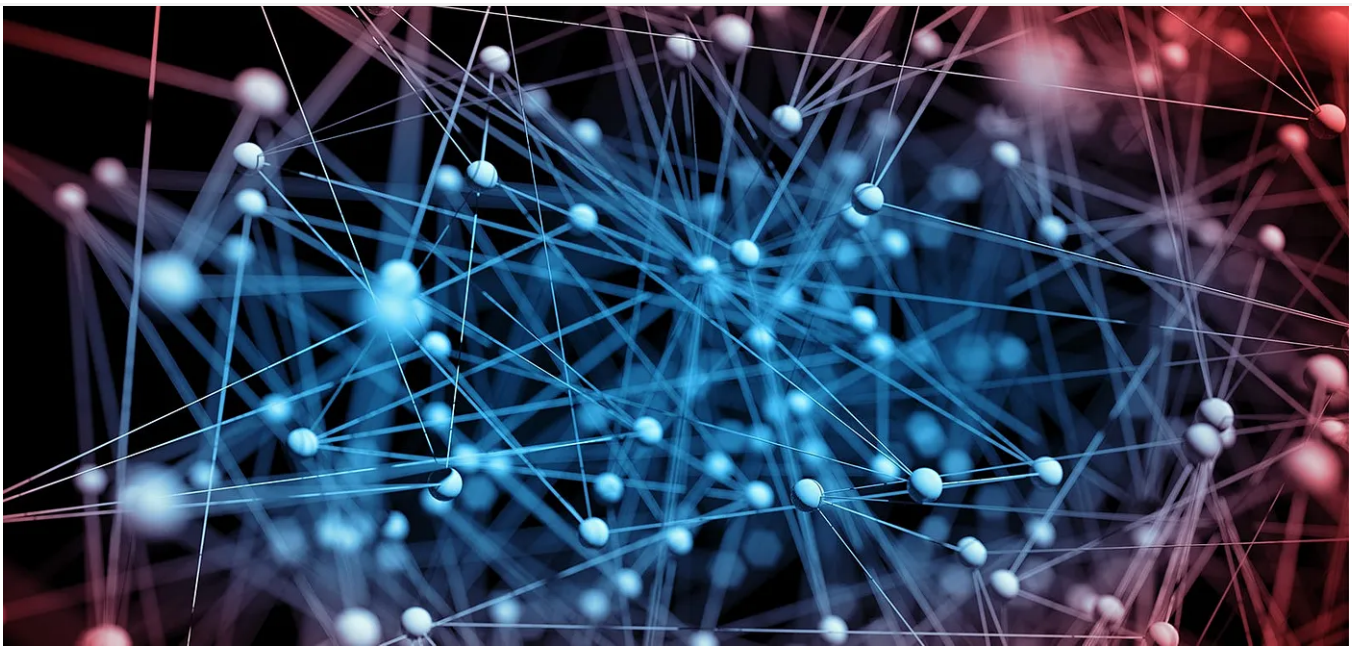
Alireza Dehbozorgi

10 min read · Just now

Listen Share More

Computational Theories of Cognition- Part I: Connectionism



https://biology.mit.edu/faculty-and-research/areas-of-research/computational-biology/

1. Connections models

Connectionst models were inspired by research into how computation works in the brain, and particularly the observation that large, densely connected networks of relatively simple processing elements can solve some complex tasks fairly easily in a modest number of sequential steps. Subsequent work has produced models of cognition with a distinctive flavor. *Processing* is characterized by patterns of activation across simple processing units connected together into complex networks. Knowledge is stored in the strength of the connections between units. It is for this

reason that this approach to understanding cognition has gained the name of *connectionism*.

Tons of literature has made it apparent that the field has entered the third age of *artificial neural network* research.

- **Background**

Over the last forty years, connectionist modeling has formed an influential approach to the computational study of cognition. It is distinguished by its appeal to principles of neural computation to inspire the primitives that are included in its cognitive level models. Also known as artificial neural network (ANN) or parallel distributed processing (PDP) models (see McClelland 2003, among others), connectionism has been applied to a diverse range of cognitive abilities, including models of memory, attention, perception, action, language, concept formation, and reasoning.

