

Francisco Perez-Sorrosal

Principal Research Engineer

N/A
94123 San Francisco, CA (U.S.)

📱 Ask through LinkedIn

✉ Ask through LinkedIn

🌐 <http://francisco-perez-sorrosal.github.io/>

in [fperezsorrosal](#)

🐙 [francisco-perez-sorrosal](#)

🌐 Full CV

"Any man could, if he were so inclined, be the sculptor of his own brain." – Santiago Ramón y Cajal

Profile and Goals

I am a reliable, goal-oriented, and detail-focused professional with extensive experience in research and software engineering across both academic and industry settings, specializing in machine learning, artificial intelligence, and distributed systems. As a Research Engineer, I am curious and love exploring and implementing with cutting-edge technologies, particularly in AI/ML, to tackle complex challenges and foster innovation. As a Software Engineer, I can design, develop, and ensure the maintainability of robust and scalable big data applications, whether on-prem or cloud, always prioritizing high availability, scalability, performance, and reliability. I am adaptable, capable of taking initiative and leading efforts with a strategic approach when needed, or collaborating as a communicative, supportive, and empathetic team player to achieve shared goals. I thrive in open, dynamic, and multidisciplinary environments, integrating visionary and practical insights to deliver impactful results while empowering others and fostering collaboration.

Professional Experience

2025–Now **Strategic Advisor on AI**, *Independent*, **Focus**: Agentic AI & LLM Apps.

- **Agentic AI**: Deep-dive research of multi-agent systems, *LLM orchestration*, and autonomous reasoning frameworks; evaluated cutting-edge frameworks including *AutoGen*, *CrewAI*, *OpenAI* and custom agentic workflows.
- **Strategic Advisory**: Provided technical guidance to early-stage startups (*Enigma*, *Wasmer*) on AI strategy, model selection, and production requirements; advised on *Agentic AI* integration patterns and AI architecture design.
- **Continuous Learning**: maintained cutting-edge expertise in rapidly evolving field.

2023–2024 **Principal Research Engineer**, *Knowledge Graph Science @ Yahoo Inc.*, Sunnyvale, **Roles:** ML/AI Expert, MLOps.

- **Dataset Quality Improvement (2024):** **Goal** – Since the labels of most entity pairs were generated in a semi-supervised manner, develop an LLM-based solution to validate and enhance the labeling accuracy of the dev and test datasets. **Achievement** – Implemented a generative Entity Matching solution with Llama 2 to identify and fix (by re-labeling) misclassified examples in the dev/test sets.
- **Entity Reconciliation Model Training (2023–2024):** Fine-tuned transformer-based language models on heterogeneously labeled datasets, integrating editorially curated data with automatically labeled data through entity linking via common identifiers. This approach aimed to deduplicate and reconcile meta-entities related to People and Creative Works. **Goal** – Improve current production metrics on both categories. **Achievements** – (1) Trained and optimized a unified, type-agnostic LM model for entity matching/reconciliation, replacing tens of type-specific production models and simplifying the maintenance of the reconciliation stage in the Knowledge Graph; (2) Metric improvements: overall increases of 22.23% and 16.11% in precision and recall for Person, and 5% and 4% for Creative Works; (3) Reduced model size by 33%, lowering storage requirements and improving efficiency without compromising accuracy.
- **Entity Reconciliation Pipeline (2023):** Built pipelines for distributed sampling, preprocessing, and transformation of large-scale entity data extracted from various complementary sources that accurately reflect user search behavior. Developed an automated pipeline for training entity reconciliation models for YK, from data curation to inference testing. **Achievements** – (1) Devised a solution to reduce the number of models maintained in production from ~30 heterogeneous (heuristic-based, SVMs, Tree-based) to just 4, which encompass the main 4 meta-entity types in YK; (2) Expanded entity type coverage to 20 new types of people entities (e.g., directors, producers, composers, writers, politicians) and 32 new types of creative works (e.g., visual works, theatrical works, critic's reviews, periodicals, poems).
- **Entity Reconciliation Inference (2023):** **Goal** – Design and architect a PoC solution for cloud-based inference. **Achievement** – Implemented a Ray Serve-based solution to deploy new YK models in the cloud, optimizing for cost efficiency by evaluating CPU and GPU-based approaches.

Jun–Dec 2022 **Principal Research Engineer**, *Mail @ Yahoo*, Sunnyvale, **Roles:** ML/AI Expert.

