

# Shramay Palta

UNIVERSITY OF MARYLAND, COLLEGE PARK

4108, Brendan Iribe Center for Computer Science and Engineering

+1 (443) 858-4922 | ✉ spalta@umd.edu | 🏠 shramay-palta.github.io | 🌐 shramay-palta | in shramay-palta | 🐦 @PaltaShramay | 📧 Shramay Palta

## Summary

I am broadly interested in the areas of **Natural Language Processing (NLP)** and **Computational Linguistics**, with a focus on **Commonsense Reasoning**, **Multilinguality** and **Bias and Fairness in NLP**.

## Education

### University of Maryland College Park

College Park, Maryland, USA

DOCTOR OF PHILOSOPHY, COMPUTER SCIENCE. **GPA: 4.00/4.00**

AUGUST 2021-PRESENT

Advisor: **Professor Rachel Rudinger**.

### Birla Institute of Technology and Science, Pilani (BITS Pilani)

Pilani, India

BACHELOR OF ENGINEERING, ELECTRICAL AND ELECTRONICS ENGINEERING . **FIRST DIVISION**

AUGUST 2017-MAY 2021

Thesis Supervisor: **Dr. Ashok Agrawala (UMD)** and **Dr. Navneet Gupta (BITS)**.

## Research and Work Experience

### Computational Linguistics and Information Processing (CLIP) Lab

College Park, Maryland, USA

GRADUATE STUDENT UNDER PROF. RACHEL RUDINGER

NOVEMBER 2021 - PRESENT

- Investigating different forms of **biases** in **NLP models and datasets**.

### Human-Data Interaction Group, University of Maryland

College Park, Maryland, USA

GRADUATE RESEARCH ASSISTANT UNDER PROF. LEO ZHICHENG LIU

SEPTEMBER 2021 - MARCH 2022

- Using **Natural Language Processing** Techniques to harvest design feedback from visualization comments on social media platforms like Reddit.

### Maryland Information and Network Dynamics (MIND) Lab, University of Maryland

College Park, Maryland, USA

RESEARCH ASSISTANT UNDER PROF. ASHOK AGRAWALA (UNDERGRADUATE RESEARCH THESIS)

MAY 2020 - FEBRUARY 2021

- Analyzing the spread of **COVID-19 and Flu virus** on campus using location and breathing data collected from **Spire Tags**. Part of the PROMETHEUS Project in collaboration with the School of Public Health.

### Global Health Centre, Graduate Institute of International and Development Studies

Geneva, Switzerland

RESEARCH INTERN UNDER DR. AMANDEEP GILL, EXECUTIVE DIRECTOR, UNSG'S PANEL ON DIGITAL COOPERATION

MAY 2020 - OCTOBER 2020

- Researched the role of **micro-narratives** as proxy variables to fill in missing data and used **natural language techniques** to study the social, health, and impacts of the **COVID-19** crisis on various sections of the society.

### TurnoutNow LLC

Lancaster, Pennsylvania, USA

DATA SCIENCE INTERN

MAY 2019-JULY 2019

- Using **real-time location data** from **IoT BLE Beacons** and **natural language generation** tools with live data connections to generate narratives for end users.

## Publications

### Investigating Information Inconsistency in Multilingual Open-Domain Question Answering – *Arxiv preprint*

**SHRAMAY PALTA**, HAOZHE AN, YIFAN YANG, SHUAIYI HUANG, MAHARSHI GOR

## Activities

**Reviewer:** EACL 2023: Ethical and Sustainable NLP track.

**Reviewer:** EMNLP 2022: Sentiment Analysis, Stylistic Analysis, and Argument Mining track.

**Member:** Graduate Admissions Committee, Department of Computer Science, University of Maryland.

## Projects

### Analyzing Inconsistencies in Multilingual Open-Domain QA

College Park, Maryland

WORK WITH PROF. JORDAN BOYD-GRABER

MARCH 2022- PRESENT

- Investigating whether **multilingual question answering** can potentially expose users to unreliable information through **cultural differences**, **divergent national laws**, or **uneven resources**.
- Analyzing if **different retriever models** present different passages—and answers—given the same question in different languages different multilingual QA datasets.
- Different answers potentially reveal valuable information about per-language **resources disparity**, and **linguistic variation**.

