GUNIKA DHINGRA

LinkedIn: gunika-dhingra-678a201ba
Github: github.com/gunika-dhingra
Blog: 12gunika

Email: 12gunika@gmail.com
Portfolio :gunika.app
Twitter: gunikadhingra

Profile

Diligent student who demonstrates complex problem-solving skills, competent in a challenging environment. A creative thinker, proficient in data science, artificial intelligence and machine learning with experience in entrepreneurship. Committed to continually advancing the horizons of knowledge of relevant technologies.

Education

- **Arizona State University** — Arizona, United States
  MS Data Science, Analytics and Engineering — *2024-2026*
  CSE 511: Data Processing at Scale,
  CSE 575: Statistical Machine Learning,
  IEE 579: Time Series Analysis/Forecasting

- **Bennett University** — India
  Bachelor of Technology - Computer Science Engineering — *2020-2024*
  Specialization - Data Science
  Minor - Entrepreneurship
  CGPA - Major: 9.29/10.00, Minor: 9.75/10.00

- **St. Joseph's Academy** — India
  Senior Secondary School — *2018-2020*
  Subjects - Physics, Chemistry, Maths, Computer Science, English
  Percentage - Class 11: 94.1 % Class 12: 94%

- **Seth Anandram Jaipuria School** — India
  Senior School — *2005-2018*
  Subjects - Physics, Chemistry, Maths, Computer Science, English
  Class 9: 9.2/10.0 CGPA Class 10: 97%

Experience

- **AI/ML Research Intern** — Remote
  Georgia Institute of Technology — *Jan 2024 - July 2024*
  - Collaborated with Dr. Vijay K. Madisetti, Mr. Arshdeep Bahga, and a research team to publish research papers in peer-reviewed journal, focused on Large Language Models (LLMs) in the domains of utility, privacy and security.
  - Researched and developed advanced techniques to enhance privacy preservation in Large Language Models (LLMs), resulting in a 4-8% reduction in data leakage while maintaining high utility for targeted tasks.
  - Conducted in-depth analysis of privacy leakage in Large Language Models (LLMs) by implementing 6 advanced black-box attacks, revealing potential risks to sensitive information and personally identifiable information (PII).

- **Data Analyst-Trainee** — Remote
  BizInsights Inc. — *Jan 2024 - April 2024*
  - Collaborated on a critical data quality assurance project for a Fortune 500 pharmaceutical company, analyzing healthcare datasets from 3 warehouses with millions of rows, resulting in improved data integrity and reliability.
  - Streamlined data quality processes by identifying and eliminating non-essential checks, reducing system latency by 60% and greatly enhancing user experience for 50,000+ daily application users across multiple platforms.
  - Implemented and fine-tuned new essential data quality checks, directly improving reporting accuracy and data reliability for 8 downstream teams, strengthening overall data governance and supporting critical business decisions.

- **AI/ML Research Intern** — Atlanta, United States
  Georgia Institute of Technology — *Aug 2023 - Dec 2023*
  - Engaged in research activities and projects within the institute, contributing to the research initiatives.
  - Co-authored and published a research paper in SCIRP journal(Jan'2024), titled 'Smaller & Smarter: Score-Driven Network Chaining of Smaller Language Models'
  - Attended audit courses for select ongoing campus courses, enhancing academic knowledge and interdisciplinary perspectives.
  Courses attended: Big Data Systems and Advanced Applications, Human and Machine Learning, Cybersecutrity Practicum

Publications

- **Protecting LLMs against Privacy Attacks While Preserving Utility** - *Journal of Information Security, Vol.15 No.4, October 2024*
  Analyzed and addressed privacy concerns in large language models (LLMs) related to the leakage of Personally Identifiable Information (PII) during training or fine-tuning. Proposed and evaluated the use of Targeted Catastrophic Forgetting (TCF), an iterative pipeline to minimize the reproduction of sensitive information in black-box attack scenarios. Demonstrated that TCF effectively reduces PII leakage, such as in autocompletion attacks, while maintaining the context and utility of the target application.