- **Kamino Project - Mail Classification System:** **Goal** – Upgrade legacy production models to a new generation of deep learning-based, small distilled models optimized for large-scale deployment. These models were designed to classify incoming mail using a newly developed multilabel taxonomy tailored to the mail team's specific use cases. **Approach** – Developed a deep learning pipeline using Hugging Face to train/evaluate models based on the Teacher/Student knowledge distillation framework. Starting with limited human-labeled training data, a large and complex teacher model was created to achieve high accuracy. This model was then used to generate teacher-labeled data for training lightweight student models optimized for deployment. Two types of student models were produced: (1) online models, optimized for speed and suitable for real-time classification tasks, and (2) offline models, designed to include richer information for improved accuracy in non-real-time applications. **Achievements** – The resulting student models for an expanded taxonomy, more than doubling the number of deployable categories in the taxonomy while improving performance on existing categories in production by 3.1-5.9%.

Apr-Dec 2022 **Technical Advisor**, *Digital Transformation Office @ Yahoo*, Sunnyvale, **Roles:** ML/AI Expert, Strategy.

- **AI/ML Working Group - AI/ML Future Strategy Definition in the Cloud:** As Yahoo transitioned its infrastructure from on-premises data centers to a fully cloud-based environment, scientists and research engineers required modern tools to streamline their workflows. These tools needed to facilitate dataset and model sharing, ensure reproducibility of experiments, enable rapid experimentation and iteration on new models, and support seamless publishing and deployment into pre-production/production environments. **Goal** – Specifically, the Model Development subgroup was dedicated to establishing the requirements, standards, tools, and frameworks needed to scale ML/DL applications effectively across the company. **Achievements** – (1) Led the discussions on the Model Development subgroup, although participated actively in the remaining three (Data Management, Model Management, and Model Serving); (2) Delivered a set of recommendations (prioritized and categorized in topics) and tasks to do (internal processes, external relations, tools, etc.) to guarantee that ML/DL practitioners could work with less friction in the new cloud environments.

2021–2022 **Principal Research Engineer**, *Ads @ Yahoo*, Sunnyvale, **Roles:** ML/AI Expert, Strategy.

- **Contextual Targeting Solution:** In the Ads platform, as we were moving towards a cookie-less world, the ability to track users' online signals for behavioral targeting would be drastically reduced, making contextual targeting an appealing alternative for advertising platforms. Using our experience in hierarchical multilabel classification in other contexts, we helped the Ads team to use it in category-based contextual targeting at Yahoo. We proposed and implemented a multilingual model that can accurately classify web pages into a hierarchical taxonomy (specifically, the Yahoo Interest Categories taxonomy) without crawling their content. **Goal** – Transfer the knowledge and experience in taxonomy-based multi-label classification to the Ads team and build a platform for training/evaluating the models. **Achievements** – (1) Helped the Ads team to develop a pipeline for training models for the task at hand; (2) *Multilingual taxonomic web page classification for contextual targeting at Yahoo* paper (ACM SIGKDD 2022) **7 Citations**.

2018–2022 **Principal Research Engineer**, *Science and Content Platform @ Verizon Media/Oath*, Sunnyvale, **Roles:** ML/DL Engineer, Research Engineer.

- **Deep Learning-based Multi-label Classification (2020-2021):** **Goal** – Modernize the production models and infrastructure using new deep learning models. **Achievements** – (1) Developed a modern pipeline based on HuggingFace Transformers to train/evaluate multi-label, multi-class and binary classification problems; (2) Built a tool for plugin taxonomies for the multi-label configurations; (3) Trained new models obtaining significantly better metrics over the ones in production (+10% over baseline); (4) Developed a serving pipeline based on NVIDIA's Triton Server to deploy in production.
- **Scalable Few-shot Classification with Parallel Prefix Conditioning (2021):** **Goal** – Modify a transformer architecture (e.g. BERT) to achieve fast and effective few-shot classification by prefixing multiple category labels to the input. **Method** – Class representations are encoded in parallel but produce independent binary labels for each input through a shared classification output layer. **Achievements** – (1) Conducted experiments on the DBPedia dataset demonstrating improved few-shot performance over standard multi-class classifiers and a speedup over binarized formulations; (2) Further analysis showed that the approach could be scaled to a large number of categories and may hold promise for zero-shot learning of unseen categories.
- **Clickbait Classifier based on Transformers (2019):** **Goal** – Improve the current SVM-based model in production using a BERT-based model. **Achievements** – (1) Proposed an integration strategy for serving the model in production using the existing pipeline; (2) The F1 metric was improved over 5% over current classifier in production; (3) PoC of a generative-based clickbait classification approach using Google's T5.
- **Hierarchical Transfer Learning for Multi-label Text Classification (2018):** **Goal** – Propose a novel transfer learning based strategy where binary classifiers at lower levels in a hierarchy of classes are initialized using parameters of the parent classifier and subsequently fine-tuned on the child categories for the classification task. **Achievement** – Paper published in ACL 2019: *Hierarchical transfer learning for multi-label text classification*. In ACL, 2019 **127 Citations**.