While many of these models seek to capture adult function, connectionism places an emphasis on learning internal representations. This has led to an increasing focus on developmental phenomena and the origins of knowledge. Although, at its heart, connectionism comprises a set of computational formalisms, it has spurred vigorous theoretical debate regarding the nature of cognition. Some theorists have reacted by dismissing connectionism as mere implementation of preexisting verbal theories of cognition, while others have viewed it as a candidate to replace the Classical *Computational Theory of Mind* and as carrying profound implications for the way human knowledge is acquired and represented; still others have viewed connectionism as a sub-class of statistical models involved in universal function approximation and data clustering.

- **Key Properties of Connectionist Models**

Connectionism starts with the following inspiration from neural systems: computations will be carried out by a set of simple processing units operating in parallel and affecting each other's activation states via a network of weighted connections. Rumelhart, Hinton, and McClelland (1986) identified seven key features that would define a general framework for connectionist processing. The seven features are as follows:

1- The set of processing units μ_i. In a cognitive model, these may be intended to represent individual concepts (such as letters or words), or they may simply be

abstract elements over which meaningful patterns can be defined. Processing units are often distinguished into input, output, and hidden units. In associative networks, input and output units have states that are defined by the task being modeled (at least during training), while hidden units are free parameters whose states may be determined as necessary by the learning algorithm.

2- A state of activation (*a*) at a given time (*t)*. The state of a set of units is usually represented by a vector of real numbers *a(t)*. These may be binary or continuous numbers, bounded or unbounded. A frequent assumption is that the activation level of simple processing units will vary continuously between the values 0 and 1.

3- A pattern of connectivity. The strength of the connection between any two units will determine the extent to which the activation state of one unit can affect the activation state of another unit at a subsequent time point. The strength of the connections between unit *i* and unit *j* can be represented by a matrix *W* of weight values — w(ij). Multiple matrices may be specified for a given network if there are connections of different types. For example, one matrix may specify excitatory connections between units and a second may specify inhibitory connections. Potentially, the weight matrix allows every unit to be connected to every other unit in the network. Typically, units are arranged into layers (e.g., input, hidden, output) and layers of units are fully connected to each other. For example, in a three-layer feedforward architecture where activation passes in a single direction from input to output, the input layer would be fully connected to the hidden layer and the hidden layer would be fully connected to the output layer.

4- A rule for propagating activation states throughout the network. This rule takes the vector —w(ij)— of output values for the processing units sending activation and combines it with the connectivity matrix *W* to produce a summed or net input into each receiving unit. The net input to a receiving unit is produced by multiplying the vector and matrix together.

5- An activation rule to specify how the net inputs to a given unit are combined to produce its new activation state.

For instance, *F* might be a threshold so that the unit becomes active only if the net input exceeds a given value. Other possibilities include linear, Gaussian, and sigmoid functions, depending on the network type. Sigmoid is perhaps the most common, operating as a smoothed threshold function that is also differentiable. It is often

important that the activation function be differentiable because learning seeks to improve a performance metric that is assessed via the activation state while learning itself can only operate on the connection weights. The effect of weight changes on the performance metric therefore depends to some extent on the activation function, and the learning algorithm encodes this fact by including the derivative of that function.

6- algorithm for modifying the patterns of connectivity as a function of experience. Virtually all learning rules for PDP models can be considered a variant of the Hebbian learning rule (Hebb, 1949). The essential idea is that a weight between two units should be altered in proportion to the units' correlated activity. For example, if a unit ui receives input from another unit uj , then if both are highly active, the weight — from μ_j to μ_i should be strengthened.

7- A representation of the environment with respect to the system. This is assumed to consist of a set of externally provided events or a function for generating such events. An event may be a single pattern, such as a visual input; an ensemble of related patterns, such as the spelling of a word and its corresponding sound and/or meaning; or a sequence of inputs, such as the words in a sentence. A range of policies have been used for specifying the order of presentation of the patterns, including sweeping through the full set to random sampling with replacement. The selection of patterns to present may vary over the course of training but is often fixed. Where a target output is linked to each input, this is usually assumed to be simultaneously available.

