

# Massimo Daul

*New York University Mathematics Undergrad*

+1 (207) 664-8852 | [mmd9604@nyu.edu](mailto:mmd9604@nyu.edu) | [linkedin.com/in/massimodaul/](https://www.linkedin.com/in/massimodaul/) | <https://github.com/MassimoDaul>

## SUMMARY

---

I'm a fourth-year Shelby Davis Scholar at New York University, pursuing a joint major in Mathematics and Computer Science. My research focuses on computational tools for low-resource languages, where I work to make language models more inclusive.

## TECHNICAL SKILLS

---

**Languages:** Python, Java, C++

**Frameworks:** PyTorch, TensorFlow, Keras, Scikit-Learn, Pandas, Seaborn

**Relevant Courses:** Linear Algebra, Linguistics in Latin America, Combinatorics, Analysis, Data Structures, Operating Systems, Fundamentals of Machine Learning, Algebra, Theory of Probability, Bilingualism

## EXPERIENCE

---

### Computational Intern - Low-Resource Language Documentation

06/2024 - current

*Endangered Languages in Contact in Istria and Kvarner*

*New York City, USA*

- Headed project for fine-tuned custom Wav2Vec2-Bert models for ASR and Forced-Alignment of low-resource languages
- Developed machine learning approaches for part-of-speech tagging, speaker diarization, and segmentation to be used in low-resource corpus creation

### Computational Intern - Addiction Research

06/2019 - 12/2023

*Jackson Laboratory*

*Bar Harbor, USA*

- Designed neural networks for gait detection and pose analysis in mice to further addiction research.

### Computational Intern - Small Business Development

06/2023 - 08/2023

*Profit Decoder*

*New York City, USA*

- Developed transparent economic data analysis solutions for a start-up supporting small businesses

## PUBLICATIONS

---

### Linguistically Informed Tokenization Improves ASR for Underresourced Languages

Accepted for presentation at Field Matters (EACL), First Author

### Comparing Phonological Feature Sets for Low-Resource ASR

Presented at the Annual Meeting on Phonology; under review for conference proceedings publication, Second Author

### Comparison of Neural Network and GMM-HMM-Based Forced Alignment Models on Istro-Venetian

Manuscript in preparation for extended conference submission, Co-Author

### Creating a Multilingual Spoken Corpus for Language Documentation and Research

Presented at *ICLDC*, Co-Author

## EDUCATION

---

### Collegio Del Mondo Unito Dell'Adriatico

*International Baccalaureate*

Duino, Italy

*September 2021 - May 2022*

### New York University

*Mathematics-Computer Science joint degree*

New York City, USA

*September 2022 - Present*

## LANGUAGES

---

**English, Italian:** Fluent    **Spanish, Portuguese:** Conversational