CRAIG:

Hi, I'm Craig Smith and this is Eye on AI.

This episode is sponsored by [Paperspace](https://www.paperspace.com/), which provides the tools for end-to-end management of machine learning workflows. It comes with dedicated GPUs in the cloud. Check them out at paperspace.com

There has been a raging debate in the past few years between the symbolists and the connectionists about the future of artificial intelligence. The symbolists say that traditional, explainable, logic-based approaches still hold tremendous promise while the connectionists say that the power of deep learning, for all its current opacity and narrow application, holds the key to more general forms of machine intelligence. This week, I speak with David Cox, IBM Director of the [MIT-IBM Watson AI Lab](https://mitibmwatsonailab.mit.edu), which is blending the two traditions in what they call neuro-symbolic AI in hopes to move AI forward. He spoke about a new video dataset his team has developed for teaching machines to reason about causal relationships. I hope you find the conversation as intriguing as I did.

CRAIG:

Can you tell me a little bit about yourself? I'm always kind of curious how people ended up in such a rarefied world.

DAVID ([00:24](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=24.07)):

Well, I got here by accident. There was no plan. I just stumbled through my life until I got here. I'm a recovering academic, so I was a professor at Harvard for 10 years and I actually started my career as a neuroscientist. I was actually studying the brains of monkeys and we were trying to understand how the brain works. But along the way, it was the right time to start thinking about using GPUs, which were not general purpose computing devices at the time, to start thinking about how we would model brains. And then that led me down the rabbit hole to eventually a career in AI.

DAVID ([00:57](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=57.85)):

Now, I run this somewhat unique lab, IBM together with MIT where we're really taking seriously the idea of, you know, how do we go to the next level with AI, not just apply AI to applications but really ask where in the space of applications can't we go today because we just don't have the right techniques and what we need to do.

CHORD

I'm at MIT, I did my PhD at MIT, so it's a little bit coming back to my roots, but work with a lot of people in the department of brain and cognitive sciences like Josh Tenenbaum, who's really a cognitive scientist but who also crosses over into AI and there's just such a rich interplay of ideas between AI and cognitive science and neuroscience.

DAVID ([01:45](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=105.31)):

It's exciting.

CHORD

DAVID ([36:10](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2170.68)):

The broader agenda of neuro symbolic AI is really about how do we take these kind of two great traditions of AI. And you know, one that's, you know, enjoyed a resurgence lately. So neural networks, artificial neural networks have now been kind of rebranded as deep learning and you know, they've become very powerful. They were enabled by, you know, the digitalization of the world. We have lots of data, we have lots of compute and that that's what made the fundamental techniques, which were largely worked out in the 80s and even before to really work at scale together with symbolic AI. And you know, the idea that we have kind of symbols and manipulation of those symbols to combine them together. And we're really interested in all the different ways you can combine those things. So, the, the first piece that I think is important is simply just the recognition that there are symbols and the symbols are important, right?

DAVID ([36:59](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2219.04)):

That there are dyed in the wool neural network, you know, devotees like say Yoshua Bengio will say things like, Hey, what we really need to do is find low dimensional, sparse, causal, you know, you know, or representations that reflect the causal relationships in the world. They're, they're resistant to use the word symbol. But when you get right down to it, if you're talking about reducing the world to a very, very low dimensional, you know, set of entities that have relationships to each other that are causally related. Very quickly what you're describing, whether you want to say the word or not is symbols, right? Like the, you know, if, if it, if it walks like a symbol and a quacks, like a symbol you know, I, I think we should, we should all except that it's a symbol. You know, one of the problems with symbolic AI has always been, well, you, if you assume that the symbols are there from the start, how did you get them?

DAVID ([37:54](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2274.21)):

Where did they come from? You know, like, you know, you kind of assumed away the hard part, which was getting from the messiness of the real world, like a picture to, to a symbol or getting from the messiness of a, of a natural language question to, to a symbolic representation like a program. You know, that's the starting point, but there wasn't a good answer. Now there's one central, you know, one piece of this is, well, neural networks it turns out are actually a great candidate for getting you from the world of the messiness of the real world and the images and the messiness of natural language to symbols. So, in many ways that that's one piece of the puzzle is like, Hey, this is a great way of extracting these symbols. Now, once you have those symbols, what's some of the visual question answering work?

