CASC Newsletter | Vol 15 March 2025

In This Issue:

From the Director

Contact: Jeff Hittinger

"Once we rid ourselves of traditional thinking we can get on with creating the future." – *Jimmy Bertrand*

He may have been talking about music, but jazz drummer Jimmy Bertrand was certainly revealing a deeper truth: Innovation comes from thinking beyond what is and instead imagining what is possible. In this issue of the CASC Newsletter, we delve into ways in which machine learning (ML) and its applications are changing how science is done. Data-driven methods are here to stay and are augmenting the already formidable tools in our computational science toolbox. The future isn't data-driven *or* theory-driven approaches—it's data-driven *and* theory-driven methods used together leveraging the best of both. We are at an exciting time where artificial intelligence (AI) is becoming a genuine game-changer in how we do our work, and CASC researchers are exploring and defining this new paradigm.

A common theme throughout this issue is the incorporation of advanced ML techniques into complex scientific workflows, from dynamic computed tomography (CT) to scalable Gaussian processes (GPs) and autonomous multiscale (AMS) simulations. In this edition, we explore several groundbreaking projects: "Machine Learning Under the Hood of Dynamic Computed Tomography" discusses the use of implicit neural representations (INR) to improve high-fidelity, 4D CT reconstruction. "MuyGPs: A Scalable and

Approximate Gaussian Process Framework" describes a novel approach to GPs that significantly reduces computational costs, making this powerful technique viable for larger datasets, such as those coming from astronomical observations.

"Safe and Trustworthy AI" highlights efforts to ensure—as we incorporate large language models (LLMs) and other ML techniques into our technical workflows and simulation codes—the safety and security of advanced AI systems throughout their lifecycle. Finally, "Autonomous Multiscale Simulations – Embedded Machine Learning for Smart Simulations" showcases the development of autonomously improving data-driven surrogate models to accelerate multiphysics simulations, enhancing both speed and reliability. These innovations not only enhance the accuracy and efficiency of scientific research but also pave the way for new methodologies in data visualization, uncertainty quantification (UQ), and AI safety.

Collaborations | Machine Learning Under the Hood of Dynamic Computed Tomography

Contact: Andrew Gillette and Hyojin Kim

CT plays an important role in nondestructive evaluation (NDE) across LLNL's diverse mission areas such as material characterization, additive manufacturing (AM), weapon component inspection, transportation security, and clinical diagnosis. Recent NDE advancements have led to the pursuit of more challenging CT imaging, including reconstruction from projection data with limited angular ranges and dynamic 4D CT for moving objects, characterized as ill-posed inverse problems.

To tackle these intricate CT problems, CASC researcher Hyojin Kim and his team have developed two distinct directions: 1) data-driven ML approaches, such as the conditional diffusion-based method for limited-angle CT [1] and 2) training-free reconstruction approaches leveraging INR for dynamic 4D CT of deformable objects [2].

In dynamic 4D time-space CT imaging, the object moves or deforms over time while X-ray projections are acquired from multiple angles, in contrast to most conventional CT setups where objects are static during imaging. Existing analytic (filtered back projection) and iterative methods produce poorly reconstructed images with severe artifacts and blurry edges, especially when the object changes considerably. The team on Kim's recent LDRD effort (22-ERD-032) demonstrated a significant advance in the INR-based approach for addressing dynamic 4D CT reconstruction. The new approach utilizes distributed network training across several compute nodes and GPUs incorporating continuous forward CT models. Unlike existing INR methods, which involve forward and back propagation through dense voxel grids of the entire object time-space coordinates, their method uses a small

subset of object coordinates. This approach leads to significantly reduced memory and computing resource requirements, enabling high-fidelity 4D reconstruction of extremely large CT data sizes, as shown in Figure 1.

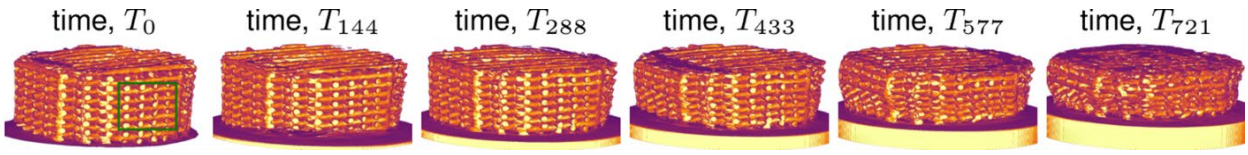

Figure 1: *Reconstructed 4D CT of real data using our INR method. We obtained CT data using the Deben stage to analyze the compression behavior of a 3D-printed lattice structure over time under the mechanical loads of the Deben stage. The volume dimension of each 3D data is 1024x1024x400 and the number of time steps is 722, which corresponds to the number of projections.*

With the data collection process established, a new challenge presented itself: How to efficiently visualize and analyze the data encoded by the INR representation? Producing snapshots like those shown in Figure 1 is a process of visual trial-and-error, typically done by adjusting parameters of the visualization and observing the response in real-time, using software such as ParaView or VisIt. Interactivity requires loading the full dataset into memory and calling well-honed algorithms for tasks like view adjustment, isosurface extraction, and setting opacity. However, when the data is on the order of gigabytes, running these algorithms is only feasible on GPU nodes, in parallel, or both, all of which are difficult.