- 2017–2018 **Senior Research Engineer**, *Content Ingestion Platform (CAP) @ Yahoo Inc.*, Sunnyvale, **Roles**: ML/DL Engineer.
- **DL-based Multi-class Classification for Content Ingestion (2018)**: Experimented with LSTM/GRU-based models for multi-class classification using TF/Keras. Explored transfer learning techniques for multi-label text classification.
 - **Machine Learning Content Classification Pipeline (2017)**: Built a pipeline for easy data ingestion, model training and evaluation based on SVMs for the Sieve platform.
 - **ML Pipeline Scalability (2017)**: Built k8s-make, a tool and a workflow-based framework to harness the compute-power in Yahoo's on-prem clusters to deploy a tailor-made Kubernetes cluster to parallelize the training of our SVM-based pipeline in a simple way.
- 2015–2017 **Senior Research Engineer**, *Sieve @ Yahoo Inc.*, Sunnyvale, **Roles**: Distributed Systems Expert, Research Engineer.
- **Twitter Firehose: At-scale Ingestion Streaming System for Tweets (2017)**: Reimplemented the new Twitter API (v2.0) in the firehose. Re-architected previous solution to be more performant yet keeping backwards compatibility. Collaborated with the Sports team for the productization of the new firehose.
 - **Scalable Content Ingestion Platform (2016)**: Transferred Omid as full open-source project into the Apache Software Foundation (ASF). Continued supporting the Omid transaction manager project in production. *Omid* presented at Hadoop Summit, 2016, San Jose, CA (USA). *Omid, reloaded: scalable and Highly-Available transaction processing* paper. In USENIX FAST, 2017 **20 Citations**.
 - **Scalable Content Ingestion Platform (2015)**: Worked at the multi-tenant content ingestion platform at Search organization. Added High Availability to the Omid Transaction Manager for HBase. Scaled Omid in multi-core architectures. Supported the Omid transaction manager in the production infrastructure of the content ingestion platform. Started exploring the transfer of Omid as open-source project to the Apache Software Foundation.
- 2012–2015 **Research Engineer**, *Scalable Computing Group @ Yahoo Labs.*, Spain, **Roles**: Distributed Systems Expert, Research Software Engineer.
- **Omid - Transaction Manager for Big Datastores (2014)**: Re-architected the original codebase and implemented new features. Supported the Omid integration in Sieve project (Yahoo's large scale Internet content ingestion platform). Presented Omid and its poster at Yahoo's internal technical conference (Techpulse Conf. 2014).
 - **Edentity/Pachiderm - Ubiquitous Content Access and Management (2014)**: Ubiquitous content access and management of personal data (images, video) held on 3rd party services (Flickr, GDrive, Dropbox). Designed and implemented a scalable synchronization data module. Defined and implemented a RESTful API for each module defined (User Registration, Search, Content Fetching).
 - **RiddLR - Percolator-like Incremental Processing System (2013)**: Prototyped the framework and built a multi-stage example application on top of it. Presented RiddLR and its poster at Yahoo's Techpulse Conf. 2013.
 - **CumuloNimbo - 7th European Framework Programme (FP7-257993) (2012–2013)**: Designed and implemented a prototype of an incremental processing system for Big Datastores. Added durability guarantees to HBase through BookKeeper. Represented Yahoo in project meetings and evaluation sessions.
- 2011–2012 **Software Architect**, *Lumata*, Spain, **Roles**: Distributed Systems Expert, Architect.
- **Giddra Project - SONY Socialife Application (2012)**: High Scalable Big-Data Backend Platform for web and social content ingestion and aggregation. Contributed to the architectural definition of the platform. Did coordination over multiple teams. Defined the REST API for allowing clients to access/use the backend.
- 2010–2011 **Software Architect/Engineer**, *Freelance, Greenhouse Cooperative*, Spain.
- Analysis, design and Ruby/Rails implementation of a web application to manage the different domains of a greenhouse farm. LoC \simeq 20,000.