Two points are of note in the translation between PDP network and cognitive model. First, a representational scheme must be defined to map between the cognitive domain of interest and a set of vectors depicting the relevant informational states or mappings for that domain. Second, in many cases, connectionist models are addressed to aspects of higher-level cognition, where it is assumed that the information of relevance is more abstract than sensory or motor codes. This has meant that the models often leave out details of the transduction of sensory and motor signals, using input and output representations that are already somewhat abstract. The same principles at work in higher-level cognition are also held to be at work in perceptual and motor systems, and indeed there is also considerable connectionist work addressing issues of perception and action.

- **Knowledge vs. Processing**

One area where connectionism has changed the basic nature of theorizing is memory. According to the old model of memory based on the classical computational metaphor, the information in long-term memory (e.g., on the hard disk) has to be moved into working memory (the CPU) for it to be operated on, and the long-term memories are laid down via a domain-general buffer of short-term memory (RAM). In this type of system, then, long-term memory is separated from processing. It is relatively easy to shift informational content between different systems, back and forth between central processing and short and long-term stores. Computation is predicated on variables: the same binary string can readily be instantiated in different memory registers or encoded onto a permanent medium.

By contrast, knowledge is hard to move about in connectionist networks because it is encoded in the weights. For example, in the past tense model, knowledge of the past tense rule "add –ed" is distributed across the weight matrix of the connections between input and output layers. The difficulty in portability of knowledge is inherent in the principles of connectionism Hebbian learning alters connection strengths to reinforce desirable activation states in connected units, tying knowledge to structure. If the foundational premise is that knowledge will be very difficult to move about in the human information processing system, what kind of cognitive architecture results? There are four main themes.

1. It is necessary to distinguish between two different ways in which knowledge can be encoded: active and latent representations (Munakata & McClelland, 2003). Latent knowledge corresponds to the information stored in the connection weights from accumulated experience. By contrast, active knowledge is information contained in the current activation states of the system. Clearly the two are related, since the activation states are constrained by the connection weights. But, particularly in recurrent networks, there can be subtle differences. Active states contain a trace of recent events (how things are at the moment) while latent knowledge represents a history of experience (how things tend to be). Differences in the ability to maintain the active states (e.g., in the strength of recurrent circuits) can produce errors in behavior where the system lapses into more typical ways of behaving (Morton & Munakata, 2002; Munakata, 1998).

2. If information does need to be moved around the system, for example from a more instance-based (episodic) system to a more general (semantic) system, this will require special structures and special (potentially time consuming) processes. Thus McClelland, McNaughton, and O'Reilly (1995) proposed a

dialogue between separate stores in the hippocampus and neocortex to gradually transfer knowledge from episodic to semantic memory (see O'Reilly, Bhattacharyya, Howard, & Ketza, 2011). For example, French, Ans, and Rousset (2001) proposed a special method to transfer knowledge between the two memory systems: internally generated noise produces "pseudo-patterns" from one system that contain the central tendencies of its knowledge; the second memory system is then trained with this extracted knowledge to effect the transfer.

3. Information will be processed in the same substrate where it is stored. Therefore, long-term memories will be active structures and will perform computations on content. An external strategic control system plays the role of differentially activating the knowledge in this long-term system that is relevant to the current context. In anatomical terms, this distinction broadly corresponds to frontal/anterior (strategic control) and posterior (long-term) cortex, with posterior cortex comprising a suite of content-specific processing systems. The design means, somewhat counter-intuitively, that the control system has no content. Rather, the control system contains placeholders that serve to activate different regions of the long-term system. The control system may contain plans (sequences of placeholders) and it may be involved in learning abstract concepts (using a placeholder to temporarily co-activate previously unrelated portions of long-term knowledge while Hebbian learning builds an association between them) but it does not contain content in the sense of a domain-general working memory. The study of frontal systems then becomes an exploration of the activation dynamics of these placeholders and their involvement in learning (see, e.g., work by Botvinick & Cohen, 2014; Davelaar & Usher, 2002; Haarmann & Usher, 2001; O'Reilly, Braver, & Cohen, 1999; Usher & McClelland, 2001).