Fortunately, devising visualization methods for INR-based data is the focus of work by CASC researcher Andrew Gillette as part of a DOE ASCR award in the Data Visualization for Scientific Discovery, Decision-Making, and Communication portfolio. In conjunction with collaborators at Vanderbilt University, the University of Arizona, and a graduate student intern from Cornell University, Gillette is developing algorithms to query INR values in an adaptive fashion, based on the level of detail required in a specified region of inputs. For instance, in the image shown in Figure 2, the regions of space near the surface of the object (yellow) require a finer mesh resolution for visualization; initial experiments suggest these regions can be detected automatically, using an algorithm that acts directly on the weight matrices of the INR.

While INRs have received an explosion of interest in recent years from the computer graphics community for rendering tasks, their use in the realm of scientific visualization is quite nascent. By dovetailing new data collection modalities with new data processing

algorithms, their aim is to demonstrate how state-of-the-art advances in ML can directly enable new methodologies for scientific experimentation.

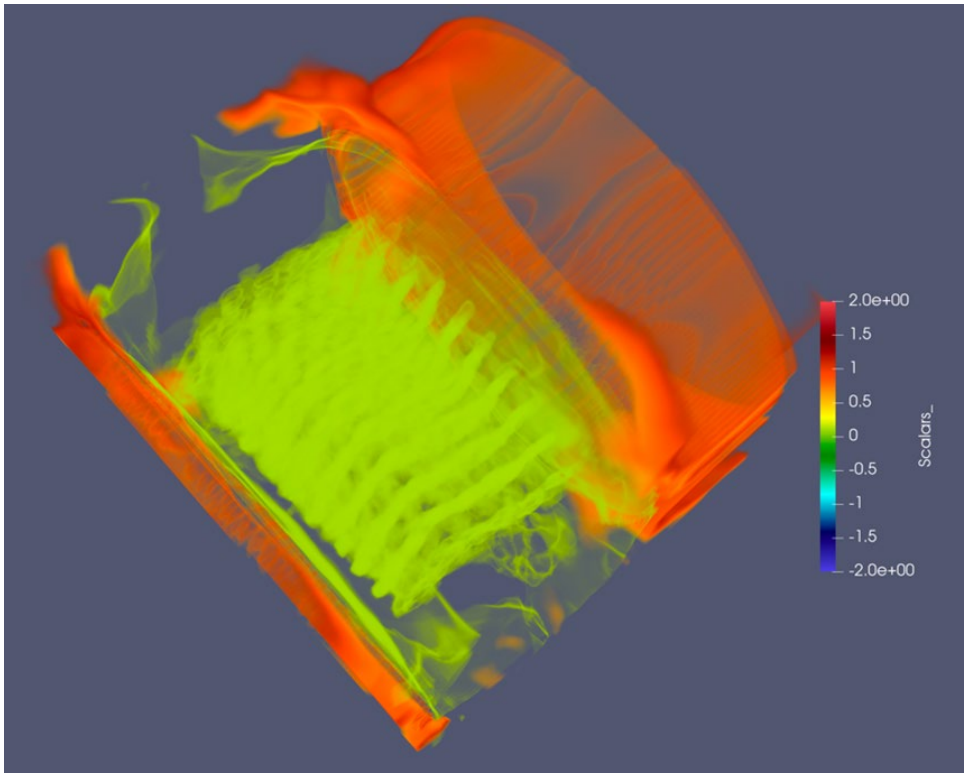

Figure 2: *A down-sampled timeslice of the CT data, visualized in the commercial software ParaView, with a custom opacity function applied. Data-guided adaptive sampling for this type of data would complete the computational pipeline from data collection to interactive visualization for scientific analysis.*

[1] J. Liu, R. Anirudh, J. Thiagarajan, *et al.* "DOLCE: A Model-Based Probabilistic Diffusion Framework for Limited-Angle CT Reconstruction." *International Conference on Computer Vision (ICCV)*, 2023.

[2] A. Reed, H. Kim, R. Anirudh, *et al.* "Dynamic CT Reconstruction from Limited Views with Implicit Neural Representations and Parametric Motion Fields." *International Conference on Computer Vision (ICCV)*, 2021.

Lab Impact | MuyGPs: A Scalable and Approximate Gaussian Process Framework

Contact: Min Priest

GPs are widely used statistical models that can learn nonlinear relationships between feature and response variables by treating them, in effect, as realizations from an infinite-dimensional joint Gaussian distribution. GPs are especially useful in settings that require a precise understanding of prediction uncertainties, such as weather and climate modeling, multiscale physics simulations, and the emulation of computer experiments. However, the utility of conventional GPs is inversely proportional to data size. On one hand, GPs are notoriously sample-efficient compared to other conventional ML methods, meaning that they are very useful when observation data is very sparse, but on the other hand, training and predicting with a GP require inverting a matrix that is square in the number of training data, which has onerous quadratic memory and cubic time complexities. Former LLNL researcher Amanda Muyskens, along with CASC research Min Priest and team, introduced MuyGPs, a scalable, approximate GP framework suitable for the distributed memory environment on HPC platforms to address this shortcoming while maintaining the high-quality UQ inherent to GPs [1].

