ADITHYA BHASKAR

Princeton University, NJ, USA | Computer Science

+1 (609) 721-0839 github.com/testzer0
adithyabcse@gmail.com ab4197@princeton.edu
955 Alexander Road FL-2, Princeton Junction, NJ 08550
adithyabh.github.io

Education

2023-*Ongoing*	Ph.D. in Computer Science, Princeton University, USA *Advised by Prof. Danqi Chen* (Area of focus : Natural Language Processing)	—
2019-23	Bachelor of Technology in Computer Science and Engineering (Honors), IIT Bombay, India	9.67/10.00
2017-19	High School, Central Board of Secondary Education, India	97.2/100.0
2017	Senior Secondary School, Central Board of Secondary Education, India	98.0/100.0

Scholastic Achievements

- 2023 Was awarded the **Thomas Dooie Class of 1974 Research Award** for my Bachelor's Thesis at IIT Bombay.
- 2020 Was awarded an **AP** grade for exemplary performance in the course **Data Structures and Algorithms**.
- 2019 Secured **All India Rank 18** in JEE Advanced 2019 out of **240 thousand** candidates.
- 2019 Achieved **All India Rank 114** in JEE Mains 2019 out of **1.1 million** candidates.
- 2018 Secured a position in the **top 39 ranks** in the **Indian National Physics Olympiad** and was **invited to the Orientation-cum-Selection-Camp in Physics** held in May-June 2018.
- 2018 Secured a position in the **top 49 ranks** in the **Indian National Chemistry Olympiad** and was **invited to the Orientation-cum-Selection-Camp in Chemistry** held in May-June 2018.
- 2016 Among the **39 students** to clear the **Indian National Mathematical Olympiad**, becoming **one of the youngest** to **ever** be invited to the Orientation-cum-Selection-Camp in Mathematics **aged 14**.

Publications

- 2023 Benchmarking and Improving Text-to-SQL Generation under Ambiguity, *EMNLP 2023 (Main)*
  Adithya Bhaskar*, Tushar Tomar*, Ashutosh Sathe and Sunita Sarawagi
- 2023 Prompted Opinion Summarization with GPT-3.5, *ACL 2023 (Findings)*
  Adithya Bhaskar, Alexander R. Fabbri, and Greg Durrett
- 2023 Performance Bounds for LASSO under Multiplicative Noise : Applications to Pooled RT-PCR Testing, *Signal Processing, Vol. 214*
  Richeek Das, Aaron Jerry Ninan, **Adithya Bhaskar**, and Ajit Rajwade

Internships

UT Austin, Summer 2022 — **Research Intern, Natural Language Processing, USA**
Guide : Prof. Greg Durrett
Very Large Language Models for Multi-Document Summarization

- Developed metrics to measure **factuality**, **faithfulness** and **specificity** (whether it is correct, prefers major viewpoints, and is not too generic) for a summary of multi-document text such as hotel reviews.
- Utilized the above along with an n-gram abstractiveness metric to benchmark **GPT-3.5**, and showed that simple hierarchical summarization of large text performs poorly on faithfulness and specificity.
- Investigated various **pre-clustering** and **pre-summarization** methods and illustrated that pre-summarization with a pretrained keyword-based extractive model improves correctness, faithfulness and specificity, while only marginally affecting abstractiveness.

Uppsala University, Summer 2021 — **Research Intern, Formal Verification,**
Guide : Prof. Parosh Abdulla
Model Checking for Programs Running under the ARMv8 Memory Model

- Developed a **model** and **simulator** for programs running under the ARMv8 memory model, and demonstrated the equivalence of the model to the ARM specification using **7500+** litmus tests.
- Used **Context Bounded Model Checking** to perform **State Reachability Analysis** for programs under the ARMv8 memory model, achieving **up to an order of magnitude** of speedup over existing checkers.

Key Projects

Robust Models
Spring 2023

Bachelor's Project, Natural Language Processing, Guide : Prof. Sunita Sarawagi
Automated data augmentation for robustness.

- Demonstrated that training a Text-to-SQL model on partially masked (underspecified) inputs leads to **diversity** in the model outputs, including in **columns/tables, string literals, integers**, and **aggregates**.
- Filtered the outputs by model probabilities relative to the output with the unmasked question.
- Furnished questions for the generated queries via an SQL-to-Text model. Data augmentation with the pairs led to increases in accuracy on the SPIDER dataset and the robustness benchmark, Dr. SPIDER.
- The increases exceeded those obtained by augmenting with Dr. SPIDER style perturbations.

Text-to-SQL
Fall 2022

Bachelor's Project, Natural Language Processing, Guide : Prof. Sunita Sarawagi
The "did you mean?" functionality in Text-to-SQL conversion.
Objective : Presenting alternatives diverse in meaning but structurally similar to the user's question.

