INSTITUT Louis Bachelier LABS

# NEWSLETTER AI & TECH

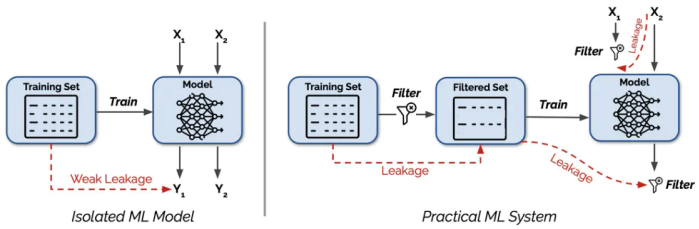
10/10/2023

## Announcement

Hello there! 🌞 This October, we've got a treat for you. You're currently reading our general ML-themed newsletter, and in just two weeks, we'll be sending out a special issue focused entirely on LLMs. Stay tuned and enjoy your reading journey!

## TRENDS

- [Paper] Privacy side channels in machine learning systems
  In the domain of machine learning, models typically form just an aspect of a broader system. While it's recognized that unprotected models can unintentionally reveal training data, the prevailing discussion mainly centers around the models. A recent seminal study broadens this scope, exploring the privacy risks associated with auxiliary system components. The research unveils a range of side-channel attacks targeting these peripheral elements, revealing a level of private information disclosure that exceeds that of attacks solely targeting the models.

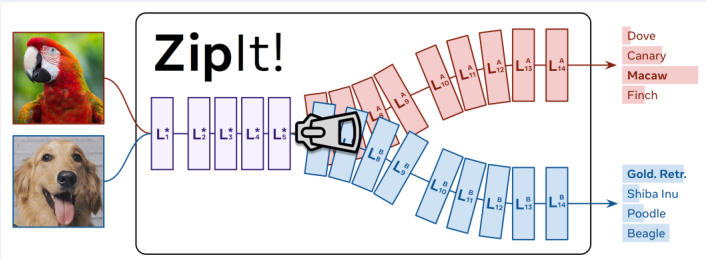



- [Blog] Beyond Words: How Multimodal Embeddings Elevate eBay's Product Recommendations
  In its quest for a flawless purchasing experience, eBay focuses on elevating the quality of listings, with a keen eye on imagery and textual content. Historically, inadequate images could misrepresent products, potentially leading to buyer dissatisfaction. The distinct storage of text and image embeddings posed challenges in creating a cohesive recommendation mechanism. Addressing this, eBay has crafted a sophisticated system that melds varied modalities and bespoke modules, such as image-text discordance detection and the integration of triplet loss with TransH. This refined approach bolsters eBay's recommendation engine, leading to an approximate 15% uptick in buyer interaction.

## STATE OF THE ART

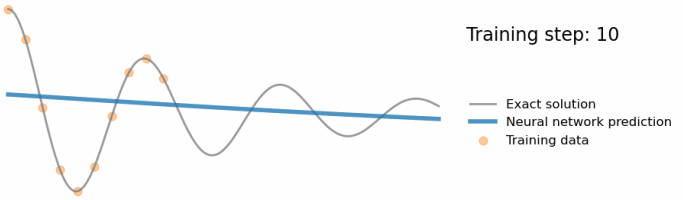
- [Paper] DOMINO: Discovering systematic errors with cross-modal embeddings
  The paper highlights the challenges machine learning models face in accurately processing specific data slices in high-dimensional inputs. Through the introduction of "Domino", an innovative slice discovery method employing cross-modal embeddings, the authors demonstrate its proficiency in slice detection and its pioneering ability to generate natural language descriptions for these segments.
- [Paper] ZipIt! Merging Models from Different Tasks without Training
  The paper presents "ZipIt!", an advanced method for merging distinct deep visual recognition models into a unified multi-task

model. Traditional techniques falter with separate tasks, but "ZipIt!" offers a nuanced "zip" operation and supports selective merging, enhancing performance by 20-60% over prior methods.




## MISCELLANEOUS

- [Blog] So, what is a physics-informed neural network?
  Machine learning has witnessed an ascension in prominence within the scientific domain. Yet, a salient inquiry emerges: do these algorithms truly grasp the scientific intricacies they aim to unravel? This treatise introduces physics-informed neural networks, an avant-garde approach that adeptly integrates fundamental physical principles into the machine learning paradigm.