4. The connectionist perspective on memory alters the conception of domain generality in processing systems. It is unlikely that there are any domain-general processing systems that serve as a "Jack of all trades," i.e., that can move between representing the content of multiple domains. However, there may be domain-general systems that are involved in modulating many disparate processes without taking on the content of those systems, either via direct connectivity or through the regional modulation of neurotransmitter levels. This type of general system might be called one with "a finger in every pie." Meanwhile, short-term or working memory (as exemplified by the active representations contained in the recurrent loop of a network) is likely to exist as a devolved panoply of discrete

systems, each with its own content-specific loop. For example, research in the neuropsychology of language tends to support the existence of separate working memories for phonological, semantic, and syntactic information (MacDonald & Christiansen, 2002). And one might expect recurrent loops in the prefrontal cortex to maintain information about current goal states and positions in task sequences. From a connectionist perspective, therefore, and in contrast to traditional cognitive theory, there is no such thing as working memory as a general mechanism; rather it is a content-specific activity carried out in multiple systems.

- **Deep Neural Networks for Cognitive Modeling**

Deep neural networks have provided a step change in the performance of artificial intelligence systems for visual object recognition and natural language processing. Do they provide the basis for better cognitive models? As a case study, a number of researchers have explored whether the representations developed in the respective hidden unit layers of deep neural networks of visual object recognition accord to the types of representation found in the hierarchy of neural areas in the ventral pathway of vision in the inferior temporal cortex (e.g., Kriegeskorte, 2015; Yamins et al., 2014). Such a comparison is made possible by assessing the representational similarity between activity produced by a range of images of objects (faces, places, animals, tools, etc.), either in functional magnetic resonance imaging data of human participants or in the hidden unit activation levels of the trained neural network. The sequence of lower level features (edges), intermediate level features (contours), and high-level features (objects) is found both in neural areas and in network layers moving further from the input, suggesting similar computations are taking place. However, in other respects, these deep neural networks are not human-like: in the face of noise, their performance declines in nonhumanlike ways, suggesting over-fitting to the training data or the absence of crucial human-like architectural constraints; and at best, current models are capturing bottom-up, feedforward aspects of visual processing, not the top-down expectation-based influences enabled by bidirectional connectivity (Kriegeskorte, 2015; Storrs & Kriegeskorte, 2019).

. . .

Thank you so much for your kind and attention! Stay tuned for more!

Alireza Dehbozorgi

Email

https://www.linkedin.com/in/alireza-dehbozorgi-8055702a/

Twitter: @BDehbozorgi83

Cognition Computation Intelligence Linguistics Data Science

Edit profile

Written by Alireza Dehbozorgi

42 Followers

I'm a linguist and AI researcher interested in mathematical/computational approaches to both human and formal languages. Twitter: @BDehbozorgi83

