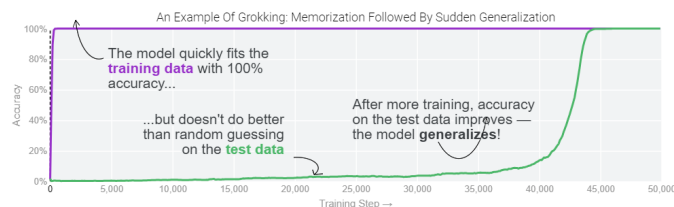


TRENDS

- [Paper] [Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks](#)
The paper explores the abstract reasoning capabilities of large language models (LLMs) by testing them on modified ("counterfactual") tasks. While LLMs show some level of abstract reasoning, their performance drops significantly on these counterfactual tasks, indicating a reliance on task-specific procedures. The study calls for a more nuanced understanding of what LLMs can and cannot do.
- [Paper] [Lost in the Middle: How Language Models Use Long Contexts](#)
The paper examines how well language models (LMs) utilize long input contexts, focusing on multi-document question answering and key-value retrieval tasks. The study finds that LMs perform best when relevant information is at the beginning or end of the context, but struggle with information in the middle. Performance also drops as context length increases, even for models designed for long contexts. The research offers new insights into LMs' context utilization and suggests evaluation protocols for future models.
- [Blog] [Do Machine Learning Models Memorize or Generalize?](#)
The paper highlights the 2021 discovery of "grokking," where tiny models suddenly transition from memorizing to generalizing on new data after long training periods. This phenomenon has sparked considerable research interest.

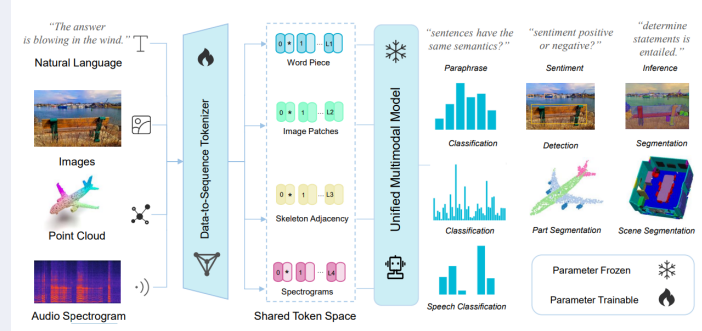


- [Vlog] [LIQUID NEURAL NETWORKS DO MORE WITH LESS](#)
This video blog, or Vlog for short, presents a machine vision system that uses a "liquid neural network" to steer a car and maintain its lane. Remarkably, the system performs well with just 19 neurons, far fewer than typical large models. An attention map shows that the system focuses on aspects of the visual field much like a human driver.

STATE OF THE ART

- [blog] [Innovative Recommendation Applications Using Two Tower Embeddings at Uber](#)
In 2022, Uber's ML team revamped their Michelangelo platform with Two-Tower Embeddings to enhance recommendation systems. The update boosts user experience, cuts computing time, and improves scalability and performance.
- [Paper] [Less is More: Parameter-Free Text Classification with Gzip](#)
The paper introduces a simple, lightweight method for text classification that combines gzip compression with a k-nearest-neighbor classifier. It performs competitively without training and outperforms BERT in out-of-distribution and few-shot scenarios.
- [Paper] [Meta-Transformer: A Unified Framework for Multimodal Learning](#)

The paper introduces Meta-Transformer, a framework designed for multimodal learning across various types of data like text, images, and audio. Using a frozen encoder, it maps different modalities into a shared token space without requiring paired multimodal training data. The framework can handle a wide range of tasks and data types, offering a unified approach for multimodal intelligence.



MISCELLANEOUS

- [Paper] [Birth of a Transformer: A Memory Viewpoint](#)
The paper investigates how transformer-based language models manage stored knowledge and adapt to new context. Using synthetic bigram distributions, the study uncovers that these models quickly learn global patterns but develop in-context understanding more slowly. The research also highlights the role of weight matrices as associative memories and offers theoretical insights on the learning process.
 - [Blog] [Active Sampling: Data Selection for Efficient Model Training](#)
The article introduces active sampling techniques for handling large datasets in machine learning. These methods select crucial data points to optimize task performance, especially useful in high-data, low-budget scenarios.
-
- [Paper] [OCR-free Document Understanding Transformer](#)
The paper presents Donut, an OCR-free model for Visual Document Understanding (VDU). Donut overcomes the limitations of OCR-based methods, offering better speed and accuracy. It uses a simple Transformer architecture and includes a synthetic data generator for flexible pre-training.

LATEST RELEASES

- [Package overhaul] [Keras core](#)
Keras Core is a rewritten codebase designed to run on multiple frameworks like TensorFlow, JAX, and PyTorch, offering improved performance and broader compatibility. It allows for resource utilization across different ecosystems, supports various data pipelines, and provides seamless integration with native workflows, setting the stage for the release of Keras 3.0 in Fall 2023.
- [Minor release] [Scipy v1.11.0](#)
The latest SciPy release features significant improvements in sparse array API, including a new base class and 64-bit index support. It adds new tools in scipy.stats for survival analysis and hypothesis testing, introduces a function for quasi-Monte Carlo integration, and extends linear algebra functions to accept nD-arrays. An axes argument has also been added to ndimage functions for better image data analysis.
- [Minor release] [Pandas v2.1.0](#)
The major updates in the pandas 2.1.0 release include making PyArrow a required dependency and avoiding NumPy object dtype for strings by default. It also introduces DataFrame reductions that preserve extension dtypes, a new DataFrame.map() method,

support for ExtensionArrays, and a new implementation of DataFrame.stack().

- [Minor release] [Tensorflow v2.14.0](#)
The TensorFlow 2.14 release removes Python 3.8 support and introduces new EagerTensor and SymbolicTensor classes. It adds an optional CUDA installation method for Linux and enables JIT-compiled kernels for large tensors on GPUs. The release also features experimental support for tf.lite models larger than 2GB and new APIs in `tf.experimental.dtensor` and `tf.experimental.strict_mode`.

EVENTS

Upcoming Conferences

- [Conference] [ECML PKDD](#)
The upcoming 18th edition of the ECML PKDD conference, taking place from September 18-22, 2023, promises to be a landmark event in the fields of machine learning and data mining. We encourage you to visit the conference website to delve into cutting-edge developments in areas such as predictive modeling and knowledge discovery, and to catch the standout moments from this exceptional gathering.
- [Conference] [RECSYS](#)
The forthcoming 17th edition of the RecSys conference, scheduled for September 18-22, 2023, serves as a pivotal occasion in the realm of recommender systems. We invite you to browse the conference website to uncover the latest innovations in recommendation algorithms, user experience, and system evaluation, and to experience the key highlights from this distinguished event.

Past Conferences Worth Checking Out

- [Conference] [SIGGRAPH](#)
The landmark 50th SIGGRAPH conference took place from August 6-10, 2023, and showcased cutting-edge developments in digital art and interactive techniques. If you missed it, you can visit the conference website to check out the latest innovations that were presented during this premier event in the world of computer graphics.
- [Conference] [IJCAI](#)
The 19th-25th of August, 2023 marked the staging of IJCAI, an influential event in the field of artificial intelligence. If you missed it, you can visit the conference website for insights into what was presented at this leading summit on AI research and developments.
- [Conference] [KDD](#)
The KDD conference, a premier event in data science, was held from August 6-10, 2023. If you missed it, you can visit the conference website for an overview of what was presented at this key gathering on knowledge discovery and data mining.

Merci !