The computational cost of conventional GPs comes from the assumption that predictions are conditioned on all the training data, while the central insight of MuyGPs is that, in practice, most of this covariance is limited to a few observations. Accordingly, MuyGPs sparsifies prediction computations by conditioning them only on a few training data points, such as their nearest neighbors. MuyGPs also avoids costly determinant evaluations by training model parameters using batched leave-one-out cross-validation as an objective function. This approach results in a training and prediction process that is linear in the size of data, an enormous improvement over conventional GPs. Moreover, MuyGPs also outperforms other approximate scalable GP methods in terms of both runtime and accuracy.

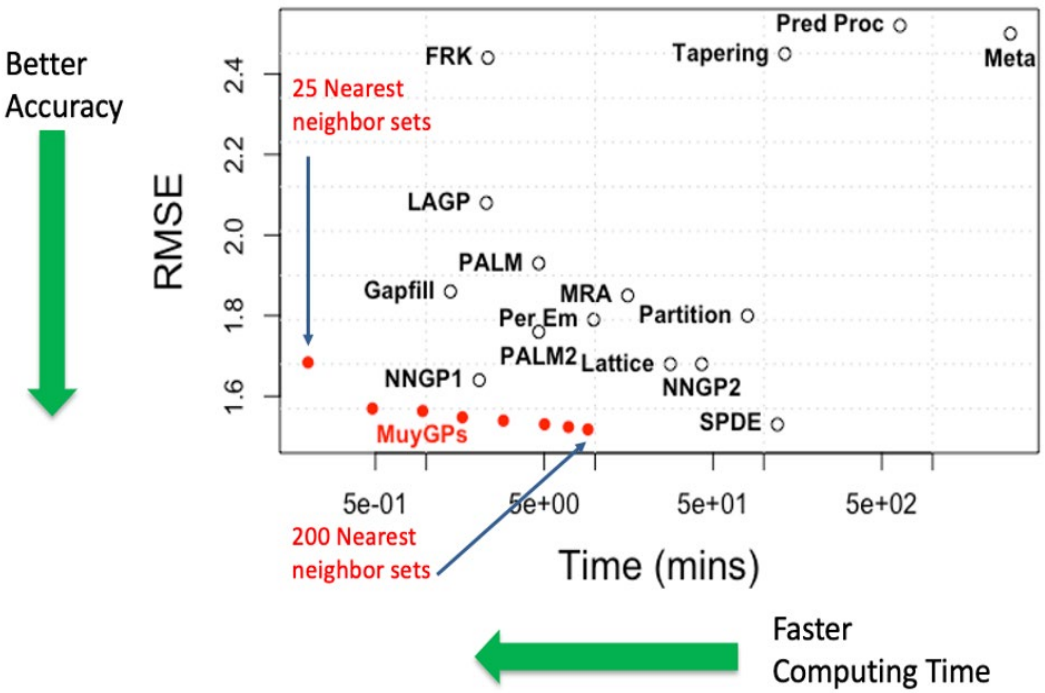

Figure 3: *MuyGPs runtime versus root mean squared error performance on a benchmark climate problem with 100k training and 70k testing data points compared to other approximate scalable GP methods. While there is a tradeoff due to nearest neighbor size, MuyGPs is superlative along both axes.*

Furthermore, MuyGPs is designed to scale to distributed memory systems for accelerated training on enormous datasets. The overwhelming majority of necessary communication occurs in the construction of neighborhood sets for each batch or prediction point, while optimization iterations are fast and merely require an allreduce of the value of the objective function at each step. Additionally, the Python implementation of MuyGPyS is written with multiple backends from conventional NumPy to hardware accelerated computation using JAX, PyTorch, or MPI, allowing for >1000x speedups [2]. These backends are written using a shared API, so that simple codes written and tested on laptops can run on enormous problems on HPC systems with little or no change to the source code. Additionally, MuyGPyS empirically optimizes using much less time, energy, and specialized equipment than state-of-the-art neural architectures like LLMs, which have become increasingly burdensome to training.