DAVID ([38:39](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2319.53)):

You know, the Josh's gained it, but people on my team well some of the power of that is, you know, if you have something that's fundamentally symbolic, like, you know you know, simple logical operations and predicates if it's, if you have the symbols, why not just solve it symbolic. But you know, like it's, it's, we have tools for doing that. They're incredibly powerful, incredibly precise. They're fundamentally transacting in the units that we as humans like to operate in, we like to operate in terms of symbols. Like I, I want, I want you to talk about the entities in an image and I want you know, to talk about the relationships and I want you use logical statements about them. That's, that's the, that's the language of my thought at some level. So, if I'm going to understand a machine and understand this decision, then that's a very powerful way to do it.

DAVID ([39:27](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2367.46)):

Now there's another place where, and you know, I should say also, there's an awful lot of different kinds of symbolic tools we have from, from logic for, you know, everything from, you know, theorem proving to you know, first order and higher order logics that are, they're very powerful. We have things like planning. So, planning is a, is a symbolic method for you know, deciding a series of actions that can get you from, from state you're in to a goal you want to get to. That's been around for a long time. Very rich field. So, if we can get from using neural networks to get from the, from the natural, from the world that we start with our actual inputs to a symbolic world, we have this wealth of tools that we can, we can use. And that's, that's powerful. But then there's another piece where we can also use neural networks to help out with the symbolic parts.

DAVID ([40:15](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2415.88)):

And this is a part where I think there's some debate over whether we should, you know, if you accept that symbols are the right things for us to be operating in terms of, well, maybe we should build neural network systems that can manipulate those symbols and can do logic you know, maybe even can do planning and things like that. And, and so there is a field of research that we engage in as well about saying how do we use neural networks to, to help do those kinds of symbolic operations. So, you know, don't use the old symbolic methods, but build neural networks that can do similar things. And, and that's powerful. And then there's ways of augmenting, you know, many symbolic methods are fundamentally about search. So how do you search through a set of different possibilities to find the right answer in a precise way?

DAVID ([40:59](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2459.74)):

In many cases, neural networks can also be used to give those symbolic search algorithms a sense of intuition, such that they can get to the answer faster or more attractively. So I think where we are really is, you know there's a very simple setup which is about getting from the world to symbols and then there's all kinds of different ways in which neural networks can be used to augment and mix with ideas from symbolic AI and that mixing, I think it's anybody's guests, which parts will emerge as being essential, but I think fundamentally having the richness of representation that symbols give us, the relationships, logic, positionality, those things are here to stay no matter what. And then really, it's about how do we use this, this sort of amazing toolbox that we have developed over decades, you know, from the two different traditions and combine them and recombine them in ways that are, that are sort of maximally powerful.

MUSIC

CRAIG ([05:58](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=358.81)):

Can you talk a little bit about this work of the CLEVRER data set, how that built on the CLEVRER data set and the experiments that you've been doing?

DAVID ([06:30](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=390.03)):

Yeah. So, [CLEVRER](http://clevrer.csail.mit.edu/) is about moving the ball forward in terms of sort of common-sense questions we can answer about, about video and, and the relationships with different objects. It descends from a data set called CLEVRER. CLEVRER was designed in many ways to illustrate a problem with deep learning. You have a very simple setting where you have objects like simple objects, like cubes and cylinders and spheres and the, the dataset was generated. So you decide where the objects are. You render a picture, you know, just like rendering a picture in a Pixar movie. And then because you generated that image, you can also ask all kinds of relational questions about it. Like, you know, what is the color of the cylinder that's to the right of the sphere or how many are there an equal number of metallic things and large things and you know the answer because you generated the dataset.

DAVID ([07:24](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=444.42)):

Now the interesting thing about those kinds of questions is even though those are trivial for us to answer, even a child can answer questions about these kinds of relationships. To train an end to end neural network - and end-to-end is watchword for deep learning. You really want to go from what you have to what you want or whatever you want and you put a neural network in between and you get out of the way, right? You don't put anything in the middle and then you train that with lots and lots of data and you get your answer. The problem is with datasets like CLEVRER, you need to have enormous amounts of training data to get them to train in the end to end way with, with any level of accuracy. And even if you have like 800,000 examples, say you're still not going to get 100% accuracy going from a question and a picture to what's the answer to that question.

DAVID ([08:15](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=495.35)):

And then the reason for it is interesting and that has to do, we think, with this fact that what's really happening is that there's a compositional element to it, right? So a deep learning system is trying to learn, you know, it's fundamentally a nonlinear function approximator which is just a fancy way of saying you're mapping from what you have to what you want. And it can be a complex relationship, but there's really a distribution of things. And for deep networks to work, you have to kind of sample all the corners of the space to some degree. And it will generalize. They generalize surprisingly well, the things that are close to what they've seen before. But you don't cover the space of all the possible combinations and permutations they can really struggle. And what CLEVRER really does is it encapsulates a case where there really are many, many, many, many, you know, an explosion of the number of possible combinations and permutations and neural networks struggle to learn and you need to have huge, huge amounts of supervision to get it done.

