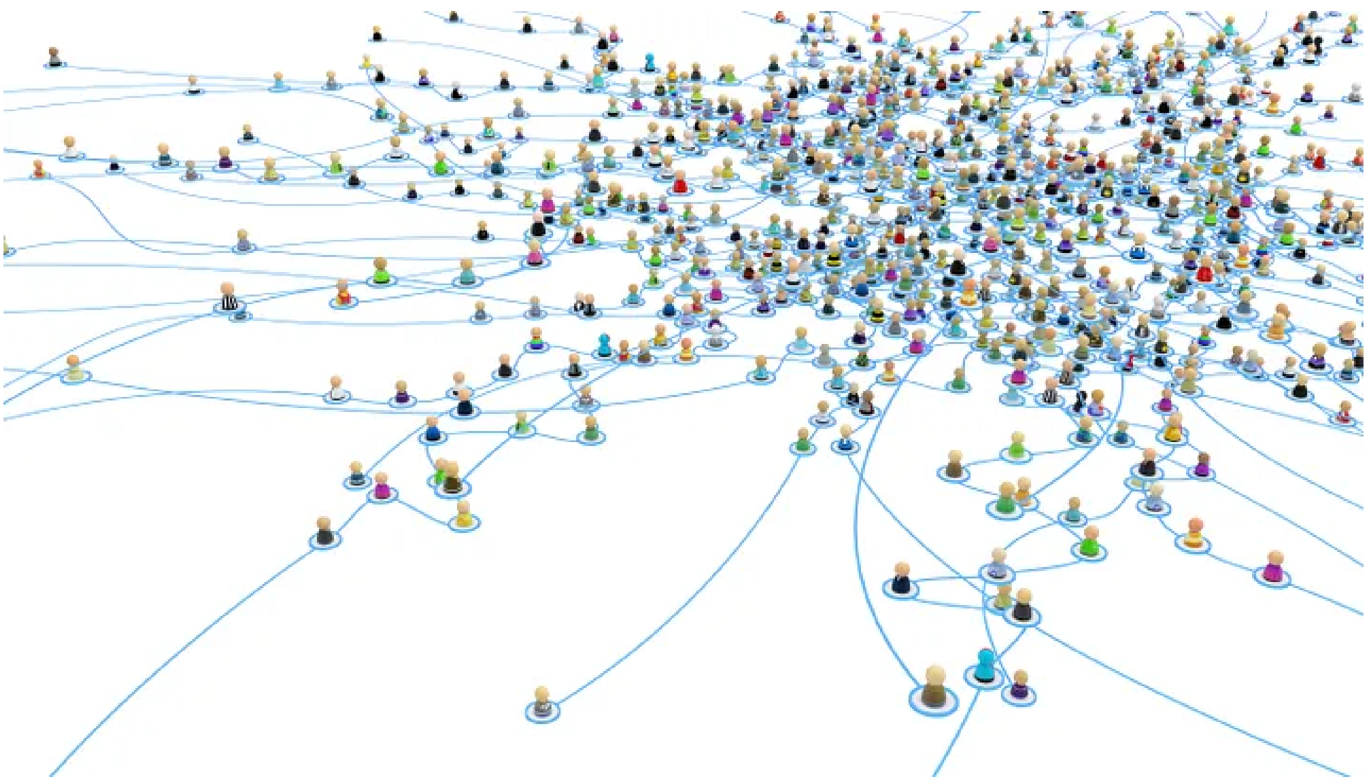
Semantics Rules. Last Ontology Part 2

John Ball · Follow
Published in Pat Inc · 8 min read · Jul 2, 2022



Are meaning-based relationships the right model for human-like language understanding? In semiotics, the study of signs, the combination of meanings enables capabilities for common sense and generalization, unlike combining just signs. Image Adobe Stock.

Semantics — i.e. meaning — rules our languages and is needed for human-like Natural Language Understanding (NLU).

