

# ADITHYA BHASKAR

Princeton University, NJ, USA | Computer Science

☎ +91 9869195425\*      🌐 [github.com/testzer0](https://github.com/testzer0)  
@ [adithyabcse@gmail.com](mailto:adithyabcse@gmail.com)    @ [ab4197@princeton.edu](mailto:ab4197@princeton.edu)  
📍 2E Malayagiri, Anushaktinagar, Mumbai, India\*  
🏠 [adithyabh.github.io](https://adithyabh.github.io)  
\*Liable to change soon.

## 🎓 EDUCATION

—	Ph.D. in Computer Science, Princeton University, USA	—
	<i>Advised by Prof. Danqi Chen</i> (Area of focus : Natural Language Processing)	
2023	Bachelor of Technology in Computer Science and Engineering (Honors), IIT Bombay, India	9.67/10.00
2019	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

2020	Was awarded an <b>AP</b> grade for exemplary performance in the course <b>Data Structures and Algorithms</b> .
2019	Secured <b>All India Rank 18</b> in JEE Advanced 2019 out of <b>240 thousand</b> candidates.
2019	Achieved <b>All India Rank 114</b> in JEE Mains 2019 out of <b>1.1 million</b> candidates.
2018	Secured a position in the <b>top 39 ranks</b> in the <b>Indian National Physics Olympiad</b> and was <b>invited to the Orientation-cum-Selection-Camp in Physics</b> held in May-June 2018.
2018	Secured a position in the <b>top 49 ranks</b> in the <b>Indian National Chemistry Olympiad</b> and was <b>invited to the Orientation-cum-Selection-Camp in Chemistry</b> held in May-June 2018.
2016	Among the <b>39 students</b> to clear the <b>Indian National Mathematical Olympiad</b> , becoming <b>one of the youngest to ever</b> be invited to the Orientation-cum-Selection-Camp in Mathematics <b>aged 14</b> .

## 📄 PUBLICATIONS

2023	Benchmarking and Improving Text-to-SQL Generation under Ambiguity, <i>[REDACTED]</i> <i>Conference name redacted due to double-blind review</i> <b>Adithya Bhaskar</b> , Tushar Tomar, Ashutosh Sathe and Sunita Sarawagi
2023	Prompted Opinion Summarization with GPT-3.5, <i>ACL 2023 (Findings)</i> <b>Adithya Bhaskar</b> , Alexander R. Fabbri, and Greg Durrett
2023	Performance Bounds for LASSO under Multiplicative Noise : Applications to Pooled RT-PCR Testing, <i>arXiv preprint, arXiv:2305.17429</i> Richeek Das, Aaron Jerry Ninan, <b>Adithya Bhaskar</b> , and Ajit Rajwade

## 👛 INTERNSHIPS

UT Austin Summer 2022	<b>Research Intern, NATURAL LANGUAGE PROCESSING, USA</b> Guide : Prof. Greg Durrett Very Large Language Models for Multi-Document Summarization <ul style="list-style-type: none"><li>➤ Developed metrics to measure <b>factuality</b>, <b>faithfulness</b> and <b>specificity</b> (whether it is correct, prefers major viewpoints, and is not too generic) for a summary of multi-document text such as hotel reviews.</li><li>➤ Utilized the above along with an <math>n</math>-gram abstractiveness metric to benchmark <b>GPT-3.5</b>, and showed that simple hierarchical summarization of large text performs poorly on faithfulness and specificity.</li><li>➤ Investigated various <b>pre-clustering</b> and <b>pre-summarization</b> methods and illustrated that pre-summarization with a pretrained keyword-based extractive model improves correctness, faithfulness and specificity, while only marginally affecting abstractiveness.</li></ul>
Uppsala University Summer 2021	<b>Research Intern, FORMAL VERIFICATION,</b> Guide : Prof. Parosh Abdulla Model Checking for Programs Running under the ARMv8 Memory Model <ul style="list-style-type: none"><li>➤ Developed a <b>model</b> and <b>simulator</b> for programs running under the ARMv8 memory model, and demonstrated the equivalence of the model to the ARM specification using <b>7500+</b> litmus tests.</li><li>➤ Used <b>Context Bounded Model Checking</b> to perform <b>State Reachability Analysis</b> for programs under the ARMv8 memory model, achieving <b>up to an order of magnitude</b> of speedup over existing checkers.</li></ul>

