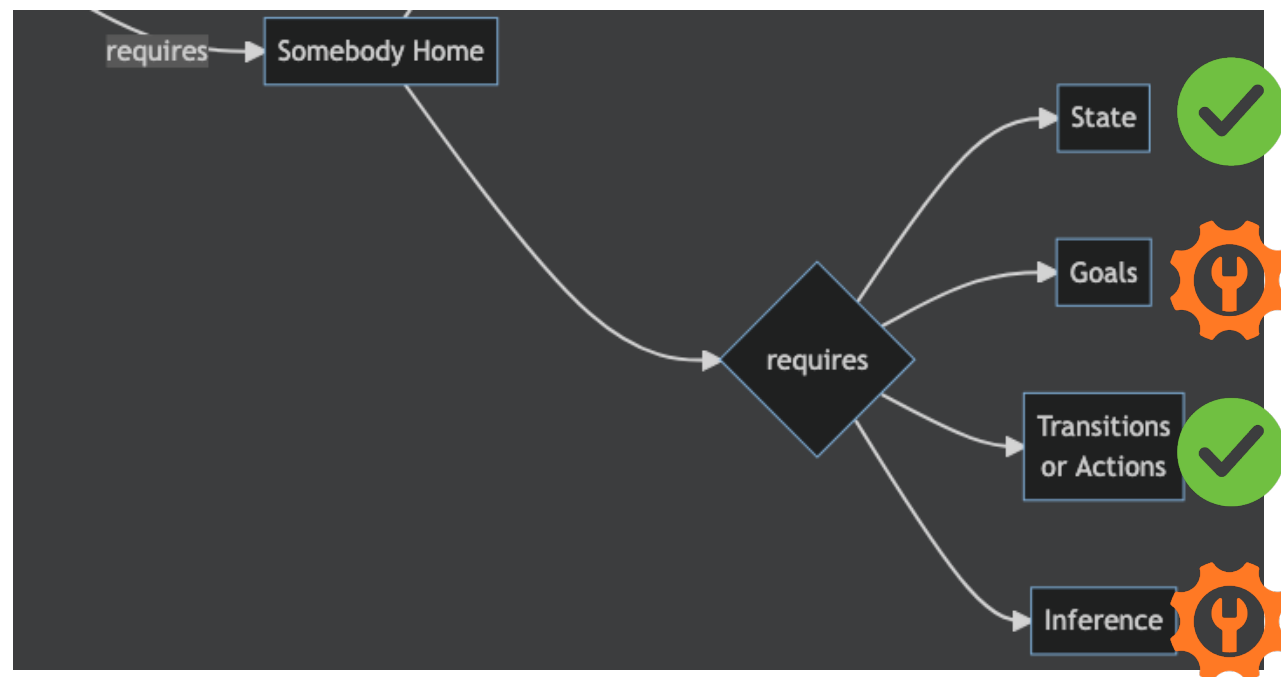
Decision transformers

<https://huggingface.co/blog/decision-transformers>



<https://deumbra.com/2022/08/rllib-for-deep-hierarchical-multiagent-reinforcement-learning/>

Imitation Learning from Video with Transformers



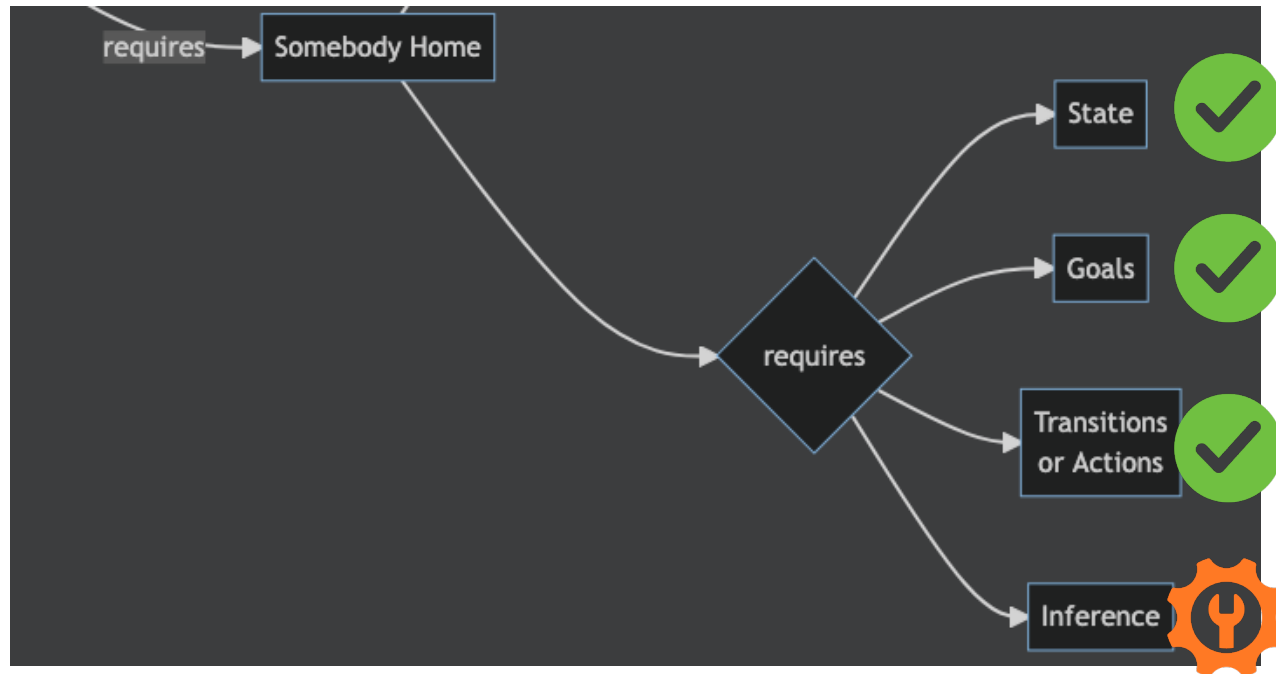
Using robots in the real or simulated world provides State.

Goals are weak since it is imitation learning.

We have transitions/actions that happen in the real world.

Also limited due to imitation learning

Reinforcement learning trained in in simulation



Using robots in the real or simulated world provides State.

Goals are more real with reinforcement learning

We have transitions/actions that happen in the real world.

This will be the challenge to generalize

How to build someone to talk to using the subsymbolic method



Have it watch videos with dialog



Give it reward for getting to good states in a simulation of our world, where actions include words

Conclusion

When we build an AI with somebody home, it will have goals.

It will talk with us to achieve those goals.

It may not be familiar and chatty like ChatGPT.

It will be a little alien and sometimes be hard to understand.

But it will be more reliable because it understands the conversation, and it will be more trustworthy because the mistakes it makes will make sense to us.

It will be somebody worth talking to!