More from Alireza Dehbozorgi

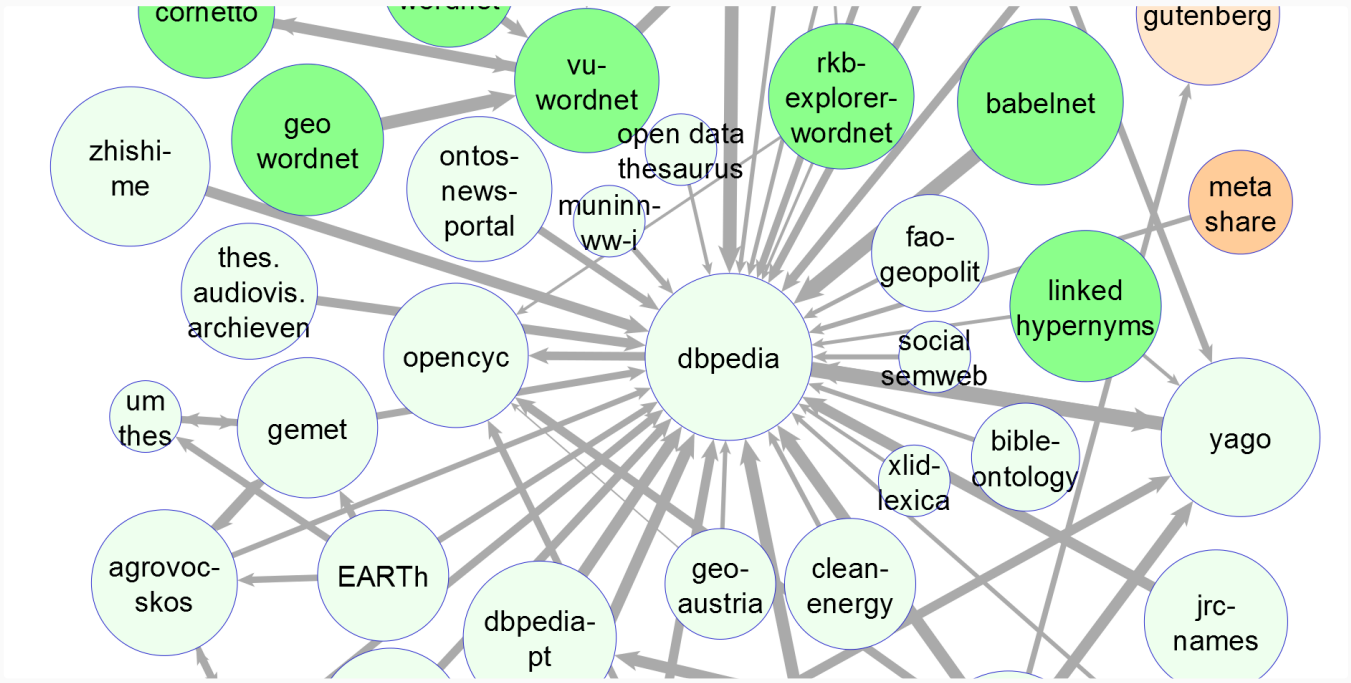
Alireza Dehbozorgi

A Survey of theories of linguistic meaning.

1- Cognitive Semantics: The linguistic representation (see also Adger 2022) of conceptual structure is the central concern of the…

10 min read · Apr 15

8 1



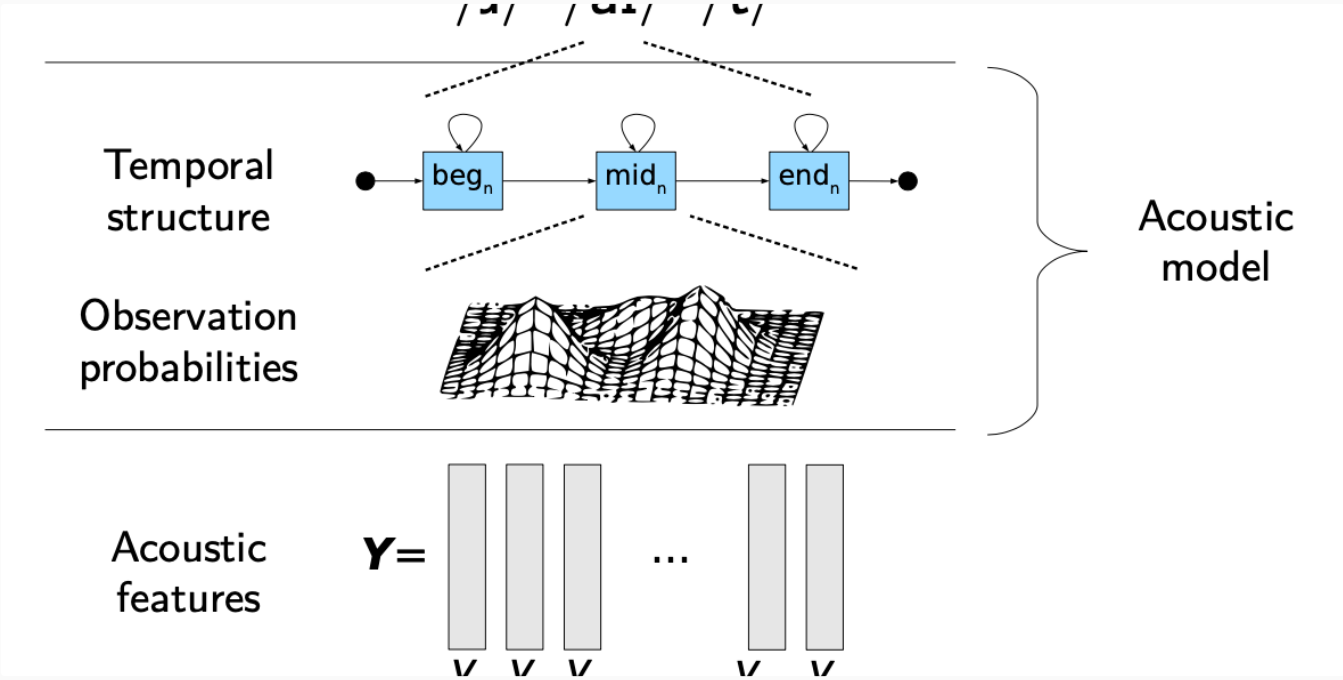
Alireza Dehbozorgi

Linguistics, Ontology Engineering, and Databases: The Power of Structured Data

Language is a complex and dynamic system, with countless variations and nuances that can make it difficult to analyze and understand…

