Kaike Ping

Homepage: https://people.cs.vt.edu/pkk/
Email : pkk@vt.edu

Education

- **Virginia Tech** — Blacksburg, VA
  ***PhD** Candidate in Computer Science; Advisor: Eugenia Rho* — *Aug 2022 - Expected May 2027*
- **Southern Medical University** — Guangzhou, China
  ***Master** of Medicine in Biostatistics; Advisor: Pingyan Chen* — *Sep 2014 - Jul 2017*
  ***Bachelor** of Science in Applied Statistics* — *Sep 2010 - Jul 2014*

Experience

- **Virginia Tech** — Blacksburg, VA, USA
  PhD Candidate — *Aug 2022 - Present*
  - **Statistical Modeling (R)**: Led large-scale quantitative research on online user behavior, surveying $1,100+$ participants and employing advanced statistical modeling in R (e.g., Structural Equation Modeling, Mixed-Effects Models) to analyze motivations and barriers related to social media counterspeech.
  - **Small-Sample NLP Evaluation**: Developed high-performance NLP models under limited data conditions, achieving $\sim$87% predictive accuracy for counterspeech efficacy. This involved engineering and evaluating transformer-based models (e.g., RoBERTa) against state-of-the-art LLMs (GPT, Claude, Gemini) on a specialized dataset of 600+ curated text samples.
  - **Research Publication Impact**: Published key findings from this research in premier Human-Computer Interaction (HCI) academic venues, including *ACM ToCHI* (journal) and *CSCW* (conference).
- **Guizhou Center for Disease Control and Prevention (CDC)** — Guiyang, Guizhou, China
  Assistant Research Fellow — *Nov 2019 - Jul 2022*
  - **Early COVID-19 Research**: Pioneered early-phase COVID-19 epidemiological analysis, being among the first global researchers to estimate its incubation period using Accelerated Failure Time models and the effective reproduction number via the Wallinga & Teunis method.
  - **Publication Impact**: Published these novel findings in *The Journal of Infection in Developing Countries*, a work subsequently cited multiple times, shaping early understanding of the pandemic.
  - **Workflow Engineering & Automation**: Engineered an automated R and SQL workflow for ingesting, cleaning, and transforming weekly surveillance data (3.8M pop.), culminating in the automated generation of epidemiological summaries with key public health metrics.
  - **Text Data Processing**: Enhanced data extraction capabilities by fine-tuning a compact NLP model (*bert-small*) to parse and quantify information from textual epidemiological reports, standardizing inputs for an Excel-based centralized data repository.
  - **Pandemic Risk Assessment**: Developed and validated an Autoregressive Distributed Lag (ARDL) model in R, leveraging 177 weeks of internet search trend data (correlating with 173K pediatric HFMD cases) to forecast incidence rates. The model was integrated into the provincial early warning system for proactive public health interventions.
- **Contract Research Organization of CLT Inc** — Guangzhou, Guangdong, China
  Software Engineer — *Jul 2017 - Jun 2019*
  - **System Design & Development**: Led design and development as chief programmer for core clinical trial systems (Central Randomization & EDC), defining database architecture (MySQL), data validation rules, and system functionalities using PHP, JavaScript, HTML/CSS.
  - **Central Randomization System**: Engineered a web-based Central Randomization System deployed in 20+ multi-center clinical trials, featuring allocation concealment algorithms tailored to trial-specific statistical needs and enabling automated data export to SAS datasets for subsequent analysis.
  - **Electronic Data Capture System**: Developed a scalable Electronic Data Capture (EDC) SaaS platform managing clinical trial data for 100K+ patients, significantly improving real-time data access and data management efficiency for research teams.
- **National Clinical Research Center for Kidney Disease** — Guangzhou, Guangdong, China
  Research Intern — *Nov 2015 - Feb 2017*
  - **CRF Design & Data Standardization**: Contributed to a large-scale chronic kidney disease cohort study by co-designing Case Report Forms (CRFs), focusing on rigorous data standardization (defining variable formats, units, and ranges) to ensure high-quality data collection for subsequent statistical analysis.

Teaching and Mentorship

- **Virginia Tech** — Blacksburg, VA
  Teaching assistant — *Fall 2022 - Spring 2025*
  - **CS2104**: Intro to Problem Solving in CS
  - **CS3604**: Professionalism in Computing
  - **CS3754**: Cloud Software Development
  - **CS5024**: Ethics & Professionalism in CS
- **Guizhou Center for Disease Control and Prevention** — Guiyang, China
  Master Thesis Reviewer — *Spring 2021 & 2022*
  - Reviewed and provided feedback on master theses on topics related to biostatistics
- **Southern Medical University** — Guangzhou, China
  Instructor — *Fall 2015 & 2016*
  - **SPSS Application (master course)**: Delivered 50% of lectures and designed the final exam.
  - **Biostatistics (undergraduate course)**: Delivered 30% of lectures, designed two assignments, and the final exam.

