

Philipp Leitner

Publication List

TOP 10 PUBLICATIONS

1. R. Khojah, M. Mohamad, **P. Leitner**, and F. Gomes: *Beyond Code Generation: An Observational Study of ChatGPT Usage in Software Engineering Practice*, in Proceedings of ACM FSE, 2024.
One of the first large-scale observational studies of how software practitioners actually use LLM-based tools in their daily work. The findings reveal how engineers interact with AI systems in realistic settings—including failure modes, workarounds, and trust calibration—providing empirical grounding for designing more effective and trustworthy AI-based development tools. The study methodology also serves as a template for evaluating AI systems in situ.
2. R. Khojah, F. Gomes, M. Mohamad, and **P. Leitner**: *The Impact of Prompt Programming on Function-Level Code Generation*, IEEE Transactions on Software Engineering (TSE), 2025.
Systematically investigates how prompt design choices affect the quality, correctness, and reliability of LLM-generated code at the function level. The controlled experimental design isolates the effect of different prompt engineering strategies on code generation outcomes. The results provide actionable engineering guidelines for building reliable AI-based code generation systems and integrating them into software development pipelines with predictable quality.
3. R. Khojah, M. Mohamad, L. Erlenhov, F. Gomes, and **P. Leitner**: *LLM Company Policies and Policy Implications in Software Organizations*, IEEE Software, 2025.
Investigates how software organizations govern the adoption and use of LLM-based tools through interviews and document analysis, revealing significant policy gaps and organizational challenges. The findings are directly relevant to engineering responsible AI deployment in industry settings, identifying the need for systematic governance frameworks that address intellectual property, code quality, and security concerns when integrating AI systems into engineering workflows.
4. H. Zhang, M. Alhanahnah, F. Ahmed, D. Fatih, **P. Leitner**, and A. Ali-Eldin: *Machine Learning Systems are Bloaty and Vulnerable*, in Proceedings of ACM SIGMETRICS, 2024.
The first systematic study demonstrating that production ML systems contain significant software bloat, which increases attack surface, resource consumption, and deployment complexity. Through analysis of popular ML frameworks and their dependency chains, the paper exposes fundamental engineering quality challenges in how AI/ML systems are built and deployed today, and quantifies the concrete security and efficiency costs of current ML engineering practices.
5. H. Zhang, M. Alhanahnah, **P. Leitner**, and A. Ali-Eldin: *BLAFS: A Bloat-Aware Container File System*, in Proceedings of ACM SoCC, 2025.
Presents a novel container file system that transparently reduces software bloat in containerized deployments without requiring changes to application code or container images. Given that AI/ML systems are predominantly deployed as containers and suffer from particularly severe bloat (as shown in our SIGMETRICS paper), BLAFS directly addresses the infrastructure challenges of operating AI systems efficiently and securely at scale, reducing image sizes and attack surface.
6. M. Jangali, Y. Tang, N. Alexandersson, **P. Leitner**, J. Yang, and W. Shang: *Automated Generation and Evaluation of JMH Microbenchmark Suites from Unit Tests*, IEEE Transactions on Software Engineering (TSE), 49(4), 2023.
Presents an automated approach to generate performance benchmarks from existing test suites, enabling continuous performance regression detection without manual benchmark authoring effort. This performance testing methodology is essential for engineering AI systems, where model serving and inference pipelines must meet strict latency and throughput requirements, and where

performance regressions introduced by framework updates or model changes need to be caught early in the development cycle.

7. L. Elenhov, F. Gomes, and **P. Leitner**: *An Empirical Study of Bots in Software Development — Characteristics and Challenges from a Practitioner's Perspective*, in Proceedings of ACM ESEC/FSE, 2020.

The first large-scale practitioner study of automated bot usage in software development, characterizing how developers perceive, trust, and interact with autonomous agents in their workflows. The findings on developer trust, notification fatigue, and the balance between automation and human control provide foundational insights for engineering AI-powered development assistants and understanding the socio-technical dynamics of human-AI collaboration in software teams.

8. J. Cito, **P. Leitner**, M. Rinard, and H. C. Gall: *Interactive Production Performance Feedback in the IDE*, in Proceedings of ACM/IEEE ICSE, 2019.