When I incorporated PAT in 2006, the goal was to prove Patom theory (PT) to be an effective brain model. PT is my brain theory developed to explain what brains do, rather than how they do it — a high level model to follow the approach that was taken to explore atomic theory (where John Dalton studied **chemical** effects of theoretical elements — saying there are different kinds of atoms). Studying how brains use patterns removes the focus on low-level observations like a neuron's spike in preference to what makes brain elements contribute to observation.

While investigating toolsets to prove it, my focus moved to human language because the transducers needed (vision/cameras, audio/microphones and synthesizers, and motion/motors) were too dissimilar to their biological equivalents for ready use. Mpeg video compression is unlike the human eye.

But language is different, because it can be simulated with text in & out for conversation. The toolsets available to help all performed some function, but for only one language, and often by introducing unrecoverable errors such as when a probable result was selected over the correct one.

First, to solve the problem of parsing in language, many unsolved problems existed — especially the ambiguity of the scientific model. My next question was, once the unsolved problem of parsing was done: "what is needed to control conversation?"

The answer was to control meaning and to avoid the current issue of toolsets that work in a pipeline, where the pipeline is forced to make decisions that it cannot make at that level because the resolution of ambiguity can't be done until later.

Pipeline Problem Example:

"I saw her duck." If you **had** to choose the meaning, maybe you saw a duck that she owned. But that would be annoying if the sentence means that she dropped her head to avoid being hit by the ball. That sentence can be unambiguous in context, such as "The line drive went straight at Beth's head. ***I saw her duck***." Or in "Lisa went to the duck convention. ***I saw her duck***."

Clearly parsing sentence by sentence leaves out the context needed in conversation. That scientific **approach** (sentence by sentence analysis) is wrong! It is also **locked into** many attempts to use techniques to solve NLU. I understand why, as dealing with context is more complex, but parsing needs to incorporate the immediate common ground (the current context or context-of-utterance/CoU — discussed in the referenced note under "Language starts with context") for human-like accuracy and enable NLU.

How semantics is the last ontology — full integration with language

Previously I explained why semantics is the last ontology we will ever need and today I'll extend that to explain how conversation uses meaning to ***identify*** what we are talking about.

Think ahead a decade and imagine a world where our machines use **speech** to interact with us, with typing and other modalities being a distant second choice.

Let's consider (a) what we are learning about semantics, for use in the Next Generation of NLU systems and then (b) think through what we can remember that supports Patom theory.

Selecting Meaning from Something

What's bigger? New York or Los Angeles?

That's a question most of us can answer and it illustrates how the brain works, as does most language.

