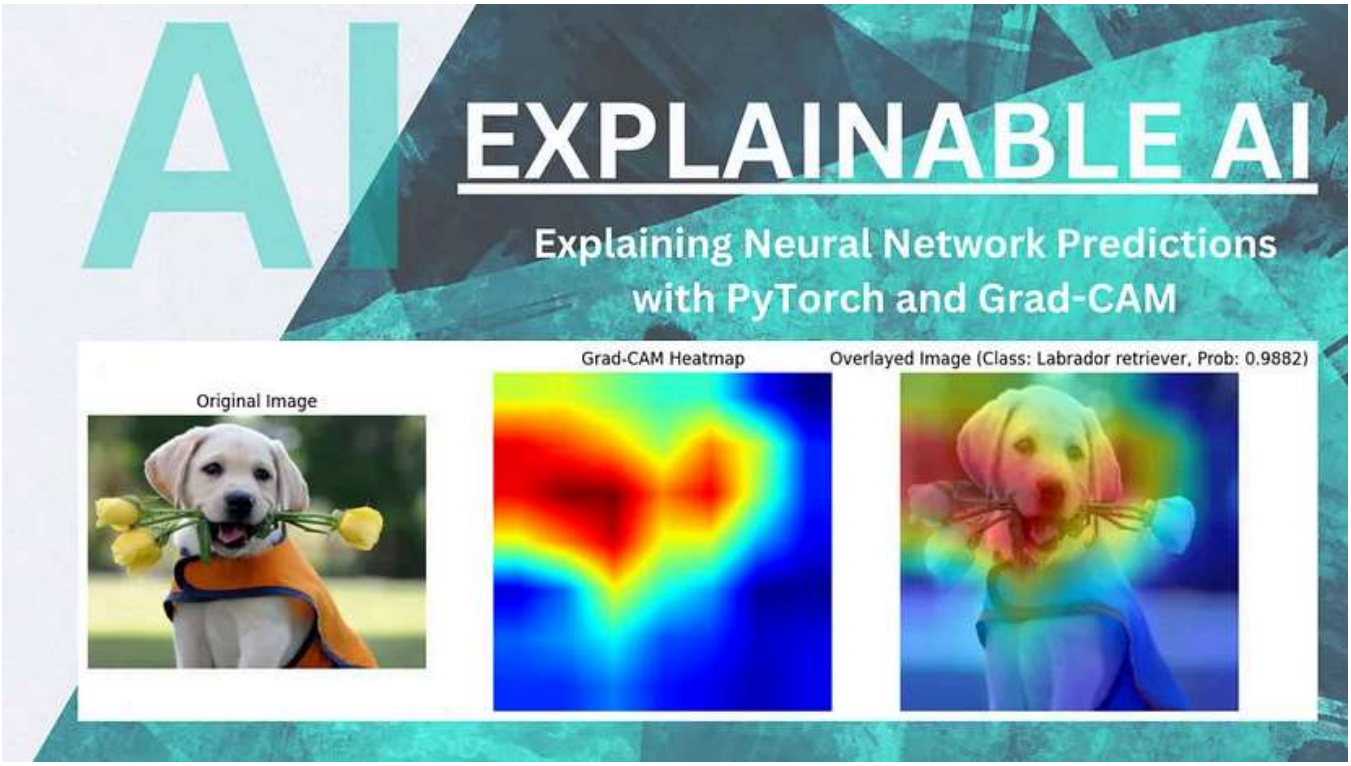
Explainable AI with PyTorch and Grad-CAM

Sawera Khadium · Follow
2 min read · Feb 5, 2024



Introduction:

As artificial intelligence (AI) systems become more prevalent, understanding their decision-making processes is crucial. Explainable AI (XAI) aims to make machine learning models interpretable and transparent. In this article, I'll explore an end-to-end project that uses PyTorch and the Gradient-weighted Class Activation Mapping (Grad-CAM) technique to create visual explanations for neural network predictions on the ImageNet dataset.

For in depth practical explanation you can visit this colab notebook.

What is Grad-CAM?

Grad-CAM is a technique used to visualize the importance of regions in an input image for a specific class prediction. It generates heatmaps by combining the gradient information flowing into the last convolutional layer of the CNN with the activation of that layer. These heatmaps help understand which parts of the input image contribute the most to the model's final prediction.

