msb.ai

Stewardship of Software for Scientific and High Performance Computing with GURU

We propose a new way to accomplish the continued advancement of DOE's ecosystem of scientific software for exascale computing through a shared software stewardship program. A purpose-built autonomous system offers a new opportunity to comprehensively achieve DOE, ASCR, and ECP goals.

MSBAI (dba for Microsurgeonbot Inc.) has created a 'Universal Interface for Simulation called GURU - a cognitive AI assistant that enables untrained users to use specialized simulation software. MSBAI submitted an SBIR Phase I proposal to the Department of Energy (DOE) focused on expanding the utilization of software funded by Advanced Scientific Computing Research (ASCR). MSBAI is responding to this RFI to build upon this foundation to answer a broader spectrum of DOE needs.

Live demonstration of:

1) Multiple Simulation Applications
2) AI Learning Engine that enables GURU to scale to more applications
Video: https://youtu.be/tcB7KB0OKk8

Potential Scope:

- *Training:* Requiring humans to learn a multitude of disparate software interfaces is a waste of time and prohibits mass adoption of these powerful tools and High Performance Computing (HPC).
- *Workforce support:* The point above represents a barrier to enabling a diverse workforce to use these tools and HPC.
- *Infrastructure:* GURU's architecture makes it easy to host from the broadest range of systems. Our modularity and interoperability is a key here.
- *Curation:* Our procedure learning engine is modular, enabling continuous updates to respond to evolving requirements. The symbolic layer of our hierarchical AI learning engine enables the enforcement of accepted standards.
- *Maintaining situational awareness:* GURU centralizes the deployment of all packages from one common client portal, and therefore offers the greatest opportunity for integration that has ever existed. For a given client / institution / group resources such as publications and other information can be made commonly available to all users. Software dependencies can be established once for the entire organization, facility deployment requirements can be satisfied once for the entire organization, and collection of relevant user data becomes straightforward.
- *Shared engineering resources:* The centralized portal enables testing, diagnosis and deployment to be done once for the entire organization. The hierarchical Learning Engine's Program Synthesis layer will minimize the time and resources required to adjust for new compilers, system software and platform versions, and changing package requirements.
- *Project support:* GURU's 'ARTIST' learning engine consists of a hierarchy of Artificial Intelligence methods originally developed to minimize the time and resources required to learn new simulation software. This system offers an entirely new way to achieve efficient and timely support for continued development, emerging hardware and software, comply with best practices, and provide updates with desired new features.

Response to DOE Requests:

1. Software dependencies and requirements for scientific application development and/or research in computer science and applied mathematics relevant to DOE's mission priorities:

We have developed a language based program synthesis capability that enables us to write configuration files and API instructions for a broad range of applications. We support an increasing number of packages in the following three simulation verticals: i) Computer Aided Engineering / Digital Engineering, ii) Trajectory simulation and mission planning, iii) Virtual World immersive training scenario generation. The greatest R&D risks involve too high of a focus on custom work rather than focusing the developments on maturing the system's ability to achieve the flexibility needed to autonomously support dependencies.

2. Practices related to the security and integrity of software and data:

We are a DoD contractor and we are enrolled in the Cybersecurity Maturity Model Certification (CMMC) program to ensure we have adopted the best system architecture and practices to ensure security and integrity of software and data.

3. Infrastructure requirements for software development for scientific and high-performance computing:

We have our own on-prem, and commercial cloud resources. We are also a long time user at the Oak Ridge Leadership Computing Facility and our goal is to continue to grow and mature this engagement and our intent is to deliver GURU in a form that will benefit this and other DOE facilities. We have an increasing number of stakeholders interested in fusing both simulation and experimentation and we would like to gain access to the Spallation Neutron Source and Manufacturing Demonstration Facility at Oak Ridge National Laboratory as well as other DOE facilities as we grow GURU's ability to autonomously set up and run simulations, as well as perform experimental data analysis, and perform comprehensive workflows at scale relevant to testing, development, and computing facilities.