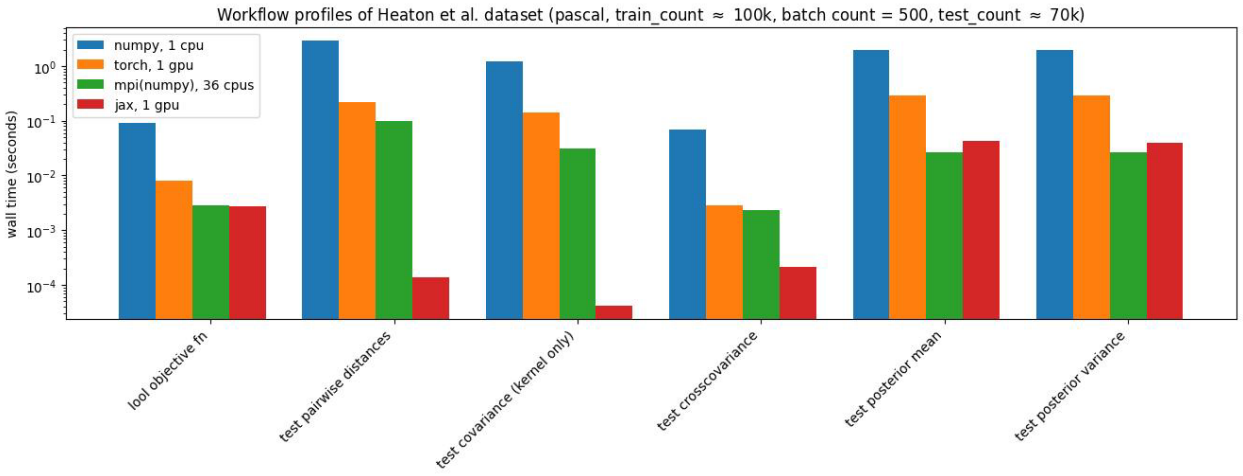

Figure 4: *MuyGPyS runtime performance on a single node of the Pascal supercomputer using different backends.*

MuyGPs was developed to model weak gravitational lensing, which are small distortions of the shape of distant galaxies observed by astronomers and caused by intervening mass, e.g., black holes. While these distortions are generally so small as to be impossible to detect by analyzing a single galaxy point source, nearby sources are distorted in systematically aligned way, which allows for the statistical measurement of black holes in the universe. However, realizing such a statistical model using the depth of field available in next-generation telescope facilities requires evaluating a GP on billions of points—a clear impossibility without a technology like MuyGPs.

The research team has also applied MuyGPs to many different use cases of interest to LLNL. Within computational astronomy, they used MuyGPs to discriminate between stars and galaxies [3], as well as to classify “blends” of both overlapping stars and overlapping galaxies, the populations of which are important parameters of models of the universe [4]. For space domain awareness, they used MuyGPs to detect closely spaced or overlapping space objects from telescope measurements [5]. More recently, they are working on methods to model patterns of life for orbiting space objects [6,7].

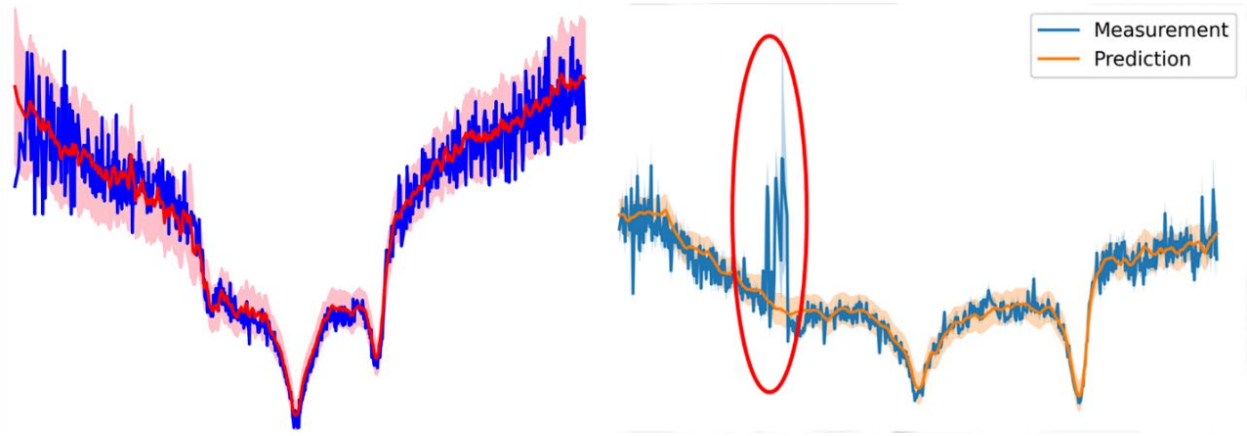

Figure 5: *MuyGPyS modeling UQ and identifying likely maneuvers of space objects.*

[1] A. Muyskens, B. Priest, I. Goumiri, and M. Schneider. "MuyGPs: Scalable Gaussian Process Hyperparameter Estimation Using Local Cross-Validation." *arXiv preprint arXiv:2104.14581*, 2021.

[2] B. Priest, A. Dunton, I. Goumiri, A. Andrews, and A. Muyskens. "MuyGPyS." *GitHub Repository*, github.com/LLNL/MuyGPyS, 2024.

[3] A. Muyskens, I. Goumiri, B. Priest, *et al.* "Star–Galaxy Image Separation with Computationally Efficient Gaussian Process Classification." *The Astronomical Journal* 163(4), 148, 2022.

[4] J. Buchanan, M. Schneider, R. Armstrong, *et al.* "Gaussian Process Classification for Galaxy Blend Identification in LSST." *The Astrophysical Journal* 924(2), 94, 2022.