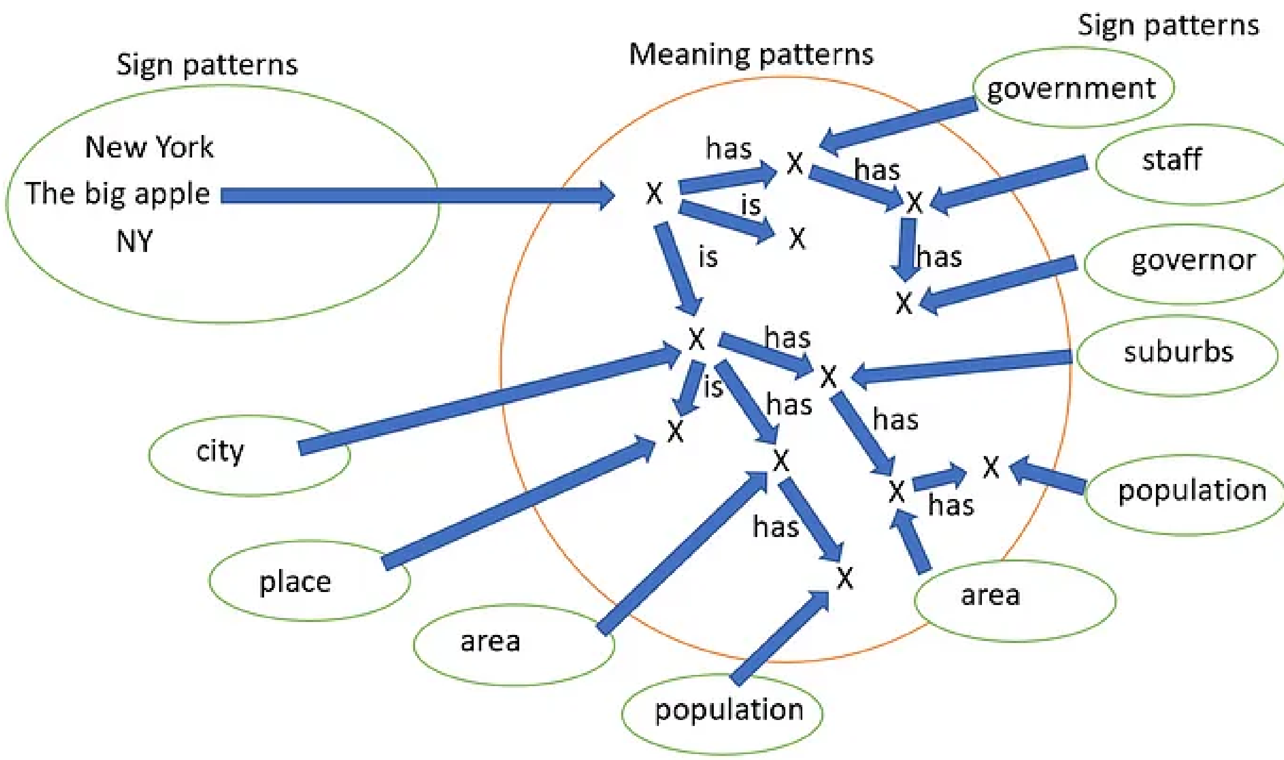


Figure 1. An example of a simple, meaning-based network for a city, where its semantic elements (is/has are basic hypernym/meronym relations) give the referent a number of additional attributes. Here, NY has a city with a population and an area, and it has suburbs that have an area and population.

Also, "what's bigger" is a question that doesn't really make sense because of its ambiguity. Figure 1 above illustrates how the semiotics model enables a rich set of associations to answer many comparative questions. The resolution of meaning is left with context, not some kind of educated guess. Is size referring to population, area of the city or its suburbs or both? Or the number of staff in its government. Or …

If the answer you give is NY, you may support that by the number of people, or the wealth per person, or the size of the suburbs… And you can answer the same way for LA. What's going on?

That interaction shows how our brain deals with two key elements in our repository of understanding (a super knowledge graph or SKG). First, we can generalize about things like properties of referents/things. And second, we can represent an awful lot of information in our small brain.

Encyclopaedic Knowledge in Conversation

You know the guy who saved millions of lives in World War Two by helping to break the Enigma Code with the Bombe? Who was that?

That was Alan Turing. But if you knew the answer, your brain would have needed to have stored **huge numbers** of statements with *signs* in English as is a starting point in machine-learning models, or at least **one** semantic relationship.

How does our brain solve the problem to recognize such language?

I define a super knowledge graph (SKG) as a lossless approach to recording context, the set of active elements within a conversation (known as the Context of Utterance or CoU in one modern model). Technically, immediate common ground (ICG) a kind of short-term memory, identifies the CoU which is then stored into general common ground, a kind of long-term memory.

That's a lot of knowledge to manage, without loss and while still retaining instant access to it.

Representation — The NER Compromise

The representation of Alan Turing is shown below in Figure 2.