Using PyTorch and Grad-CAM

I have used PyTorch, a popular deep learning framework, to load a pre-trained ResNet-50 model and preprocess input images. Then implemented Grad-CAM to generate visual explanations for the model's predictions. The project involves visualizing the original image, the Grad-CAM heatmap, and an overlay of the heatmap on the original image.

Gaining Insights

By analyzing the Grad-CAM visualizations, it becomes clear that which regions of the input image have the most significant impact on the model's decision. This is invaluable in gaining insights into the model's decision-making process, identifying potential biases or errors, and improving trust in AI systems.

Improving Model Performance

While Grad-CAM helps us understand our model's predictions, it does not inherently improve the model's performance. To achieve better performance, consider fine-tuning the model on a more diverse dataset, use techniques like data augmentation, or explore more advanced architectures.

Conclusion:

Explainable AI techniques, like Grad-CAM, plays a vital role in understanding and interpreting the decision-making processes of AI systems. By implementing this end-to-end project using PyTorch and Grad-CAM, we can visualize the impact of input image regions on the model's predictions, gain valuable insights, and work towards more transparent and trustworthy AI systems.

Responsible Ai | Explainable Ai | Grad Cam | Deep Learning | AI

Follow

Written by Sawera Khadium

193 Followers

AI Engineer specializing in Python, with a passion for NLP, Computer Vision, and Data Analytics. Over 4 years of experience leading high-impact projects in AI.

More from Sawera Khadium

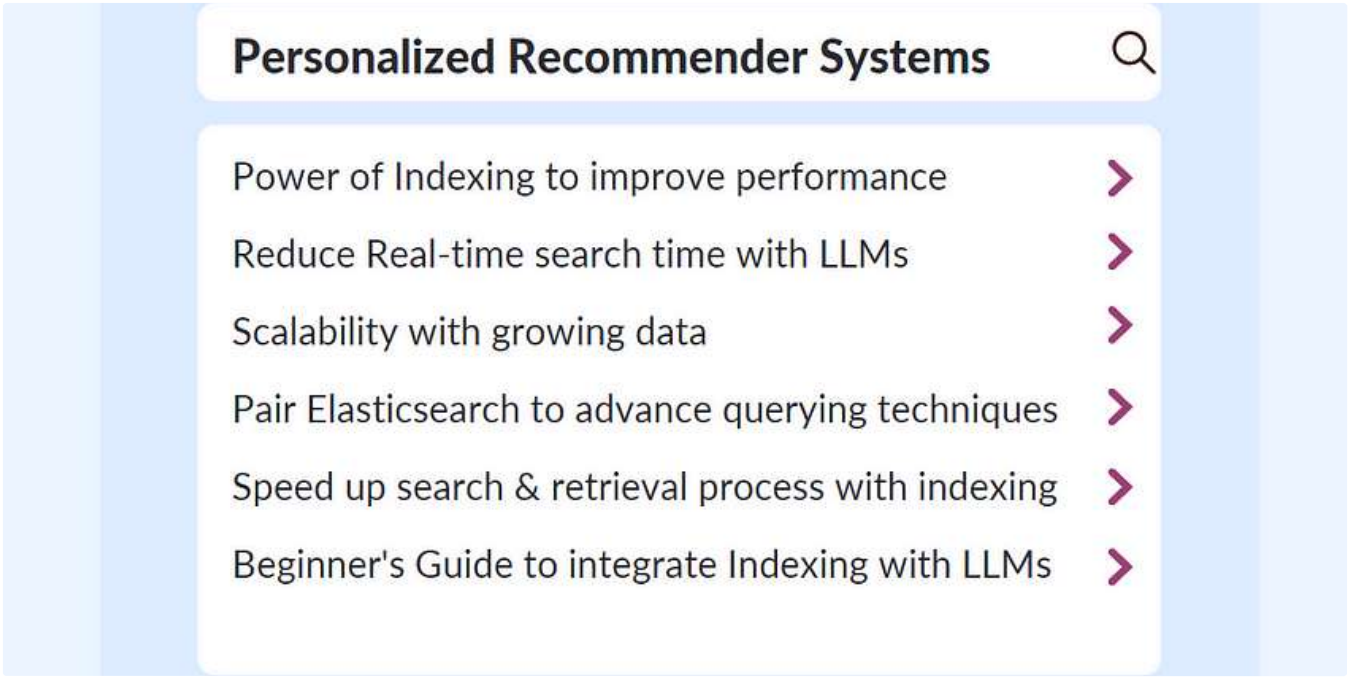
Sawera Khadium

Is Hosting Your Own LLM Cheaper than OpenAI?