  Link to the publication

- **Evaluating Privacy Leakage and Memorization Attacks on Large Language Models (LLMs) in Generative AI Applications**- *Journal of Software Engineering & Applications, Vol.17 No.5, May 2024*
  This paper addresses privacy concerns in Generative AI applications utilizing large language models (LLMs). It explores various black-box attacks by potential adversaries and examines the extent and nature of Personally Identifiable Information (PII) leakage and memorization. The study analyzes the impact of model size, architecture, and attack types, including PII leakage-focused and memorization-focused attacks, providing a comprehensive understanding of LLM vulnerabilities and attack effectiveness through detailed evaluative metrics. Link to the publication

- **Smaller & Smarter: Score-Driven Network Chaining of Smaller Language Models** - *Journal of Software Engineering & Applications, Vol.17 No.1, January 2024*
  Employed research activities in the domain of LLMs, finance and arithmetic. Addressed the shortcomings of existing models, while working with Meta's Llama2. Collaborated with Prof. Vijay Madisetti at Georgia Institute of Technology, independent researcher Arshdeep Bahga, and a fellow intern to conduct this study. The proposed network chaining not only streamlined processes by reducing costs by 55% but also enhanced model accuracy by 5-10%. *Link to the publication*

- **Employing deep learning models to financial news data and historical data for efficient prediction of stock prices**-*Cyber Tech Publications*
  This paper involved research on deep learning models for efficient stock price prediction using financial news and historical data. It addressed limitations of previous approaches and emphasized the importance of accurate prediction models in the volatile stock market. Additionally, it even explored market mood's impact on stock prices and aimed to contribute to more precise algorithms. Link to the publication

Projects

- **BELL**: *hereforyou.app*
  Developed Bell, a **women's health and mental well-being platform**, leveraging **OpenAI's ChatGPT and LangChain** framework as the core tech stack. The platform focuses on countering online misinformation, promoting authentic discussions, and ensuring user anonymity. Committed to empowering women's wellness, Bell provides a secure, community-driven space for open dialogue on health-related concerns.

- **StockFi**:
  Developed and executed a **deep learning** project to **predict stock prices** using historical market data, news sentiment, and technical indicators. The study focused on overcoming previous model limitations and demonstrating the impact of market mood on stock prices. Results contributed to the development of **more accurate prediction algorithms** and highlighted the potential for deep learning in financial forecasting.

- **HatchNet**:
  Developed a comprehensive **resource hub** for entrepreneurs, offering connections to incubators and real-time news updates. Engineered analytical tools for incubators to monitor the progress of associated startups and assess the potential of new applicants. Created a sophisticated dashboard for investors to evaluate potential investments, featuring streamlined data visualization. Implemented an **NLP** model to curate up-to-date news, ensuring entrepreneurs receive the latest industry insights.Key responsibilities even included **UI development.**

- **SCRIVI**:
  Designed and developed a **text analyzer** tool that generates detailed reports on written content. Utilizing **NLP** techniques such as stop-word removal, sentiment analysis, and paraphrasing, the tool provides insights into the text's structure and sentiment. The report includes key information such as the number of sentences, sentiment analysis, and a brief summary. The tool aims to improve the overall quality of written content by highlighting areas for improvement.

- **Ethereum Trend Analyzer and Price Forecastor**:
  Conducted a project that utilized **big data tools** and **machine learning techniques** to forecast the price **trend of Ethereum cryptocurrency**. The project involved analyzing Ethereum's historical data, including its value since its inception, to effectively predict its future price. Techniques such as ElasticNet Regression and time series analysis using Meta's Prophet Model were employed to analyze the data. The project aimed to provide insights into Ethereum's price movements and contribute to the development of accurate cryptocurrency forecasting models.

- **GlamFinder**:
  Designed and implemented a **beauty product Recommender System** that used **collaborative filtering** to recommend personalized products based on user preferences and past purchases. The project involved data collection, preprocessing, and model development to improve user experience and business sales..

- **Eclair**:
  Contributed to the development of Eclair, a **decentralized NFT marketplace** built on the CELO Blockchain. The platform enables artists and creators to sell and trade unique digital assets in a secure and transparent manner, while also rewarding users based on their activity score. The project involved extensive research on blockchain technology, NFTs, and their potential applications. Key responsibilities involved designing the **UI/UX, project management** and **research** in the domain.

- **CRYPTOGRAM**:
  Developed CryptoGram, **a social media app for NFTs** that fetched data from multiple marketplaces to generate an explore page for users. The app showcased NFTs similar to Instagram and provided necessary details about the collectables.

Skills

- **Languages**: Python, Structured Query Language(SQL), C++, Linux, Scala, Hive Query Language(HQL)
- **Frameworks/Libraries**: PostgreSQL, Scikit-Learn, NumPy, Pandas, Matplotlib, NLTK, Seaborn, Beautiful Soup, PySpark, TensorFlow, Keras, PyTorch, Transformers, OpenAI, Hadoop, SparkSQL
- **Data Science/ Machine Learning**: ML Techniques, Data Visualization, Data Collection, Data Analysis, Big Data Analytics, Generative AI, Natural Language Processing(NLP), Statistical Analysis, Data Processing at Scale, Time Series Analysis & Forecasting, Statistical Machine Learning, A/B Testing
- **Large Language Models(LLMs)**: LLaMA(Meta), Gemini(Google), Mistral(Mistral AI), GPT(OpenAI)
- **Algorithms:** Classification, Regression, Clustering, Transformers, Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Convolutional Neural Networks (CNNs), Decision Trees
- **Tools**: VSCode, Google Colab, Microsoft SQL Server, Jupyter Notebook, Microsoft Office, Apache Spark, Snowflake, GitHub, Docker, IntelliJ IDEA, Looker, Runpod, Hadoop, Azure DevOps, GCP, NoSQL

