

MUHAMMAD ALI GULZAR

220 Gilbert St. Room 4106
Blacksburg, VA, 24061

Assistant Professor
Department of Computer Science
Virginia Tech

gulzar@cs.vt.edu
(540) 231 0851

1 Research Interests

My research vision is to build systems that improve developer productivity through automated debugging and testing for emerging applications. My research delivers integral insights to transfer traditional software engineering techniques to the next generation of software in emerging domains such as ML/AI, computational/data science, web, and big data.

2 Education

Ph.D.	University of California, Los Angeles Computer Science <i>Interactive and Automated Debugging for Big Data Analytics</i>	2014 - 2020
B.S.	Lahore University of Management Sciences Computer Science	2010 - 2014

3 Employment History

08/2020 - Current	Assistant Professor , Computer Science, Virginia Tech
03/2024 - Current	Visiting Academic , Amazon Web Services
06/2019 - 09/2019	Software Engineering Intern , Google, Inc.
06/2018 - 09/2018	Software Engineering Tools and Infrastructure Intern , Google, Inc.
06/2016 - 09/2016	Summer Research Assistant , NEC Labs America, Princeton, NJ.

4 Awards and Grants

• Commonwealth Cybersecurity Grant on Accessible Web, Virginia Tech	2024
• 4-VA Grant on Securing Software with Forensic Logging, Virginia Tech	2024
• Rising Start Faculty Award, Computer Science Department, Virginia Tech	2022
• NSF Medium:Reinventing Fuzz Testing for Data and Compute Intensive Systems Total. \$900K. Sole PI at VT. VT's Share: \$323,982	2021 - 2025
• Indonesia-US Research Collaboration in Open Digital Technology. Total. \$60K	2022 - 2025
• Outstanding Computer Science Graduating Ph.D. Student	2020
• Northrop Grumman Outstanding Computer Science Graduate Student Research Award	2020
• NSF i-Corps, Amount: \$50K, Role: Entrepreneurial Lead	2018
• Google Ph.D. Fellowship 2017-20 (with 3rd-year extension)	2017 - 2020
• Gold medal at ACM Student Research Competition at ICSE'18	2018
• Best Paper Award, VLDB 2016.	2016

5 Professional Activity

Program Committee

IEEE/ACM International Conference on Software Engineering (ICSE '25) 2025.
ACM Foundations of Software Engineering (ESEC/FSE '25) 2025
IEEE/ACM International Conference on Automated Software Engineering (ASE '24) 2024.

ACM Foundations of Software Engineering (ESEC/FSE '24) 2024
 IEEE/ACM International Conference on Software Engineering (ICSE '24) 2024.
 IEEE/ACM International Conference on AI Engineering – SE for AI (CAIN'24) 2024.
 ACM Symposium of Cloud Computing (SoCC '23) 2023
 IEEE/ACM International Conference on AI Engineering – SE for AI (CAIN'23) 2023.
 ACM Foundations of Software Engineering, Demo (ESEC/FSE '23) 2023
 ACM Foundations of Software Engineering, Artifact Evaluation (ESEC/FSE '23) 2023
 ACM Foundations of Software Engineering (ESEC/FSE '22) 2022
 ACM Foundations of Software Engineering Demonstration Track (ESEC/FSE '22) 2022
 International Workshop on Testing Database Systems (DBTest '22) co-located with SIGMOD 2022
 International Conference on AI Engineering – SE for AI (CAIN'22)2022
 ACM Symposium of Cloud Computing (SoCC '22) 2022
 The Theory and Practice of Provenance (TaPP'21) 2021
 ACM Symposium of Cloud Computing (SoCC '21) 2021

Grant Proposal Panels

NSF 2024
 The Department of Energy (DOE) (2024) NSF 2022
 The Department of Energy (DOE) (2021)
 Virginia Tech Presidential Postdoctoral Fellowships

Journal Reviewer

IEEE Transaction on Software Engineering, TSE (since 2018)
 ACM Transactions on Software Engineering and Methodology, TOSEM (since 2019)
 IEEE Transactions on Dependable and Secure Computing, TDSC (since 2021)

Departmental Service

Graduate Program Committee, 2023-24
 Department-wide Ph.D. Qualifier Exam Organizer, Spring 2024
 Software Engineering Ph.D. Qualifier Exam Chair, Spring 2024
 Faculty Search Committee, 2022-23
 Software Engineering Ph.D. Qualifier Exam Chair, Spring 2023
 Graduate Admissions Committee, 2020-21