- 2003–2010 **Software Architect/Research Engineer**, *School of CS at Univ. Politécnica de Madrid (UPM)*, Spain, **Roles**: Distributed Systems Expert, Software Architect, Software Engineer.
I participated in European and national research projects that included the analysis, design, implementation and testing of different kinds of middleware architectures and applications.
- **NEXOF-RA: NESSI Open Framework - Reference Architecture (2008–2010)**: 7th European Framework Programme (FP7-216446). Contributed to a reference architecture (RA) for a European service platform. Built the specification of a set of architectural patterns for non-functional attributes and analysis of how to integrate cloud platforms in the RA.
 - **Highly Scalable Platform for the Construction of Dependable and Ubiquitous Services (2007–2010)**: Spanish Ministry of Education and Science (TIN2007-67353-C02). Did analysis and tests of consistency problems that arose in end-user applications when second-level caches (e.g. Coherence, JBoss cache) are combined with object persistence mechanisms (e.g. Hibernate).
 - **AUTOMAN: Autonomic Management of Grid-Based Enterprise Services (2006–2007)**: Did the integration of self-configuration and self-repair properties of autonomic computing in the core of a cloud platform at INRIA (France).
 - **High Performance Distributed Systems (2006–2009)**: Community of Madrid (S-0505/TIC/000285). Developed a high-available and scalable service for the JOnAS J(2)EE application server. It provided high availability for critical applications deployed in application server clusters, scaling-out the cluster when overloaded. UPM & BULL signed a pre-agreement to include it in the commercial version. LoC Java (HA&S Service) \simeq 4,000.
 - **S4ALL (Services for All) (2005–2007)**: 5th European Framework Programme (IST-2001-37126). Developed a high-available service for the JOnAS application server maintained by Bull SAS (France). Available since v.4.8. LoC Java (JOnAS) \simeq 150,000.
 - **AUTONOMIC: Autonomic, Dependable and Middleware for Scalable, Distributed, Ubiquitous and Highly Available e-Services (2004–2007)**: Spanish Ministry of Education and Science (TIN2004-07474-C02-01). Built an open-source reference implementation of the WS-CAF specification for adding transactions to SOAP Web Services. LoC \simeq 50,000.
 - **ADAPT: Middleware for Adaptive and Composable Distributed Components (2002–2005)**: EUREKA/ITEA project (Label 04025). Implemented a transactional-aware replication architecture for stateful EJBs for JBoss. LoC Java \simeq 10,600. Open-sourced the implementation of the Activity Service specification to add advanced transactions models to J2EE. LoC Java \simeq 10,000.
- 2001–2003 **Lecturer**, *School of CS at Universidad Pontificia de Salamanca (Madrid Campus)*, Spain, **Roles**: Lecturer.
- Courses on Operating Systems and Programming (C and Pascal).
- 2000–2001 **Systems Administrator**, *School of CS at Universidad Pontificia de Salamanca (Madrid Campus)*, Spain, **Roles**: Systems Administrator.
- Managed and maintained UNIX/Linux servers and Windows workstations: task automation, security.
- 1999 **Quality Analyst**, *Meta 4 S.A. (now Cegid)*, Madrid, Spain, **Roles**: Quality Analyst.
- Tested the database connection modules of Meta4's ERP suite (now Cegid), gaining hands-on experience with multiple DBMSs, including Oracle, Microsoft SQL Server, Informix, and Sybase, as well as JDBC (Java Database Connectivity). My responsibilities included configuring database connections via JDBC, planning and executing tests, analyzing results, and reporting bugs to ensure system reliability and performance.

Patents

- 2025 **18/512,871**, *Systems and methods for automatically adding text content to generated images*, App.
- 2024 **18/365,941**, *Method and system for webpage classification and content delivery*, App.

Academic Research Experience

2003–2011 **Researcher**, *School of CS at Universidad Politécnica de Madrid (UPM)*, Spain, **Roles**: PhD. Candidate/Researcher.