- [Package] Carton : Run any ML model from any programming language
  Carton emerges as a versatile tool, propelling the execution of Machine Learning (ML) models across diverse programming environments. It encapsulates models, retaining their originality, while ensuring compatibility with specified frameworks. The ingenuity lies in its framework-agnostic approach, enabling seamless model operation irrespective of the underlying ML framework. Boasting a low overhead and broad platform support, Carton not only simplifies model deployment but accelerates the iterative processes of ML experimentation. Unlike ONNX, which transmutes models, Carton retains the model's native framework during execution, thus obviating potential conversion pitfalls and fostering expeditious deployment and iteration.
- [Paper] On the Strategyproofness of the Geometric Median
  This scholarly analysis delves into the strategyproofness of the geometric median in high-dimensional spaces, crucial for robust machine learning and content moderation. Although traditionally robust against malicious voter minorities, the geometric median's strategyproofness in high dimensions is not thoroughly explored. The authors find it not entirely strategyproof in high dimensions, yet it approximates strategyproofness with ample voter numbers. The manuscript also explores the effect of voter preferences across dimensions on strategyproofness and considers using skewed geometric medians to enhance strategyproofness, despite inherent limitations. This provides insight into reconciling high-dimensional disagreements through the geometric median.
- [Blog] Your Features Are Important? It Doesn't Mean They Are Good
  The insightful blog post sheds light on the often misunderstood idea of "feature importance" in machine learning, which is crucial for

explaining model behaviors. It clarifies the mistake of equating "importance" with "beneficial" when it comes to features, explaining that an important feature, showcased through the accidental inclusion of 'Customer ID', may negatively affect the model's ability to predict unseen data, even though it significantly contributes to predictions. The post also introduces a clear distinction between 'Prediction Contribution' and 'Error Contribution', aiming to accurately measure these aspects to enhance understanding and potential improvement of predictive models.

## LATEST RELEASES

- [Minor releases] Python 3.12
  Python 3.12 introduces several noteworthy enhancements that will be beneficial for developers. Some of these include improved error messages offering valuable suggestions, enriched f-strings functionality thanks to Python's PEG parser, optimization features like inlined comprehensions for a performance boost, a novel syntax for type variables aiding in annotating generics, and the inclusion of a robust perf profiler on Linux for better performance analysis


- [Minor release] Matplotlib 3.8.0
  In the recent Matplotlib 3.8.0 release, key enhancements include the introduction of Type Hints, and a suite of Plotting and Annotation upgrades like customizable antialiasing for text, and configurable legend shadows. A new class, PolyQuadMesh, has been introduced for drawing quadrilateral meshes, adding a fresh capability. Layout improvements like a new public method for modifying Legend location and a `pad_inches="layout"` option for `savefig` have been added, making it more user-friendly. Moreover, 3D plotting saw advancements like the ability to specify tick and axis label positions, enriching the 3D visualization experience.
- [Minor release] Seaborn 0.13.0
  The Seaborn version 0.13.0 release introduced significant enhancements, notably a comprehensive revamp of the categorical plotting functions, enriching them with new capabilities and better aligning their API with the rest of the library. Additionally, the release provided provisional support for alternative dataframe libraries like Polars, and introduced a new theme and display configuration system for objects.Plot, alongside many smaller bug fixes and enhancements.
- [Minor release] Pytorch 2.1
  The PyTorch 2.1 release unveils substantial upgrades like automatic dynamic shape support via torch.compile, enabling streamlined saving and loading of distributed training jobs with torch.distributed.checkpoint, and broadening torch.compile support for the NumPy API. Performance enhancements such as CPU inductor improvements, AVX512 support, and scaled-dot-product-attention support are also featured. Additionally, a prototype of torch.export is introduced, delivering a solid full-graph capture mechanism, alongside a torch.export-based quantization, reflecting ongoing endeavors to bolster PyTorch's functionalities.

## EVENTS

- [Conference] ICCV
  The prestigious International Conference on Computer Vision (ICCV) 2023 took place from October 2-6 at the Paris Convention Center in Paris, France. Being a paramount event in the computer vision domain, ICCV 2023 showcased a wide array of innovative research and cutting-edge technologies over these five days. If you missed attending, the conference website will serve as a rich resource to catch up on the groundbreaking insights shared in the realm of computer vision.

*Merci !*