Outreach Service

Student Transition Engineering Program (STEP), Mentor, Summer 2023
 TechGirls, Invited Speaker, Summer 2023
 C-Tech² - Computers and Technology at Virginia Tech, Invited Speaker. Summer 2022
 Black Engineering Excellence at Virginia Tech (BEE VT), Invited Speaker. Summer 2022

6 Teaching Activity

CS 3304 Comparative Languages, Computer Science, Virginia Tech	Spring '24.
CS 6704 Advacned Topics in Software Engineering, Computer Science, Virginia Tech	Fall '23.
CS 5614 Big Data Engineering, Computer Science, Virginia Tech	Spring '23.
CS 3304 Comparative Languages, Computer Science, Virginia Tech	Fall '22.
CS 4974 Independent Study, Computer Science, Virginia Tech	Spring '22.
CS 3304 Comparative Languages, Computer Science, Virginia Tech	Spring '22.
CS 6704 Advanced Topic on Software Engineering, Computer Science, Virginia Tech	Fall '21.
CS 5614 Big Data Engineering, Computer Science, Virginia Tech	Spring '21.
CS 130 Software Engineering, TA, Computer Science, UCLA	Spring '17 and Fall '15.

7 Selected Publications

7.1 Research Track Publications

- Blocking Tracking JavaScript at the Function Granularity. Abdul Haddi Amjad, Shaour Munir, Zubair Shafiq, and Muhammad Ali Gulzar. The 31st ACM Conference on Computer and Communications Security. **CCS 2024**.
- Human-in-the-Loop Synthetic Text Data Inspection with Provenance Tracking. Hong Jin Kang, Fabrice Harel-Canada, Muhammad Ali Gulzar, Nanyun Peng, and Miryung Kim. Annual Conference of the North American Chapter of the Association for Computational Linguistics 2024. **NAACL 2024**.
- DeSQL: Interactive Debugging of SQL in Data-Intensive Scalable Computing. Sabaat Haroon, Chris Brown, and Muhammad Ali Gulzar. The ACM International Conference on the Foundations of Software Engineering. **FSE 2024**.
- Co-Dependence Aware Fuzzing for Dataflow-based Big Data Analytics. Ahmad Humayun, Miryung Kim, and Muhammad Ali Gulzar. ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. **FSE 2023**.
- NaturalFuzz: Natural Input Generation for Big Data Analytics. Ahmad Humayun, Yaoxuan Wu, Miryung Kim, and Muhammad Ali Gulzar. The 38th IEEE/ACM International Conference on Automated Software Engineering. **ASE 2023**.
- FedDebug: Systematic Debugging for Federated Learning Applications. Waris Gill, Ali Anwar, and Muhammad Ali Gulzar. The ACM/IEEE 45th International Conference on Software Engineering. **ICSE 2023**.
- Blocking JavaScript without Breaking the Web: An Empirical Investigation Abdul Haddi Amjad, Zubair Shafiq, and Muhammad Ali Gulzar. The 23rd Privacy Enhancing Technologies Symposium. **PETS 2023**
- Detecting Build Conflicts in Software Merge for Java Programs via Static Analysis. Sheikh Towqir, Bowen Shen, Muhammad Ali Gulzar, and Na Meng. The 37th IEEE/ACM International Conference on Automated Software Engineering. **ASE 2022**.
- A Characterization Study of Merge Conflicts in Java Projects. Bowen Shen, Muhammad Ali Gulzar, Fei He, Na Meng. The Transactions on Software Engineering and Methodology. **TOSEM 2022**.
- Sibylvariant Transformations for Robust Text Classification. Fabrice Harel-Canada, Muhammad Ali Gulzar, Nanyun Peng, Miryung Kim. The Proceedings of the Conference of the 60th Annual Meeting of the Association for Computational Linguistics **ACL Findings 2022**.
- TrackerSift: Untangling Mixed Tracking and Functional Web Resources. Abdul Haddi Amjad, Danial Saleem, Muhammad Ali Gulzar, Zubair Shafiq, Fareed Zaffar. The proceedings of the 21st ACM Internet Measurement Conference. **ACM IMC 2021**.
- OptDebug: Fault-Inducing Operation Isolation for Dataflow Applications. Muhammad Ali Gulzar, Miryung Kim. The proceedings of the 12th ACM Symposium on Cloud Computing. **SoCC 2021**.
- Is neuron coverage a meaningful measure for testing deep neural networks?. Fabrice Harel-Canada, Lingxiao Wang, Muhammad Ali Gulzar, Quanquan Gu, Miryung Kim. The proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. **ESEC/FSE 2020**
- BigFuzz: efficient fuzz testing for data analytics using framework abstraction. Qian Zhang, Jiyuan Wang, Muhammad Ali Gulzar, Rohan Padhye, Miryung Kim. The proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering. **ASE 2020**
- Influence-based provenance for dataflow applications with taint propagation. Jason Teoh, Muhammad Ali Gulzar, and Miryung Kim. The proceedings of the 11th ACM Symposium on Cloud Computing. **SoCC 2019**
- HeteroRefactor: Refactoring for Heterogeneous Computing with FPGA. Jason Lau*, Aishwarya Sivaraman*, Qian Zhang*, Muhammad Ali Gulzar Jason Cong, Miryung Kim. The proceedings of the 42nd International Conference on Software Engineering. **ICSE 2020**
- White-Box Testing of Big Data Analytics with Complex User-Defined Functions. Muhammad Ali Gulzar, Shaghayegh Mardani, Madan Musuvathi, Miryung Kim. The proceedings of the

