

curriculum vitae of
Christian Bock, PhD

🌐 christian.bock.ml ✉ chrs.bock@gmail.com 📞 +1 206 468 7460
in christian-bock-mlcb 📄 google scholar 🐙 Github

LLM-BASED AGENTS · DEEP LEARNING
LARGE-SCALE MODEL TRAINING · LARGE LANGUAGE MODELS
ML FOR HEALTHCARE · TOPOLOGICAL DATA ANALYSIS

WORK EXPERIENCE

- Jul. 2025 – present Amazon Web Services, Inc. SEATTLE, USA
Senior Applied Scientist
- Leading research and development of lightweight guide models post-trained via SFT, RL, and test-time adaptation to dynamically steer large coding models, reducing cost while improving code generation quality
 - Directing team of applied scientists as technical lead, serving as senior author on neurosymbolic code localization and prompt optimization research
 - Spearheaded creation and publication of SWE-PolyBench, a novel multi-language benchmark with 2,110 repository-level coding tasks across 21 repositories, achieving 1.2k+ Hugging Face downloads and wide adoption within AWS for evaluating autonomous coding agents and foundation models
- Nov. 2022 – Jun. 2025 Amazon Web Services, Inc. SEATTLE, USA
Applied Scientist
- Developed and led evaluation of Amazon Q for Software Development, an LLM-powered software development agent, spearheading features for autonomous multi-file code changes, bug fixes, and unit test generation in IDE workspaces
 - Directed comprehensive evaluation efforts including submissions to industry benchmarks like SWE-bench
 - Enabled large-scale, distributed fine-tuning and inference of foundation models (T5, BART, BERT) on AWS's custom Trainium hardware by contributing to the compiler codebase and collaborating with compiler engineers to onboard models to the XLA framework
 - Partnered with Hugging Face engineers and contributed to their open-source codebase to improve XLA-based distributed training efficiency and accessibility
- Dec. 2021 – Jun. 2022 Machine Learning Architects Basel, GmbH BASEL, SWITZERLAND
Lead Healthcare & Solutions Architect
- Led strategic initiatives at the intersection of technology consulting, machine learning, and healthcare, driving innovation in medical ML applications
 - Defining the company's healthcare portfolio strategy and architected and executed a pilot study evaluating the feasibility of using ML to analyze data from the pneumology department of a university hospital
- Nov. 2020 – Feb. 2021 Amazon Web Services, Inc. VIRTUAL
Applied Scientist Intern
- Developed an unsupervised and interpretable online anomaly detection model for univariate and multivariate time series by enhancing the state space formulation of Gaussian processes
 - Improved model robustness through a novel heuristic that skips Kalman updates during anomaly encounters, achieving competitive performance on standard benchmarks with superior computational efficiency
- Oct. 2014 – Sep. 2016 SAP SE (SAP Innovation Center) WALLDORF, GERMANY
Working Student
- Worked for the SAP Foundation for Health team and contributed to front-end and back-end development utilizing technologies such as **jQuery**, **d3.js**, **SAPUI5**, **SQL**, and **JavaScript**
 - Implemented the full stack of a modularized interactive genome browser and complex visualizations collaborating with academic partners such as the American Society of Clinical Oncology

 EDUCATION

- Oct. 2017 – Oct. 2021 **Ph.D.** in Machine Learning and Healthcare ETH ZURICH, SWITZERLAND
 I developed and utilized methods for the classification of real-world **biomedical time-series** such as ECG signals and ICU data. Furthermore, I advanced and applied methods from **topological data analysis** for machine learning tasks such as graph classification and to improve our understanding of deep learning and neuroscience. These methods extend to **statistical work** on pioneering significant pattern mining for time-series data.
 Dissertation title: “Motifs and Manifolds – Statistical and Topological Machine Learning for Characterising and Classifying Biomedical Time Series”.
- Apr. 2015 – Mar. 2017 **M.Sc.** in Biomedical Informatics (valedictorian, final grade 1.1/A) HEIDELBERG UNIVERSITY, GERMANY
 Thesis title: “Analysis and Classification of Pathogenic Amino Acid Substitutions in Predicted Protein Structures and Special Classes of Proteins”. Authored at the University of Washington, Seattle.
- Oct. 2011 – Mar. 2015 **B.Sc.** in Biomedical Informatics (top of class, final grade 1.4/A) HEIDELBERG UNIVERSITY, GERMANY
 Thesis title: “A Feasibility Study of a Home-Based Sensor System for Older Adults, using the Microsoft HomeOS Platform”. Authored at University of Washington, Seattle.

 PUBLICATIONS (H-INDEX: 12)

†denotes equal contribution.