DAVID ([09:18](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=558.09)):

Now what the team with Josh Tenenbaum together to with our own Chuang Gan did was to add in a dash of symbolic logic, which gives it that compositionality. So you're sort of building in from the ground up that Hey, we're going to have a system that has an, has a rich expressive way of dealing with that composition reality. So you don't have to see every combination and permutation to make things work. And when you do that, a bunch of magical things happen. All of a sudden you need a fraction of the amount of data. You can get by with 1% of the data and do as well as many deep learning approaches. If you have just 10% of the amount of data you can just completely solve the problem. And of course you were previously the, it was designed to be hard for neural networks by themselves to solve.

DAVID ([09:59](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=599.7)):

If you add a dash of symbolic processing, all of a sudden you've got it. You can get effectively perfect performance. And then the last piece, which is really magical is you have something that's intrinsically explainable because you're going from a question to a symbolic program which operates on, on the symbolic representation of the data. And you just walk through and say, did this get the right answer for the right reason. And then now that CLEVRER is solved what CLEVRER is, is an extension say, well let's take it to the next step. You know, a big part of human reasoning is not just being able to identify what's there. They know basic sort of descriptive questions like is the cylinder to the left of the cube? But you want to say, Hey, if the ball rolls into the scene and bumps into something and the cube moves, why did the cube move? Can we, can we explain the reasons why things happen? Can we attribute causes and effects. Can we do counterfactuals can we say what would have happened if the, if the ball hadn't been there, what would have happened instead? And then we were really getting into increasingly more and more sophisticated ways of being able to reason about the world. And I think this is ultimately a direction that's going to bring a lot of value for, for real world applications of AI.

CHORD

CRAIG ([11:12](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=672.37)):

Yeah. Yeah. Well that's fascinating. Can you walk me through one example and, and where the neural nets are being applied, where the symbolic logic is being applied.

DAVID ([11:24](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=684.35)):

Maybe it'll be easier if we start with CLEVRER and then we can move. We can work up to CLEVRER. So, so the setup of the problem is we have a, we have a picture with all these objects in it and we have a question, are there an equal number of metallic things and large things? Straightforward enough? And you know, the answer depending on the picture is yes or no. Right? So in the traditional deep learning context, what you would do is you'd take what you have. You know, I have a question which is natural language. And I have a picture and going to put them into a neural network, I'm going to feed that through. I'm, and I'm going to map it to what I want, which is the answer. Yes. And I'm going to turn that crank a jillion times, right? That's the standard end to end model.

DAVID ([12:03](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=723.88)):

What the team did for CLEVRER which I think was CLEVRER, was to take you know, let's break apart that question. Are there an equal number of metallic things and large things? Well, you can take the question and we can put it into a recurrent neural network. You know, they're good at processing natural language. It's the best tools we have available. So let's use them for their strengths. But instead of transforming that sentence into say another language, you know, neural translation, something we know how to do or instead of rating it, you know, for its sentiment or whatever, you know, whatever. Let's instead translate the sentence, the sequence of words into a program which is going to give a set of steps. And if you look at the question, are there an equal number of metallic things and large things? Okay, well there's a series of steps we need to look at the scene.

DAVID ([12:53](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=773.77)):

You know, it's asking about metallic things. Okay, let's, let's use our visual system and now we're going to use another neural network. When you use a convolutional neural network, cause we know that those are great at doing vision. And so we've, we processed the question, we know it's asking about something about large. We filter for the large objects and then it's asking us something about the metallic objects. We filter for the metallic objects and then we ask if they're equal? Are there an equal number? And that's a, that's a logical operation fundamentally. You're going to compare this number and compare that number. So what we do is we translate the question into a program. We run the program on the output of a convolutional neural network, but that de-renders the image into a structured representation of what's in the scene. And then we step through this symbolic program.

DAVID ([13:35](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=815.44)):

And the magic of this thing is that the whole thing is trained jointly. So it is going to learn, it's not just like we're going to bolt pre-trained neural networks onto a symbolic process system, but rather we're going to combine the things together and we're going to use reinforcement learning such that the computer vision part learns to better parse the world into objects by virtue of being part of the system. When it gets the right answer, it gets reward and that, you know, encourages whatever the neural network was doing. And then we're also going to encourage, you know, the translation of the sentence into a program that we're going to run. And again, because that's trained together with this reinforcement learning approach, the, the neural network part learns to do a better job of turning the natural language into a program by virtue of being part of this joint system together.