### Food and Culture in Commonsense NLP

College Park, Maryland

WORK WITH PROF. RACHEL RUDINGER AND PROF. ANTOINE BOSSELT

NOVEMBER 2021- PRESENT

- Investigating modern-day **commonsense reasoning** NLP Models and datasets like **Delphi**, **COMET**, **ATOMIC**, and **CommonsenseQA** to determine if they have an implicit or explicit **cultural bias** baked into them.
- Preparing a **dataset** of test bed questions that can be used to stress test these models using examples of cultural and social norms, material and physical differences to prove such biases.
- Testing multiple models like **BERT-base**, **BERT-Large**, **RoBERTa** to determine the effect on performance by measuring how the answer varies when the cultural context is explicitly or implicitly specified in the question and when it is not.

## Detecting Dietary Activity with Wearable Earphones

College Park, Maryland

WORK WITH PROF. NIRUPAM ROY

SEPTEMBER 2021- DECEMBER 2021

- Developed a technique to detect the types of food being consumed using **eSense**, a consumer wireless earphone device by Bell Labs.
- Used data from the microphone to **detect chewing activity** and classify the food being chewed as **solid, liquid or semi-solid**.
- Choosing **Convolutional Neural Networks (CNNs)** as the feature extractor, used standard **Conv2d** layers and a standard training and evaluation procedure to train the models both with and without the **Mel Spectrogram**.

## Economic Psychology: Stock Market Prediction using BERT

BITS Pilani, India

WORK WITH PROF. RAJNEESH CHOUBISA

JANUARY 2021- MAY 2021

- Implemented a **neural network** for utilizing information in **SEC 8-K forms** for predicting the movement of the **S&P 500 index**.
- Used **BERT** for capturing the contextual information in the form of two methods: **Masked Language Modeling (MLM)** and **Next Sentence Predicting (NSP)**.

## Darknet Insights using R and Python

BITS Pilani, India

WORK WITH PROF. VISHAL GUPTA

AUGUST 2019-DECEMBER 2019

- Used information from DNS queries to predict a **DDoS attack** from Darknet data from **Center for Applied Internet Data Analysis (CAIDA)** supercomputer servers of the **University of California, San Diego (UCSD)**.
- Implemented **Python scripts for feature extraction** (like TTL, IP length, Packet Count etc.) and **CAIDA's internal tool, corsaro** for large scale analysis of trace data.
- Used **vector quantization algorithms including K-means and EM** on the extracted features to predict DDoS attacks.

## Skills

<b>Languages</b>	Python, R, SQL, Linux/Unix shell, Java, C++, C, Assembly Language.
<b>Tools</b>	Pandas, NumPy, NLTK, spacy, Keras, TensorFlow, PyTorch, scikit-learn, Matplotlib, Jupyter, Git, $\text{\LaTeX}$ , MATLAB, MySQL.
<b>Key Courses</b>	CommonSense Reasoning and Natural Language Understanding, Natural Language Processing, How and Why AI Answers Questions, Human AI Interaction, Advanced Numerical Optimization.
<b>OS</b>	MacOS, Linux, Windows, FreeBSD.

## Achievements

<b>2022</b>	<b>Dean's Fellowship Award:</b> Awarded the Graduate School Dean's Fellowship Award for outstanding academic achievement.
<b>2021</b>	<b>Dean's Fellowship and Chair's Fellowship Award:</b> Awarded the Graduate School Dean's Fellowship and the Chair's Fellowship Award for outstanding academic achievement.
<b>2015</b>	<b>National Merit Holder:</b> One of the <b>top 0.1%</b> scorers across India in the board examinations conducted by Central Board of Secondary Education; Received Letter of Honour from the HRD Minister, Govt. of India.
<b>2010</b>	<b>World Robot Olympiad:</b> Represented India at the World Robot Olympiad held in Manila, Philippines and managed to secure a <b>world rank of 31</b> in my category.
<b>2010</b>	<b>Indian Robot Olympiad:</b> Awarded the <b>1st Runners Up Award</b> in my category for the north chapter.