3 min read · 5 days ago

3 1



Alireza Dehbozorgi

The possible acoustic correlates of logical calculi in language and speech

The study of logic and language has a long history, dating back to ancient Greece. In recent years, researchers have been exploring the…

3 min read · 5 days ago

10 1

Alireza Dehbozorgi

The English translation of Dr. Andrés Hohendahl's article on #llms, etc.

Andrés Hohendahl: "The AI algorithms are not applicable to/ don't work for Spanish, as it is a much more complex language."

6 min read · Apr 17

 3  1 

See all from Alireza Dehbozorgi

Recommended from Medium

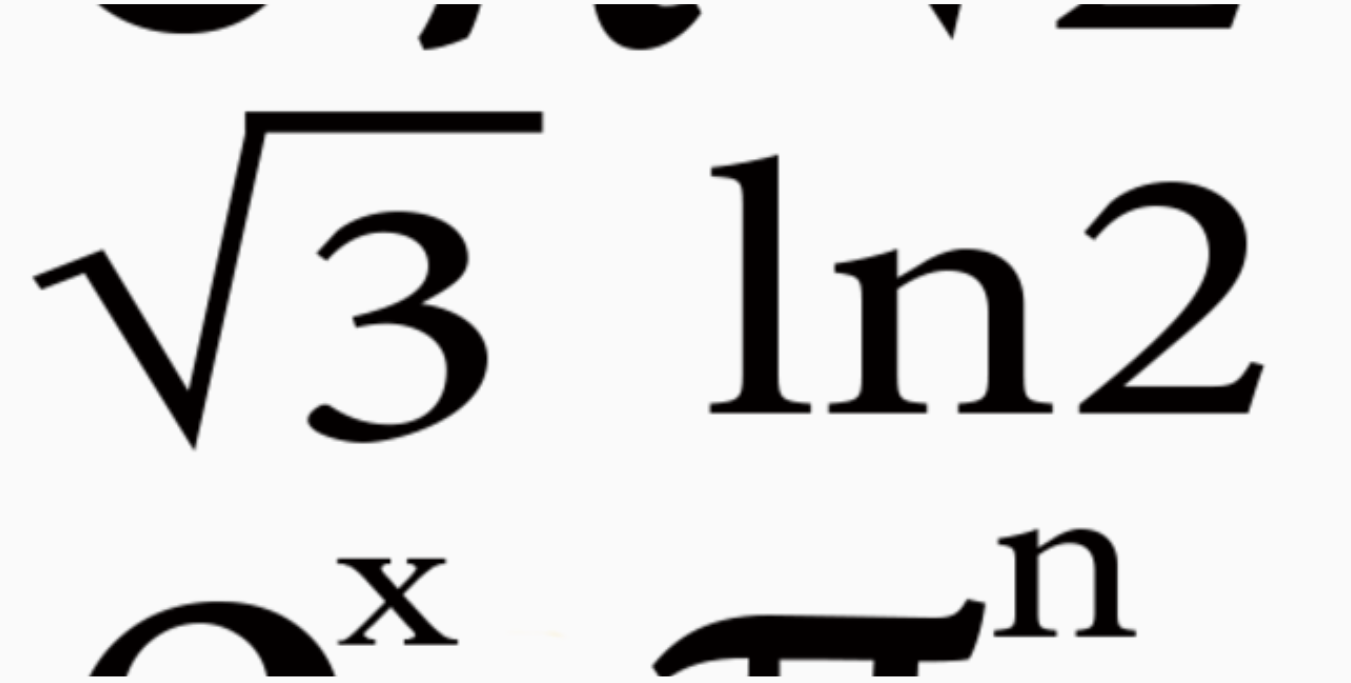
Kenny Minker in The New Climate.