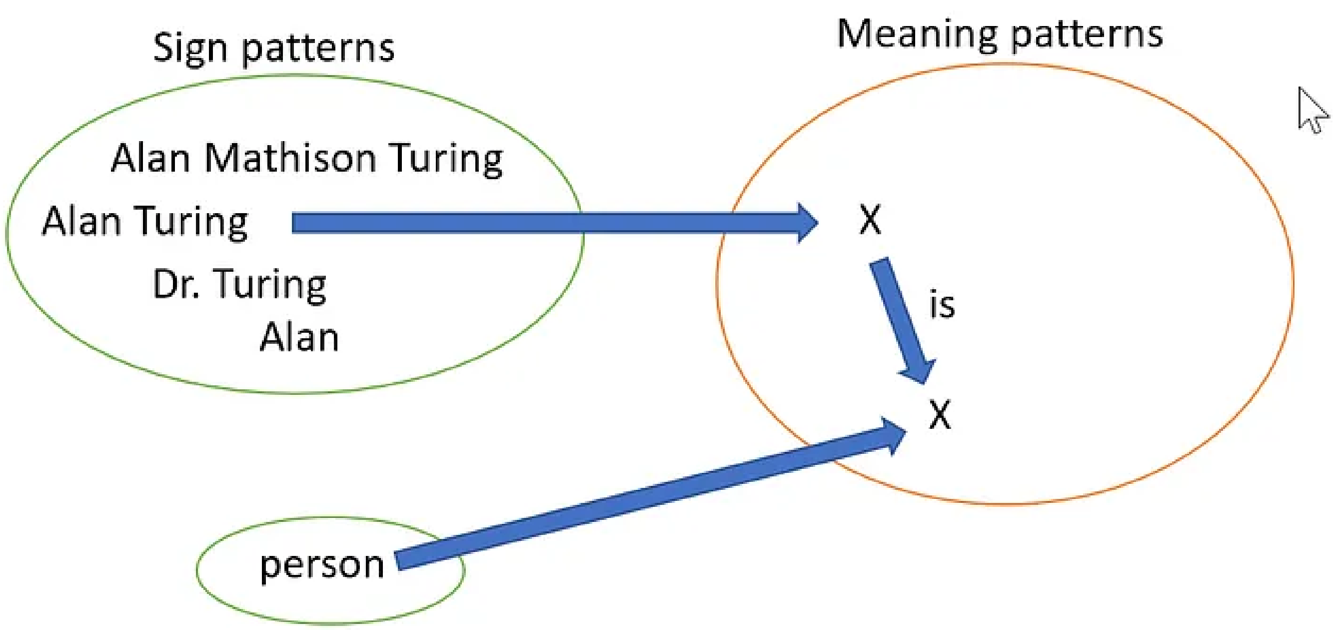


Figure 2. Initial Representation of Alan Turing — a KIND OF (hypernym) person. This is like the NER model.

Notice that the recognition of a name can include pattern matching of the signs — 'Turing' or 'Alan M. Turing' or 'Dr. Turing.' They are just three of a wide-range of possible patterns that match. Depending on the context/ICG, Alan may be sufficient for unambiguous recognition, or even Dr. Alan Mathison Turing may be **in**adequate. The narrowing down of a referent correctly is a key aspect of language involving questions and answers. It is something that is needed for a lossless SKG since errors in such identification results in data integrity loss.

At this point, we have a representation of meaning that is aligned with today's Named Entity Recognition (NER) systems. NER systems provide categorisation of names — e.g. person, location, organization, and money. The weakness of this approach is that many objects are ambiguous.

In “New York declared a war on drugs” the phrase ‘New York’ is a place that has a government whose leadership made a public statement. The place didn’t say anything!

The point is that NER is a ***compromise*** that is only correctly resolved when the meaning is identified in context. The category comes from the actual referent itself, not a statistical analysis of a collection of signs (like a corpus). If we are to understand a thing (the referent of the sign), a richer model of meaning is needed with more than just a selection of some arbitrary elements (the NER resolution set of elements).

What would that be?

Fortunately, there is something we can use to replace NER with a human-like capability. It is the interpretant in the semiotics model. There is a more colloquial term … **Meaning.**

Using Meaning instead of NER

Meaning is far better. We want to convert what it means to understand Alan Turing to understand the earlier reference: “the guy who saved millions of lives in World War Two by helping to break the Enigma Code with the Bombe.”

The historians among you who have looked at AI from the time it was named in 1956 will see this as the relationship between dictionary entries and encyclopaedic knowledge. Normally, we think of dictionaries as defining words and phrases. And then perhaps, adding a definition in the source language, its parts-of-speech and its pronunciation and etymology. We think of encyclopedias as explaining historical events and other factual information in a particular language.

