

Yankai Jiang

📞 (+1)773-997-6553

🔗 lukejyk.github.io/

📍 140, Fenway, MA 02115, USA

Education

Northeastern University

Ph.D. in Computer Engineering

- **Advisor:** Prof. Devesh Tiwari

- **Research Interests:** Sustainable Computing, High Performance Computing, Cloud Computing

Boston, MA

Sept. 2023 – Present

Northwestern University

M.S. in Computer Engineering

Evanston, IL

Sept. 2021 – Jun. 2023

Xi'an Jiaotong University

B.E. in Automation (Youth Program, equivalent to Honors Program)

Xi'an, China

Aug. 2016 – Jun. 2020

Research Experience

Goodwill Lab, Northeastern University, Boston, MA

Sept. 2023 – Present

- *Research Assistant*, advised by Prof. Devesh Tiwari

- Designing open-source solutions for constructing sustainable large-scale data centers and HPC systems, and cloud computing systems: heterogeneous hardware execution to optimize performance and energy in HPC.

Prescience Lab, Northwestern University, Evanston, IL

Mar. 2022 – Jun. 2023

- *Research Assistant*, advised by Prof. Peter Dinda

- Ported the Parallel Standard ML language, Maple, to the Nautilus aerokernel and developed serialization and deserialization capabilities within Maple to enable distributed computation.

PSEC Lab, Northwestern University, Evanston, IL

Mar. 2022 – Oct. 2022

- *Research Assistant*, advised by Prof. Yueqi Chen and Prof. Xinyu Xing

- Reproduced 100+ vulnerabilities and analyzed the vulnerable objects in the Linux kernel based on proof-of-concepts (PoC) generated from Syzkaller to evaluate the effectiveness of HotBPF ↗.

Cybersecurity R&D Lab, National University of Singapore, Singapore

Jul. 2019 – Aug. 2019

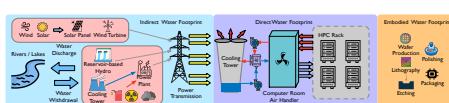
- *Summer Intern*, advised by Prof. Ee-Chien Chang

- Designed two attacking scenarios to exploit vulnerabilities (e.g., CVE-2019-12735, CVE-2019-13272) in Linux and constructed benign and malicious audit log datasets to evaluate the Watson ↗ (NDSS 2021).

Awards

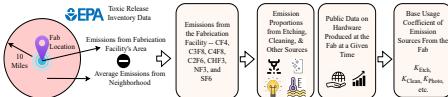
- ACM PPoPP 2025 Travel Award 2025
- IEEE HPEC 2024 Outstanding Paper Award Nomination 2024
- MVAPICH User Group (MUG) Conference Travel Award 2023, 2024, 2025
- Honorary Graduate of Qian Xuesen College 2020

Research Publications



Water Footprint Modeling, Characterization, and Analysis Toward Water-aware HPC System Design and Operations

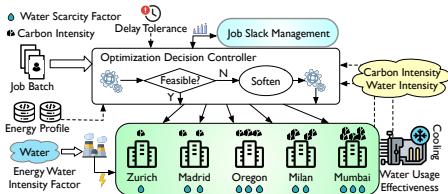
Yankai Jiang, Rohan Basu Roy, Raghavendra Kanakagiri, Devesh Tiwari
2025 *The 38th IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2025)*



ForgetMeNot: Modeling and Analyzing the Impact of Forever Chemicals in Designing Sustainable Computing Systems

Rohan Basu Roy, Raghavendra Kanakagiri, **Yankai Jiang**, Devesh Tiwari
2025 ACM Special Interest Group on Measurement and Evaluation (SIGMETRICS 2025)

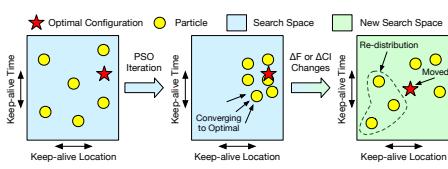
[Paper [🔗](#)] [Open-source Artifact [🔗](#)]