- **Ph.D. Thesis: Middleware for High Available and Scalable Multi-Tier and Service-Oriented Architectures (2003–2009)**: Advisors: Prof. Marta Patiño-Martínez and Prof. Ricardo Jiménez-Péris (UPM). **Specialization** – *Distributed Systems, Transactional Systems, Scalability, High Availability, SOAs*. Developed a brand-new approach to provide high availability and scalability to multi-tier architectures by combining snapshot-isolation and an innovative vertical replication approach.
- **Research Internship at SARDES Research Group at INRIA Grenoble, France (2007)**: Advisor: Prof. Sara Bouchenak. Integration of self-configuration and self-repair properties of autonomic computing in the core of a cloud platform.
- **Publications in Top Conferences and Journals**: (1) *Elastic SI-Cache: Consistent and Scalable Caching in Multi-Tier Architectures*. In VLDB Journal, 2011 **44 Citations**; (2) *Scalability Evaluation of the Replication Support of JOnAS, an Industrial J2EE Application Server*. In EDCC Conf., Valencia (Spain), 2010 **10 Citations**; (3) *A System of Architectural Patterns for Scalable, Consistent and Highly Available Multi-tier Service Oriented Infrastructure*. In Architecting Dependable Systems VI, Springer, 2009 **13 Citations**; (4) *Consistent and Scalable Cache Replication for Multi-tier J2EE Applications*. In ACM/IFIP/USENIX Middleware Conf., CA (USA), 2007 **43 Citations**; (5) *WS-Replication: A Framework for Highly Available Web Services*. In ACM WWW Conf., Edinburgh, 2006 **230 Citations**; (6) *Highly Available Long Running Transactions and Activities for J2EE Applications*. In IEEE ICDCS Conf., Lisbon (Portugal), 2006 **28 Citations**; (7) *ZenFlow: A Visual Tool for Web Service Composition*. In IEEE VL/HCC Conf., Dallas (USA), 2005 **53 Citations**.

Skills

AI/ML & Data Science

AI/ML Agentic AI Fmwks (OpenAI, CrewAI, etc.), HF Transformers, PyTorch, Scikit-learn.
Areas Gen AI, NLP, Multi-label Classification, Knowledge Graphs.
Data Pandas, Dask, SQL, Vector Databases, Large-scale Data Pipeline Development.
MLOps Ray, NVIDIA Triton Server, Model Deployment & Monitoring.

Programming & Engineering

Tools Agentic Tools (Cursor, Claude Code), Git, GH Actions, GitLab CI, IntelliJ, VSCode.
Code Python, Java, Rust, Go.
Cloud AWS, Google Cloud, Docker, Kubernetes.
Dev RESTful APIs, Distributed Systems, Software Architecture, Unit Testing.

Data & Infrastructure

DB Relational, NoSQL, Vector Databases, HBase.
BigData Distributed Processing, Highly Available and Scalable System Design, Perf. Tuning.

Architecture, Design & Processes

Processes Architecture and Software Design, Design patterns, RESTful APIs, UML, Agile

Courses

- 2025 **Claude with Amazon Bedrock**, *Anthropic Education*, Certificate.
- 2025 **Claude Code: A Highly Agentic Coding Assistant**, *DeepLearning.AI*, Certificate.
- 2025 **Model Context Protocol: Advanced Topics**, *Anthropic Education*, Certificate.
- 2025 **Introduction to Model Context Protocol**, *Anthropic Education*, Certificate.
- 2025 **Claude Code in Action**, *Anthropic Education*, Certificate.
- 2025 **Claude with the Anthropic API**, *Anthropic Education*, Certificate.