DAVID ([14:19](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=859.21)):

And when you put those systems together, then you, you can, it's basically the system doesn't have to learn every possible, you know, combination and permutation of this object thing next to that object. You can leverage concepts in the composition of these different parts in, in a very powerful way. And that's ultimately, I think the magic of neuro symbolic systems.

CHORD

CRAIG ([14:40](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=880.21)):

What data set are you using on the semantic side that gives definitions to specific words like metallic.

DAVID ([14:51](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=891.38)):

Okay, so that, that's a great question. So in the first neuro symbolic visual question answering paper, basically what I just described, a lot of those concepts were hard coded, so you just needed to know there was something called metallic and that was predefined. Now of course, that's obviously not where we want to stop because you know, one of the, one of the, one of the concerns about symbolic processing is you know, how do you, how do you build up lots and lots and lots of symbolic knowledge.

DAVID ([15:19](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=919.88)):

You'd like to be able to learn that from the world. So, the next paper that that team produced, again, this is a collaboration between Josh Tenenbaum and some folks in my group and some other folks in other places. They developed something called the neuro symbolic concept learner. And this is a system now that can learn to acquire new concepts. So it's using neural networks and using you know, it's using tools of deep learning to be able to more flexibly learn, you know, sort of imagine new values of the, of the property of color. You know, we might have, you know, we know about red and we know about green, but we don't know about blue yet or we invent a new color, you know, aquamarine that it hasn't seen before. It can learn from experience to, to incorporate those new concepts and then still be able to use symbolic reasoning on top of these new concepts.

DAVID ([16:09](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=969.8)):

So they are increasingly making the system more and more and more flexible. And I think that's actually a really interesting thing because that's kind of our roadmap at some level. We were, we want to move from systems that require lots of human knowledge and human hand engineering and move towards increasingly more and more autonomous systems where they can learn new concepts autonomously. They can learn new tasks, autonomously. And I think this is a big piece of the unsupervised learning drive, is we want to increasingly have systems that don't require huge amounts of careful handcrafting and development by, you know, frankly very expensive people who have, you know, trained for a long time to twiddle the parameters. We really want to have systems that are much more flexible and fluid that can drop into an environment. They can learn new concepts as they need.

DAVID ([16:56](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1016.9)):

They can learn new tasks and, and increasingly build on what's been done before. And that's really the progression of what we've been doing from the early visual question and answering work with CLEVRER to the neuro symbolic concept learner that the team did to things like CLE RER, which is pushing the envelope. You know, basically CLEVRER is solved. So let's move beyond that and start saying, can we look at systems that have dynamics? They actually have physics to them and, and start to learn intuitions about that and be able to reason about things, be able to reason about counterfactuals, you know, things that didn't happen. What would happen if something had been different? And that's kind of, you know, increasingly, an increasing level of sophistication that I think as a field AI needs to drive towards if we're really going to have the impact we want to have on everyone's daily lives, you know, solving hard problems. A lot of these hard problems have that quality of being much more like puzzles that we need to figure out than simple mappings between what we, what we have and what we want.

CHORD

CRAIG ([17:56](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1076.31)):

Yeah. And I have questions about CLEVRER, but on this, this idea of building up from a limited labeled data set, is there an estimate of the minimum amount of labeled data to, to start a system? I mean, I'm just thinking, you know, a child learns a certain amount of language from, it's parents or it's, it's family, but then starts inferring language exactly in the way that you say, you know, the mother asks for something and the child looks around and knows the names of all these other objects, but it sees something that it doesn't know the name to and it infers that that's what the mother is asking for. So is there sort of a minimum amount of data that's required to get this started? And then, you know, in an imaginary world you would have the system that would run forever and you know, look around and would gradually build up a body of conceptual knowledge?

DAVID ([19:02](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1142.33)):

Yeah, I mean, I don't know if we have any way of saying theoretically what the minimum is, but we do have something like an existence proof of a, of a bound. We, we know how much it takes for a child to learn about the world. We know, we know how development progresses and how children learn some of these things. You know, that we use the word common sense to refer to a lot of this, you know, like the, our, our basic underlying knowledge of, you know, the affordances of how things relate to each other in the world of basic physics. Things that we never write down. Actually, this is one of the, one of the big problems with purely supervised approaches. You could imagine ingesting all the texts of the world to, you know, to try and learn, you know, everything.