2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. **ESEC/FSE 2019**.

- PerfDebug: Performance Debugging of Computation Skew in Dataflow Systems. Jason Teoh, Muhammad Ali Gulzar, Harry Xu, Miryung Kim. The proceedings of the ACM Symposium on Cloud Computing. **SoCC 2019**
- LogLens: A Real-Time Log Analysis System. Biplob Debnath, Mohiuddin Solaimani, Muhammad Ali Gulzar, Nipun Arora, Cristian Lumezanu, Jianwu Xu, Bo Zong, Hui Zhang, Guofei Jiang, Latifur Khan. 2018 IEEE 38th International Conference on Distributed Computing Systems. **ICDCS 2018**.
- Adding data provenance support to Apache Spark. Matteo Interlandi, Ari Ekmekji, Kshitij Shah, Muhammad Ali Gulzar, Sai Tetali, Miryung Kim, Todd Millstein, Tyson Condie. The VLDB Journal — The International Journal on Very Large Data Bases. **VLDB Journal 2018**.
- Automated Debugging in Data-Intensive Scalable Computing. Muhammad Ali Gulzar, Matteo Interlandi, Xueyuan Han, Mingda Li, Tyson Condie, Miryung Kim. The proceedings of the 2017 Symposium on Cloud Computing. **SoCC 2017**.
- BigDebug: Debugging Primitives for Interactive Big Data Processing in Spark. Muhammad Ali Gulzar, Matteo Interlandi, Seunghyun Yoo, Sai Tetali, Tyson Condie, Todd Millstein, Miryung Kim. The proceedings of the 38th International Conference on Software Engineering. **ICSE 2016**.
- Optimizing Interactive Development of Data-Intensive Applications. Matteo Interlandi, Sai Tetali, Muhammad Ali Gulzar, Joseph Noor, Tyson Condie, Miryung Kim, Todd Millstein. The proceedings of the Seventh ACM Symposium on Cloud Computing. **SoCC 2016**.
- Titian: Data Provenance Support in Spark. Matteo Interlandi, Kshitij Shah, Sai Tetali, Muhammad Ali Gulzar, Seunghyun Yoo, Miryung Kim, Todd Millstein, Tyson Condie. The proceedings of the VLDB Endowment, Volume 9, Issue 3. **VLDB 2016**.

7.2 Workshops









- FedDefender: Backdoor Attack Defense in Federated Learning. Waris Gill, Ali Anwar, Muhammad Ali Gulzar. The 1st International Workshop on Dependability and Trustworthiness of Safety-Critical Systems with Machine Learned Components. **SE4SafeML 2023**
- Towards a Serverless Bioinformatics Cyberinfrastructure Pipeline. Shunyu David Yao, Muhammad Ali Gulzar, Liqing Zhang, and Ali R. Butt. In Proceedings of the 1st Workshop on High-Performance Serverless Computing. **HiPS 2021**.
- Perception and Practices of Differential Testing. Muhammad Ali Gulzar, Yongkang Zhu, and Xiaofeng Han. In Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice. **ICSE SEIP 2019**.
- Interactive Debugging for Big Data Analytics. Muhammad Ali Gulzar, Xueyuan Han, Matteo Interlandi, Shaghayegh Mardani, Sai Deep Tetali, Todd Millstein, and Miryung Kim. In 8th USENIX Workshop on Hot Topics in Cloud Computing. **HotCloud 16**.