- 2025 [LLMs as Operating Systems: Agent Memory \(Letta\)](#), *DeepLearning.AI*, Certificate.
- 2025 [ACP: Agent Communication Protocol](#), *DeepLearning.AI*, Certificate.
- 2025 [Fundamentals of Agents](#), *HuggingFace*.
- 2025 [Building toward Computer Use with Anthropic](#), *DeepLearning.AI*, Certificate.
- 2025 [MCP: Build Rich-Context AI Apps with Anthropic](#), *DeepLearning.AI*, Certificate.
- 2025 [Attention in Transformers: Concepts and Code in PyTorch](#), *DeepLearning.AI*, Certificate.
- 2025 [How Transformer LLMs Work](#), *DeepLearning.AI*, Certificate.
- 2025 [Practical Multi AI Agents and Advanced Use Cases with crewAI](#), *DeepLearning.AI*, Certificate.
- 2024 [Multi AI Agent Systems with CrewAI](#), *DeepLearning.AI*, Certificate.
- 2024 [Knowledge Graphs for RAG](#), *DeepLearning.AI*, Certificate.
- 2024 [AI Agentic Design Patterns with AutoGen](#), *DeepLearning.AI*, Certificate.
- 2024 [Functions, Tools and Agents with LangChain](#), *DeepLearning.AI*, Certificate.
- 2024 [Improving Accuracy of LLM Applications](#), *DeepLearning.AI*, Certificate.
- 2024 [LangChain for LLM Application Development](#), *DeepLearning.AI*, Certificate.
- 2024 [AI Agents in LangGraph](#), *DeepLearning.AI*, Certificate.
- 2024 [RAG Developer Bootcamp](#), *Pinecone + Anyscale*, In-Person, San Francisco. Bootcamp
- 2023 [LLM102x: Large Language Models: Foundation Models from the Ground Up](#), *edX/Databricks*, Certificate.
- 2023 [LLM101x: Large Language Models: Application through Production](#), *edX/Databricks*, Certificate.
- 2023 [LLM Bootcamp](#), *FSDL*, In-Person, South San Francisco, CA (USA).
- 2023 [Yahoo Leadership Program](#), The Forem, Remote.
- 2022 [Full Stack Deep Learning 22'Edition](#), *FSDL*, Course.
- 2022 [Machine Learning Specialization \(3 Courses\)](#), *Coursera/DeepLearning.AI/Stanford*, Certificate.
- 2021 [Full Stack Deep Learning 21'Edition](#), *FSDL*, Course.
- 2019 [Machine Learning Specialization \(4 Courses\)](#), *Coursera/Univ. Washington*, Certificate.
- 2018 [Mathematics for Machine Learning Specialization \(3 Courses\)](#), *Coursera/Imperial College London*, Certificate.
- 2018 [Deep Learning Specialization \(5 Courses\)](#), *Coursera/DeepLearning.AI*, Certificate.
- 2011 [Certificate of Training, Advanced Scala](#), *Typesafe*, Switzerland.
- 2011 [Certificate of Training, Scala](#), *Typesafe*, Switzerland.
- 2011 [Course in Business Administration and Economics \(268 hours\)](#), *Funded by Community of Madrid*, Spain.

Education

- 2003–2009 [Ph.D. in CS](#), *School of CS at Universidad Politécnica de Madrid (UPM)*, Spain.
- 2004 [Postgraduate Certificate in Education](#), *Educational Sciences Institute at Universidad Complutense de Madrid (UCM)*, Spain.
- 1994–2001 [B.Eng & M.Sc. in CS](#), *School of CS at Universidad Pontificia de Salamanca (Madrid Campus)*, Spain.

Leadership & Communication

Mentoring & Team Leadership

- Research Mentored junior researchers across multiple teams; led cross-functional product, engineering, and strategy discussions; enabled knowledge transfer from research to production engineering.
- Academia Taught undergraduate courses in [specific subject/area], developing curriculum and guiding students through core concepts and practical applications.

Technical Communication

- Misc. Presented research findings at both international and national conferences, effectively communicating complex ideas to diverse audiences. Delivered project updates in internal meetings within corporate environments. Authored 10+ peer-reviewed publications with extensive citations.

Professional Service

- Apache Committer and maintainer for *Apache Omid*; conference reviewer; active open-source maintainer of *mdbook-bib* (Bibliography management tool).

Languages

- | | | | |
|---------|---------------------------------|---------|---------------------------------|
| Spanish | <i>Mother tongue</i> | English | <i>Professional proficiency</i> |
| French | <i>Intermediate proficiency</i> | Catalan | <i>Intermediate proficiency</i> |

Other Activities Related to Computer Science

Open Source Project Maintainer.

- *mdbook-bib*: Rust-based bibliography management plugin for mdBook (official Rust documentation system). Active maintenance and community engagement with significant adoption. Crates Info

Committer in the Apache Software Foundation.

- *Apache Omid project*, a high-performant and scalable Transaction Manager for HBase

Reviewer in International Academic Conferences.