But the Next Generation of NER merges both of these together in an SKG, super knowledge graph, representation. These resources are best created with automation tools, as SKG entries take a while to render manually, and are error-prone, since few people have expert knowledge of tense, modality, aspect and so on, with the connection between predicates and their selectional restrictions within context/ICG.

CoU adds dynamically by the recognition of new context and its insertion into the SKG. And the Role and Reference Grammar (RRG)'s linking algorithm for any language converts signs into meaning. So based on language-independent meaning, we have an SKG produced automatically.

Sound too good to be true?

This is just reflecting what humans appear to do under their brain's control.

Implementation

An SKG implementation, creating structure from meaning, can be seen in Figure 3 below.

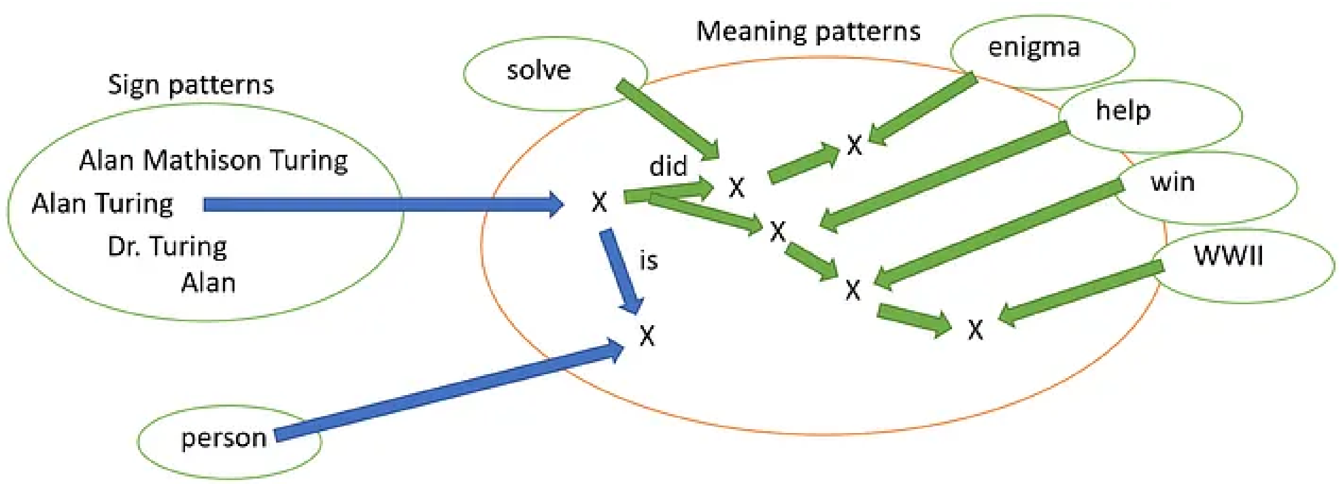


Figure 3. A view of a sign sequence matched to resolve 'Alan Turing'. Note meaning is not represented here beyond its pattern-atom.

Through experience, you can imagine that the Alan Turing Patom is recognized in some conversation. And through the use of that Patom, additional associations are added: “he did solve Enigma” and “he helped to win WWII”. The representations of those meanings through a semantic representation aren’t shown in Figure 3, but they are just SKG elements.

Note that the representation above can’t just be a sequence, as seen with the arrows, as there is a juncture there (helped to win…), multiple predicates (‘help’, ‘win’, ‘solve’) and direct relationships to Turing, as well as potential generalization with any person (e.g. “Some people may have solved the Enigma code,” and “Some people may have helped to win WWII,” and “WWII was won”).

By hearing: “Who solved the Enigma code and helped to win WWII?” the system matches the green elements. Through the Patom theory concept of linkset intersection, only elements consistent with those two facts remain: leaving, in many people’s *SKG*s, Dr. Turing as the only match.

The architecture of Patom theory is bidirectional. By storing the meaning of signs in conversation allows the generation of that meaning into a sequence of signs with phrases. Similarly, the recognition of patterns in meaning allows the resolution of the experience with any relevant experiences. That’s the technique that requires no “intelligence” beyond the pattern matching, but even without intelligence, there is great power in the use of the semanticist categories for conversation (as we do at PAT) and here, now, in referent identification through encyclopaedic knowledge matching.