DAVID ([19:48](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1188.65)):

And the problem is that actually we don't write down a lot of things. Like when Gary Marcus gives a great example of if you're hammering a nail into the floor, what's the orientation of the nail? And it's like, well we know what floors are, we know what hammering looks like. We can picture in our mind's eye what that would look like. And we know that the nail is vertical, but you're not going to find - with the exception fact that Gary's now written about that - you're not going to find that kind of thing written anywhere. You're not going to be able to search Wikipedia and you know, find the article on hammering nails into floors. And what's the orientation of the nail? Because that's just, that's something we take for granted. And that base level of knowledge is something that we don't really have incorporated into AI.

DAVID ([20:26](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1226.93)):

And we don't really have a great story yet about how we're going to do it. But this is really important than someone that a lot of us have of us are working on. In fact, there's a DARPA program called machine common sense that we're participating in together with MIT and Stanford and Harvard. And one of the interesting properties there a apropos, the question of, so what's the minimum required, you know, experience to acquire some of these, some of this knowledge? One of the interesting requirements of that program is that they actually have a team of evaluators that the government has hired to evaluate how well we're doing. And it includes a child psychologist, developmental psychologist, and our own team was required to, and we, we actually have Liz Spelke, Rebecca Saxe and Tomer Ullman. And you know, people who actually do experiments with human babies and children to learn about how they develop over time some of these common sense notions.

DAVID ([21:18](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1278.161)):

And, and what we're doing is, is then building AI agents that they live in a simulated environment and are able to acquire some of these common sense notions. So you know, I think it's an interesting intersection then of cognitive science and developmental science in particular and AI where we're sort of asking, you know, can we get our AI systems to the point where with comparable amounts of exposure to the world, can they acquire that same baseline level of common sense knowledge upon which you can build the scaffolding of ever increasing knowledge and conceptual understanding and frameworks and sort of mental models, which then ultimately I think give rise to the kind of intelligence we really need if we're going to move to the next level with AI.

MUSIC

DAVID ([01:54](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=114.25)):

I think where everyone's going though and where the field needs to go is to reduce the dependency on so much labeled training data. Because if you look at it from an application standpoint, imagine you have rivers of data flowing through your, through your enterprise and you need to like build a dam at great expense to extract value from it. And the problem is there just aren't that many rivers. So there are always going to be cases where you rely on past experience. Pre-Trained models on image net is a great example where you can use those features that were trained in one way with lots of data for some other purpose. So I think there's going to be elements of that, but we just need to dramatically reduce the amount of actual supervision you need. A lot of what we do, I should say is focused around how do you dramatically reduce the amount of supervision that these systems need. So that's very near and dear to our hearts.

CRAIG ([02:49](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=169.43)):

Yeah, Lacun is doing this thing with video and predicting futures with video with no supervision or no labeled data, it's just training the system on millions of YouTube videos so that it develops an understanding of the world.

DAVID ([03:13](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=193.21)):

When I was back when I was at Harvard, we developed something called the PredNet, which is exactly in the same frame and then the same model where it's, you can think of it as being sort of self supervised or they, you know, it's, it's you, you've taken a video stream and you're trying to predict what the next frame is going to look like. And that's a very powerful way to learn about, about the structure of the world. And, and we do that too. I mean ultimately, you know, you're going to want systems that can take, you know, mixtures or you know, if you have supervision, it's, it's, it's good to, you know, you want to have it, you, you want to be able to learn about structure by just you know, observing streams of data. And increasingly, one of the things that neuro symbolic potentially gives you also is to be able to sort of incorporate external knowledge as well.

DAVID ([03:56](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=236.5)):

So, you know, if a human could write it down, you hopefully can take that into account as well. You know, Yann had this idea, this, this picture of the cake where most of it was supervised and then the supervision was sort of the icing on the cake and the reinforcement was the cherry on top. And I think that kind of maybe antagonize the reinforcement learning people a little bit, but I think the spirit is sort of right. If you look at a human, where do we mostly learn from? It's very rare that we get somebody telling us, you know, or showing us flashcards or something. You know, we, we observe the world around us. We get lots of exposure like that. Very occasionally we get rewards. Like something good happens. I get a reward and I can try and figure out why I got that reward and optimize it.

