

SHIVAM CHAUHAN

Data & Applied Scientist II

✉ oshivam33@gmail.com | 📞 +91 8979344068

EDUCATION

- 2020 **Indian Institute of Science, Bengaluru** (*Masters in Computer Science, Machine Learning*)
2017 **College of Technology, Pantnagar** (*B.Tech in Electronics and Communication Engineering*)

SUMMARY

Data and Applied Scientist with 3 years of experience in the Natural Language Processing (NLP) domain. As an integral part of the machine learning algorithms team, contributed significantly to the global ads recommendation system at Microsoft.

EXPERIENCE

MICROSOFT

Data and Applied Scientist II at Microsoft Bengaluru, India

JULY'20 - AUG'23

Team: MICROSOFT ADVERTISING (BING ADS)

- **Multilingual large language models (LLM) into BingAds Search system:**
 - Trained and integrated **LLM (TwinBERT, Uni-Retriever, etc.)** model into the BingAds search system, transitioning from static dictionaries to dynamically retrieve TextAds for user queries.
 - Introduced a clustering approach to enhance the density of ads embeddings (training transformer-based models from scratch on search ads data), resulting in a minimum 20% increase in the proportion of ads served by LLM models as opposed to dictionaries.
 - Enhanced online search system capacity and automating processes to seamlessly incorporate new language models into the online architecture.
- Worked with a team of three **in ownership of recommendation systems, quality models, and system capacity** for emerging models. Spearheaded the planning and implementation of new features across these domains.
- **Ensemble of Models:** Demonstrated independent expertise in training and integrating ensemble of DNN models (Uni-Retriever: Fine-Tuning encoder and decoder again to simulate search Ads data) into the BingAds search system, effectively screening out undesirable ads produced.
 - Collaborated with Microsoft Research to leverage various Ad attributes, such as quick back rate and click-through rate, with the aim of enhancing Ads Quality and driving improved conversion rates.
 - Achieved a noteworthy reduction of at least 30% in the Bad Match Ratio (BMR) for key revenue markets for BingAds, concurrently expediting the onboarding process of new markets into the BingAds platform.
- Enhanced the training infrastructure, pipelines, and data storage by transitioning from manually managed offline machines to Azure directory, facilitating swift and efficient utilization by DNN models.
- Collaborated on the **EMEA and APAC clusters within the BingAds search system**, and played a key role in ensuring the smooth integration of new markets, including Japan and Middle Eastern countries.

MICROSOFT

Data and Applied Scientist Intern at Microsoft Bengaluru, India

JUNE'19 - AUG'19

Team: MICROSOFT ADVERTISING (BING ADS)

- Product Specification Extraction **SPECTRE Self-Supervision Model**
 - Created an end-to-end NLP system for automatic product specification extraction in Product Ads, enhancing the user search experience. Mined product attributes from e-commerce web pages and matched them with user queries to better understand their intent.

AWARDS & ACHIEVEMENTS

- 2022 Runner-up in **Microsoft FY2022-Q3** for the **Poster Presentation** on "Clustering Ads Embedding to Improve Density with Controlled BMR."
- 2021 **Microsoft FY2021-Q2 Excellence in Innovation Award** for the "Introduction of retriever models to the BingAds search system".
- 2018 Secured **All India Rank - 14** in **GATE Computer Science** among 1,00,013 Candidates.

TECHNICAL SKILLS

Programming Languages	Python, C, C++
Deep Learning Libraries	PyTorch, Keras
Others	Search and Recommendation Systems, Microsoft Azure, NumPy, Pandas, Scikit-learn, SQL, \LaTeX , Git, A/B Testing