  PhD Thesis Reviewer — *Spring 2015 & 2016*
  - Reviewed and provided feedback on the statistical sections of PhD theses

Honors and Awards

- `Oct 2021`: COVID-19 Experts Committee Member of Guizhou Province, China
- `Sep 2020`: Consortium author of Preterm Birth Prediction Challenge in Cell Rep Med
- `Oct 2015`: Silver medal, the 13th Challenge Cup Competition of Science Achievement in China
- `Jul 2014`: Best Poster Award, China Biostatistics Conference 2014
- `May 2011`: Honorable Mention, the 9th ACM-ICPC Guangdong Collegiate Programming Contest

Skills Summary

- **Statistical Modeling**: Regression analysis, Time series analysis (e.g., ARDL), Survival analysis (e.g., AFT models)
- **Data Visualization**: ggplot2 (R), Matplotlib/Seaborn (Python), Adobe Illustrator
- **Machine Learning**: NLP (e.g., Transformer models, HuggingFace), Predictive modeling, Classification, Clustering
- **Programming Languages**: Python, R, SQL, PyTorch, HuggingFace
- **Web Development**: HTML, CSS, JavaScript, PHP, MySQL
- **Developer Tools**: Git, GitHub, Docker
- **Biostatistics & Epidemiology**: Epidemiological Study Design (e.g., RCTs, Cohort Studies, Retrospective/Prospective Analyses), Surveillance Data Analysis, Application of Biostatistical Models, Public Health Principles

Selected Research Publications

[1] **Kaike Ping**, Anisha Kumar, Xiaohan Ding, et al. "Behind the Counter: Exploring the Motivations and Barriers of Online Counterspeech Writing". In: *ACM Trans. Comput.-Hum. Interact.* (May 2025).

[2] **Kaike Ping**, James Hawdon, and Eugenia H Rho. "Perceiving and Countering Hate: The Role of Identity in Online Responses". In: *Proc. ACM Hum.-Comput. Interact.* 9.2, CSCW147 (May 2025), pp. 1–28. DOI: 10.1145/3711045.

[3] **Kaike Ping**, Mingyu Lei, Yun Gou, et al. "Epidemiologic Characteristics of COVID-19 in Guizhou Province, China". In: *The Journal of Infection in Developing Countries* 15.03 (Mar. 31, 2021), pp. 389–397. ISSN: 1972-2680. DOI: 10.3855/jidc.12818.

[4] **Kaike Ping**, Mingyu Lei, Yun Gou, et al. "Epidemiological characteristics of coronavirus disease 2019 in Guizhou, 2020". In: *Disease Surveillance* 36.6 (June 30, 2021), pp. 581–586. ISSN: 1003-9961. DOI: 10.3784/jbjc.202103100109.

[5] Buse Carik, **Kaike Ping**, Xiaohan Ding, et al. "Exploring Large Language Models Through a Neurodivergent Lens: Use, Challenges, Community-Driven Workarounds, and Concerns". In: *Proc. ACM Hum.-Comput. Interact.* 9.1, GROUP15 (Jan. 2025), pp. 1–28. DOI: 10.1145/3701194.

[6] Li-Yen Yang, **Kaike Ping**, Yunan Luo, and Andrew C. McShan. "BioDolphin as a Comprehensive Database of Lipid–Protein Binding Interactions". In: *Communications Chemistry* 7.1 (Dec. 2024), pp. 1–11. DOI: 10.1038/s42004-024-01384-z.

[7] Yun Gou, **Kaike Ping**, Mingyu Lei, et al. "Initial Clinical Characteristics of 146 Patients with COVID-19 Reported in Guizhou Province, China: A Survival Analysis". In: *The Journal of Infection in Developing Countries* 16.01 (Jan. 31, 2022), pp. 32–40. ISSN: 1972-2680. DOI: 10.3855/jidc.15027.

[8] Chongyang Duan, Yingshu Cao, Yong Liu, Lizhi Zhou, **Kaike Ping**, et al. "A New Preprocedure Risk Score for Predicting Contrast-Induced Acute Kidney Injury". In: *Canadian Journal of Cardiology* 33.6 (June 1, 2017), pp. 714–723. ISSN: 0828-282X. DOI: 10.1016/j.cjca.2017.01.015.

[9] Shaoqun Zhang, Ji Qi, Lei Zhang, Chao Chen, Shubhro Mondal, **Kaike Ping**, et al. "Cervical Rotatory Manipulation Decreases Uniaxial Tensile Properties of Rabbit Atherosclerotic Internal Carotid Artery". In: *Evidence-Based Complementary and Alternative Medicine* 2017 (Feb. 16, 2017), e5189356. ISSN: 1741-427X. DOI: 10.1155/2017/5189356.

[10] Jiao Huang, Liying Wang, Chun Yu, Zhaobing Liu, Guanghong Yang, Ziyao Lan, Zhongfa Tao, **Kaike Ping**, et al. "Response and Assessment of the Effectiveness of the Countermeasures for a COVID-19 Outbreak — Guizhou Province, China, March 2022". In: *China CDC Weekly* 4.30 (July 29, 2022), pp. 655–659. ISSN: 2096-7071. DOI: 10.46234/ccdcw2022.141.