Japan’s Government Wants a Baby Boom—What Am I Missing?

Population decline presents challenges—and opportunities.

· 8 min read · 5 days ago

1.3K 23



Kasper Müller in Cantor’s Paradise

The 7 Most Important Mathematical Constants

The beauty of π, the magic of ϕ, the mystery of γ, and the powers of e

· 13 min read · 6 days ago

761 21

Tim Andersen, Ph.D. in The Infinite Universe

The universe may be alive

But we might never know

· 4 min read · 4 days ago

1.2K 19

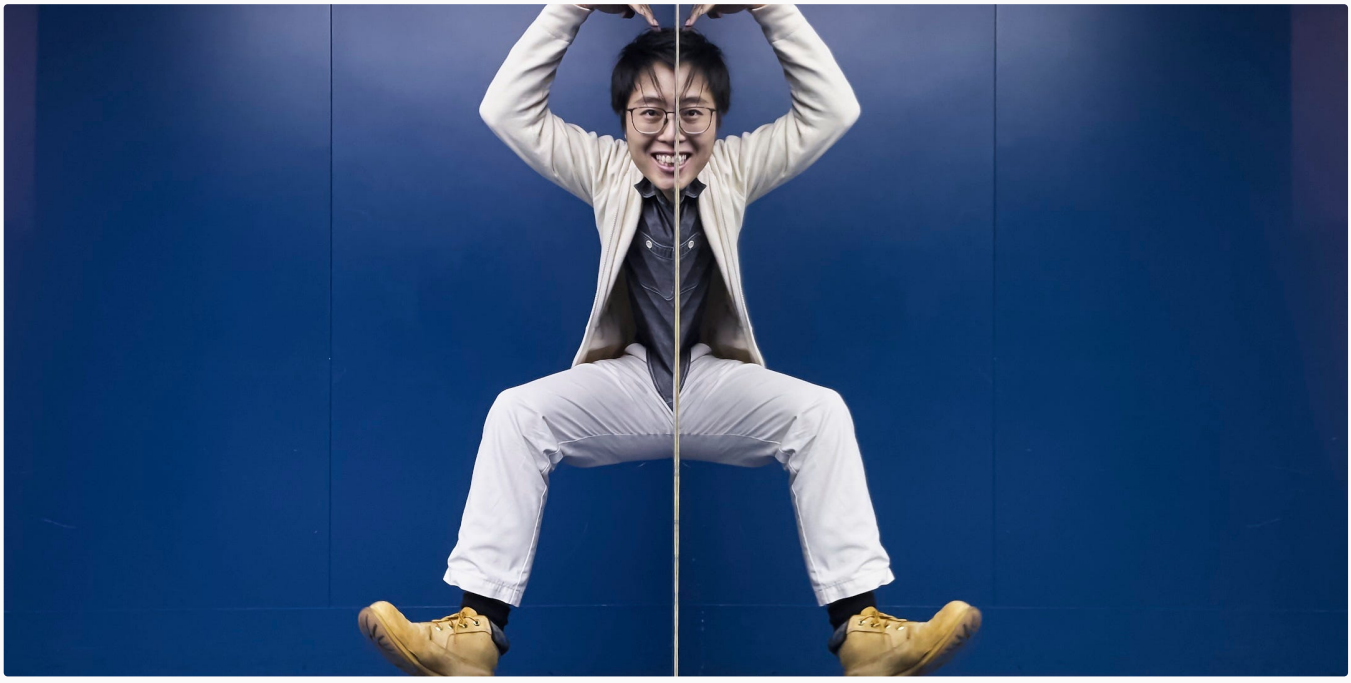
Cassie Kozyrkov

How AI is Evolving

A prediction for the next decade

✨ · 3 min read · 6 days ago

1.2K 27



Ethan Siegel in Starts With A Bang!

Why do mirrors flip left-and-right but not up-and-down?

If you look into a mirror, you'll notice that left-and-right are reversed, but up-and-down is preserved. The reason isn't what you think.

· 10 min read · 6 days ago

863 19

Cory Doctorow

How Amazon makes everything you buy more expensive, no matter where you buy it

Most Favored Nation is my least favorite scam.

· 10 min read · Apr 25

2K 26

See more recommendations