4. Developing and maintaining community software:

We have focused a large percentage of our R&D into GURU's 'ARTIST' learning engine to be able to autonomously maintain and generate code in order to scale up the development/maintenance of community code. The ARTIST hierarchical learning engine has been built to autonomously control and update simulation software packages with an ability to enforce best practices, unlike a general program synthesis system like Github's Copilot (using more than transformers trained from existing corpuses of language and computer code). This is an area where future larger R&D phases offers an opportunity to grow the role of ARTIST to ensure development and maintenance can be scaled up and provide full anticipated user and facility needs.

5. Challenges in building a diverse workforce and maintaining an inclusive professional environment:

There has always been a limit to the number of specialized personnel available to become experts in a variety of solver algorithms, HPC methods, etc. and the number of users who should benefit from these capabilities are orders of magnitude larger. While scaling up the number of people who can be educated in these fields is important to do, the likely feasible scaleup of such new people receiving such training would not be anywhere near the order of magnitude of need. The opportunity of GURU to supplement this need and radically increase the access of general users to simulation and HPC is obvious. This is a fundamental reason we created GURU.

6. Requirements, barriers, and challenges to technology transfer, and building communities around software projects, including forming consortia and other non-profit organizations:

We expect to reach a level of maturity in the ARTIST learning engine, when third parties such as Educational Institutions and Non-Profits will be able to create their own 'agents' that learn new simulation software packages, specialized design/processing/analysis workflow skills to autonomously perform workflows most important to those institutions. We have an opportunity to maximize this new way to achieve inclusivity and equity by making scientific and high-performance-computing software accessible to all, and new funding opportunities focused on this goal would enable us to place more resources and emphasis on achieving this important goal.

7. Overall scope of the stewardship effort:

There is an obvious role for an autonomous system to scale up a great deal of the required activities identified. Explicitly identifying roles (and funding opportunities) for such autonomous system development and maturation should be considered important here.

8. Management and oversight structure of the stewardship effort:

GURU offers DOE labs, Educational Institutions, Nonprofits, an ability to collaborate on a single platform for integrating, orchestrating, and running a broad range of packages and systems, while facilitating the individual needs and focuses of each site. There will be obvious opportunities for organizational bodies to be built to represent these stakeholders and integrate their needs, and coordinate them when they align. DOE will clearly play an important role in this oversight in a manner similar to the way NASA oversees satellite and astronaut launches provided by commercial companies such as SpaceX. A project should be created to demonstrate the deployment of GURU across multiple facilities (e.g., OLCF + SNS + UT).

9. Assessment and criteria for success for the stewardship effort:

DOE should conduct an assessment of the types of workflows most desired to be scaled up, with magnitudes. For example, if it is desired to enable people other than - the developers of ASCR funded packages and highly trained experts at DOE labs - to make regular use of such ASCR packages then DOE should devise a matrix of goals for: i) factor increase in number of users (e.g., 100x, 1000x), ii) profile of the new users - including goals for inclusivity and equity, iii) goals for new systems the users in ii should be able to run on - including gov't systems, on-prem, and commercial cloud. These metrics should clearly identify the level of massive scale up that is implied in this RFI so that the goal is explicit.

10. Other:

Key elements of our proposed solution to achieving the extraordinarily important goals identified herein are summarized below.

Description of GURU

MSBAI created a 'Universal Interface for Simulation called GURU - a cognitive AI assistant that enables untrained users to utilize and manage software ecosystems, including the ability to run specialized simulation software such as those funded by ASCR. GURU enables any user to quickly and easily set up, run, and post-process/visualize simulations or other data in support of project curation, project support, workforce support, and shared engineering resources as outlined in this RFI. Mass-adoption will be achieved by removing the 'expertise-barrier' to: learning simulation software, setting up simulations, and deploying these compute jobs to clusters, commercial cloud, and government systems.

