

Vasisht Duddu

• ✉ vasisht.duddu@uwaterloo.ca • 🌐 <https://vasishtduddu.github.io> •

RESEARCH INTERESTS

My goal is to **enhance trust in machine learning systems** by

- systematic evaluation of risks to security, privacy, fairness, and transparency; and designing practical defenses
- designing technical mechanisms for accountability (e.g., regulatory compliance, safety, reliability)

Additional Links: 📄 → Paper | 🏆 → Certificate | 📄 → Poster | 🏆 → Award | 📰 → News | 📁 → Code

EDUCATION

UNIVERSITY OF WATERLOO

SEP'22-

DOCTOR OF PHILOSOPHY (PH.D.), COMPUTER SCIENCE | SUPERVISOR: N. ASOKAN (SECURE SYSTEMS GROUP)

ONTARIO, CANADA

- Research: Trust and Accountability in Machine Learning (tentative)
 - 🏆 IBM PhD Fellowship (2024) 📄
 - 🏆 Distinguished Paper Award @ IEEE Symposium on Security and Privacy (2024) 📄
 - 🏆 Mastercard Cybersecurity and Privacy Excellence Graduate Scholarship (2024) 📄 🏆
 - 🏆 David R. Cheriton Graduate Scholarship (2024-2026)
- GPA: 92/100

UNIVERSITY OF WATERLOO

SEP'20-AUG'22

MASTER OF MATHEMATICS (MMATH), COMPUTER SCIENCE | SUPERVISOR: N. ASOKAN (SECURE SYSTEMS GROUP)

ONTARIO, CANADA

- Research: *Towards Effective Measurement of Membership Privacy Risk for Machine Learning Models*
 - Thesis Committee: Florian Kerschbaum, Xi He
 - Technical Report: “SHAPr: An Efficient and Versatile Membership Privacy Risk Scores for Machine Learning Models” 📄
 - 🏆 International Master’s Award of Excellence (2020-22)
- GPA: 90.5/100

INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY (IIIT)

AUG'15-DEC'19

BACHELOR OF TECHNOLOGY (B.TECH.), ELECTRONICS AND COMMUNICATION ENGINEERING

DELHI, INDIA

- Research: *Fault Tolerant Neural Networks in Adversarial and Benign Settings* (Collaboration: Prof. Valentina E. Balas)
 - Research Paper: “Towards Enhancing Fault Tolerance of Deep Neural Networks” [C13]
 - Research Paper: “Fault Tolerance of Neural Networks in Adversarial Settings” [C12]
 - 🏆 Dean’s Award for Innovation Research & Development (2019)



RESEARCH INTERNSHIPS


- **INRIA Privatics Lab** **France**
Research Affiliate. *Privacy in Machine Learning* (Mentor: Prof. Antoine Boutet) 2020-22
- **National University of Singapore** **Singapore**
Research Intern. *Efficient Privacy Preserving Deep Learning* (Mentor: Prof. Reza Shokri) May-Dec'18
- **Indian Institute of Technology, Kharagpur** **India**
Research Intern. *Machine Learning Model Extraction Attacks* (Mentor: Prof. Debasis Samanta) May-Jul'17



INVITED TALKS

- **SoK: Unintended Interactions among Machine Learning Defenses and Risks.**
 - *University of Toronto*, hosted by Nicolas Papernot **July'24**
- **SHAPr: An Efficient and Versatile Membership privacy Metric for Machine Learning**
 - *Huawei*, hosted by Qiongxu Li **Nov'22**

2025

C1 Laminator: Verifiable ML Property Cards using Hardware-assisted Attestations.  



Vasisht Duddu, Oskari Järvinen, Lachlan J. Gunn, N. Asokan
[ACM Conference on Data and Application Security and Privacy \(CODASPY\).](#)
[Poster@IEEE Symposium on Security and Privacy \(S&P\). 2024.](#) 

C2 Espresso: Robust Concept Filtering in Text-to-Image Models.  