[5] K. Pruett, N. McNaughton, and M. Schneider. "Closely Spaced Object Classification Using MuyGPyS." *Proceedings of the Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, 2023.

[6] I. Goumiri, A. Dunton, A. Muyskens, *et al.* "Light Curve Completion and Forecasting Using Fast and Scalable Gaussian Processes (MuyGPs)." *arXiv preprint arXiv:2208.14592*, 2022.

[7] I. Goumiri, A. Muyskens, B. Priest, and R. Armstrong. "Light Curve Forecasting and Anomaly Detection Using Scalable, Anisotropic, and Heteroscedastic Gaussian Process Models." *Proceedings of the Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, 2023.

Advancing the Discipline | Safe and Trustworthy AI

Contact: Bhavya Kailkhura

Advancements in AI, particularly LLMs, offer significant potential for accelerating scientific research and bolstering national security. However, their blind adoption presents serious risks, especially in mission-critical areas such as CBRN (chemical, biological, radiological, nuclear), cyber security, and critical infrastructure. LLNL's "AI Safety" team, led by CASC researcher Bhavya Kailkhura, is collaborating with Turing Award winner Prof. Yoshua Bengio and other experts on designing theoretical foundations and practical tools for ensuring advanced AI systems are safe and secure.

They are adopting a holistic approach to safety and security by addressing the entire AI lifecycle—from data preparation (acquisition, curation, and processing) to model development (design, training, validation) to deployment (integration, monitoring, and maintenance). This comprehensive strategy ensures that safety and security are integral at every stage, mitigating potential risks and building trust throughout the AI lifecycle. To support this, they are making fundamental advancements in three key research areas:

- **Comprehensive Validation:** This area focuses on developing rigorous methodologies to assess the correctness and vulnerabilities of AI models before their deployment. The goal is to create approaches for automated testing and analysis that can efficiently identify issues, ensuring that models are scrutinized under comprehensive conditions despite the computational challenges posed by their scale. Some of their early works have contributed to popular benchmarks (TrustLLM [1], MLCommons AI Safety [2], GTBench Reasoning [3]) and identified previously unknown vulnerabilities of LLMs. For example, they showed that model compression can create hidden security vulnerabilities [10]; popular finetuning methods like LoRA can amplify data biases [11]; and scaling up model sizes cannot solve robustness [12] and reasoning flaws [3] of existing AI models.

