MAGAZINE

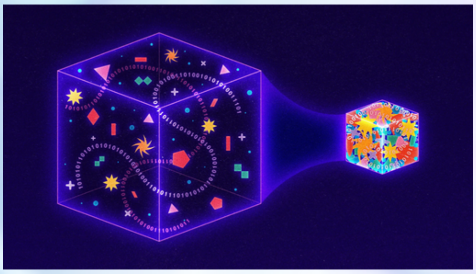
ISSUE NO: 141
September 18, 2023

Byte Quest

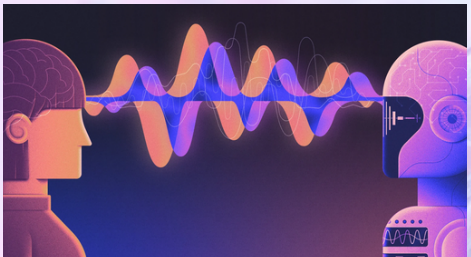
Department of
CSE

CHATBOTS DON'T KNOW WHAT STUFF ISN'T

SECRET MESSAGES CAN HIDE IN AI-GENERATED MEDIA



DATA COMPRESSION DRIVES THE INTERNET. .



SOME NEURAL NETWORKS LEARN LANGUAGE LIKE HUMANS

Department Vision

To be a center for academic excellence in the field of Computer Science and Engineering education to enable graduates to be ethical and competent professionals.

Department Mission

To enable students to develop logic and problem solving approach that will help build their careers in the innovative field of computing and provide creative solutions for the benefit of society.

FACULTY COORDINATORS

DR. BHARGAVI PEDDIREDDY
(ASSOCIATE PROFESSOR)
S. KOMAL KAUR
(ASST. PROFESSOR)

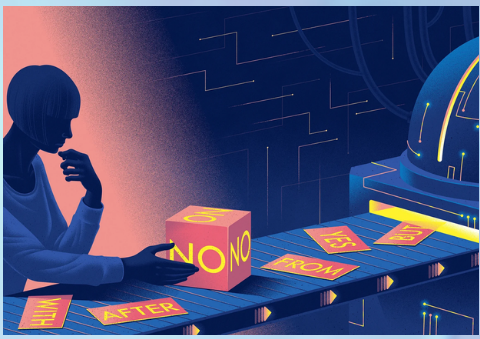
STUDENT COORDINATORS

VAMSI (3/4) CSE C
SPOORTHI (3/4) CSE C

Byte Quest

CHATBOTS DON'T KNOW WHAT STUFF ISN'T

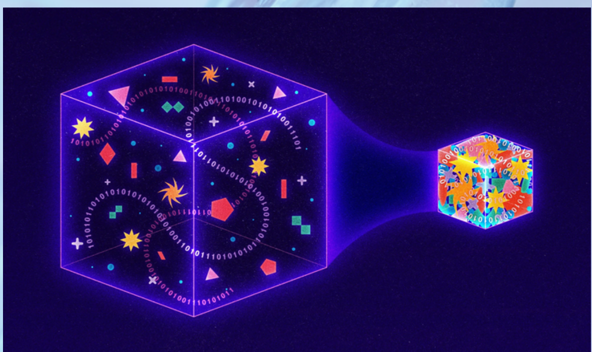
Nora Kassner, a researcher, questioned the genuine understanding of language by large language models (LLMs) like BERT, discovering their inability to comprehend negation. Despite significant advancements in LLMs' size and capability, their struggles with nuanced logic involving words like "not" persisted. Unlike humans, LLMs process language mathematically, focusing on predictions rather than genuine understanding. Their reliance on mathematical weights and limited exposure to negative statements hindered their grasp of negation, evident in responses where negations were misunderstood or overlooked.



