🏠 Baltimore, MD
✉ jshinm@gmail.com

# Jong M. Shin

in linkedin.com/in/jshinm
github.com/jshinm

## Professional Experience

**Software Engineer — AI/ML** ............................................................................... **Nov 2021 - Present**
*Johns Hopkins University - Whiting School of Engineering, Baltimore, MD*

- Leads a team of 4 AI researchers and 20+ engineering students and oversees development of 6 ML **python** packages (tree-based ensemble models, GMM models) to be integrated into **scikit-learn** and **scikit-image**
- Wrote 30+ python tutorials and unit tests for **CI/CD** using **parallel processing** and **GPU utilization**
- Maintains and publishes research software in an open source setting in GitHub organization
- Published 2 preprints on AI/ML subjects as a co-author (google scholar)
- Presented at 2022 NAISys conference on machine-human intelligence and machine behavior (poster)

**Data Scientist | Machine Learning Engineer** ......................................................... **Jan 2022 — Present**
*U.S. Bureau of Labor Statistics, Washington, DC*

- Retrained and optimized the transformer-based **NLP model** using 2.4M records achieving 81% accuracy
- Deployed newly updated **natural language processing (NLP)** model using **scikit-learn, TensorFlow, PyTorch, Huggingface** through **CI/CD** from GitLab to Linux GPU server
- Integrated pre-trained **transformer** (BERT, RoBERTa) for lemmatization and **transfer-learning** (connecting to the custom ResNet) into the NLP pipeline and increased overall mF1 score by 30%
- Prepared 300K unstructured records of text data via **feature engineering**/pre-processing/resampling
- Rewrote auxiliary **python** analytic software for **machine learning** pipeline (ML) to pull data from **Oracle** data warehouse to support 10+ economists

**Research Assistant** ............................................................................................. **Sep 2020 — Nov 2021**
*Johns Hopkins University - Whiting School of Engineering, Baltimore, MD*

- Studied the extrapolative behavior of various ML algorithms implemented using **python** and **scikit-learn**
- Built a production-grade **web app** and **REST API** to conduct human behavioral experiments using **HTML, Javascript** and **Flask**, and backend database management using **SQLAlchemy** hosted on heroku

**Research Intern** ................................................................................................. **May 2021 — Jul 2021**
*Microsoft Research, Redmond, WA*

- Developed a web application that automates the process of causal inference using **React and Typescript** to power front-end and **python Flask** to service back-end in the context of human trafficking and COVID-19
- Built and designed an **end-to-end** pipeline of **causal inference** using **Python DoWhy** package to deploy a research application product branded as ShowWhy while leading 6 software engineers

**Data Science Intern** ........................................................................................... **Jan 2021 — May 2021**
*Johns Hopkins University School of Medicine, Baltimore, MD*

- Developed a data preparation pipeline for JHU COVID-19 initiative using **probabilistic linkage** methods to provide backend database management in **R**
- Built a data analytic pipeline to parse hospital datasets via **causal inference** methods to provide a clinical data science guideline for COVID-19 vaccination distribution

**Data Science Consultant** ..................................................................................... **Dec 2020 — May 2021**
*MindX, Bethesda, MD*

- Built an analytical pipeline to parse out biometric signals from the **multivariate time-series** hologram signal datasets in **python** using **scikit-learn** and **TensorFlow**
- Cleaned and pre-processed real-world multimodal datasets using **pandas** and **hyppo**

**Post baccalaureate IRTA research fellow** ............................................................... **Nov 2018 — May 2021**
*The National Institutes of Health, Bethesda, MD*

- Built a prediction pipeline for HIV detection by HIV antibody titer in **Python** using **pandas** and **scikit-learn**
- Conducted parametric/nonparametric multivariate linear regression analysis on the national big data (metabolomics, proteomics) using **Python** and **R**

**Graduate Researcher** ............................................................................................ **Jun 2017 - Oct 2018**
*Virginia Commonwealth University, Richmond, VA*

