Title:

Not for Scale but for (Biological) Discovery: Small AI

Abstract:

For a time, it seemed that progress in AI could be achieved through scale alone. Ever-larger language models fueled optimism, with early results suggesting that training over biological sequences could capture intricate molecular structures and functional properties. Yet, critical evaluations of benchmark tasks tempered these claims. Data was both blamed as the problem and framed as the solution—but at the intersection of AI and scientific discovery, we have always known that there will never be enough data. After all, how can discrete observations fully capture the continuous, physics-driven phenomena?

This is a talk about small AI. It presents an alternative vision: leveraging small, specialized AI models that integrate biological priors, expert knowledge, and physical constraints to achieve state-of-the-art performance in proteomics, genomics, and cheminformatics. Within an academic lab where exceeding 1B parameters is a milestone (and a triumph), we focus on strategies that do not rely on data volume. Grounded, structured models informed by domain expertise maintain predictive accuracy while reducing dependency on massive datasets. Physics-derived representations and expert-in-the-loop refinement enable robust performance in data-scarce environments. Aligning AI with scientific intuition and instructing models through human feedback ensures that small AI systems provide relevant, actionable insights for biologists and chemists working at the bench. This is small AI designed not for scale, but for discovery—precision-driven, expert-aligned, and purpose-built to accelerate biological innovation.

Bio (limited to 100 words)

Dr. Amarda Shehu is a Professor in the Department of Computer Science and Associate Dean for Research in the College of Engineering and Computing at George Mason University, where she is also Vice President and Chief AI Officer. Amarda is an IEEE Senior Member, a Fellow of the American Institute for Medical and Biological Engineering, and a Member of the Virginia Academy of Science, Engineering, and Medicine. Her research bridges foundational AI and AI-enabled scientific discovery. Amarda served as 2019-2022 NSF Program Director in the Information and Intelligent Systems Division of the Directorate for Computer and Information Science and Engineering.