# Yinkai Wang

E-mail: ywang88@gmu.edu | Phone: +86-18650146958
Personal Website: https://yinkaiw.github.io/

## EDUCATION

**George Mason University, the US** **08/2018-08/2021**
Bachelor of Science in *Computer Science*

- Honors/Awards: Dean's List (2018-2020)

**Huaqiao University, China** **08/2017-08/2022**
Bachelor of Science in *Computer Science*

## PUBLICATIONS

- Yuanqi Du, **Yinkai Wang**, Fardina Alam, Yuanjie Lu, Xiaojie Guo, Liang Zhao, Amarda Shehu. Deep Latent-Variable Models for Controllable Molecule Generation. Research Paper for IEEE BIBM 2021.
- **Yinkai Wang***, Kaiyi Guan*, Aowei Ding*, Yuanqi Du. Ensemble Machine Learning System for Student Academic Performance Prediction. Educational Data Mining (EDM) 2021, Workshop for Undergraduates (W4U).
- Fahim Faisal, **Yinkai Wang**, Antonis Anastasopoulos. Dataset Geography: Mapping Language Data to Language Users. Research paper for ACL 202. (In submission)
- **Yinkai Wang**, Antonis Anastasopoulos. On the Cross-Lingual Consistency of Named Entity Recognition Models. Student Abstract for AAAI-UC. (In submission)

## RESEARCH EXPERIENCES

**Multilingual Geospatial Language Expression Discovery** **05/2021-Present**
*Research Assistant (Advisor: Dr. Antonios Anastasopoulos)* *Fairfax, US*

- Developed a system trained by parallel multilingual datasets. Used mBert model+aligned to predict the consistency of multilingual datasets.
- Defined cross-lingual consistency as the desirable property that two parallel sentences in two languages, which should in principle use the same-named entities are actually tagged with the same named entities. Proved the significance of consistency in a NER task.
- Focused on the notion of cross-lingual consistency for multilingual NER models and showed the importance of parallel data in multiple languages to evaluate NER models.

**Predicting Minimum Inhibitory Concentration for Quaternary Ammonium Compounds w/ Machining Learning** **09/2021-Present**
*Research Assistant (Advisor: Dr. Amarda Shehu)* *Fairfax, US*

- Created three settings based on the 70 features of ~450 Quaternary Ammonium Compounds. Used machine learning models to predict four properties of molecules based on these settings.
- Predicted four Minimum Inhibitory Concentration values with ten regression machine learning models. Used feature selection to analyze the best features settings for every property.

**Deep Latent-Variable Models for Controllable Molecule Generation** **08/2021-09/2021**
*Research Assistant (Advisor: Dr. Amarda Shehu)* *Fairfax, US*

- Proposed several deep latent-variable models to generate small molecules with desired molecular properties.
- Made the models operated under supervised, disentangled representation learning and leverage both graph representation learning to learn inherent constraints in the chemical space and inductive bias to connect chemical and biological space.
- Completed the evaluation that the models are a promising step in controllable molecule generation in support of cheminformatics, drug discovery, and other application settings.

**Ensemble Machine Learning System for Student Academic Performance Prediction** **01/2021-05/2021**

*Researcher* *Beijing, China*

- Used an ensemble machine learning system to predict students' final grades according to students' performance in different aspects.
- Created a system for teachers to know students' learning conditions during COVID-19 pandemic.
- Proposed a stable, effective and accurate model that consists of two components: the ensemble feature engineering module and the ensemble prediction module.

## SELECTED PROFESSIONAL EXPERIENCE

**Peking University** **11/2021-Present**

*Researcher in VDIG lab*

- Worked on object detection, a computer vision technique for locating instances of objects in images or videos. Produced meaningful results like the concept of an objection by object detection algorithms.
- Used self-supervised learning (SSL) and transformer structure to detect the object of images.

**Bytedance** **04/2021-07/2021**

*Intern in DevEco*

- Focused on the base of the host app of android, which had a coupling relationship with most of the apps from ByteDance.
- Created a mock setting environment implement to help QA test, which improved the efficiency of testing and publishing greatly.
- Made a great and comfortable environment for all the developers who were developing microapp on ByteDance.

**Google Smartphone Decimeter Challenge in Google Kaggle** **06/2021-08/2021**

*Team member*

- Designed data cleaning, preprocessing, data analysis, model selection, result evaluation and visualization pipeline.
- Mastered real-world data science challenge with machine learning pipeline and team collaboration.
- Ranked top 20% on the Google Smartphone Decimeter Challenge.

## SKILLS

- **Programming skills**: Python, Java, C, Kotlin, MIPS, Julia
- **Language**: Chinese (native), English
- **Hobbies**: Basketball, Movies
- **Research Interests**: Machine Learning, Deep Learning, AI for Science, Deep Graph Learning, Natural Language Processing