- *IEEE International Symposium on Reliable Distributed Systems (SRDS)*, 2008 and 2009
- *IEEE International conference in Distributed Computing Systems (ICDCS)*, 2009
- *International Conference on Parallel and Distributed Computing (Europar)*, 2007 and 2010
- *ACM Symposium on Applied Computing (SAC)*, 2008 and 2010
- *International Conference on Service-Oriented Computing (ICSOC)*, 2009
- *EDBTs International Workshop on Data Management in Peer-to-peer Systems (DAMAP)*, 2009
- *International Workshop on Assurance in Distributed Systems and Networks (ADSN)*, 2010

Speaker/Attendee in Academic/Technical Conferences.

- *Agentic AI Summit*, Aug 2nd, 2025, Berkeley, CA (USA)
- *AI Engineer World's Fair*, Jun 3rd-5th, 2025, San Francisco, CA (USA)
- *MLSys 2025 conference*, May 11th-15th, 2025, Santa Clara, CA (USA)
- *Neurips 2024 conference*, Dec 10th-15th, 2024, Vancouver, BC (Canada)
- *ICML 2024 conference*, Jul 21st-27th, 2024, Vienna (Austria)
- *AI Engineer World's Fair*, Jun 25th-27th, 2024, San Francisco, CA (USA)
- *MLSys 2024 conference*, May 13th-16th, 2024, Santa Clara, CA (USA)
- *W&B Fully Connected Conference 2024, The Era of Generative AI*, Apr 17th-18th, 2024, San Francisco, CA (USA)
- *Neurips 2023 conference*, Dec 10th, 2023, New Orleans, LA (USA)
- *Amazon Re:Invent 2023*, Nov 27th-30th, 2023, Las Vegas, NV (USA)
- *Ray Summit 2023*, Sep 18th-20th, 2023, San Francisco, CA (USA)
- *ICML 2023 conference*, Jul 23rd-29th, 2023, Honolulu, Hawaii (USA)
- *MLSys 2023 conference*, Jun 4th-8th, 2023, Miami, FL (USA)
- *Ray Summit 2022*, Aug 23rd-24th, 2022, San Francisco, CA (USA)
- *MLSys 2022 conference*, Aug 31st-Sept 3rd, 2022, Santa Clara, CA (USA)
- *ACL 2019 conference (Speaker)*, Jul 28th-Aug 2nd, 2019, Florence (Italy)
- *@Scale conference*, Aug 31st, 2016, San Jose, CA (USA)
- *Hadoop Summit (Speaker)*, Jun 28th-30th, 2016, San Jose, CA (USA)
- *@Scale conference*, Sep 14th, 2015, San Jose, CA (USA)
- *Hadoop Summit Europe*, Apr 15th-16th, 2015, Brussels (Belgium)
- *Hadoop Summit Europe*, Apr 2nd-3rd, 2014, Amsterdam (Netherlands)
- *NoSQL Matters Conf.*, Nov 21st-22nd, 2014, Barcelona (Spain)
- *NoSQL Matters Conf.*, Nov 29th-30th, 2013, Barcelona (Spain)
- *NoSQL Matters Conf.*, Oct 6th, 2012, Barcelona (Spain)
- *IEEE International Symposium on Reliable Distributed Systems (SRDS)*, Oct 4th-7th, 2011, Madrid (Spain)
- *World Wide Web Conference*, Apr 20th-24th, 2009, Madrid (Spain)
- *Spanish Conference on Concurrency and Distributed Systems*, Sep 13th-16th, 2005, Granada (Spain)
- *ObjectWebCon'05*, Jan 17th-20th, 2005, Lyon (France)
- *ObjectWeb's Workshop on Transactions*, Feb 23rd-24th, 2004, Grenoble (France)
- *ObjectWeb's Architecture Meeting*, Jan 13th-15th, 2004, Sevilla (Spain)

Member.

- *Apache Software Foundation (ASF)*, <https://www.apache.org/>
- *ObjectWeb Consortium*, <http://www.ow2.org>
- *Java Community Process program (JCP)*, <http://jcp.org>

Hobbies and Interests

Beyond my core professional focus, but some way related to it, I have a strong interest in the intersection of fields such as neuroscience, psychology, decision-making, cognitive sciences, learning techniques, behavioral economics, and philosophy, as they offer valuable insights into human behavior, intelligence, and how the brain works (and why!) I am deeply fascinated by the mechanisms of human thought, learning, and behavior, and how these insights can inform and inspire advancements in AI and our daily lives and wellbeing in general.

In my personal life, I enjoy staying active through activities like running, mountain biking, yoga, cold-plunging, and playing tennis, which not only keep me physically fit but also provide a sense of balance and focus.