MICHELE MERLER

Research Scientist — Foundation Model Evaluation & Multimodal AI

New York, NY, USA • +1 (646) 510-1702 • michele.merler@gmail.com

www.michelemerler.com • Google Scholar • LinkedIn • X

SUMMARY

Principal Research Scientist at IBM Research AI with 14+ years building, training, and evaluating machine learning systems across language, vision, speech, and audio. Core contributor to the Granite Code Models (open foundation models for code), leading data curation, post-training for specialized use cases, and evaluation; created benchmarks (ScarfBench, DiF) and co-led award-recognized large-scale empirical evaluations of LLM behavior and failure modes. Deep multimodal expertise, from Emmy Award-winning audio-visual-text highlight curation deployed in production at global sporting events, to interpretable speech biomarkers for ALS and visually-supervised cross-lingual LLM transfer. 50 publications, 20 U.S. patents; Area Chair, WACV 2027, CVPR 2026, ECCV 2024, ACM Multimedia 2016-17. Senior Member of the IEEE.

EXPERIENCE

IBM T.J. Watson Research Center — IBM Research AI *Yorktown Heights, NY*

Principal Research Scientist *June 2025 – present*

- **Benchmark design:** contribution to creation of ScarfBench, a benchmark for evaluating LLM-driven cross-framework application migration in enterprise Java (2026).
- **Agent steering & interpretability:** co-developed CRANE, constrained reasoning injection for code agents via nullspace editing of model activations (2026).
- **LLM evaluation & post-training:** lead research on how code LLMs acquire and retain multi-task capabilities, comparing data mixing against model merging strategies, with systematic evaluation across enterprise code tasks (LLM4Code @ ICSE 2026).
- **Human-AI evaluation in the wild:** co-led one of the first large-scale empirical studies of AI coding assistant usage, effects, and requirements in enterprise settings (LLM4Code @ ICSE 2026).

Senior Research Scientist *June 2022 – June 2025*

- **Granite Code Models:** core contributor to IBM's open foundation models for code — owned data preparation and curation pipelines, lead post-training of specialized checkpoints for enterprise Java use cases, and built evaluation harnesses.
- **LLM failure-mode analysis:** contributed to "Lost in Translation" (ICSE 2024), a systematic study and taxonomy of bugs introduced by LLMs translating code across languages — Pat Goldberg Best Paper Honorable Mention; findings informed model training and benchmark design.
- **Multimodal health AI:** built interpretable, attention-based speech models for clinical assessment of ALS dysarthria from remotely collected audio, validated against clinical scales (npj Digital Medicine 2025; Best Paper, IEEE ICDH 2024).
- **Efficient models:** task-agnostic distillation and neural architecture search for compressing transformer LMs while preserving capability (EMNLP 2023).

Research Staff Member *May 2012 – June 2022*

- **Multimodal foundation for sports highlights:** co-invented production system fusing audio (crowd/commentator excitement), visual (action, players), and text signals to auto-curate broadcast highlights; deployed at Wimbledon, the US Open, and the Masters — Technology & Engineering Emmy Award (2023); IEEE Trans. on Multimedia 2018.
- **Vision-language:** cross-lingual LLM transfer via visually-derived supervision for low-resource languages (ACM Multimedia 2023); large-scale NAS for vision and language models (AAAI 2021).
- **Benchmarks & datasets:** co-authored Diversity in Faces; 1st place, ImageCLEF Medical Modality Classification (2012, 2013) with multimodal visual+text fusion; semantics-aware fine-grained food recognition (ACM Multimedia 2016).

Earlier *2004 – 2012*

- Graduate Research Assistant, Columbia University (video understanding, presentation-video indexing); research internships at IBM Research (complex video event detection; CVPR 2009) and UC San Diego / Calit2 (GroZi grocery-recognition dataset).

SELECTED PUBLICATIONS

Full list of publications on Google Scholar.

[1] Mingzhi Zhu, **Michele Merler**, Raju Pavuluri, Stacy Patterson. "CRANE: Constrained Reasoning Injection for Code Agents via Nullspace Editing". *arXiv preprint arXiv:2605.14084*, 2026.

[2] Anudeep Pavuluri, B. McGinn, A. Saxena, G. Safta, S. Tamilselvam, Raju Pavuluri, **Michele Merler**, et al.. "ScarfBench: A Benchmark for Cross-Framework Application Migration in Enterprise Java". *arXiv preprint arXiv:2605.06754*, 2026.

[3] Mingzhi Zhu, Boris Sobolev, Rahul Krishna, Raju Pavuluri, Stacy Patterson, **Michele Merler**. "Multi-task Code LLMs: Data Mix or Model Merge?". *LLM4Code Workshop at ICSE*, 2026.

[4] Maja Vukovic, Rangeet Pan, Tin Kam Ho, Rahul Krishna, Raju Pavuluri, **Michele Merler**. "Usage, Effects and Requirements for AI Coding Assistants in the Enterprise: An Empirical Study". *LLM4Code Workshop at ICSE*, 2026.