Robust Models Spring 2023	<p><b>Bachelor's Project, NATURAL LANGUAGE PROCESSING, Guide : Prof. Sunita Sarawagi</b></p> <p>Automated data augmentation for robustness.</p> <ul style="list-style-type: none"> <li>&gt; Demonstrated that training a Text-to-SQL model on partially masked (underspecified) inputs leads to <b>diversity</b> in the model outputs, including in <b>columns/tables, string literals, integers, and aggregates</b>.</li> <li>&gt; Filtered the outputs by model probabilities relative to the output with the unmasked question.</li> <li>&gt; 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.</li> <li>&gt; The increases exceeded those obtained by augmenting with Dr. SPIDER style perturbations.</li> </ul>
Text-to-SQL Fall 2022	<p><b>Bachelor's Project, NATURAL LANGUAGE PROCESSING, Guide : Prof. Sunita Sarawagi</b></p> <p>The “did you mean?” functionality in Text-to-SQL conversion.</p> <p><i>Objective</i> : Presenting alternatives diverse in meaning but structurally similar to the user's question.</p> <ul style="list-style-type: none"> <li>&gt; Illustrated the inadequacy of plain Beam Search to elicit diversity from Text-to-SQL models.</li> <li>&gt; Trained an <b>alignment model</b> to align words of the prompt with the SQL in an <b>online</b> fashion, which were used to drive a <b>split beam search</b> on the Text-to-SQL model, extracting meaningful alternatives.</li> <li>&gt; Trained a <b>faithfulness classifier</b> using <b>contrastive learning</b> to identify and mask spans of the text incompatible with the SQL alternatives, which were filled in by a <b>masked reconstruction model</b>.</li> <li>&gt; Demonstrated the utility of the approach on the <b>SPIDER</b> and <b>WikiSQL</b> datasets.</li> </ul>
Group Testing Fall 2022	<p><b>R&amp;D Project, COMPRESSIVE SENSING, Guide : Prof. Ajit Rajwade</b></p> <p>Applying compressive sensing to improve COVID-19 Group Testing.</p> <ul style="list-style-type: none"> <li>&gt; Proved the theoretical applicability of Compressive Sensing with Weighted LASSO for <b>any general noise model</b> with an asymptotically well-defined Moment-Generating-Function.</li> <li>&gt; Derived values of the optimal weights for the case of <b>Multiplicative Gaussian Noise</b> as in RT-PCR tests.</li> <li>&gt; Demonstrated improvements in <b>sensitivity, specificity, MCC</b> and <b>RMSE</b> by <b>Monte Carlo Simulations</b>.</li> </ul>
Text-to-Speech Summer 2022	<p><b>Passion Project, NLP &amp; ASR, Reference : Grad-TTS (ICML 2021)</b></p> <p>Text-to-Speech Conversion Using Gaussian Diffusion Models</p> <ul style="list-style-type: none"> <li>&gt; Wrote a transformer with <b>learned and windowed self-attention layers</b> along with <b>feedforward layers</b> to encode input text, and a simple CNN to expand each frame to a predicted duration.</li> <li>&gt; Used <b>Probabilistic Diffusion Models</b> to gradually enhance pure noise into target audio given input text, by solving a <b>Time-Reversed Stochastic Differential Equation</b>.</li> <li>&gt; Utilized a <b>UNet-style network</b> to predict terms of the differential equation using the encoded text.</li> </ul>
Reverse Dictionary Spring 2022	<p><b>Course Project, NATURAL LANGUAGE PROCESSING, Guide : Prof. Pushpak Bhattacharya</b></p> <p>Identifying Words From Their Meanings</p> <ul style="list-style-type: none"> <li>&gt; Built a <b>BERT + Multi-BiLSTM</b> model to map a given statement to the word closest in meaning to it.</li> <li>&gt; Wrote a <b>crawler</b> to curate a <b>high-quality dataset</b> of <b>89930</b> word-meaning pairs from online sources.</li> <li>&gt; Utilized the distance between <b>GloVe</b> word vectors as a loss function to partially reward synonyms.</li> <li>&gt; Obtained a top-1 accuracy of <b>48.71%</b> and a top-100 accuracy of <b>65.62%</b>, exceeding the State of The Art by <b>5.4x</b> and <b>1.13x</b>, respectively.</li> </ul>
Speech Denoising Spring 2022	<p><b>Course Project, AUTOMATIC SPEECH RECOGNITION, Guide : Prof. Preethi Jyothi</b></p> <p>Speech Denoising Using GANs [Reference : MetricGAN (ICML 2019)]</p> <ul style="list-style-type: none"> <li>&gt; Modified and utilized the <b>Microsoft Noisy Speech Dataset</b> to generate <b>over 30 hours</b> of noisy speech and corresponding clean speech with noises from <b>over 100</b> sources.</li> <li>&gt; Utilized the Discriminator in the GAN framework to differentially approximate a non-differentiable loss function such as Short-Term-Objective-Intelligibility, and the Generator to produce clean speech.</li> <li>&gt; Identified <b>novel failure modes</b> and demonstrated that <b>data augmentation</b> is an effective remedy.</li> <li>&gt; Demonstrated generalization to background music removal despite it not being in the training pool.</li> </ul>
Sparse Recovery Spring 2021	<p><b>Course Project, ADVANCED IMAGE PROCESSING, Guide : Prof. Ajit Rajwade</b></p> <p>Presented performance bounds on <math>l_1</math>-minimization based compressive sensing that are simultaneously</p> <ul style="list-style-type: none"> <li>&gt; easily <b>verifiable</b>,</li> <li>&gt; upto an <b>order of magnitude tighter</b> than bounds based on the Restricted Isometry Property, and</li> <li>&gt; <b>applicable to wider range of matrices</b> than both RIP and Mutual Coherence based bounds.</li> </ul>

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	L <sup>A</sup> T <sub>E</sub> X, Git, FastAPI, Cypher, Spark

## ✔ CERTIFICATIONS

2020	Deep Learning Specialization (Coursera) <ul style="list-style-type: none"><li>&gt; 5 Courses By Prof. Andrew Ng</li></ul>
2019	Codechef Advanced Certification

## ⚽ EXTRA CURRICULARS

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