Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
DOI: 10.11159/mvml25.002

Human AI - How Big Data is Big Enough?

KC Santosh
AI Research Lab (https://ai-research-lab.org)
Department of Computer Science, The University of South Dakota
Email: kc.santosh@usd.edu

How large should datasets be before machine learning scientists can begin meaningful work? Continuous learning has always been the most promising approach, as even a few new samples can significantly alter a model's trajectory [1,2]. Can we afford to wait years to gather large datasets, especially when such delays could limit critical testing opportunities? This dilemma reflects what we encountered during the COVID-19 pandemic—when rapid, data-driven responses were vital. As future epidemics loom, **human-in-the-loop machine learning** becomes not just useful but essential for timely and effective public healthcare [3,4,5]. And as we harness AI for social good, **what about its carbon footprint** [6]? While tech giants train models on massive datasets, we must also confront the **environmental cost** of such efforts. Now is the time to champion **sustainable computing practices**, including **green AI**, and integrate **human-AI collaboration** at the core. Consider another frontier: **dark matter –** should we wait for annotations before building models, or begin iterative learning like humans do? Just as people learn continuously from birth, so too must AI evolve ethically [7], transparently [8], and sustainably. Green computing, guided by **responsible AI governance**, is crucial for developing AI systems that not only solve problems but also uphold societal and environmental well-being through **real-world experimentation** and **interpretability [7,8,9]**.

References

1) A Vettoruzzo, MR Bouguelia, J Vanschoren, T Rognvaldsson and KC Santosh. Advances and Challenges in Meta-Learning: A Technical Review, IEEE Transactions on Pattern Analysis & Machine Intelligence (2024) URL: https://doi.org/10.1109/TPAMI.2024.3357847
2) A Jain, SR Dubey, SK Singh, KC Santosh, BB Chaudhuri: Non-Uniform Illumination Attack for Fooling Convolutional Neural Networks. IEEE Transactions on Artificial Intelligence (2025) URL: https://doi.org/10.1109/TAI.2025.3549396
3) KC Santosh and S Nakarmi: Active learning to minimize the risk of future epidemics, eISBN. 978-981-99-7441-2, SpringerBriefs in Applied Sciences and Technology, Springer (2023) URL: https://link.springer.com/book/9789819974412
4) P Singh, R Rizk, KC Santosh: PATL: Pool-Based Active Twin Learner from Oracle with Imitation Learning for Early Epidemic Detection, IEEE Conference on Artificial Intelligence, (2025) URL: https://doi.org/10.1109/CAI64502.2025.00102
5) Suprim Nakarmi and KC Santosh. Active Learning to Minimize the Risk from Future Epidemics, IEEE Conference on Artificial Intelligence (2023), pp. 329–330, URL: https://doi.org/10.1109/CAI54212.2023.00145
6) KC Santosh R Rizk, and S Bajracharya: Cracking the machine learning code: technicality or innovation?, Studies in Computational Intelligence, Springer Nature (2024) URL: https://doi.org/10.1007/978-981-97-2720-9
7) KC Santosh and Casey Wall: Artificial Intelligence, Explainability, and Ethical Issues – Applied Biometrics, Springer-Briefs in Applied Sciences and Technology, Springer (2022)
URL: https://doi.org/10.1007/978-981-19-3935-8

8) Casey Wall, Longwei Wang, Rodrigue Rizk, KC Santosh: Winsor-CAM: Human-Tunable Visual Explanations from Deep Networks via Layer-Wise Winsorization, IEEE Transactions on Pattern Analysis & Machine Intelligence (2025, under review) URL: https://arxiv.org/abs/2507.10846
9) L Wang, I Uddin, X Qin, Y Zhou, KC Santosh: Explainability-Driven Defense: Grad-CAM-Guided Model Refinement Against Adversarial Threats, AAAI Summer Symposium Series (Dubai, 2025) URL: https://ojs.aaai.org/index.php/AAAI-SS/article/view/36024/38179