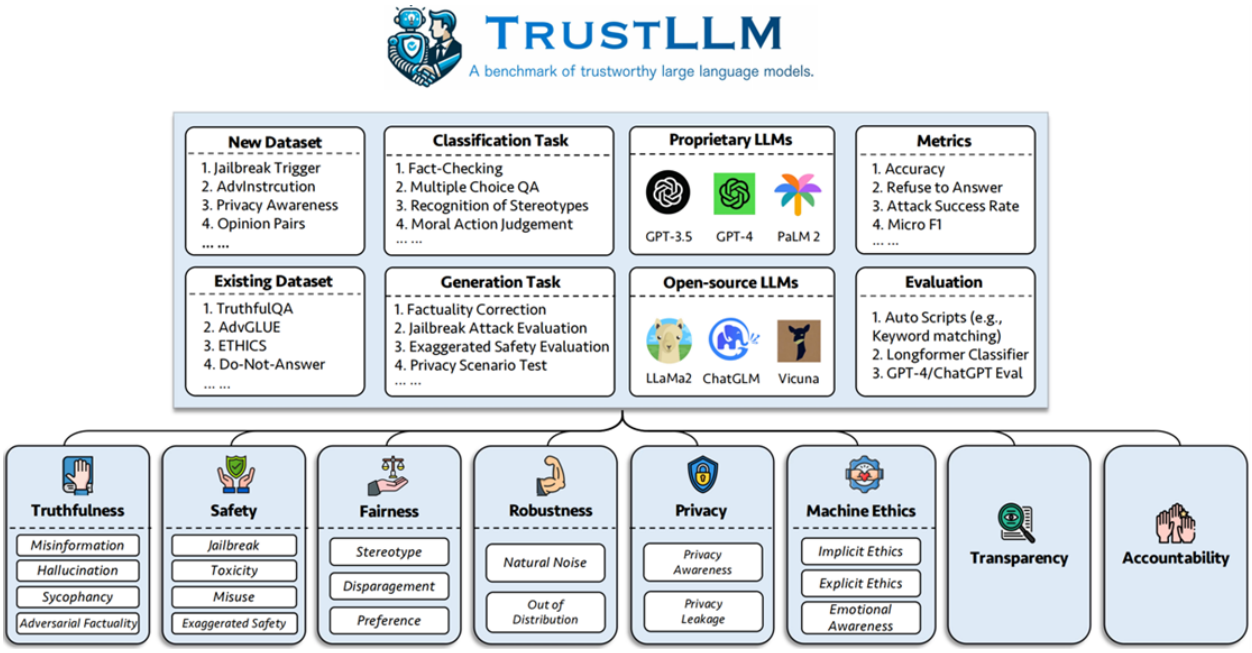

Figure 6: *Schematic of safety benchmarks in TrustLLM [1].*

- **Provably Safe Design:** This area aims to establish design principles that guarantee safety properties from the ground up. By leveraging formal methods, they seek to create AI architectures that are provably resistant to harmful behaviors, thus minimizing risks associated with unforeseen consequences during deployment. Furthermore, improving the reasoning ability of these models is also expected to improve the robustness to adversaries or unseen situations. Some of their contributions have significantly improved instruction-following [4], mathematical reasoning [5], planning [6], and adversarial robustness [7] abilities of models like LLMs and vision language models. Their past work on providing robustness guarantees [8] has won the International Verification of Neural Networks Competition (VNN-COMP) for three consecutive years and is currently being scaled to foundation models.

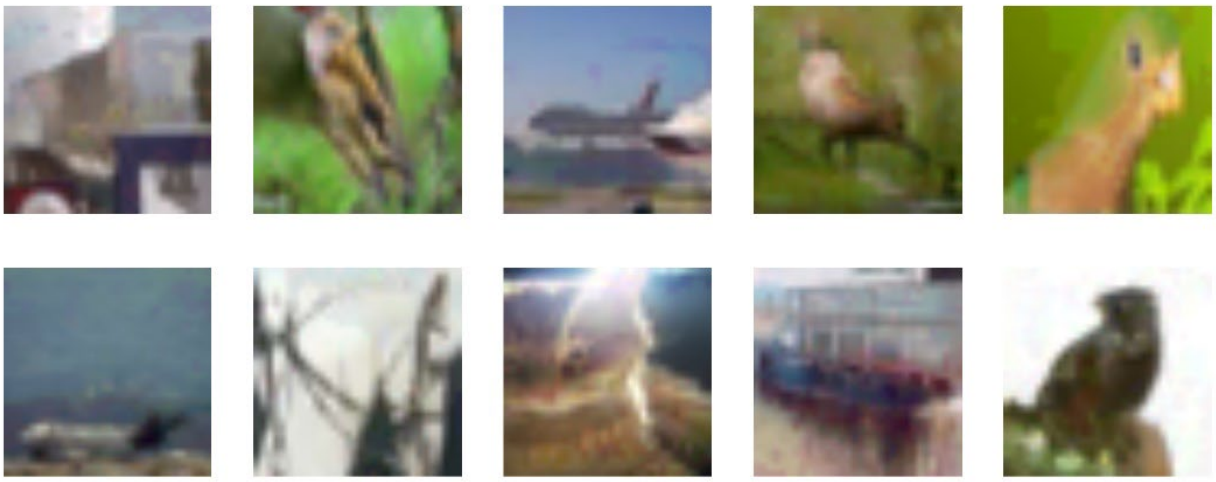

Figure 7: *Subtle attacks on deep neural nets transfer to humans, showing brittleness in human visual perception [7]. Quiz: t.co/UuductMLWR*

- **Robust Monitoring:** This area emphasizes the creation of dynamic oversight mechanisms that continuously assess AI behavior in real-time. The challenge here is twofold: 1) implementing monitoring systems that can analyze vast amounts of data without introducing latency, and 2) detecting subtle anomalies in the behavior of sophisticated models. They are innovating ways to ensure prompt interventions and adaptations to maintain system integrity and alignment with safety goals, even as AI systems continue to learn and evolve. Their work on quantifying confidence in LLM generations [9] topped the LM-Polygraph UQ4LLM benchmark (github.com/IINemo/lm-polygraph). Calibrated confidence provided by their method helps in accurately detecting hallucinations (or fact checking).

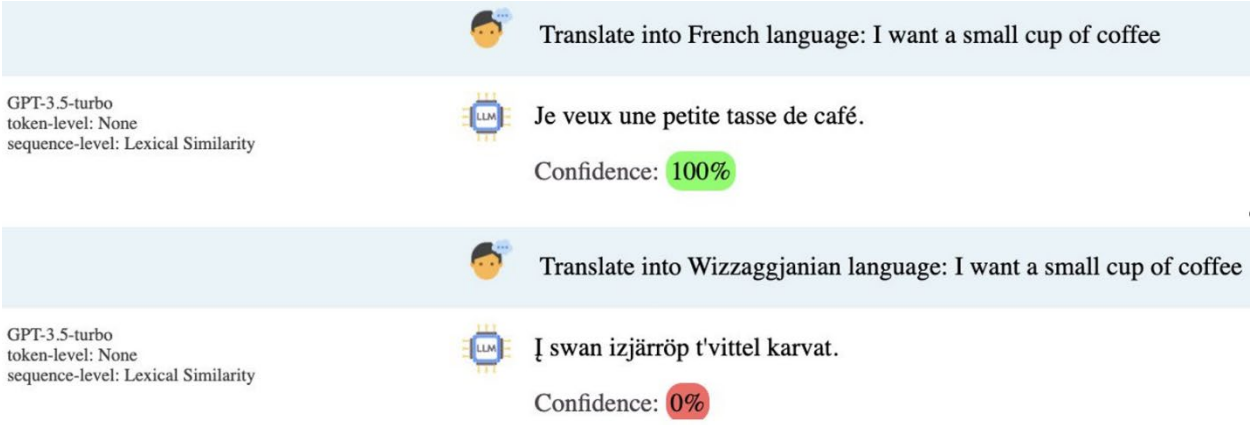

Figure 8: *Calibrated confidence in LLM generation can help with mitigating hallucinations [9].*