Obviously, it is a prerequisite to be able to convert language into its meaning (such as with the RRG linking algorithm) and also to convert meaning into a target language (such as with the RRG linking algorithm), but given that, the representation of encyclopaedic knowledge is a good way to represent language-independent meaning.

Summary

The world of linguistics has been dominated by a model based on syntax-first parsing for a long time, certainly since the 1930s. But before that, semiotics put the case for both signs (words and fixed phrases) and meaning. By bringing back the older models, there is wide scope for a new generation of systems that exploit NLU.

Today, many systems involving language advocate the use of Named Entity Recognition to identify fragments of meaningful text sequences. “Qantas” would be identified, perhaps, as a company. “Alan Turing” could be identified as person.

But the case needs to be made for the integration of resources like dictionaries and encyclopaedias, probably to language-specific repositories to handle the recognition and generation requirements, along with language-independent repositories to solve the limitations of today’s NLU systems as outlined previously.

The ability to store meaning in a consistent way that relates to the specific things in the world promises new human-like capabilities that will usher in a new world of possibilities for NLU systems.

Linguistics | Machine Learning | Naturallanguageprocessing | Artificial Intelligence | Knowledge

More from the list: "NLP"

Curated by Himanshu Birla

Jon Gi... in Towards Data ...

Characteristics of Word Embeddings

· 11 min read · Sep 4, 2021

Jon Gi... in Towards Data ...

The Word2vec Hyperparameters

· 6 min read · Sep 3, 2021

Jon Gi... in

The Word2ve

· 15 min rea

View list

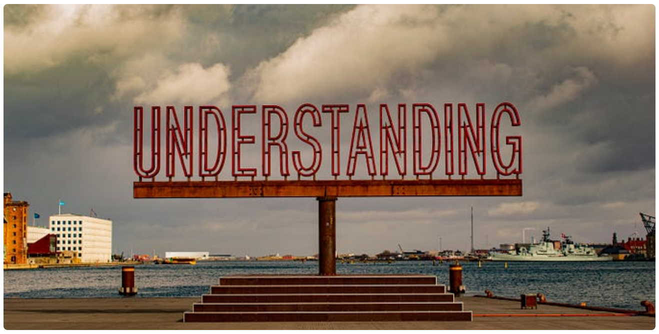
Written by John Ball

Follow

1.7K Followers · Editor for Pat Inc

I'm a cognitive scientist working on NLU (Natural Language Understanding) systems based on RRG (Role and Reference Grammar). A mouthful, I know!

More from John Ball and Pat Inc



John Ball in Pat Inc

2023: The Year of Understanding Theory

Cybernetics was a name before AI was coined, and it is probably a better label for…

· 16 min read · Jan 5

28 4

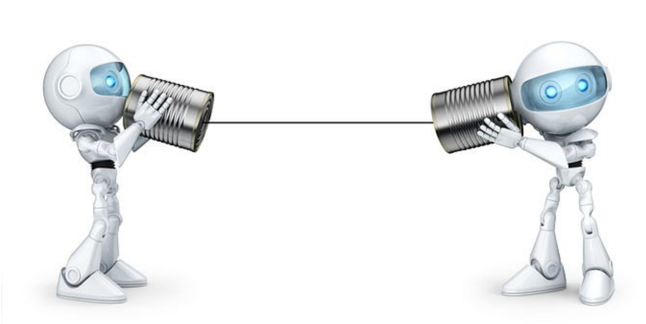
John Ball in Pat Inc

NLU and Consciousness

Natural Language Understanding (NLU) doesn't rely on consciousness, nor do most…

· 10 min read · Oct 28, 2022

99 7



John Ball in Pat Inc

Fixing Automatic Speech Recognition (ASR) with NLU

Why doesn't speech recognition operate at human-levels?

· 10 min read · Oct 22, 2022

94 2

John Ball

66 words handles 50% of language!?