Anudeep Das, *Vasisht Duddu*, Rui Zhang, N. Asokan
[ACM Conference on Data and Application Security and Privacy \(CODASPY\).](#)

2024

C3 SoK: Unintended Interactions among Machine Learning Defenses and Risks  

Vasisht Duddu, Sebastian Szyller, N. Asokan
[IEEE Symposium on Security and Privacy \(S&P\)](#) [ **Distinguished Paper Award**] 

C4 GrOVe: Ownership Verification of Graph Neural Networks using Embeddings  

Asim Waheed, *Vasisht Duddu*, N. Asokan
[IEEE Symposium on Security and Privacy \(S&P\)](#)

C5 Attesting Distributional Properties of Training Data  

Vasisht Duddu, Anudeep Das, Nora Khayata, Hossein Yalame, Thomas Schneider, N. Asokan
[European Symposium on Research in Computer Security \(ESORICS\)](#)

C6 On the Alignment of Group Fairness and Attribute Privacy 

Jan Aalmoes, *Vasisht Duddu*, Antoine Boutet
[International Web Information Systems Engineering Conference \(WISE\)](#)

2023

C7 Comprehension from Chaos: What Users Understand and Expect from Private Computation  

Bailey Kacsmar, *Vasisht Duddu*, Kyle Tilbury, Blase Ur, Florian Kerschbaum
[ACM Conference on Computer and Communications Security \(CCS\)](#)

2022

C8 Inferring Sensitive Attributes from Model Explanations  

Vasisht Duddu, Antoine Boutet
[ACM International Conference on Information and Knowledge Management \(CIKM\)](#)

C9 Towards Privacy Aware Deep Learning for Embedded Systems  

Vasisht Duddu, Antoine Boutet, Virat Shejwalkar
[ACM Symposium On Applied Computing \(SAC\)](#)
[NeurIPS Workshop on Privacy Preserving Machine Learning - PriML and PPML Joint Edition \(2020\)](#)

2021

C10 Good Artists Copy, Great Artists Steal: Model Extraction Attacks Against Image Translation GANs 

Sebastian Szyller, *Vasisht Duddu*, Tommi Gröndahl, N. Asokan
[Technical Report \(ArXiv\)](#)



2020



C11 Quantifying Privacy Leakage in Graph Embedding  

Vasisht Duddu, Antoine Boutet, Virat Shejwalkar
[EAI International Conference on Mobile and Ubiquitous Systems \(MobiQuitous\)](#)
[NeurIPS Workshop on Privacy Preserving Machine Learning - PriML and PPML Joint Edition](#)


C12 Fault Tolerance of Neural Networks in Adversarial Settings  

Vasisht Duddu, Rajesh Pillai, D. Vijay Rao, Valentina E. Balas
[Journal of Intelligent and Fuzzy Systems](#)


C13 Towards Enhancing Fault Tolerance in Neural Networks  
Vasisht Duddu, D. Vijay Rao, Valentina E. Balas
EAI International Conference on Mobile and Ubiquitous Systems (MobiQuitous)


C14 Quantifying (Hyper) Parameter Leakage in Machine Learning  
Vasisht Duddu, D. Vijay Rao
IEEE International Conference on Multimedia Big Data (BigMM)



2018

C15 Stealing Neural Networks via Timing Side Channels 
Vasisht Duddu, Debasis Samanta, D. Vijay Rao, Valentina E. Balas (2018)
Accepted to International Conference on Privacy, Security, and Trust (PST). 2019. [Withdrew (No funding)]
AIVillage @ DEFCON 27, Las Vegas, NV, USA. 2019.


PRE-PRINTS

P1 Position: Contextual Integrity Washing for Language Models 
Yan Shvartzshnaider, *Vasisht Duddu*
Under review.




P2 Combining Machine Learning Defenses without Conflicts. 
Vasisht Duddu, Rui Zhang, N. Asokan
Under review.

P3 Investigating Privacy Bias in Training Data of Language Models. 
Yan Shvartzshnaider, *Vasisht Duddu*
Under review.
AAAI Workshop on Privacy-Preserving Artificial Intelligence (AAAI-PPAI). 2025.  **Oral Presentation**

SOFTWARE FRAMEWORK

S1 Amulet: A Library for Evaluating Interactions among Machine Learning Defenses and Risks 
Supported by Intel for technology transfer. 2024.
Contributors: *Vasisht Duddu*, Rui Zhang, Asim Waheed
Maintainers: Asim Waheed, Sebastian Szlyler (Intel)

MENTORING

-
- **Anudeep Das** (MMath Thesis), *Espresso: Robust Concept Filtering in Text-to-Image Models*  **2024**
 - **Erin Li** (URA), *Quantifying Training Data Copying in Graph Generative Models* **Fall'23**
 - **Asim Waheed** (MMath Thesis), *On Using Embeddings for Ownership Verification of Graph Neural Networks*  **2021-23**
 - **Anudeep Das** (URA+URF), *Attesting Distributional Properties of Machine Learning Training Data*  **Fall'22**