Pioneers data-driven developer tooling that integrates production runtime performance data directly into the development environment, enabling developers to reason about real-world system behavior while writing code. The approach establishes engineering patterns for building intelligent, context-aware developer tools that bridge the gap between development and production—a paradigm now central to AI-powered IDEs and coding assistants that leverage runtime context for recommendations.

9. C. Laaber, J. Scheuner, and **P. Leitner**: *Software Microbenchmarking in the Cloud. How Bad is it Really?*, Empirical Software Engineering (EMSE), 24(4), 2019.

Comprehensively quantifies how cloud performance variability affects the reliability and reproducibility of software benchmarking results, identifying which benchmark configurations yield stable measurements and which do not. Since AI model training and inference predominantly run on cloud infrastructure, these findings are essential for engineering trustworthy performance evaluation pipelines for AI systems—ensuring that reported model latency, throughput, and cost metrics are meaningful and reproducible.

10. **P. Leitner** and J. Cito: *Patterns in the Chaos — a Study of Performance Variation and Predictability in Public IaaS Clouds*, ACM Transactions on Internet Technology (TOIT), 16(3), 2016.

Provides a systematic analysis of performance predictability in public cloud environments through long-running empirical measurements, identifying patterns, sources of variation, and predictive models for infrastructure performance. As AI systems increasingly depend on cloud infrastructure for both training and serving workloads, understanding and predicting infrastructure performance variability is a prerequisite for engineering AI systems with reliable quality-of-service guarantees and accurate cost estimation.

ALL RELEVANT PUBLICATIONS 2018–2026

Peer-reviewed original articles (journals)

- S. Talluri, S. Eismann, J. Scheuner, A. Iosup, **P. Leitner**, C. Abad, and E. van Eyk: *Let's Trace It: Fine-Grained Serverless Benchmarking for Synchronous and Asynchronous Applications*, Future Generation Computer Systems (FGCS), 2025.
- R. Khojah, M. Mohamad, L. Elenhov, F. Gomes, and **P. Leitner**: *LLM Company Policies and Policy Implications in Software Organizations*, IEEE Software, 2025.
- R. Khojah, F. Gomes, M. Mohamad, and **P. Leitner**: *The Impact of Prompt Programming on Function-Level Code Generation*, IEEE Transactions on Software Engineering (TSE), 2025.
- P. Samoa, L. Aronsson, A. Longa, **P. Leitner**, and M. H. Chehreghani: *A Unified Active Learning Framework for Annotating Graph Data for Regression Tasks*, Engineering Applications of Artificial Intelligence (EAAI), 138, 2024.

- H. M. Ayas, R. Hebig, and **P. Leitner**: *An Empirical Investigation on the Competences and Roles of Practitioners in Microservices-Based Architectures*, Journal of Systems and Software (JSS), 213, 2024.
- H. M. Ayas, **P. Leitner**, and R. Hebig: *An Empirical Study of the Systemic and Technical Migration Towards Microservices*, Empirical Software Engineering (EMSE), 28(4), 2023.
- M. Grambow, D. Kovalev, C. Laaber, **P. Leitner**, and D. Bermbach: *Using Microbenchmark Suites to Detect Application Performance Changes*, IEEE Transactions on Cloud Computing (TCC), 11(3), 2023.
- M. Jangali, Y. Tang, N. Alexandersson, **P. Leitner**, J. Yang, and W. Shang: *Automated Generation and Evaluation of JMH Microbenchmark Suites from Unit Tests*, IEEE Transactions on Software Engineering (TSE), 49(4), 2023.
- F. Markusse, **P. Leitner**, and A. Serebrenik: *Using Benchmarking Bots for Continuous Performance Assessment*, IEEE Software, 39(5), 2022.
- L. Erlenhov, F. Gomes, and **P. Leitner**: *Dependency Management Bots in Open-Source Systems—Prevalence and Adoption*, PeerJ Computer Science, 8:e849, 2022.
- C. Laaber, **P. Leitner**, and H. Gall: *Applying Test Case Prioritization to Software Microbenchmarks*, Empirical Software Engineering (EMSE), 26, 133, 2021.
- M. Grambow, C. Laaber, **P. Leitner**, and D. Bermbach: *Using Application Benchmark Call Graphs to Quantify and Improve the Practical Relevance of Microbenchmark Suites*, PeerJ Computer Science, 7:e548, 2021.
- Y. Guo and **P. Leitner**: *Studying the Impact of CI on Pull Request Delivery Time in Open Source Projects — a Conceptual Replication*, PeerJ Computer Science, 2019.
- D. Costa, C. P. Bezemer, **P. Leitner**, and A. Andrzejak: *What’s Wrong With My Benchmark Results? Studying Bad Practices in JMH Benchmarks*, IEEE Transactions on Software Engineering (TSE), 2019.
- B. Varghese, **P. Leitner**, S. Ray, K. Chard, et al.: *Cloud Futurology*, IEEE Computer, 52(9), 2019.
- C. Laaber, J. Scheuner, and **P. Leitner**: *Software Microbenchmarking in the Cloud. How Bad is it Really?*, Empirical Software Engineering (EMSE), 24(4), 2019.
- **P. Leitner**, E. Wittern, J. Spillner, and W. Hummer: *A Mixed-Method Empirical Study of Function-as-a-Service Software Development in Industrial Practice*, Journal of Systems and Software (JSS), 149, 2019.
- G. Schermann, J. Cito, **P. Leitner**, U. Zdun, and H. Gall: *We’re Doing It Live: A Multi-Method Empirical Study on Continuous Experimentation*, Information and Software Technology (IST), 99, 2018.
- G. Schermann, J. Cito, and **P. Leitner**: *Continuous Experimentation — Challenges, Implementation Techniques, and Current Research*, IEEE Software, 35(2), 2018.