[5] **Michele Merler**, Carla Agurto, Julian Peller, Esteban Roitberg, Alan Taitz, Marcos A. Trevisan, Indu Navar, James D. Berry, Ernest Fraenkel, Lyle W. Ostrow, Guillermo A. Cecchi, Raquel Norel. "Clinical assessment and interpretation of dysarthria in ALS using attention based deep learning AI models". *npj Digital Medicine*, 2025.

[6] Mayank Mishra, Matt Stallone, Gaoyuan Zhang, Yikang Shen, Aditya Prasad, Adriana Meza Soria, **Michele Merler**, Parameswaran Selvam, Saptha Surendran, Shivdeep Singh, Manish Sethi, Xuan-Hong Dang, Pengyuan Li, Kun-Lung Wu, Syed Zawad, Andrew Coleman, Matthew White, Mark Lewis, Raju Pavuluri, Yan Koyfman, Boris Lublinsky, Maximilien de Bayser, Ibrahim Abdelaziz, Kinjal Basu, Mayank Agarwal, Yi Zhou, Chris Johnson, Aanchal Goyal, Hima Patel, Yousaf Shah, Petros Zerfos, Heiko Ludwig, Asim Munawar, Maxwell Crouse, Pavan Kapanipathi, Shweta Salaria, Bob Calio, Sophia Wen, Seetharami Seelam, Brian Belgodere, Carlos Fonseca, Amith Singhee, Nirmit Desai, David D. Cox, Ruchir Puri, Rameswar Panda. "Granite Code Models: A Family of Open Foundation Models for Code Intelligence". *arXiv preprint arXiv:2405.04324*, 2024.

[7] Rangeet Pan, Ali Reza Ibrahimzada, Rahul Krishna, Divya Sankar, Lambert Pouguem Wassi, **Michele Merler**, Boris Sobolev, Raju Pavuluri, Saurabh Sinha, Reyhaneh Jabbarvand. "Lost in translation: A study of bugs introduced by large language models while translating code". *International Conference on Software Engineering (ICSE)*, 2024.

[8] Masayasu Muraoka, Bishwaranjan Bhattacharjee, **Michele Merler**, Graeme Blackwood, Yulong Li, Yang Zhao. "Cross-Lingual Transfer of Large Language Model by Visually-Derived Supervision Toward Low-Resource Languages". *ACM Multimedia*, 2023.

[9] Takuma Udagawa, Aashka Trivedi, **Michele Merler**, Bishwaranjan Bhattacharjee. "A Comparative Analysis of Task-Agnostic Distillation Methods for Compressing Transformer Language Models". *EMNLP*, 2023.

[10] **Michele Merler**, Dhiraj Joshi, Quoc-Bao Nguyen, Stephen Hammer, John Kent, Jinjun Xiong, Minh N. Do, John R. Smith, Rogerio S. Feris. "Automatic Curation of Sport Highlights using Multimodal Excitement Features". *IEEE Transactions on Multimedia*, 2018.

SELECTED HONORS

- Pat Goldberg Best Paper Honorable Mention — "Lost in Translation", LLM code-translation bug study (2025)
- Best Paper Award, IEEE ICDH — early ALS biomarker identification from remote speech (2024)
- Technology & Engineering Emmy Award — AI/ML curation of sports highlights (2023)
- Outstanding Reviewer, CVPR (2021)
- 1st Place, ImageCLEF Medical Modality Classification (2012, 2013)
- Multiple IBM Outstanding Technical Achievement, Corporate, and Research Division Awards (2012–2023)

SERVICE & COMMUNITY

Area Chair: WACV 2027, CVPR 2026, ECCV 2024.

Associate Editor: IEEE Transactions on Multimedia (2021–2023).

Workshop Program Chair: CVsports @ CVPR (2020–2026); Fair, Data Efficient and Trusted CV @ CVPR (2020–2025); Bias Estimation in Face Analytics @ ECCV 2018 / CVPR 2019.

Teaching: Adjunct Instructor, Applied Generative AI, Columbia University School of Professional Studies (Fall 2026, upcoming).

Memberships: Senior Member, IEEE; IEEE Computer Society (CS) & TCPAMI; IEEE Circuits and Systems Society (CAS); New York Academy of Sciences (NYAS).

Reviewer since 2008: CVPR, ECCV, ICCV, NeurIPS, ICML, AAAI, ACM Multimedia, IEEE T-PAMI, IEEE T-MM.

Mentored 6 PhD/MS research interns.

EDUCATION & SKILLS

Ph.D. & M.S., Computer Science — Columbia University (2013, 2008).

M.Eng. & B.Eng., Telecommunications Engineering — University of Trento, Summa cum Laude (2007, 2004).

Skills: Python, PyTorch, C/C++, OpenCV; LLM data curation, pre-/post-training, model merging, distillation, NAS, evaluation & benchmarking; multimodal (vision, speech, audio, text) modeling.

Languages: Italian (native), English (fluent), Spanish (intermediate).