Currently, close collaboration is taking place with subject matter experts across different application domains to define and assess what constitutes risk and how to evaluate and mitigate it. This domain-informed approach addresses the unique challenges posed by various mission-critical applications. The techniques devised will be versatile and integrated into a range of ongoing LLNL efforts using LLMs in applications, such as bioassurance, materials discovery, and HPC code translation. Their efforts aim to lay the foundation for next-generation mission-critical AI systems that can provide provable assurances against vulnerabilities and risks, thus enhancing national security and fostering trust in AI across the DOE/NNSA application space.

Validation and Benchmarking:

[1] Y. Huang, L. Sun, H. Wang, *et al*. "Position: TrustLLM: Trustworthiness in Large Language Models." In *International Conference on Machine Learning*, pp. 20166-20270. PMLR, 2024.

[2] ML Commons AI Safety Team. "Introducing v0. 5 of the AI Safety Benchmark from MLCommons," 2024.

[3] J. Duan, R. Zhang, J. Diffenderfer, *et al*. "GTBench: Uncovering the Strategic Reasoning Limitations of LLMs via Game-Theoretic Evaluations." *arXiv preprint arXiv:2402.12348*, 2024.

Improving AI Capabilities:

[4] N. Jain, P.-Y. Chiang, Y. Wen, *et al*. "NEFTune: Noisy Embeddings Improve Instruction Finetuning." In *The Twelfth International Conference on Learning Representations*, 2023.

[5] S. McLeish, A. Bansal, A. Stein, *et al*. "Transformers Can Do Arithmetic with the Right Embeddings." *arXiv e-prints: arXiv-2405*, 2024.

[6] J. Duan, S. Wang, J. Diffenderfer, *et al*. "ReTA: Recursively Thinking Ahead to Improve the Strategic Reasoning of Large Language Models." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 2232-2246, 2024.

[7] B. R. Bartoldson, J. Diffenderfer, K. Parasyris, and B. Kailkhura. "Adversarial Robustness Limits via Scaling-Law and Human-Alignment Studies." In *Forty-First International Conference on Machine Learning*, 2024.

Monitoring and UQ:

[8] K. Xu, Z. Shi, H. Zhang, *et al*. "Automatic Perturbation Analysis for Scalable Certified Robustness and Beyond." *Advances in Neural Information Processing Systems* 33: 1129-1141, 2020.

[9] J. Duan, H. Cheng, S. Wang, *et al*. "Shifting Attention to Relevance: Towards the Predictive Uncertainty Quantification of Free-Form Large Language Models." In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5050-5063, 2024.

Cautionary Tales:

[10] J. Hong, J. Duan, C. Zhang, *et al*. "Decoding Compressed Trust: Scrutinizing the Trustworthiness of Efficient LLMs Under Compression." In *Forty-First International Conference on Machine Learning*, 2024.

[11] S. Das, M. Romanelli, C. Tran, *et al*. "Low-Rank Finetuning for LLMs: A Fairness Perspective." *arXiv preprint arXiv:2405.18572*, 2024.

[12] E. Debenedetti, Z. Wan, M. Andriushchenko, *et al*. "Scaling compute is not all you need for adversarial robustness." *arXiv preprint arXiv:2312.13131*, 2023.

Machine Learning & Applications | Autonomous Multiscale Simulations – Embedded Machine Learning for Smart Simulations

Contact: Jayram Thathachar

The integration of AMS simulation systems represents a significant leap forward in computational science, blending deep learning with HPC to enhance simulation efficiency and accuracy. This innovative approach, developed by CASC researchers Timo Bremer and Jayram Thathachar and team, under an LDRD-SI project initiated in FY22, seeks to unify simulations across different scales and dimensions, leveraging advanced ML techniques to optimize and accelerate multiphysics simulations traditionally hindered by subscale computational models.

At the core of many multiphysics simulations are subscale models that resolve complex phenomena such as equations of state and chemical kinetics. These models are computationally expensive and often become bottlenecks when integrated into larger systems, such as molecular dynamics codes. To address this, the AMS initiative utilizes

deep learning to create surrogate models, which are trained on commonly used physics packages to replicate and replace traditional computational models. This shift not only promises substantial improvements in simulation speed but also introduces challenges in ensuring the accuracy and reliability of these surrogate models, particularly in unforeseen scenarios.