Peer-reviewed conference contributions

- H. Zhang, M. Alhanahnah, **P. Leitner**, and A. Ali-Eldin: *BLAFS: A Bloat-Aware Container File System*, in Proceedings of ACM SoCC, 2025.
- R. Khojah, M. Mohamad, **P. Leitner**, and F. Gomes: *Beyond Code Generation: An Observational Study of ChatGPT Usage in Software Engineering Practice*, in Proceedings of ACM FSE, 2024.
- H. M. Ayas, R. Hebig, and **P. Leitner**: *The Roles, Responsibilities, and Skills of Engineers in the Era of Microservices-Based Architectures*, in Proceedings of CHASE, 2024.
- A. Johansson, C. Holmberg, F. Gomes, and **P. Leitner**: *The Impact of Compiler Warnings on Code Quality in C++ Projects*, in Proceedings of IEEE/ACM ICPC, 2024.
- H. Zhang, M. Alhanahnah, F. Ahmed, D. Fatih, **P. Leitner**, and A. Ali-Eldin: *Machine Learn-*

ing Systems are Bloated and Vulnerable, in Proceedings of ACM SIGMETRICS, 2024.

- P. Samoaa, L. Aronsson, **P. Leitner**, and M. H. Chehreghani: *Batch Mode Deep Active Learning for Regression on Graph Data*, in Proceedings of IEEE Big Data, 2023.
- J. Scheuner, R. Deng, J.-P. Steghöfer, and **P. Leitner**: *CrossFit: Fine-grained Benchmarking of Serverless Application Performance across Cloud Providers*, in Proceedings of IEEE/ACM UCC, 2022.
- H. Samoaa, A. Longa, M. Mohamad, M. H. Chehreghani, and **P. Leitner**: *TEP-GNN: Accurate Execution Time Prediction of Functional Tests using Graph Neural Networks*, in Proceedings of PROFES, 2022.
- H. M. Ayas, F. Gomes, and **P. Leitner**: *Empirical Analysis of Microservices Systems Using Consumer-Driven Contract Testing*, in Proceedings of Euromicro SEAA, 2022.
- H. M. Ayas, **P. Leitner**, and R. Hebig: *A Grounded Theory Study on the Migration Journey Towards Microservices*, in Proceedings of PROFES, 2021.
- H. M. Ayas, R. Hebig, and **P. Leitner**: *Facing the Giant: a Grounded Theory Study of Decision-Making in Microservices Migrations*, in Proceedings of ACM/IEEE ESEM, 2021.
- P. Samoaa and **P. Leitner**: *An Exploratory Study of the Impact of Parameterization on JMH Measurement Results in Open-Source Projects*, in Proceedings of ACM/SPEC ICPE, 2021.
- C. Laaber, S. Würsten, H. C. Gall, and **P. Leitner**: *Dynamically Reconfiguring Software Microbenchmarks: Reducing Execution Time Without Sacrificing Result Quality*, in Proceedings of ACM ESEC/FSE, 2020.
- L. Erlenhov, F. Gomes, and **P. Leitner**: *An Empirical Study of Bots in Software Development — Characteristics and Challenges from a Practitioner’s Perspective*, in Proceedings of ACM ESEC/FSE, 2020.
- G. Schermann, F. Oliveira, E. Wittern, and **P. Leitner**: *Topology-aware Continuous Experimentation in Microservice-based Applications*, in Proceedings of ICSOC, 2020.
- J. Cito, **P. Leitner**, M. Rinard, and H. C. Gall: *Interactive Production Performance Feedback in the IDE*, in Proceedings of ACM/IEEE ICSE, 2019.
- V. Holmqvist, J. Nilsfors, and **P. Leitner**: *Cachematic — Automatic Invalidation in Application-Level Caching Systems*, in Proceedings of ACM/SPEC ICPE (Industry Track), 2019.
- G. Schermann and **P. Leitner**: *Search-Based Scheduling of Experiments in Continuous Deployment*, in Proceedings of IEEE ICSME, 2018.
- C. Laaber and **P. Leitner**: *An Evaluation of Open-Source Software Microbenchmark Suites for Continuous Performance Assessment*, in Proceedings of ACM MSR, 2018.