- Illustrated the inadequacy of plain Beam Search to elicit diversity from Text-to-SQL models.
- Trained an **alignment model** to align words of the prompt with the SQL in an **online** fashion, which were used to drive a **split beam search** on the Text-to-SQL model, extracting meaningful alternatives.
- Trained a **faithfulness classifier** using **contrastive learning** to identify and mask spans of the text incompatible with the SQL alternatives, which were filled in by a **masked reconstruction model**.
- Demonstrated the utility of the approach on the **SPIDER** and **WikiSQL** datasets.

Group Testing
Fall 2022

R&D Project, Compressive Sensing, Guide : Prof. Ajit Rajwade
Applying compressive sensing to improve COVID-19 Group Testing.

- Proved the theoretical applicability of Compressive Sensing with Weighted LASSO for **any general noise model** with an asymptotically well-defined Moment-Generating-Function.
- Derived values of the optimal weights for the case of **Multiplicative Gaussian Noise** as in RT-PCR tests.
- Demonstrated improvements in **sensitivity**, **specificity**, **MCC** and **RMSE** by **Monte Carlo Simulations**.

Text-to-Speech
Summer 2022

Passion Project, NLP & ASR, Reference : Grad-TTS (ICML 2021)
Text-to-Speech Conversion Using Guassian Diffusion Models

- Wrote a transformer with **learned and windowed self-attention layers** along with **feedforward layers** to encode input text, and a simple CNN to expand each frame to a predicted duration.
- Used **Probabilistic Diffusion Models** to gradually enhance pure noise into target audio given input text, by solving a **Time-Reversed Stochaistic Differential Equation**.
- Utilized a **UNet-style network** to predict terms of the differential equation using the encoded text.

Reverse Dictionary
Spring 2022

Course Project, Natural Language Processing, Guide : Prof. Pushpak Bhattacharya
Identifying Words From Their Meanings

- Built a **BERT + Multi-BiLSTM** model to map a given statement to the word closest in meaning to it.
- Wrote a **crawler** to curate a **high-quality dataset** of **89930** word-meaning pairs from online sources.
- Utilized the distance between **GloVE** word vectors as a loss function to partially reward synonyms.
- Obtained a top-1 accuracy of **48.71%** and a top-100 accuracy of **65.62%**, exceeding the State of The Art by **5.4x** and **1.13x**, respectively.

Speech Denoising
Spring 2022

Course Project, Automatic Speech Recognition, Guide : Prof. Preethi Jyothi
Speech Denoising Using GANs [Reference : MetricGAN (ICML 2019)]

- Modified and utilizied the **Microsoft Noisy Speech Dataset** to generate **over 30 hours** of noisy speech and corresponding clean speech with noises from **over 100** sources.
- Utilized the Discriminator in the GAN framework to differentiably approximate a non-differentiable loss function such as Short-Term-Objective-Intelligibility, and the Generator to produce clean speech.
- Identified **novel failure modes** and demonstrated that **data augmentation** is an effective remedy.
- Demonstrated generalization to background music removal despite it not being in the training pool.

Sparse Recovery
Spring 2021

Course Project, Advanced Image Processing, Guide : Prof. Ajit Rajwade
Presented performance bounds on l_1-minimization based compressive sensing that are simultaneously

- easily **verifiable**,
- upto an **order of magnitude tighter** than bounds based on the Restricted Isometry Property, and
- **applicable to wider range of matrices** than both RIP and Mutual Coherence based bounds.

C Decompiler
Fall 2020

Course Project, Software Systems, Guide : Prof. Amitabha Sanyal
Recovering Code From Compiled RTL

- Built a **decompiler** to convert **Register Transfer Language** to **C** for portability across architectures.
- Utilized **lex** and **bison** to parse source code in RTL and identify program elements like **assignments**, **basic arithmetic operations**, **conditional/looping constructs**, **function calls** and **memory accesses**.
- Performed **local & global data flow analysis** and **control flow analysis** to contextualize parsed code.

Distributed DL
Summer 2020

Institute Technical Summer Project, Deep Learning & Distributed Systems,

- Developed a **Hierarchically Distributed Deep Convolutional Neural Network** in order to parallelize the workload across all computation nodes in a system.
- Trained the system on **Super High Resolution Datasets** via spatial segmentation of samples, using state-of-the-art nets such as **VGG16, ResNet, DenseNet,** and **LeNet** as the underlying Neural Network.
- Verified the approach by using the **CINIC-10** and **Retinal OCT** datasets.

</> Technical Skills

Programming	C, C++, Python, Golang, BASH, SQL, MATLAB, Java, Haskell
Web Dev	Javascript, Angular, React, HTML, CSS
Software & Tools	LaTeX, Git, FastAPI, Cypher, Spark

Certifications

2020	Deep Learning Specialization (Coursera) › 5 Courses By Prof. Andrew Ng
2019	Codechef Advanced Certification

Extra Curriculars

2022	Rated **2040** on Codeforces.
2022	A chess and badminton hobbyist.
2019	Selected for and enrolled in the **National Sports Organisation** program of **Football** at IIT Bombay.
2013-15	Studied **Sanskrit** for 3 years in school and secured a grade of **A1** in all three years.