WaterWise: Co-optimizing Carbon- and Water-Footprint Toward Environmentally Sustainable Cloud Computing

Yankai Jiang, Rohan Basu Roy, Raghavendra Kanakagiri, Devesh Tiwari
2025 The 30th ACM Symposium on Principles and Practice of Parallel Programming (PPoPP 2025)

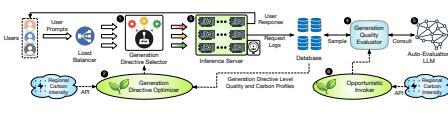
[Paper [🔗](#)] [Slides [🔗](#)] [Open-source Artifact [🔗](#)] [Bibtex [🔗](#)]



EcoLife: Carbon-Aware Serverless Function Scheduling for Sustainable Computing

Yankai Jiang, Rohan Basu Roy, Baolin Li, Devesh Tiwari
2024 The 37th IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2024)

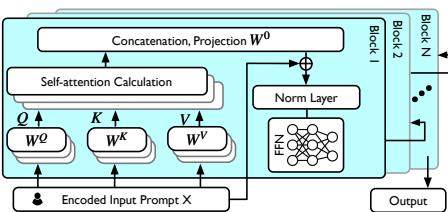
[Paper [🔗](#)] [Talk [🔗](#)] [Slides [🔗](#)] [Open-source Artifact [🔗](#)] [Press [🔗](#)] [Bibtex [🔗](#)]



Sprout: Green Generative AI with Carbon-Efficient LLM Inference

Baolin Li, **Yankai Jiang**, Vijay Gadepally, Devesh Tiwari
2024 The 29th ACL Empirical Methods in Natural Language Processing Conference (EMNLP 2024)

[Paper [🔗](#)] [Slides [🔗](#)] [Open-source Artifact [🔗](#)] [Bibtex [🔗](#)]



LLM Inference Serving: Survey of Recent Advances and Opportunities

Baolin Li, **Yankai Jiang**, Vijay Gadepally, Devesh Tiwari
2024 The 28th IEEE High Performance Extreme Computing Conference (HPEC 2024)

Outstanding Paper Nomination

[Paper [🔗](#)] [Slides [🔗](#)] [Bibtex [🔗](#)]

$$C_{\text{Footprint}} = \left(\frac{E_{\text{DRAM}} * \Delta_{\text{DRAM}}}{L_{\text{DRAM}}} + \frac{E_{\text{CPU}} * \frac{n_f}{N_{\text{CPU}}}}{L_{\text{CPU}}} + \text{Emitted CPU} \right) * T_{\text{exe}} + CI * \left(J_{\text{DRAM}} * \Delta_{\text{DRAM}} + \text{Operational DRAM} \right)$$

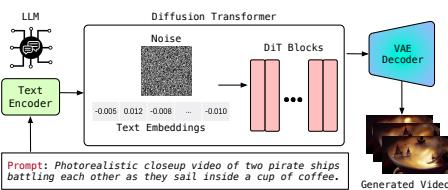
$$\text{Emitted Storage} = \left(\frac{S_f}{S_{\text{ST}}} \right) * T_{\text{exe}} + CI * \left(J_{\text{DRAM}} * \Delta_{\text{DRAM}} + J_{\text{ST}} * \frac{S_f}{S_{\text{ST}}} \right)$$

$$\text{Operational DRAM} = J_{\text{CPU}} * \frac{n_f}{N_{\text{CPU}}} + J_{\text{ST}} * \frac{S_f}{S_{\text{ST}}}$$

The Hidden Carbon Footprint of Serverless Computing

Rohan Basu Roy, Raghavendra Kanakagiri, **Yankai Jiang**, Devesh Tiwari
2024 The 15th ACM Symposium on Cloud Computing (SoCC 2024)

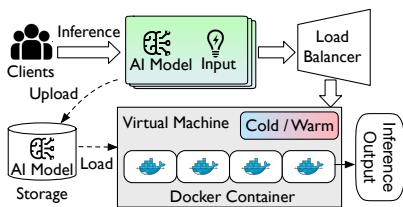
[Paper [🔗](#)] [Slides [🔗](#)] [Bibtex [🔗](#)]



