Deep Karkhanis

dkarkhan@andrew.cmu.edu ▪ +1 (412)-214-2212 ▪ linkedin.com/in/deep-karkhanis

Education

Carnegie Mellon University — **Pittsburgh, PA**
MS in Machine Learning ○ *GPA : **4.17**/4.00* — *Aug '21 - Dec '22*
- Courses: Advanced **Natural Language Processing**, Advanced Deep Learning, Machine Learning PhD, Multimodal ML

Indian Institute of Technology, Bombay (IIT Bombay) — **Mumbai, India**
B.Tech. with Honors in Computer Science and Engineering [Thesis Link] — *July '16 - May '20*
- Core GPA : **9.58**/10.0 ○ Honors GPA : **10.0**/10.0
- Minor in **Applied Statistics** and Informatics
- **Teaching Assistant** in Computer Networks, Linear Algebra, Differential Equations
- Key Courses: **Advanced Machine Learning**, Advanced Statistics, Derivative Pricing, Image Processing, Remote Sensing

Experience

Tower Research Capital LLC — **New York City, NY**
Quantitative Trading Intern | Limestone Team — *Jun '22 - Aug '22*
- Developed Machine Learning models to get **subsets of trades** which give superior signals as proven by out-of-sample testing
- Analyzed market data and microstructures to design better **alphas** & features which lead to superior **high-frequency trading**

Microsoft Research Lab — **Bangalore, India**
Microsoft Research Fellow — *Aug '20 - Aug '21*
- Upgraded Microsoft's DiskANN indexing algorithm to add semantic search in **MS Exchange** (**pending Patent** Application)
- The algorithm can host all Outlook/Exchange emails and is planned to be **shipped worldwide**. Currently deployed internally
- Improved the graph kNN algorithm to require **90% lesser** memory and achieve **>95%** recall with **half** the computational costs
- Added low-RAM (<20MB) support and supported constantly changing **large scale** datasets with **>1M** points, **>100** dimensions
- Developed prototype (kNN based) for a new **Bing Ads system** serving **all European Markets** in **all supported languages**
- Designed the **first graph ANN algorithm** which supports and optimizes for filter-based search directly in the index build phase

Co-Founder and CTO | Kwikpic.in (StartUp) — **Mumbai, India | May '20 - May '21**
- Adapted leading face recognition algorithms to work on **Indian faces** and built web-system to deliver guests their event photos
- Optimized algorithm to handle varied lighting conditions to achieve accuracy of **>98%** for indoor & **>95%** for night events

RWTH Aachen University [Link] — **Aachen, Germany**
Scientific Research Intern | Prof. Joost-Pieter Katoen — *May '19 - July '19*
- Combined concepts of Stochastic Model Checking & **Counting-SAT** to compute Bounded Reachability Probabilities in MDPs
- Designed a **10x faster** solver by formulating a succinct CNF encoding for Markov Chains using the transition probabilities *BDD*

Institute of Science and Technology (IST) Austria — **Vienna, Austria**
Scientific Research Intern | Published in ICAPS 2020 [1] — *May '18 - July '18*
- Created a Monte-Carlo planning based MEMDP solver which has **50x** speed-up and **20x** accuracy over POMDP solvers
- Exploited the sparse transitions in Multiple Environment MDPs to get faster belief updates (**O(n)** as opposed to **O(n^2)**)

Publication

- [1]Krishnendu Chatterjee, Martin Chmelík, **Deep Karkhanis**, Petr Novotný and Amélie Royer, *"Multiple-Environment Markov Decision Processes: Efficient Analysis and Applications"*, **ICAPS 2020**, Nancy, France [**AAAI** Link]

Awards and Scholastic Achievements

- Secured the **Undergraduate Research Award**, IIT Bombay for distinguished research in POMDP solvers (2019)
- Awarded the **DAAD scholarship** by the *German Federal Ministry* for pursuing advanced research in Germany (2019)
- Received the **Institute Academic Excellence Award** from the Dean for securing **Institute Rank 1** (2017)

Academic Research

Kernel Approach for Attention Sparsity in Transformers [Link] — *Carnegie Mellon University | Jan '22 - Ongoing*
- Reducing the **O(n^2)** complexity of Transformer Attention Calculation by encoding **attention masks** as union of intrinsic kernels
- Developed algorithms to generate both: fixed kernels and kernels which can be **trained directly** during transformer training.

Multimodal-Multihop Question Answering [Link] — *Carnegie Mellon University | Jan '22 - Ongoing*
- Designed the **first ever multimodal-multihop retriever** & achieved state-of-the-art corpus level retrieval accuracy on WebQA
- Used novel **visual-grounding** techniques (**X-VLM**, **CLIP**) to integrate the question in the visual embeddings of the QA reader

Code Generation from Natural Language [Link] — *Carnegie Mellon University | Aug '21 - Dec '21*
- Appended python library documentation while training the model to **beat state of the art** BLEU scores on the CoNaLa dataset
- Improved **functional correctness** of generated code by enhancing performance on the HumanEval dataset

Software & Programming Skills

C++/C, Python, MATLAB, SQL, Web Development, Java, Scheme, Android Studio, VHDL, Prolog, Clingo, Solidworks, LaTeX