Certifications

- 8/10 courses of IBM Data Science Professional Certificate
- Fundamentals of Accelerated Computing with CUDA Python *by NVIDIA*
- Specialized Models: Time Series and Survival Analysis *by IBM*
- Natural Language Processing in TensorFlow *by DeepLearning.AI*
- Social Change: How can Marketing help? *by Griffith University*
- Digital Skills: User Experience *by Accenture*
- Optimizing a Website for Google Search *by University of California, Davis*
- Elements of AI *by University of Helsinki*
- Entrepreneurship: From Business Idea to Action *by King's College London*
- Agile with Atlassian Jira *by Atlassian*
- Introduction to NLP SQL *by Kaggle*
- Python, Java, SQL *by Kaggle*

Entrepreneurial Experience

Accumulated several years of entrepreneurial experience with startups specializing in blockchain, clothing, skincare, and brand development. Additionally, attended entrepreneurship minor classes to acquire foundational knowledge, effectively applying insights in practical business scenarios.

Honors and Awards

- **Ideal Student of the Year**:

  Recipient of the St. Joseph's Academy annual award for **outstanding performance in academics and co-curricular activities.** The award recognizes a student who demonstrates exceptional leadership skills, academic excellence, and active participation in extracurricular activities. Key achievements included maintaining **high grades**, engaging in **community service**, and demonstrating **leadership skills** in various student organizations

- **French Olympiad Scholar**:

  Received the prestigious French Olympiad award for five consecutive years, demonstrating **exceptional French language skills** and **cultural knowledge of France**. The award recognizes students who excel in their ability to communicate in French and demonstrate a deep understanding of French culture and society. Key achievements included achieving high scores on the French language test and demonstrating a strong interest in French culture.

- **The Guinness World Record Event**:

  Received an award from GUVI as part of the Guinness World Record event for the **most users taking an online computer programming lesson in 24 hours**. The award recognized my **exceptional performance and achievement in the programming lesson**, demonstrating a deep understanding of computer programming concepts and skills. Key achievements included actively participating in the event and successfully completing the lesson in record time.

Positions of Responsibility

- Head of HR, *Computer Society of India at Bennett University* - **Managed the society's HR activities**, including recruiting, onboarding, and organizing events for the members.
- Public Relations and Outreach Executive, *Computer Society of India at Bennett University* - Collaborated with the team to plan and execute **marketing and outreach campaigns**, increasing the society's visibility.
- Outreach Assistant, *Zenevia - The Annual TechFest at Bennett University* - Assisted in **planning and executing marketing and outreach campaigns** to promote the event to a wider audience, resulting in increased participation.
- Organizing Committee Member, *Breathe India Fundraiser* for the Covid Relief Fund - Contributed to the planning and execution of a **successful fundraising event** for Covid-19 relief efforts.
- Nature Club Member, *St. Joseph's Academy* - Participated in various activities and **initiatives to promote environmental awareness and conservation**.
- Student Task Force Member, *Seth Anandram Jaipuria School* - Contributed to various **initiatives and events organized** by the student task force to improve the school's environment and student experience.
- Interact Club Student Representative (*associated with Rotary International, Delhi*) - Represented the club in various school activities and organized initiatives to **promote community service and volunteerism**.
- Student Volunteer, *HelpAge India* - Participated in various initiatives and events organized by HelpAge India to **support the elderly community** in terms of **community service**.

Achievements

- Letter of Appreciation by the Dean of School of Computer Science and Engineering Technology for getting a perfect 10 CGPA
- Achieved DELF A1 certification
- Won first place at 'Pitch Please', an annual startup pitching competition
- Earned medals for outstanding performance in various competitions, including Zonal Badminton, Zonal Debate, and District Debate tournaments
- Competed at the district level in debates

Test Scores

1. **Duolingo English Test : 145/160** - November, 2023
2. **GRE : 305, AWA : 4.0** - July, 2022
3. **IELTS : 7/9** - November, 2019
4. **DELF A1 : 81/100** - September, 2017

Languages

- **English** - Professional Proficiency
- **Hindi** - Native Proficiency
- **French** - Elementary Proficiency