DAVID ([04:35](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=275.5)):

But the vast majority of what we get is, is that unsupervised parts. I think that's the right, the right picture. From a practical standpoint, we want to build systems that can take, take in all different kinds of information. If you have supervision available to prime the pump, so much the better. If you can observe the world and learn the structure, so much the better. If you can build in structures and biases to the architecture that incorporate things, we just know. The structure of a convolutional neural network encapsulates that idea and makes it so you can leverage that. You don't have to, to have so many training examples learn. That's one of the great successes in deep learning, but there's lots of structure that we can know, and we can build in from the get go. And the more of those you're can incorporate, obviously that's also gonna be very powerful. The keeping with neuro symbolic for us is really this idea of symbols. What's really magical about them that your ability to flexibly compose them and recompose them and build and have more richer more structured representation, including things like programs. We do a program induction where you translate natural language, for instance, into a symbolic program and having some, some biases around the structure of what those programs can look like can be extremely powerful for augmenting systems that learn from, from observation, learn from examples.

MUSIC

CRAIG ([22:04](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1324.42)):

Yeah. And the CLEVRER dataset was how large and was that of static images or, or was that video as well?

DAVID ([22:14](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1334.8)):

So the CLEVRER or dataset is all video. I don't actually remember off the top of my head exactly how many clips we have in that. But you know, in principle it's an infinitely generatable source, right? So what these datasets are is a way of generating, you know, generating images, generating videos where, where, you know, what ground truth is, you know, having been asked these sort of relational questions and then, and counterfactual questions and scripted questions. And then back out from that, you know, you use that as a tool.

New Speaker ([22:47](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1367.051)):

The CLEVRER data set is a bunch of pictures paired with questions and their answers. You need to take an a picture and a question and say what the answer is. Right? Well, so, so in the case of CLEVRER the spirit is like you have a, you're basically running simulate, you know, simulations in the case of CLEVRER or so you have, you know, a simple physics engine that you can render images and objects.

DAVID ([23:14](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1394.47)):

So to generate a question of what would have happened if, if the, if the cube hadn't been there, you rerun the simulation, you know, without the cube there and run the physics and then you can determine those answers. So there's a little bit of an art to, to producing some, some details to how you produce a good set of good questions that, that are interesting questions. But fundamentally the, the, the spirit of these datasets is let's, let's simplify the demand greatly. Let's generate the M the, the images. So we know what the answers are. So we can generate a large set to test over. Like the counterpoint would be to do something like you'd have to do a naturalistic experiment where you collected lots of real videos in the world and you know, things happened or they didn't happen, then you'd have to have people annotate them and it's a huge amount of effort and you know,

DAVID ([24:00](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1440.921)):

In the, in the domain of visual question answering some of that work, there exists data sets that where you have static pictures and people ask questions about the relationships between those, those items and some of those same methods that we are applying to these simpler data sets like CLEVRER. We're also applying increasingly to real world datasets. But, but one of the strengths of these kinds of CLEVRER is you can have, you know, generate large data sets of really interesting questions that really kind of carve at the right joints that ask the right questions and push the field in the right direction. And currently there's no equivalent data set that would, you know, would have, you know, enough counterfactual examples or predictive questions that you could ask that would, that would be able to drive progress.

DAVID ([24:43](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1483.91)):

In general, one thing, one comment and I would make is that so much of the progress in AI has been driven by the creation of good datasets and good child's problems to organize our efforts. And I think, you know, the CLEVRER days, that's good example. We as part of the community are pushing forward this idea that really what we need to get towards are not just, you know, not just simple examples of, of taking a picture and telling what's in it, but rather answering these kinds of harder, more nuanced questions about, you know, what's going to happen next or what would have happened. And that's, I think starting to get a much richer kinds of reasoning that are going to be really valuable as we think about applications of AI in the future.

CRAIG ([27:13](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1633.39)):

So the NSCL the, the neuro symbolic concept learner can be trained on this CLEVRER data. Is that right?

DAVID ([27:26](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1646.13)):

That's right. And when the CLEVRER paper comes with a new neuro symbolic dynamic learner, which includes you know, includes models of the physics of the world. So the original question answering models didn't include any notion of time. There they were, they were, they were built to, to work on static images. Right? But, but, but you know, the way to think about it, you have a, a neural dynamics engine now that, that, that, that the system can rely on to model the interactions and the physics of different objects. And if you think about how we solve problems like this you know, if you want, if you asked a question, you know, what would happen if we had removed this ball? You know, what, what would the cubit hit? When we think about how we solve those problems, you know, introspectively, we actually imagine them, right?

DAVID ([28:16](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1696.08)):

Like the, we do a simulation in our mind and that simulation includes dynamics. So, so, so to solve this problem in this sort of neuro symbolic mold, you know, we needed to add it, you know, the team needed to add an additional piece, which was this dynamics engine, sort of think of like a little physics engine or a little game engine. You know, like the game engines that control video games, you know, that the system can use to do these simulations and then that, that's another sort of tool that's the disposal of the symbolic system for answering these questions.

