Proceedings of the 1st international Symposium on AI and Fluid Mechanics
Abstract No S5 A6

BLASTNet: Accessible community-involved big data as key-enabler for Fluids-AI

Matthias Ihme[1,2,3], Jen Zen Ho[1], Beverley Yeo[1], Bassem Akoush[1]

[1]*Department of Mechanical Engineering, Stanford University, Stanford, CA, USA*

[2]*Department of Energy Science and Engineering, Stanford University, Stanford, CA, USA*

[3]*Department of Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA*

Abstract

Machine learning and artificial intelligence for fluid dynamics rely on large datasets and massive deep learning models to predict quantities of interest. Within the broader field of fluid dynamics, large amounts of data exist in the form of high-fidelity simulation data and detailed measurements that have been accumulating. Yet, this data remains distributed and can be difficult to access. To make this data available, we present a realistic and feasible framework that combines (i) community involvement, (ii) public data repositories, (iii) models and (iv) and challenges for enabling broad access to high-fidelity data via a network-of-datasets approach. This Bearable Large Accessible Scientific Training Network-of- Datasets (BLASTNet) is consolidated on a community-hosted web-platform (https://blastnet.github.io) and is targeted towards improving accessibility to diverse scientific data for deep learning algorithms. We discuss procedures for community to contribute and utilize this database and present illustrative examples on how this database has contributed towards the development of new ML models in the context of turbulent subgrid-scale modeling and predicting transient dynamics in environmental flow applications.

Keywords: Artificial intelligence, Machine learning; Training data; Large-scale data; Foundation models