7.3 Demonstration Track

- BigTest: Symbolic Execution Based Systematic Test Generation for Apache Spark. Muhammad Ali Gulzar, Madan Musuvathi, Miryung Kim. The 42nd International Conference on Software Engineering. Research Demonstration Track. **ICSE Demo 2020**.
- BigSift: Automated Debugging of Big Data Analytics in Data-intensive Scalable Computing. Muhammad Ali Gulzar, Siman Wang, Miryung Kim. The 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Research Demonstration Track. **ESEC/FSE Demo 2018**.
- Debugging Big Data Analytics in Spark with BigDebug Muhammad Ali Gulzar, Matteo Interlandi, Tyson Condie, Miryung Kim. The Proceedings of The 2017 ACM SIGMOD/PODS Conference. Demonstration Track. **SIGMOD Demo 2017**.
- BigDebug: Interactive Debugger for Big Data Analytics in Apache Spark Muhammad Ali Gulzar, Matteo Interlandi, Tyson Condie, Miryung Kim. The proceedings of the 24th Joint

European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Research Demonstration Track. **ESEC/FSE Demo 2016.**

7.4 Software Artifacts

- NaturalFuzz. <https://github.com/SEED-VT/NaturalFuzz>
- DepFuzz.  <https://github.com/SEED-VT/DepFuzz>
- FedDebug.   <https://github.com/SEED-VT/FedDebug>
- Blocking JavaScript without Breaking the Web.  
<https://github.com/hadiamjad/Blocking-JavaScript-without-Breaking-the-Web>
- FedDefender. <https://github.com/warisgill/FedDefender>
- OptDebug. <https://github.com/maligulzar/OptDebug>
- BigFuzz. <https://github.com/qianzhanghk/BigFuzz>
- Is Neuron Coverage a Meaningful Measure for Testing Deep Neural Networks?  
<https://zenodo.org/record/4021473#.X8J8ohNKgTU>
- HeteroRefactor.  <https://github.com/heterorefactor/heterorefactor>
- BigTest. <https://github.com/maligulzar/BigTest>
- BigSift. <https://github.com/maligulzar/BigSift-Zepelin>
- BigDebug. <https://github.com/maligulzar/bigdebug>
- Titian. <https://github.com/maligulzar/bigdebug>

7.5 Patent

- Systems and Methods with a Realtime Log Analysis Framework. Biplob Debnath, Nipun Arora, Hui Zhang, Guofei Jiang, Mohiuddin Solaimani, Muhammad Ali Gulzar. US Patent App. 15/784,393. 2018.

8 Student Advising

Current Students

- Abdul Hadi Amjad, Ph.D. Student at VT. Fall 2021 - Present
Internships at Brave, Inc in Summer '22 and '23.
Talks at ACM IMC '21, PETS '23, AdSummit '21, '22, and '23, and Google.
IEEE S&P and PETS 2023 Travel Award
- Waris Gill, Ph.D. Student at VT. Fall 2021 - Present
Internships at Cisco, Inc in Summer '23 and Fall '23.
Talks at ICSE '23 and SE4SafeML '23.
NSF Travel Award to ICSE '23. Student Volunteer at FSE 2023 and ICSE 2022.
- Ahmad Humayun, Ph.D. Student at VT. Fall 2021 - Present
Talks at ASE '23 and FSE '23.
Travel Award to ASE '23. Student Volunteer at ICSE 2022.
- Sabaat Haroon, Ph.D. Student at VT. Fall 2022 - Present
- Tien K. Nguyen, Ph.D. Student at VT. Fall 2023 - Present

Graduated Students

- Satish Venkatesan, VT. Graduated March 2021
MS Thesis: Automatic Restoration and Management of Computational Notebooks, First
Job: Salesforce.

Remote Research Students

- Billy Tobin, Undergraduate Student at Washington and Lee. Spring 2022
- Mohamed Elhussiny, Undergraduate Student at Washington and Lee. Spring and Fall 2023
- Agista Hidayatillah Suparno Putri, Undergraduate Student. Telkom University. 2023-24
- Jabalnur Jabalnur, Undergraduate Student. Institut Teknologi Sepuluh Nopember 2023-24
- Muhammad Danish, Undergraduate Student at ITU. Fall 2023 - Present
- Gillian Doherty, High-School Outreach via TechGirls. Fall 2022 - Present