CRAIG ([28:47](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1727.06)):

When, a child pushes stuff off its high chair table, its learning the laws of physics. For a computer system to learn that without it explicitly being coded. It has to infer that from multiple examples. Is that the sort of thing that's, that's happening here? Or when you talk about a physics engine it does that have coded into it the laws of physics that are being applied.

DAVID ([29:18](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1758.42)):

Yeah, there's different ways you could do it. Obviously you, I mean, you, you could imagine building a system that has an innate understanding of physics and uses that to help it, you know, reason about the world. You can also imagine building systems that learn the physics. You know, they learn what happens when an object falls off the table. You can learn what happens when an object collides with another object. But I do believe that that is exactly what's happening. You know, when, when the baby's pushing thing, you know, moving objects around, you know, it may be the case that they have some, you know, native knowledge of physics. You could imagine over evolutionary timescales they might've acquired, you know you know, the, the groundwork for some of those things, making easy to learn. And, and that's you know, there's a debate between nativists and empiricists within the developmental psychology literature, sort of arguing over to what extent are these capacities, you know, are you born with them or to what extent do you learn them?

DAVID ([30:16](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1816.96)):

But leaving those, that aside, I think the broader question is this idea of, Hey, I have a world, you know, in front of me. I want to be able to achieve goals in that world. I want to be able to predict what's going to happen in that world, cause it's going to impact me. I need to build a model of that world. And that model includes things like physics. And I can, I can build that model in a couple of ways. I mean, to some extent I could be born with that model. And you know, you can imagine hard coding in some of the rules of physics. That might be one way to do it. You can imagine observing the world and learning, you know, new physics and you know, things like, you know, magnets you know, strong magnets probably didn't exist or weren't easy to find on evolutionary timescales when humans were first of evolving, you know, on the African Savannah.

DAVID ([31:02](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1862.71)):

But we were able to learn intuitions by, by working with these powerful magnets and seeing how they repel and attract and build a mental model of how that happens. So we can learn by observing and then of course we can learn by doing. And that's a big part of what the kids are doing. And, and we think that's important as well. They have the idea that you need to be able to have agents that go into the world and try things. You know, you do little experiments. Like, what happens if I drop, you know, what happens if I knock my bottle off of my high chair? What happens if I take this object and I ram it into that object? Does it pass through? You know, you can, you can imagine interaction with the world as being a very, very valuable way of very quickly refining your model.

DAVID ([31:42](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1902.97)):

And, and, and building that mental model, which then you can use for the rest of your life to predict how the world's going to unfold. Which is a valuable thing to predict how, you know, to plan how you can, you know, take actions in the world to achieve some particular goal. And you know, sort of simulate forward. If I did this, if I hadn't done that, what would happen? That's really the spirit of what CLEVRER really started or is really starting to try and, and get into.

CHORD

CRAIG ([32:09](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1929.91)):

Yeah. Is this the first dataset of its kind?

DAVID ([32:12](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1932.79)):

This is the first video reasoning dataset like this. And again, you know, this is a moving this is how the field moves. We need to find problems. You know, we, we often define problems that are well short of the final goal.

DAVID ([32:29](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1949.86)):

CLEVRER’s a great example of that. You know, it, then you solve it. You know, in CLEVRER came from a group at Stanford and, and Facebook. You know, the field solves the problem and then you ratchet up the complexity and say, okay, well now, now we really want to get at is, is some of these other ideas. And of course you can imagine this is a very impoverished simple world with simple objects, with simple directions. But you know, where we're going is much more complex, richer worlds, 3d worlds with realistic physics. But also, you know, abstract worlds as well. We want to build, build methods that aren't just about the sort of Newtonian physics and kinematics of the world, but also include things like, you know, when you, you know, when you have an organization with people, you build a mental model of how did my organization function or do you have a machine?

DAVID ([33:19](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=1999.25)):

You build them. You have an internal mental model of how that machine functions. You know, if you repair a car, you have a mental model of how the car functions and what different parts of the car do. And if something's broken, you use that internal model to think about what would you need to do to fix the car? Right? And it turns out everything we do and you really look at it, you know, a huge fraction of what we're doing in our day to day jobs is about constantly refining our mental models of the world and then using those mental models to solve problems to for our benefit. And then that's kind of a, in a nutshell, you know, almost everything we do and you know, it also encapsulates an awful lot of what we'd like AI to help us with. They could go out there form a mental model, a shared mental model, ideally with you so that it can solve problems that can show you what it's doing.