At my company, PAT, we are working to replace today's systems with ones that…

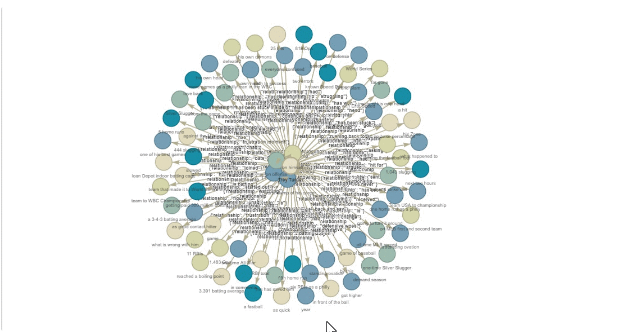
· 11 min read · Jul 6, 2022

107

See all from John Ball

See all from Pat Inc

Recommended from Medium



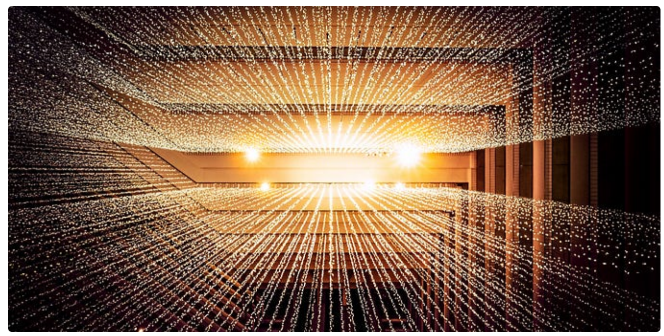
Wenqi Glantz in Better Programming

7 Query Strategies for Navigating Knowledge Graphs With...

Exploring NebulaGraph RAG Pipeline with the Philadelphia Phillies

· 17 min read · 4 days ago

501 4



Beatriz Stollnitz in Towards Data Science

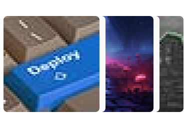
Add Your Own Data to an LLM Using Retrieval-Augmented...

Learn how to add your own proprietary data to a pre-trained LLM using a prompt-based...

21 min read · 4 days ago

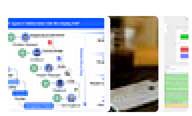
189 1

Lists



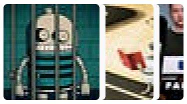
Predictive Modeling w/ Python

20 stories · 452 saves



Natural Language Processing

669 stories · 283 saves



AI Regulation

6 stories · 138 saves

ChatGPT prompts

24 stories · 459 saves

David Shapiro

A Pro's Guide to Finetuning LLMs

Large language models (LLMs) like GPT-3 and Llama have shown immense promise for...

12 min read · Sep 23

283 6

0xdevshah in AI Skunks

Knowledge Graphs: A Comprehensive Analysis

Knowledge graphs are becoming increasingly important in NLP due to their ability to mode...

8 min read · Apr 9

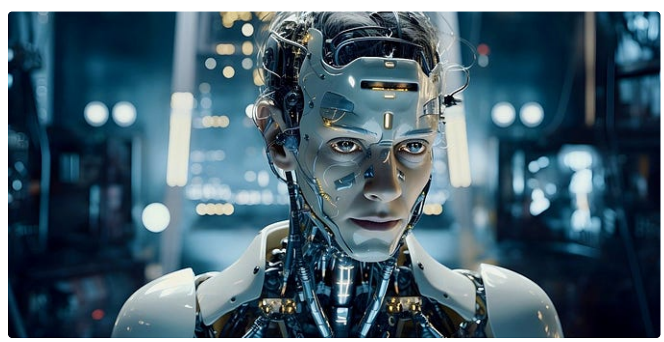
63

Will Lockett in Predict

China May Have Just Condemned The Entire World

A crucial shift in rhetoric from the superpower is deeply worrying.

· 7 min read · Sep 26



Nidhi Jain in Level Up Coding

Demystifying the Magic behind Large Language Models: The...

Step into the realm of AI architecture and uncover the designs that fuel the large...

· 10 min read · Sep 25