

ABHISHEK DAS

Applied Scientist | ML Researcher | Innovation Driver

📞 (412) 708-4892 ✉️ abhishekdas.0697@gmail.com 🔗 <https://abhishek0697.github.io/> 📍 Redmond, Washington

SUMMARY

Abhishek is an Applied Scientist specializing in large-scale machine learning systems, with expertise in LLMs, transformer optimization, and web-scale inference. Passionate about solving complex problems, he has a proven track record of developing innovative solutions in Generative & Responsible AI at Microsoft and Carnegie Mellon.

EXPERIENCE

Senior Applied Scientist

Microsoft AI

📅 08/2021 - Present 📍 Redmond, United States

- Designed and productionized ultra-low-latency vision-language model (VLM) powering **multimodal autosuggest** in Bing Visual Search, achieving **sub-100 ms** p95 latency via layer pruning and inference optimizations.
- Modernized the Related Searches recommendation stack across SERP and Image verticals with deep-learning-based **recall** and **ranking**, enhancing traffic shaping and faster experimentation.
- Conducted research on prompt engineering best-practices, and synthetic data distillation using LLMs from **GPT-4** family. Fine-tuned and deployed customized student transformer models for web-scale inference, efficiently processing several **billion queries** on a regular basis.
- Collaborated with stakeholders to identify potential misuse of generative images, and established high-quality measurement sets and **fine-grained guidelines** for sensitive categories, including religious figures, drugs, and other critical topics.
- Developed the **first prompt safety** neural network model for Bing Image Creator, significantly reducing over-blocking and leakage to prevent the generation of offensive, misinformative, or harmful AI-generated images, while maintaining user experience and creative freedom.
- Generated actionable product insights from user sessions to **identify user frustrations** and devised solutions to improve search experience. Conducted opportunity analysis for ads revenue growth and monetization.
- Progressed through **three** promotions, taking on greater tech-lead responsibilities, mentoring junior researchers, peer-reviewing internal research, rapid prototyping and shipping impactful AI solutions.

Deep Learning Researcher

Carnegie Mellon University

📅 05/2020 - 07/2021 📍 Pittsburgh, PA

- Built a fine-grained retail product image classifier with large, noisy datasets and long-tail distributions. Explored and refined advanced deep learning techniques from the literature such as CutMix, Label Smoothing, attention mechanisms, and sampling strategies to optimize state-of-the-art neural network architectures on production datasets.

EDUCATION

Master of Science in Electrical and Computer Engineering

Carnegie Mellon University

📅 08/2019 - 12/2020

- Relevant Coursework: Multimodal Machine Learning, Computer Vision

Bachelor of Engineering in Electronics and Telecommunication Engineering

University of Mumbai

📅 08/2015 - 12/2019

- Relevant Coursework: Discrete-Time Signal Processing, Image Processing

SKILLS

Prompt Engineering, Fine-Tuning & Preference Optimization for LLMs

Languages: Python, SCOPE, C#, MATLAB

Tools: PyTorch, Hugging Face, vLLM, NumPy, Linux, ONNX, Streamlit

KEY ACHIEVEMENTS



Improved Agility for Related Searches Recommendations Stack

Boosted user clicks by over **10%** and coverage by up to **50%** by refactoring and automating the pipeline enabling weekly dataset refreshes.



Enhanced AI Content Moderation

Delivered a prompt safety neural network model for DALL-E3, enabling safe launch of Bing Image Creator with **~2M** Daily Active Users.



Microsoft AI Peer Recognition Awardee

Received multiple accolades in the Microsoft AI Peer Recognition Program for **Quality, Empowerment, and Agility**.



Featured in Maharashtra Times, a leading daily newspaper in India

For innovation and creativity demonstrated in the **IoT-based Smart Refrigerator** project, showcased at the DJ Spark college fest. The solution simplifies household tasks with features such as real-time inventory tracking, mobile restocking alerts, and personalized recipe suggestions. ([link](#))

PATENTS

System and method for identifying products in a shelf management system.

U.S. Patent Application Number 17506115

Das A., Savvides M., et al. Patent Pending. Oct 2021

🔗 <https://patents.google.com/patent/US20220051179A1/>

- Designed and implemented an EfficientNet-based deep neural network classifier for n-way product categorization, enabling deployment in mobile robots equipped with camera sensors for automated inventory monitoring in commercial retail environments such as Walmart.

PUBLICATIONS

Detecting Hate Speech in Multi-modal Memes

🔗 <https://arxiv.org/abs/2012.14891>

- Devised Object detection-based Image Captioning to tackle the adversarial Benign Confounders in the challenge dataset and integrated it with a multimodal architecture comprising BERT and ResNeXt101.
- Achieved **+2%** AUROC over the best multimodal baseline Visual BERT COCO.
- Presented our research ideas in the **NeurIPS 2020** Facebook Hateful Memes Challenge session - Contributed Talks.

Multi-Image Steganography Using Deep Neural Networks

🔗 <https://arxiv.org/abs/2101.00350>

- Enhanced a Convolutional Neural Network (CNN) Encoder-Decoder architecture to embed **three** secret images within a carrier image by using feature concatenation and noise addition.