

Gaurav Maheshwari

Lille, France | mail@gauravmaheshwari.page | gauravm.gitbook.io | github.com/saist1993

ML engineer and researcher, with over 5 years of experience in developing impactful solutions in both academic and industrial settings.

Education

INRIA

Ph.D candidate in applied Machine learning

Lille, France

Expected December 2023

University of Bonn

M.Sc. in Computer Science with focus on ML – GPA: 1.4 (1.0 Max)

Bonn, Germany

May 2019

Dhirubhai Ambani Institute of Information and Communication Technology

B.Tech. in Information and Communication Technology – GPA: 8.0 (10.0 Max)

Gandhinagar, India

May 2016

Experience

Ph.D. Candidate, INRIA - Lille, France

November 2020 – Present

- Developed and validated FairGrad, a scalable gradient-based reweighting algorithm, enhancing fairness by up to 30% across diverse datasets in various ML models, including Logistic Regression, CNNs, ResNet, and BERT; published in [TMLR, 2023](#).
- Packaged and released [FairGrad via PyPI](#) in PyTorch, ensuring easy integration into diverse ML pipelines.
- Designed and implemented an adversarial learning mechanism leveraging differential privacy to privatize output of LLMs, successfully obfuscating sensitive attributes in text, and improving privacy by over 15% points; published in [Findings of EMNLP 2022](#).
- Identified and addressed limitations in existing intersectional fairness evaluation metrics. Introduced and validated, a new generalized metric enabling robust ML model assessments; accepted at [EMNLP 2023](#).
- Proposed a novel data generation mechanism improving performance by 60% in imbalanced dataset settings across multiple datasets..

Dialogue Engineer, Fraunhofer IAIS - Dresden, Germany

June 2019 – October 2020

- Orchestrated the development of dialogue systems to query documents, coordinating various stakeholders, resulting in a widely-adopted platform utilized across multiple internal projects.
- Implemented POC of a multilingual question-answering system based on LLMs, managing end-to-end responsibilities from exploratory data analysis to deployment and monitoring including CI/CD pipeline, utilizing tools like pandas, Docker, and Gradio.
- Streamlined and improved various NLP components, such as preprocessing module, entity linking, and sentiment analysis thereby boosting interoperability and accuracy by over 25%.

Research Assistant, Fraunhofer IAIS - Bonn, Germany

September 2017 – May 2019

- Surveyed and benchmarked different reading comprehension systems for semantic search enhancing selection and integration strategy.
- Co-authored a detailed [review on neural network-based Knowledge Graph Question Answering \(KGQA\)](#) methods and led tutorials on the subject at various forums, sharing expertise and facilitating learning.

Research Assistant, Smart Data Analytics - Bonn, Germany

September 2016 – August 2017

- Bootstrapped a [large-scale Text to SPARQL dataset](#) using Amazon Mechanical Turk, culminating in a 10x larger dataset.
- Refined Transformer architecture to exploit the SPARQL structure, boosting KGQA system F1 score by 8% (published at [ISWC](#)).

CoFounder, Rygbee - Gandhinagar, India

October 2014 – May 2016

- Formulated a TF-IDF and topic modeling based [approach to document similarity](#) for research paper retrieval, and deployed it at Google Compute Engine, using a combination of custom Elasticsearch plugins, and OrientDB.
- Actively participated in product design, and deployment over GCE, and pitched the product to multiple seed investors.

Skills

Tools: Python, PyTorch, NumPy, SpaCy, WandB, NLTK, Transformers (HuggingFace), LangChain, SQL, pandas, TensorFlow, Git.

Concepts: Deep learning, NLP, Fairness, Differential Privacy, Bias, Transfer learning, Information Extraction, Graph Neural Networks, Semantic Parsing, Domain Adaptation, Language Models, Visualizations, Transformers, Knowledge Graphs.

Additional

- Co-authored and presented over **15 research papers** on fairness, privacy, graph embedding, and KGQA at top-tier ML/NLP venues. Publications collectively cited over 600 times.
- Awarded Top 100 Most Influential Scholar in the field of Knowledge Engineering by Aminer.org

Selected Publications

- **Gaurav Maheshwari**, and Michael Perrot. "FairGrad: Fairness Aware Gradient Descent". *Transactions on Machine Learning Research* (2023).
- **Gaurav Maheshwari**, Aurélien Bellet, Pascal Denis, and Mikaela Keller. "Fair Without Leveling Down: A New Intersectional Fairness Definition" In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.
- **Gaurav Maheshwari**, Pascal Denis, Mikaela Keller, and Aurélien Bellet. "Fair NLP Models with Differentially Private Text Encoders." In *Findings of the Association for Computational Linguistics: Empirical Methods in Natural Language Processing 2022*.
- **Gaurav Maheshwari**, Priyansh Trivedi, Denis Lukovnikov, Nilesh Chakraborty, Asja Fischer, and Jens Lehmann. "Learning to rank query graphs for complex question answering over knowledge graphs." In *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18*, pp. 487-504. Springer International Publishing, 2019.
- Galkin, Mikhail, Priyansh Trivedi, **Gaurav Maheshwari**, Ricardo Usbeck, and Jens Lehmann. "Message Passing for Hyper-Relational Knowledge Graphs." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.
- Priyansh Trivedi, **Gaurav Maheshwari**, Mohnish Dubey, and Jens Lehmann. "LC-QuAD: A Corpus for Complex Question Answering over Knowledge Graphs." In *International Semantic Web Conference*, 2017.
- **Gaurav Maheshwari**, Priyansh Trivedi, Harshita Sahijwani, Kunal Jha, Sourish Dasgupta, and Jens Lehmann. "Simdoc: topic sequence alignment based document similarity framework." In *Proceedings of the Knowledge Capture Conference*, 2017.

Selected Courses

Master: Natural Language Processing (Seminar and Lab), Deep Learning For Visual Recognition, Probabilistic Graphical Models, Intelligent Learning and Analysis System: Machine Learning, Cognitive Robotics, Intelligent Information Systems

Bachelor: Calculus and Complex Variables, Object Oriented Programming, Introduction to Discrete Mathematics, Probability and Statistics, Operating Systems, Database Management System, Optimization, Computer Networks, Data Structures.