SELECTED PUBLICATIONS

- Rashid, M. S.[†], **Bock, C.**[†], Zhuang, Y., Buchholz, A., Esler, T., Valentin, S., ... & Callot, L. (2025). Swe-polybench: A multi-language benchmark for repository level evaluation of coding agents. arXiv preprint arXiv:2504.08703.
- Chaudhari, S., Jacob, R. T., Goswami, M., Cao, J., Rashid, S., & **Bock, C.** (2025). SpIDER: Spatially Informed Dense Embedding Retrieval for Software Issue Localization. arXiv preprint arXiv:2512.16956.
- Melcer, D., Chen, Q., Chiang, W. H., Garg, S., Garg, P., & **Bock, C.** (2025). Textual Gradients are a Flawed Metaphor for Automatic Prompt Optimization. arXiv preprint arXiv:2512.13598.
- **Bock, C.**[†], Moor, M.[†], Jutzeler, C. R., & Borgwardt, K. (2020). Machine learning for biomedical time series classification: from shapelets to deep learning. *Artificial Neural Networks*, 33-71.
- Muhamed, A., **Bock, C.**, Solanki, R., Park, Y., Wang, Y., & Huan, J. (2023, August). Training large-scale foundation models on emerging AI chips. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 5821-5822).
- **Bock, C.**, Walter, J. E., Rieck, B., Strebel, I., Rumora, K., Schaefer, I., ... & Mueller, C. (2024). Enhancing the diagnosis of functionally relevant coronary artery disease with machine learning. *Nature Communications*, 15(1), 5034.

 SKILLS

Strong knowledge of **python**¹ with a focus on data science/ML libraries such as **numpy**, **TensorFlow/Keras**, **pytorch**, **pandas**, **transformers**, **scikit-learn**.
 Strong knowledge in **LaTeX** (including creating visualizations with **TikZ**), the **git** version control system, and the **bash** shell.
 Working knowledge in **Java**, **JavaScript** (e.g., **node.js** and **d3.js**), **C++**, **SQL**, and **ABAP**.
 Working knowledge in the **SNOMEDCT** medical ontology, the **HL7 FHIR** communication standard, and programming for the **SAP HANA** database.
 German is my native language. I am proficient in both spoken and written English. I have elementary proficiency in French.

¹Example repositories: TorchMTL, WTK, P-WL, or ml-on-fhir

PRESENTATIONS AND TALKS

Aug. 2023	29 th Conference on Knowledge Discovery and Data Mining (KDD) [Tutorial] “Training Large-scale Foundation Models on Emerging AI Chips”.	LONG BEACH, CA, USA
Jun. /Jul. 2020	The Machine Learning Summer School (MLSS) [Video] “Predicting stress-induced myocardial ischemia from multi-lead ECG using multi-task learning”.	VIRTUAL
Nov. 2019	19 th IEEE International Conference on Data Mining (ICDM) [Oral] “A Wasserstein Subsequence Kernel for Time Series”.	BEIJING, CHINA
Jul. 2018	26 th Conference on Intelligent Systems for Molecular Biology (ISMB) [Oral] “Association Mapping in Biomedical Time Series via Statistically Significant Shapelet Mining”.	CHICAGO, IL, USA
Jun. 2018	Personalized Health Technologies and Translational Research Conference [Oral] “Association Mapping in Biomedical Time Series via Statistically Significant Shapelet Mining”.	ZURICH, SWITZERLAND

SERVICE TO THE SCIENTIFIC COMMUNITY

Reviewing	Conferences NeurIPS (2025, 2023, 2021, 2020), ICLR (2026, 2025, 2023, 2021, 2024), UAI (2024, 2023), AAAI (2021), MLHC (2020), ECML-PKDD (2019), ICLR Workshop on Geometrical and Topological Representation Learning (2021), NeurIPS “Deep Learning for Code in the Agentic Era” Workshop (2025), ICLR “Deep Learning for Code” Workshop (2024), NeurIPS “ML4H” Workshop (2020), NeurIPS “TDA and Beyond” Workshop (2020)	
	Journals OUP Bioinformatics (2020), IEEE Transactions on Neural Networks and Learning Systems (TNNLS) (2019), Springer Machine Learning (2019)	
Supervision	Master’s Thesis Levente Lippenszky. “Generalizing the Wasserstein Time Series Kernel Using Locality Sensitive Hashing”. Co-Supervision with the Seminar of Statistics (ETH Zurich)	
Teaching	Teaching Assistant for Data Mining II (2021) Contents of Lecture: (Kernel) Principal Component Analysis, Singular Value Decomposition, Multidimensional Scaling, Self-Organizing Maps, Transductive Learning, Cotraining, Association Rule and Graph Mining, Topic Modeling	ETH ZURICH, SWITZERLAND
	Teaching Assistant for Data Mining I (2019) Contents of the Lecture: Similarity measures and metrics, Classification algorithms (k NN, Naive Bayes, Linear Discriminant Analysis, Logistic Regression, Decision Trees, Support Vector Machines), Kernel methods, Clustering ((Kernel) k -means), DBScan, Spectral Clustering, EM Clustering, Hierarchical Clustering)	ETH ZURICH, SWITZERLAND

HONORS AND SCHOLARSHIPS

2011 – 2017	Full undergraduate and graduate scholarship	FRIEDRICH-EBERT FOUNDATION, GERMANY
2017	DREAM Challenges Student Travel Scholarship	INT. SOCIETY FOR COMP. BIOLOGY, USA
2017	Finalist for the “Wings of Excellence” Award	ST. GALLEN SYMPOSIUM, SWITZERLAND
2017	Award for graduating top of class	CHILLI GMBH, GERMANY
2016	Full travel scholarship to participate in the IPHIE Master Class	HEIDELBERG UNIVERSITY, GERMANY
2015	Award for graduating top of class	ERNST-FRANZ VOGELMANN FOUNDATION, GERMANY
2014	1 st Prize at the Startup Weekend	UNIVERSITY OF WASHINGTON, USA