COMMONLY ASKED QUESTION BY STARTUPS

3 min read · Feb 5, 2024

55



Sawera Khadium

Why Indexing is Useful for Personalized Recommender Systems

Introduction

6 min read · Feb 5, 2024

279

Sawera Khadium

Google BigQuery

Why Data Scientists Should Add Google BigQuery to Their Skillset

3 min read · Feb 5, 2024

101

Sawera Khadium

Dive into Powerful Machine Learning Frameworks

PyTorch, TensorFlow, Jax, Theano

4 min read · Feb 5, 2024

50

See all from Sawera Khadium

Recommended from Medium

Hakan Ateşli in Hepsiburada Data Science and Analytics

Explainable AI With SHAP

From Complexity to Clarity: Exploring AI Transparency with SHAP

13 min read · Dec 22, 2023

233

Shubham Bhandari in Towards AI

Explainable AI: GradCAM using PyTorch

The articles sheds light on Explainable AI. It explains and illustrates GradCAM method using PyTorch on ResNet model.

6 min read · Feb 20, 2024

57 1

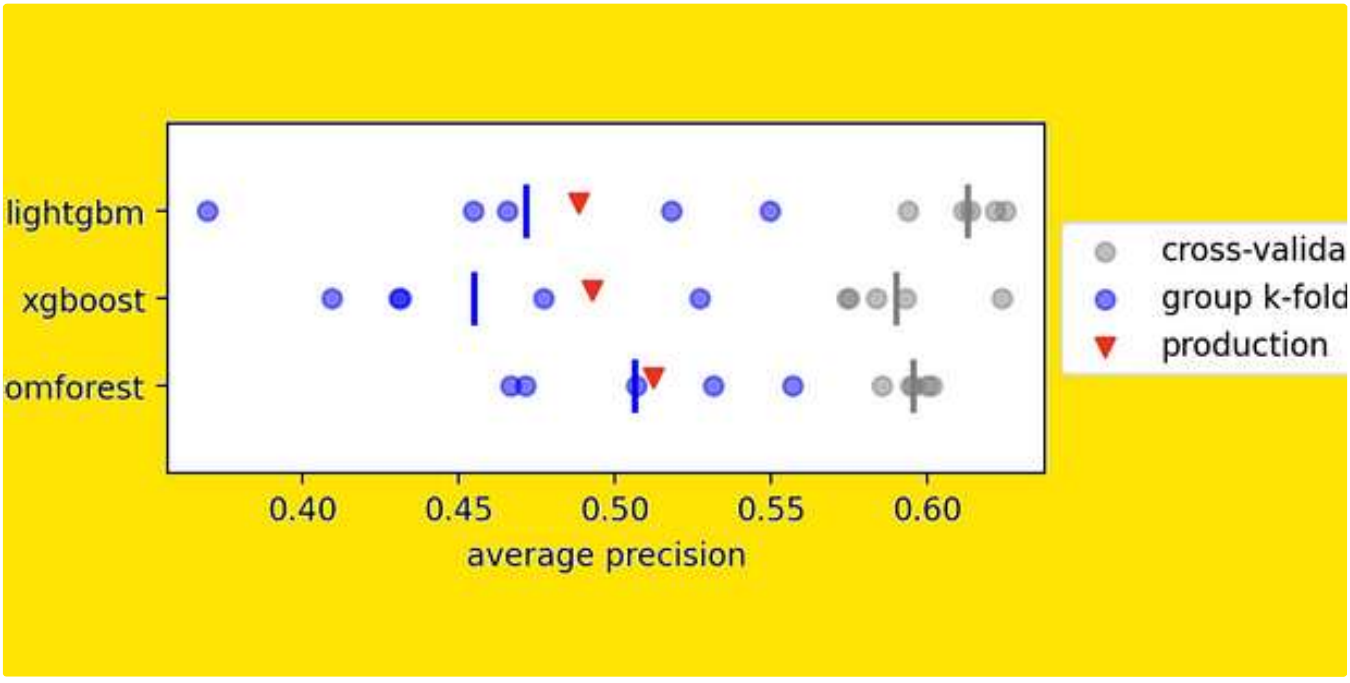
Lists

Generative AI Recommended Reading
52 stories · 891 saves

What is ChatGPT?
9 stories · 332 saves

The New Chatbots: ChatGPT, Bard, and Beyond
12 stories · 350 saves

Natural Language Processing
1335 stories · 821 saves



Samuele Mazzanti in Towards Data Science

Why You Should Never Use Cross-Validation

In real-world applications, using randomized cross-validation is always a bad choice. Here is why.

