

AI IN RESEARCH INFRASTRUCTURE AT TACC AND THE LCCF

Dan Stanzione

Executive Director, TACC

Associate Vice President for Research, UT-Austin

Austin, TX

June 2023



**DISCLAIMER: CHATGPT DID *NOT*
WRITE *ALL* OF THIS TALK.**



TACC - 2023



LEADERSHIP-CLASS
COMPUTING FACILITY

TACC

TEXAS ADVANCED COMPUTING CENTER

A QUICK TACC REMINDER

- ▶ We operate the Frontera, Stampede-2, Jetstream, and Chameleon systems for the National Science Foundation
 - ▶ Upcoming MREFC deployment of “Horizon”
- ▶ Lonestar-6 for our Texas academic and industry users.
- ▶ Altogether, ~20k servers, >1M CPU cores, 1k GPUs
- ▶ About seven billion core hours over several million jobs per year.
 - ▶ >\$150M of “cloud equivalent” compute time used last year at TACC.
- ▶ 94,000 users
- ▶ Petabytes of new data ingested and stored each month



MODERN COMPUTATIONAL SCIENCE

Simulation

Computationally query our
mathematical models of the world

Machine Learning/AI

Computationally query our
data sets

(depending on technique,
also called deep learning)

Analytics

Computationally analyze our
experiments

(driven by instruments that produce
lots of digital information)

I would argue that modern science and engineering combine all three

WE HAVE LOTS OF AI INFRASTRUCTURE...

- ▶ ~600 GPUs now – growing to >1k this year, >4k by 2025.
 - ▶ With 100s of Petabytes of solid-state high speed storage, and very fast networking.
- ▶ Containerized versions of many of the popular tools
- ▶ Expert staff support for optimization, I/O and hyperparameter tuning, etc.
- ▶ Large collections of datasets. \
- ▶ We have a growing set of AI training/workforce activities.
- ▶ Seeing an increase of hybrid/surrogate workflows in the user projects.

...AND WE ALSO CONSUME A LOT OF AI

- ▶ Our AI Code Assistant for Python/Jupyter is in beta and near release.
- ▶ Our AI user support assistant, built on LLM tech, is also in beta (provide draft answers to our user support teams for all tickets).
- ▶ Ramping up efforts in AI-Automated metadata tagging and relationship discovery for stored datasets.
- ▶ Exploring uses of AI in enhanced systems operations.
- ▶ Part of the “Trillion Parameter Generative Models Consortium”

HELPING PLAN FUTURE INVESTMENTS. . .

- ▶ On January 24th, the NAIRRTF (that would be the “National AI Research Resource Task Force”) voted officially to send the final report to the President and Congress.
- ▶ Recommends a multi-billion dollar investment in AI compute and data resources for US research and competitiveness.

Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem

An Implementation Plan for a National Artificial Intelligence Research Resource



January 2023

FRONTERA

TACC | NSF | TEXAS

