

# Saksham Singh Kushwaha

Website: <https://sakshamsingh1.github.io/>

Email : [sxk230060@utdallas.edu](mailto:sxk230060@utdallas.edu)

Mobile : +1-917-443-3706

## EDUCATION

---

- **The University of Texas at Dallas** May 2027(expected)  
*Ph.D. - Computer Science(Audio and Multimodal machine learning)* GPA: 4.0/4.0
- **New York University, Courant** May 2023  
*M.S. - Computer Science* GPA: 3.95/4.0
- **Indian Institute of Technology, Delhi** May 2018  
*B.Tech - Mathematics and Computing* GPA: 7.97/10.0

## EXPERIENCE

---

- **Dolby Laboratories | PhD Research Intern** May 2024 - Aug 2024  
*Worked in Advanced Technology Group(ATG) to develop approaches for spatial audio generation* SF, USA
  - Proposed a novel task and a flow-matching generative approach for ambisonics generation from scratch.
  - Outperformed traditional baselines on objective (FAD~ 72%) and subjective metrics (Fidelity~60%).
- **Nvidia | Deep Learning Intern** May 2022 - Aug 2022  
*Worked in Product Security team to efficiently detect anomalous user-behavior in AWS accounts* Remote, USA
  - Developed and implemented a **multi-task autoencoder** that replaced up to 10 production models.
  - Improved existing intrusion detection system by 65% (F-score) and reduced false positives by 50%
- **Zomato | Machine Learning Engineer II** July 2018 - April 2021  
*Part of Search, User personalization, and Logistics teams* Gurgaon, India
  - Improved auto-suggestion search by 10% avg. rank, 4% CTR & 2% OTR using **point-wise ML** model
  - Created DQN **RL** based rider dispatch service, improving next order time(3min) & order probability(9%)
  - Developed similar restaurants service by creating restaurants' embedding using modified **Word2Vec**

## PUBLICATIONS

---

- Audio-Visual dataset distillation (paper), (slides)  
*Saksham S. Kushwaha, Siva Vasireddy, Kai Wang, Yapeng Tian [CVPR Sight and Sound Workshop, 2024]*
- Diff-SAGe: end-to-end **spatial audio generation** using diffusion models (project page)  
*Saksham S. Kushwaha, Jianbo Ma, Mark Thomas, Yapeng Tian, Avery Bruni [Under review]*
- A **multimodal prototypical** approach for **unsupervised** sound classification (paper)  
*Saksham S. Kushwaha, Magdalena Feuentes [INTERSPEECH 2023]*
- **Sound source distance estimation** in diverse and dynamic acoustic conditions (paper)  
*Saksham S. Kushwaha, Iran Roman, Magdalena Feuentes, Juan Pablo Bello [WASPAA 2023]*
- Analyzing the effect of equal-angle spatial discretization on **sound event localization & detection** (paper)  
*Saksham S. Kushwaha, Iran R. Roman, Juan Pablo Bello [DCASE Workshop 2022]*

## ACADEMIC EXPERIENCE

---

- **Invited talks:** DL-MIR Workshop, Stanford University (slides)
- **Reviewer:** MLSP 2023, AAAI-24, ICASSP-24, NIPS-24, ACMMM-24
- **Research Assistant:** Prof. Magdalena Fuentes (multimodal deep learning) [Aug'22-May'23], Prof. Raveesh Mayya (ML for digital policy change) [Aug'21-July'22], Prof. Yapeng Tian [Aug'24-Present]
- **Teaching Assistant:** Intro to python programming (Prof. Junpei Komiyama) [Jan'22-May'22], Discrete Mathematics (Prof. Simeon Ntafos) [May'23-Aug'23], Computer Science I (Prof. Scott Dollinger) [May'23-Aug'23]