· 12 min read · 4 days ago

611 12

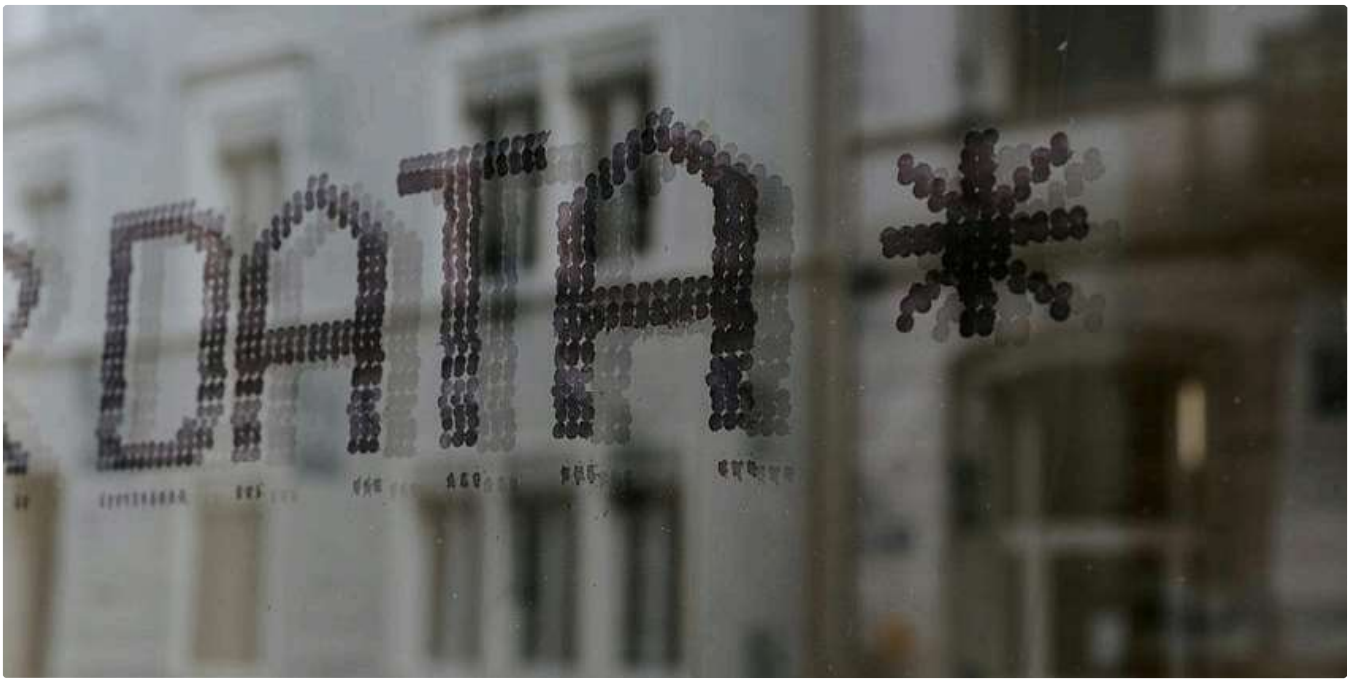
btd

Explainable AI (XAI) Tools and Libraries

Explainable AI (XAI) tools and libraries are essential components for developing, evaluating, and deploying machine learning models with...

· 3 min read · Nov 23, 2023

1 2



Gawain Chin - DS Blog

Correlation vs SHAP: Understanding Feature Importance in ML Models

Both correlation and SHAP values tell you something about the relationship between variables. but they do so in very different ways...

3 min read · Feb 18, 2024

103

Sanjana Tule in MatrixnTensors

Multimodal Visual Instruction Tuning with Phi-2 & CLIP using pytorch lightning — part 1

Introduction

5 min read · Feb 5, 2024

60 1