Carbon in Motion: Characterizing Open-Sora on the Sustainability of Generative AI for Video Generation

Baolin Li, **Yankai Jiang**, Devesh Tiwari
2024 The 3rd HotCarbon Workshop on Sustainable Computer Systems (HotCarbon 2024)

[Paper [🔗](#)] [Slides [🔗](#)] [Bibtex [🔗](#)]



Advancing Serverless Computing for Scalable AI Model Inference: Challenges and Opportunities
Li Wang*, Yankai Jiang*, Ningfang Mi
2024 The 10th International Workshop on Serverless Computing (WoSC 2024)
 * indicates equal contribution.
[\[Paper ↗\]](#) [\[Slides ↗\]](#) [\[Bibtex ↗\]](#)

Open-Source Software Artifacts

- **ForgetMeNot:** [https://doi.org/10.5281/zenodo.15123080 ↗](https://doi.org/10.5281/zenodo.15123080)
 ForgetMeNot is the first open-source tool to quantify the forever chemical compounds during the manufacturing process of computer hardware.
- **WaterWise:** [https://doi.org/10.5281/zenodo.14583915 ↗](https://doi.org/10.5281/zenodo.14583915)
 WaterWise is a novel job scheduler that uses mixed integer linear programming (MILP) to co-optimize both carbon and water footprint. WaterWise leverages delay tolerance, soft constraints, and slack management to exploit opportunities across different geographical locations.
- **EcoLife:** [https://doi.org/10.5281/zenodo.11003259 ↗](https://doi.org/10.5281/zenodo.11003259)
 EcoLife is the first carbon-aware serverless function scheduler, EcoLife, builds on the key insight of intelligently exploiting multi-generation hardware to achieve high performance and lower carbon footprint.
- **Sprout:** [https://github.com/boringlee24/EMNLP24_Sprout ↗](https://github.com/boringlee24/EMNLP24_Sprout)
 Sprout is a framework that leverages generation directives to guide the autoregressive generation process, achieving a balance between ecological sustainability and high-quality outputs.

Teaching Experience

CS340 Introduction to Computer Networks, Northwestern University
 Peer Mentor, grading homework and hosting office hours.

Evanston, IL
Winter 2023, Fall 2022

Talks and Presentations

- SIGMETRICS 2025 Conference**, Stony Brook, NY *June 2025*
 ○ Presented: ForgetMeNot: Modeling and Analyzing the Impact of Forever Chemicals in Designing Sustainable Computing Systems.
- PPoPP 2025 Conference**, Las Vegas, NV *March 2025*
 ○ Presented: WaterWise: Co-optimizing Carbon- and Water-Footprint Toward Environmentally Sustainable Cloud Computing.
- SC 2024 Conference**, Atlanta, GA *November 2024*
 ○ Presented: EcoLife: Carbon-Aware Serverless Function Scheduling for Sustainable Computing
- Green AI Summit**, Cambridge, MA *October 2024*
 ○ Presented: Carbon in Motion: Characterizing Open-Sora on the Sustainability of Generative AI for Video Generation.
- HotCarbon 2024 Workshop**, Santa Cruz, CA *July 2024*
 ○ Presented: Carbon in Motion: Characterizing Open-Sora on the Sustainability of Generative AI for Video Generation.

Professional Service

- **Conference Artifact Evaluation Committee Membership**
USENIX Conference on File and Storage Technologies (FAST), 2026.
- **Conference Artifact Evaluation Committee Membership**
IEEE/ACM International Symposium on Microarchitecture (MICRO), 2025.
- **Conference Artifact Evaluation Committee Membership**
IEEE International Symposium on Computer Architecture (ISCA), 2025.
- **Conference Review Program Committee (PC) Membership**
International Joint Conference on Artificial Intelligence (IJCAI) - Demo Track, 2025.
- **Workshop Review Program Committee (PC) Membership**
ACL Student Research Workshop, 2025.