DAVID ([34:03](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2043.29)):

And I think that's really the Holy grail for what AI needs to do. And in the next couple of decades.

MUSIC

DAVID ([34:10](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2050.4)):

A lot of these ideas very, very naturally feed into robotics, right? Because robotics is all about, you know, an agents with certain degrees or freedom interacting in the world. And you know, oftentimes one of the, one of the problems with robotics traditionally is you actually have to wait for the real world to unfold, right? Like you need to the world you know, the actual reality that we live in, the physical reality we live in has a certain clock to it, which is, is what it is. And you, you can't run the world faster than, than it then it's gonna go. So people do a lot of simulation in robotics to, to, to start to get around this and you have to take your simulation and transfer it to the real world.

DAVID ([34:49](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2089.79)):

It's hard to get supervision and having agents that can actually get out in the world and interacting with the world and efficiently. You know, efficiency is key cause you're, you're running up the clock that the world natively runs at. To be able to, to have that model, to come up with the experiments you want to do, to, to either validate them all over, find the model to ultimately to solve some sort of task. I think that's very much in the heart of, of everything we do. But I think it's much broader than just robotics too, right? I mean, you look at almost any problem we have in the world that we on our day to day lives, try and solve, think about being in things like climate change. You know, we're, we have you know, build models of the climate. We're trying to understand what's going to happen. We're going to do experiments to help refine that model. You know, that that loop of observe, interact, refine our internal model, you know, and repeat is very, very powerful. And that's ultimately, I think what we need to get into all of these different fields.

CHORD

DAVID ([41:53](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2513.72)):

Really what these neuro-symbolic question answering approaches do is they take the question which is natural language that we produce and then turns it into a symbolic set of symbolic operations program. And that program is now very precise. That's really what in the neuro symbolic approaches, they use neural networks to go from the natural language into a symbolic representation. And then use neural networks to go from the image into a symbolic representation of what's in the image. Like what objects are there, what, what properties do they have, where are they in the scene? And then what you do is you take the program that you got from the question and you run it on the symbolic representation. So you're, you're into the symbolic space and you can even do things like running simulations. You know, you have it, you have a, you know, a, a deterministic or even a stochastic program that you can run that says, I know I'm going to simulate forward what would have happened if this object wasn't there.

DAVID ([42:45](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2565.59)):

So it really, what, what the neural symbolic is about is going using neural networks to get from the real world, carving out the right joints, extracting the structure into a rich, symbolic, structured representation of the world. And then once you're there, you can then operate on it in very sophisticated ways. And CLEVRER really, you know, drives us one step closer to a world in which we can use those symbolic representations to simulate you know, both what will happen next, you know, in a predictive mode. But also, you know, counterfactuals what would have happened or what would I need to, to affect a particular kind of change. That's just the next step. I, you know, there are many more steps that come after that, but that's really the direction that we're going.

CRAIG ([43:28](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2608.79)):

Yeah, that's fascinating.

CHORD

DAVID ([43:30](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2610.56)):

I think the key idea here is that it's really about building models of the world and that that really is the unsupervised piece of this.

DAVID ([43:38](https://www.temi.com/editor/t/LXgc8ZTxGuI5rc5LVoRtpWGKc1y9SPrafdQhmYGj_KgIVP4VbeEJC4wmQXZ3KBjyvoLgcpLwg0ZsVzhWExvEStu9Y2s?loadFrom=DocumentDeeplink&ts=2618.27)):

How do you observe the world interact with the world in a way that lets you build a model that lets you either solve problems, predict what's going to happen, you know, do, do any number of things that are ultimately advantageous for us. You know, as humans we're, we're trying to, you know, you know, survive and reproduce and get food and all that. We build a model of the world by observing passively, by interacting to refine that model. What we're interested in in particular with this neuro symbolic idea, is that that model, the best kind of model we can have is a fundamentally symbolic one. It's about entities, the relationships we can, we can simulate in that symbolic world. We can you know, we can apply logic. And that's a very powerful tool to mix with all of the power that comes from, from learning. And it is, it can be certainly a fundamentally unsupervised and, and self-supervised. You know, you go out into the world, you have agents that can affect changes, that can unpick and do experiments. And we think that, that, that mixing, that, those ideas together with symbolic ideas is very, very powerful.

That's it for this week's podcast. I want to thank David for his time. If you want to learn more about neuro-symbolic computing, you can find a transcript of this episode on our website, Eye on AI. We love to hear from listeners, so feel free to contact us with comments or suggestions.

The singularity may not be near, but AI is about to change your world, so pay attention