The problem lies in how LLMs learn. They lack the innate understanding humans possess, perceiving words as mathematical associations devoid of genuine meaning. Even advancements like human-in-the-loop learning or algorithmic modifications offer partial solutions. Some proposed strategies involve adjusting training data or incorporating additional layers in language processing, yet these approaches are far from foolproof. While larger models with diverse training data might hold promise, skeptics believe a fundamental shift in how LLMs learn and understand language is necessary. This challenge hints at the complexity of replicating human-like understanding in machines and suggests that scaling models and adding more data might not be the sole solution.

DATA COMPRESSION DRIVES THE INTERNET. HERE'S HOW IT WORKS.

Python takes less number of lines for any problem. Python is very simple to learn since the syntax of Python is similar to the English language. It offers different data types like tuples, sets, dictionaries, etc. Python is also used in interviews to write code. Despite all these pros, Python imports many libraries, which take a lot of time to run, so using Python comes with a time complexity error.



Java is completely object-oriented. Java takes many lines for a simple problem. It is even hard to learn Java since the syntax is similar to C and C++. Java offers simple data types like int, float, double, etc. It is hard to write code in Java compared to Python. Java is versatile, with many features, such as memory management. Java is faster than Python.

Even though Python is good in many things, it can't replace Java in some aspects like speed, etc. Java has no competition when it comes to APIs. Simply both are significant.

SECRET MESSAGES CAN HIDE IN AI-GENERATED MEDIA

The arrest of Russian spies in 2010, working as American professionals near New York City, exposed a sophisticated spy network that utilized steganography—an art of concealing secret messages within innocuous ones—to communicate covertly. Unlike cryptography that encrypts messages, steganography aims to hide the very existence of a secret message. However, achieving perfect security, where neither human nor machine can detect the hidden message, has long been considered unattainable, particularly in human-generated text due to its complexity.

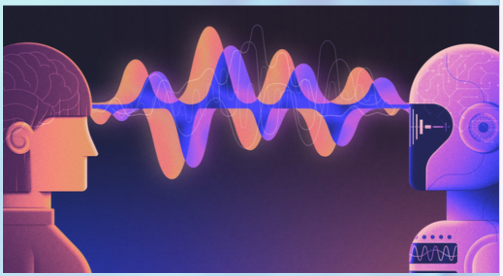
Recent advancements in large language models like ChatGPT have revolutionized the landscape. A breakthrough in steganography has been achieved for machine-generated messages—texts, images, videos, and more—promising perfect security. This achievement arises from the application of information theory and minimum entropy coupling, ensuring that the statistical properties of the hidden message remain indistinguishable from the cover text. While perfect simulation in human-generated text remains challenging due to its inherent complexity, machine-generated content, created by AI models, provides a promising avenue for achieving truly secure steganography.

This novel approach bridges the gap between theory and practicality, offering potential applications in scenarios like bypassing censorship or securely transmitting information. However, computational limitations currently restrict the size of concealed messages in real-world applications. Still, as machine-generated content becomes more prevalent, the potential for achieving perfect security in steganography grows stronger, heralding both promising and concerning implications for security and privacy in the digital age.

Byte Quest

SOME NEURAL NETWORKS LEARN LANGUAGE LIKE HUMANS

Gašper Beguš and colleagues at UC Berkeley conducted a groundbreaking study comparing human brain waves while listening to a syllable with signals from neural networks trained on the same sounds. Surprisingly, the artificial networks, built with general-purpose neurons without language-specific biases, displayed activity closely resembling human brain waves. This challenges the notion that language acquisition relies on unique innate machinery in the brain.



Using a generative adversarial network (GAN), the researchers found that the network's discriminator showed activity akin to human brain waves after training, indicating a feedback mechanism similar to caregiver influence on infant sound production. Furthermore, differences in how English and Spanish speakers processed the sound were mirrored in the English- and Spanish-trained neural networks.
These findings question established theories positing innate language capabilities and suggest that language learning might not necessitate specialized brain features, offering insights into both human and artificial language acquisition.

BROUGHT TO YOU BY

Department of
Computer Science and Engineering

Vasavi College of Engineering