- Built automated Fourier transformed signal detection program in **Matlab** for mouse behavioral experiments
- Conducted 50+ multivariate parametric/nonparametric data analysis on biomedical data for publication

# Jong M. Shin

## Skills

Python, R, SQL (MySQL, MS SQL Server), Git, MATLAB, Java, Typescript, React, HTML/CSS, Linux (Bash/Shell)

## Recent Publications

1. *Deep discriminative to kernel generative modeling* (2022), arXiv [Link]

2. *Prospective Learning: Back to the Future* (2022), arXiv [Link]

3. *Interclass GPCR heteromerization affects localization and trafficking* (2020), Science Signaling [Link]

4. *Site-Specific Incorporation of Genetically Encoded Photo-Crosslinkers Locates the Heteromeric Interface of a GPCR Complex in Living Cells* (2020), Cell Chemical Biology [Link]

5. *Fully automated head-twitch detection system for the study of 5-HT2A receptor pharmacology in vivo* (2019), Scientific Reports [Link]

## Education

**Masters of Engineering, Biomedical Engineering – Data Science** (GPA: 4.0/4.0) .................... **May 2021**
*Johns Hopkins University - Whiting School of Engineering, Baltimore, MD*

**Masters of Science, Physiology and Biophysics** (GPA: 4.0/4.0) ........................................ **August 2018**
*Virginia Commonwealth University - School of Medicine, Richmond, VA*

## Professional Projects

### Automated end-to-end causal inference application (Microsoft Research)
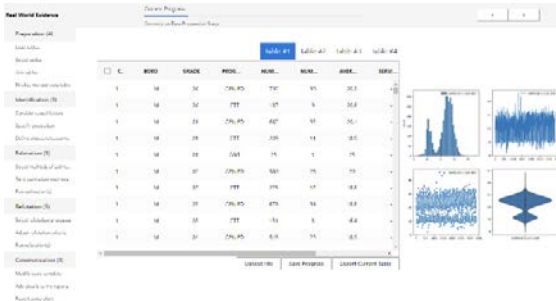


- Devised application design from a data scientist perspective and laid out the basic architecture from scratch
- Built proof-of-concept framework of the app using **Fluent-UI**, **Typescript**, and **python-flask**
- Wrote in-house python libraries and jupyter notebooks to develop the **end-to-end data science pipelines**
- Used visualization tools such as **matplotlib** and **seaborn** to generate internal statistics of the data sets within the app

### Inductive bias experiment (Johns Hopkins University - JOVO Lab)
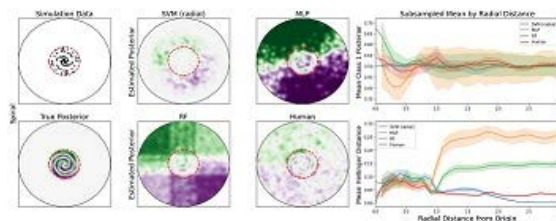


- Implemented ML models from **sklearn** and trained on nonlinear simulation data
- Generated mathematically derived posterior probability for exclusive OR and spiral dataset
- Implemented point-wise Hellinger distance and explored extrapolative behavior of ML models such as **SVM**, **DN**, **RF**
- Generated publication figures and presented experimental findings using visual tools such as **matplotlib** and **seaborn**

### Web application for human behavioral experiment (Johns Hopkins University - JOVO Lab)
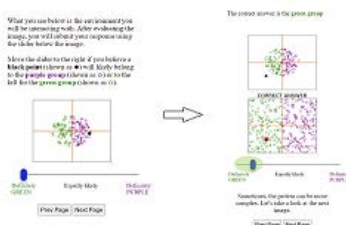


- Developed **web application** for human behavioral experiment to collect extrapolative inference pattern performed by humans
- Designed front-end using **HTML/CSS**, **JavaScript** and powered back-end using **python-Flask** and **SQLalchemy**
- Managed **SQL database** that stored over 150 participants' behavioral experiment data on heroku