

# Lalchand Pandia

Github: [lalchandpandia](#)  
website: [home-page](#)  
[lcandia@gmail.com](mailto:lcandia@gmail.com)

---

**Overview** I am a PhD student in Computer Science at the University of Chicago, specializing in Natural Language Processing (NLP), Large Language Models (LLMs), and data-centric AI. My research focuses on data selection and model robustness for efficient LLM training. Before my PhD, I spent over 5 years in industry developing large-scale distributed file systems as a Systems Architect and Engineer. I have extensive experience training and fine-tuning LLMs on multi-GPU distributed systems using PyTorch, Accelerate, and DeepSpeed. My work has been published in EMNLP and CoNLL. I am seeking a Summer 2026 Research Scientist or Applied Scientist internship.

**Publications** My research explores data efficiency for large language models.

## Ongoing Work

**Rethinking Data Selection: The Importance of Coverage over Difficulty in Generative Fine-Tuning** Under Review at ICLR 2026

## Conference Papers

**Pragmatic competence of pre-trained language models through the lens of discourse connectives** *Accepted at CoNLL 2021*  
Lalchand Pandia, Yan Cong, Allyson Ettinger

**Sorting through the noise: Testing robustness of information processing in pre-trained language models** *Accepted at EMNLP 2021*  
Lalchand Pandia, Allyson Ettinger

**Education** **University of Chicago**  
Ph.D., Computer Science *Sept 2022 - present*  
**Advisor:** Dr. Allyson Ettinger

**Indian Institute of Technology Kanpur**  
M.Tech. Computer Science and Engineering (CSE) *2015-2017*  
**Advisor:** Dr. Harish Karnick

**Kalasalingam University, Krishnankoil**  
B.Tech.(Honours) Computer Science and Engineering (CSE) *2008-2012*

**Industry experience** **Huawei Technologies India Pvt. Ltd**  
System Architect, *Dec 2018- July 2022*

- Designed and implemented a logical replication module for Huawei FileSystem, enabling seamless replication with Amazon S3, reducing manual backup.
- Developed distributed data manager for cross-cluster NAS replication, improving I/O performance by optimizing data transfer paths and reducing latency

**NetApp India Pvt. Ltd**

Member of Technical Staff II,

*July 2017- July 2018*

Focused on backup of data from NetApp storage servers to backup devices such as tape using NDMP.

**Tata Consultancy Services Pvt. Ltd**

System Engineer,

*Dec 2012- June 2015*

Worked on Java client-server applications dealing with insurance policies.

**Achievements  
and Awards**

- Ranked in top 0.3% in Graduate Aptitude Test in Engineering(GATE), 2015, Computer Science.
- Secured 2nd rank in Computer Science and Engineering Department in Bachelor of Technology, 2012.

**Skills**

- **Programming Languages:** Python, C, C++, Java
- **ML & NLP Frameworks:** PyTorch, Accelerate, Transformers (Hugging Face), NumPy, Scikit-learn
- **Distributed Multi-GPU Training:** PyTorch Distributed, Accelerate, DeepSpeed, multi-GPU training and optimization