Research review articles

- H. Samoaa, F. Bayram, P. Salza, and **P. Leitner**: *A Systematic Mapping Study of Source Code Representation for Deep Learning in Software Engineering*, IET Software, 2022.
- J. Scheuner and **P. Leitner**: *Function-as-a-Service Performance Evaluation: A Multivocal Literature Review*, Journal of Systems and Software (JSS), 170, 2020.

Peer-reviewed books and book chapters

- R. van der Mei, H. van den Berg, I. Ganchev, K. Tutschku, **P. Leitner**, P. Lassila, et al.: *State of the Art and Research Challenges in the Area of Autonomous Control for a Reliable Internet of Services*, in: Autonomous Control for a Reliable Internet of Services, Springer, 2018.

Other publications

Workshop papers, demos, and tutorials:

- R. Khojah, F. Gomes, and **P. Leitner**: *From Human-to-Human to Human-to-Bot Conversa-*

tions in Software Engineering, in Proceedings of Alware, 2024.

- J. Scheuner, M. Bertilsson, O. Grönqvist, H. Tao, H. Lagergren, J.-P. Steghöfer, and **P. Leitner**: *TriggerBench: A Performance Benchmark for Serverless Function Triggers* (Short Paper), in Proceedings of IEEE IC2E, 2022.
- J. Scheuner and **P. Leitner**: *Transpiling Applications into Optimized Serverless Orchestrations*, in Proceedings of FAS*W, 2019.
- L. Erlenhov, F. Gomes, R. Scandariato, and **P. Leitner**: *Current and Future Bots in Software Development*, in Proceedings of BotSE (co-located with ICSE), 2019.
- J. Scheuner and **P. Leitner**: *Tutorial: Performance Benchmarking of Infrastructure-as-a-Service (IaaS) Clouds with Cloud WorkBench*, in Proceedings of IEEE ICAC, 2019.
- J. Scheuner and **P. Leitner**: *Tutorial: Performance Benchmarking of Infrastructure-as-a-Service (IaaS) Clouds with Cloud WorkBench*, in Companion Proceedings of ACM/SPEC ICPE, 2019.
- J. Cito, **P. Leitner**, C. Bosshard, M. Knecht, G. Mazlami, and H. Gall: *PerformanceHat — Augmenting Source Code with Runtime Performance Traces in the IDE* (Demo), in Companion Proceedings of ACM/IEEE ICSE, 2018.
- J. Scheuner and **P. Leitner**: *A Cloud Benchmark Suite Combining Micro and Applications Benchmarks*, in Proceedings of QUDOS, 2018.

Invited and non-reviewed contributions:

- F. Markusse, **P. Leitner**, and A. Serebrenik: *Towards Continuous Performance Assessment of Java Applications With PerfBot*, in Proceedings of BotSE (invited), 2023.
- U. Zdun, E. Wittern, and **P. Leitner**: *Emerging Trends, Challenges, and Experiences in DevOps and Microservice APIs*, IEEE Software, 37(1), 2020.
- A. van Hoorn, P. Jamshidi, **P. Leitner**, and I. Weber: *Report from GI-Dagstuhl Seminar 16394: Software Performance Engineering in the DevOps World*, arXiv:1709.08951, 2018.