The novel aspect of the AMS framework lies in its autonomous operational mode, which continuously evaluates, corrects, and updates the surrogate models without human intervention. Through a newly developed UQ method for neural networks, the system assesses the accuracy of the models in real-time. If a model is deemed accurate, its output is used directly; if not, the system reverts to the conventional subscale model, gathering data to refine the surrogate model further. This loop of operation ensures that the surrogate models evolve and adapt continuously, enhancing their reliability and performance over time. In a nutshell, AMS replaces an embarrassingly parallel evaluation of some subscale package with a sophisticated facility-wide workflow.

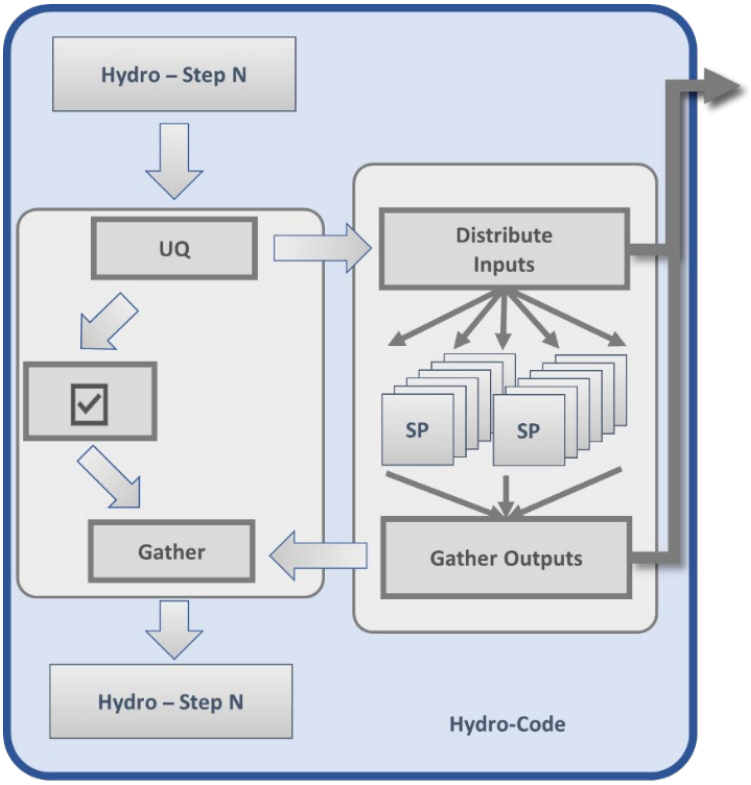

Figure 9: *A workflow example of the AMS framework for hydro-code simulations.*

The project is programmatically organized around three main thrusts: Applications, Machine Learning, and Workflow, which belies the interconnectivity required to provide an effective solution. The cross-thrust coordination includes incorporating the application requirements into the surrogate models as well as integrating UQ and continuous training in the HPC workflow. The solutions have been validated using two subscale codes with diverse characteristics: (1) CHEETAH, a thermochemical computer code designed to predict the properties of energetic materials, including their detonation velocities and energy release, and (2) CRETIN, a multidimensional non-local thermodynamic equilibrium simulation code used in plasma physics applications. The solutions involve newly developed techniques, including:

1. learnable Fourier basis to encode the geometry of the input and output feature space in CHEETAH, analogous to what is used in computer vision to accurately capture underlying "high frequency" information between the input coordinates and the output pixels;

2. a UQ method called Delta-UQ to accurately flag data and model inconsistencies, and failure detectors to categorize the data regimes into different risk levels;

3. a transformer-based encoder-decoder architecture to map CRETIN features into a space in which the correlation between the features is effectively captured using attention mechanisms;

4. physics-informed transformations on coupled emissivity and absorption features to improve the fidelity of energy calculations concerning radiation, and incorporating the mean-free path of photons to target regions of the plasma physics that are more meaningful for the physics. These techniques result in demonstrably more effective loss functions for training that go beyond the simple feature error metrics; and

5. a novel and scalable workflow architecture that manages the complex data flows and computational tasks required by the AMS system.

The AMS project not only promises to revolutionize how simulations are conducted in terms of speed and efficiency but also highlights the potential of deep learning in scientific computing. By automating the training and refinement of models, AMS allows for more dynamic and responsive simulation environments. This capability could lead to broader implications for predictive modeling in various scientific domains, potentially changing how researchers approach complex simulations, and likely set a new standard for the integration of ML and HPC in scientific research.

[1] J. Thiagarajan, R. Anirudh, V. Narayanaswamy, and P.-T. Bremer. "Single Model Uncertainty Estimation via Stochastic Data Centering." *Advances in Neural Information Processing Systems*, 35, pp. 8662-8674, 2022.

[2] J. Thiagarajan, V. Narayanaswamy, P. Trivedi, and R. Anirudh. "PAGER: A Framework for Failure Analysis of Deep Regression Models." *arXiv preprint arXiv:2309.10977*, 2023. To appear in ICML 2025.

CASC Newsletter Sign-Up

Was this newsletter link passed along to you? Or did you happen to find it on social media? Sign up to be notified of future newsletters.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-MI-872687. Edited by Ming Jiang.