

# Sid Mangalik

📍 Stony Brook, New York 11790 ✉ [mangaliks@gmail.com](mailto:mangaliks@gmail.com)  
🌐 [linkedin.com/sid-mangalik](https://www.linkedin.com/sid-mangalik) 🐙 [smangalik.github.io](https://smangalik.github.io) 🎓 [Google Scholar](#)

## Education

- Stony Brook University** August 2020 - May 2025  
*Ph.D. in Computer Science* Stony Brook, NY
- Research in NLP and Computational Social Science, focus on human-centered AI/ML
  - Advised by Dr. H. Andrew Schwartz and Dr. Ryan L. Boyd
  - President of the Computer Science Graduate Student Organization (CSGSO)
- Stony Brook University** August 2014 - May 2018  
*B.S. in Computer Science* Stony Brook, NY
- Research thesis using NLP for authorship attribution, advised by Dr. Ritwik Banerjee
  - Member of the highly selective Honors Computer Science program

## Experience

- Meta** May 2024 - Present  
*Machine Learning Engineer PhD Intern* New York, NY
- Worked with the Facebook Dating team to increase user growth and engagement metrics
  - Developed highly scalable classifiers and tools leveraging sparse neural networks in PyTorch
  - Conducted A/B testing on the production Facebook application
  - Adopted standard machine learning best practices for model deployment
- Monitaur AI** August 2019 - May 2024  
*Research Scientist* Boston, MA (Remote)
- Wrote the foundational code of a machine learning model auditing system
  - Conducted research to automatically identify harmful bias in data and models
  - Created educational materials to teach data scientists about AI explainability
  - Published materials on Fair & Safe AI, including [The AI Fundamentalists](#) podcast
- Movable Ink** May 2022 - August 2022  
*Data Scientist* New York, NY
- Developed a contextual NLP pipeline to extract sales information from marketing materials
  - Predicted click-through rate of emails using a BERT-based regressor based on text content
  - Fine-tuned T5 and BART models to automatically generate engaging email subject lines

## Technical Skills

**Languages:** Python, Java, C, JavaScript, MATLAB, R, TypeScript, Scala, GoLang, SML, Prolog, MIPS  
**Markup:** HTML, CSS, LaTeX, JSON, CSV  
**textbfDatabases:** SQL, MySQL, PostgreSQL, DB2, Snowflake, Redis  
**Technologies:** AWS, GCP, pandas, scikit-learn, PyTorch, TensorFlow, Keras, HuggingFace Jupyter, Hadoop, Spark, MapReduce, Docker, d3.js, Grafana, ReactJS, AngularJS  
**Concepts:** Natural Language Processing, Machine Learning, Data Science, Artificial Intelligence, Neural Networks

## Projects

**COVID-19 Causal Language Inference:** Research into causal inference between major events and language

**Self Beliefs:** Research into the classification and taxonomy of beliefs about the self.

**Parenting Beliefs:** Research into commonly expressed beliefs about the nature of parenting.

**Predator Detection:** Research into whether chat participants were predators with StanfordCoreNLP.

**LUSId:** Research into the effects of social demographics on scientific writing style using Mallet API.

**QA Clarification Generation:** Using GPT-2/LSTMs to generate clarification questions to improve NLP QA systems.

## Patents and Publications

---

### **Robust language-based mental health assessments in time and space through social media** | Nature Digital Medicine

- First of its kind research into creating robust mental health assessments across the United States in space and time
- Multidisciplinary project between UPenn, Stanford, Stony Brook, and Stony Brook Medicine
- Combining expertise from computer science, machine learning, linguistics, psychology, and population health

### **I slept like a baby: using human traits to characterize deceptive ChatGPT and human text** | IACT'23

- Identifying AI generated language using human traits that text generation systems fail to recreate

### **Archetypes and Entropy: Theory-Driven Extraction of Evidence for Suicide Risk** | CLPsych 2024

- Novel approach to classifying suicide risk in patients using psychologically motivated constructs and archetypes

### **Document Classification of Files on the Client Side Before Upload** | Patent Inventor

- Co-invented with Dimitri Bilenkin in December 2019 under USPTO IDF07015, reference #4375.1650000
- Method for client browser to classify documents before they are sent to the server
- Implemented with Typescript, Tensorflow, and Angular.JS

### **Distance Points Interactive Interface and Related Methods** | Patent Inventor

- Co-invented with Brian Lin in January 2020, pending under U.S. Application Number 17/156,823
- Methods for presenting credit card rewards miles in an interactive user interface
- Implemented in React.JS with python web-scraping backend