

# Zachary Webel

Data Scientist & Applied Mathematician



Washington, D.C.



@zac-webel



561-225-4251



zow2@georgetown.edu



zacharywebel

## Profile

Skilled engineer with over 8+ years working with Python, 5+ years working with Machine and Deep Learning and experience with TensorFlow, Keras, Scikit Learn, PyTorch, Gensim, NLTK, NEO4J, PostgreSQL, R, C++, Docker, CUDA, MLOps, Generative AI, Large Language Models (LLM), BigData, Google Cloud and AWS. Continuously staying updated with the latest industry trends and advancements, I am **committed to applying cutting-edge methodologies to deliver impactful solutions that address complex business challenges**.

Experienced Data Scientist and Applied Mathematician with a **passion for leveraging analytical skills to extract valuable insights from complex datasets**. Possessing a strong background in mathematics and statistics, coupled with expertise in machine learning and data mining techniques, I am adept at transforming raw data into actionable solutions. A **collaborative and innovative problem-solver**, I thrive in multidisciplinary teams, effectively communicating technical concepts to non-technical stakeholders.

## Education

2022 – 2023

### Master of Science: Health Informatics & Data Science

Georgetown University – Washington, D.C.

GPA: 4.0

Courses Taken:

- Machine Learning
- EHR Data Science
- Genomic Data Science
- Imaging Artificial Intelligence
- Advanced Health Informatics
- AI for Health Applications
- Digital Health Application Design
- Biostatistics
- Research Capstone

2019 – 2022

### Bachelor of Science: Applied & Computational Mathematics

Florida State University Honors College – Tallahassee, FL.

GPA: 3.9 – Magna Cum Laude

Courses Taken:

- Ordinary Differential Equations
- Linear Algebra
- Partial Differential Equations
- Numerical Analysis
- Math Modeling
- Complex Variables
- Discrete Mathematics
- Vector Calculus
- Calculus I, II, and III
- Economics
- Statistics
- Physics A & B

2019

### Shanghai Jiao Tong University: Machine Learning

Shanghai, China

I was a member of the Machine Learning immersion program at the Shanghai Jiao Tong University in Shanghai, China. While in the program, I took a course in Machine Learning which provided me with a deeper understanding of the theoretical foundations and practical applications of Artificial Intelligence. I also attended China Shanghai International Technology Fair.

## Experience

---

### 2023 - Current

#### Artificial Intelligence Researcher

Food and Drug Administration (FDA), Washington, D.C.

*As an Artificial Intelligence Researcher at the FDA, I built a nucleotide transformer language model from scratch to predict mutations in virus RNA coding for glycoproteins. My model was able to simulate the evolution of the virus genome to provide insight to guide viral countermeasures towards future versions of the virus before the virus has evolved to that stage.*

- Built LLM from scratch.
- Collaborated with a team of researchers and scientists to develop and refine the language model.
- Employed cutting-edge AI technologies and probability theory to analyze and predict virus mutations.

### 2023

#### Data Scientist NLP – Private Intelligence

Stealth Startup, Washington, D.C.

- Deployed 3b parameter language model summarizing scientific papers.
- Deployed several pretrained models from Huggingface – summarization, ner, token classification.
- Created a dynamic topic model to model the evolution of research topics in a target country.
- Created a knowledge graph of foreign institutions/scientists collaborating on technology advancements.
- Created a directed word embedding graph for relationship extraction.

### 2019-Current

#### Artificial Intelligence Research – Data Science Projects

##### 2023 - Generative Sequential Language Model - Virus Protein Design

- *I created a language model that generates a novel protein sequence with a third party predicted function of virus tail fiber and cell binding when an entirely random seed sequence.*

##### 2023 – Generative Language Model – Protein Design

- *I created a language model that generates a novel protein sequence with under 3% for alpha helix, turn and beta sheet desired properties.*

##### 2023 – Genetic Biomarker Discovery – Stage 2 Invasive Bladder Cancer

- *Statistical analysis on gene expression data from Stage 2 invasive bladder cancer*

##### 2023 – Tabular EHR Patient Mortality Prediction

- *I created an ensemble of shallow machine learning classifiers xgboost, random forest, and catboost to predict patient mortality given tabular EHR data.*

##### 2023 – Graduate Research Paper on Federated Machine Learning

- *I co-authored a research paper on federated machine learning in healthcare applications.*

##### 2023 – Imaging Artificial Intelligence

- *I created convolutional neural network that classified CT images from COVID patients.*
- *I created convolutional neural network that classified whole slide images as cancer or not.*

##### 2022 – Modeling Equilibria and Stability of the Lorenz Equations

- *I created a notebook that explored the dynamics, chaos, and stability of the Lorenz Equations.*

##### 2022 – System of Differential Equations Model – Population Dynamics - Optimization

- *Expanded SIR that modeled the effect of masks and vaccines during the COVID pandemic. Then optimized the system based on user parameters.*

##### 2022 – Complex Visualization

- *Visualization of the contour plot, gradient and level curves of a complex multivariable function.*

##### 2021 – Numerical Analysis Approximation – Quadrature – Interpolation – Root finding

- *I created notebooks that modeled: Orthogonal Polynomials, Least Squares, Monte Carlo Integration, Gauss-Legendre rule, Newton-Cotes quadrature, cubic splines, Hermite interpolation, Runge's function, Newton method, and Muller's method.*

##### 2020 – Probabilistic Modeling – A Neural Approach

- *I created a neural network that models expected yards gained on a football play resulting in a masked probability distribution that can be sampled to simulate a play.*

##### 2019 – Tabular Data Modeling

- *Random forest model that predicted the play call using play by play data.*
- *Random forest model that predicted the blitz package for the defense on any given nfl play.*

##### 2019 – Modeling Aerodynamics

- *Modeled the power needed to propel an aircraft forward optimized for energy minimized.*