Right now ***only 1%*** of engineers use simulation as a tool in their design toolbox. ***MSBAI will make it 100%***. Massive virtual prototype explorations are rarely done in new product development, because computer-aided engineering (CAE) packages take years to learn, and setup time for a new simulation can be hours of laborious work. We must enable a new user to set up and run thousands of models quickly to evolve virtual prototypes. The DOE has spent nearly $100 million[1,2] in taxpayer money just recently, and decades of development, to advance HPC. While many of the software packages are open source, they are difficult to use and often require an ability to work with source-code, which tends to minimize the potential user base. There is massive untapped potential in the thousands of simulation packages in existence, and the commercial cloud compute that is plentiful and affordable today. Computational physics and HPC needs to be put in the hands of every engineer to begin an American renaissance in innovation and manufacturing.

[1] https://datainnovation.org/2020/06/does-30-million-investment-in-supercomputing-software-will-help-maintain-u-s-top-spot/
[2] https://insidehpc.com/2021/07/doe-funds-28m-for-scientific-supercomputing-research-projects/

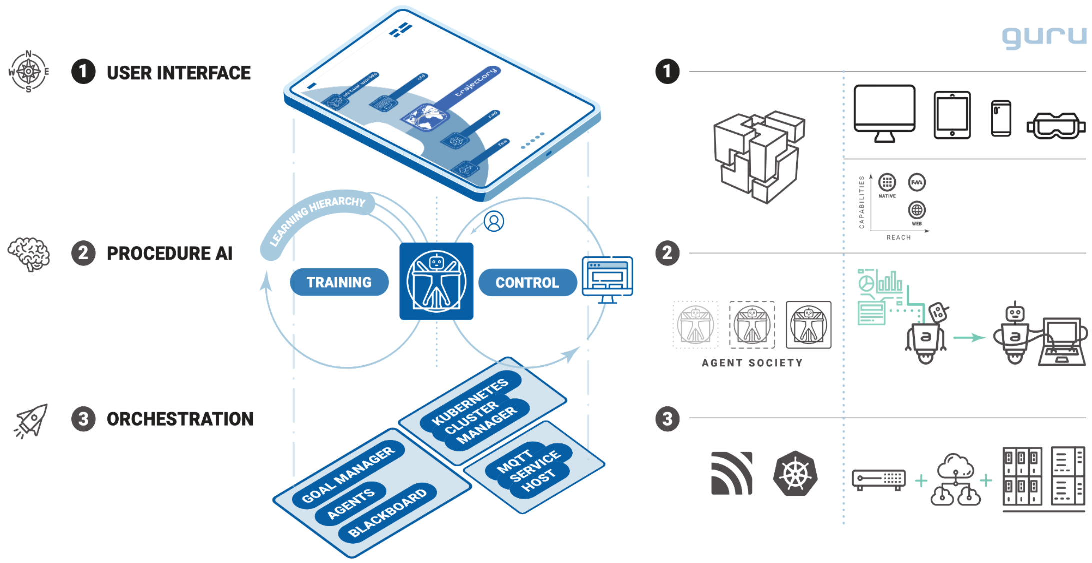

Figure 1: *GURU Enables Mass-Adoption of Simulation & HPC*
Portable modular easy-to-use interface, End-to-end procedure learning AI, Multi-platform deployment

GURU consists of a modular and portable interface that can be used on multiple devices from the browser, a Kubernetes style deployment system that utilizes the MQTT protocol to deploy to a broad range of systems, and a procedure learning & execution system that enables GURU to learn how to run a broad range of simulation packages.

As outlined in Figure 1, GURU's baseline technology is composed of three different levels to optimize the experience and usability of the data suites available to the user:

1. GURU utilizes an intuitive, easy-to-use interface that allows users to easily access and manipulate data on different platforms.
2. GURU's learning engine, Automated Reinforced-learning-guided Tree-based hybrid-Intelligence Synthesis Trainer (ARTIST), leverages a modern procedure-learning AI hierarchy to access the targeted software.
3. GURU adapts and runs the software to fit its operational environment (private servers, HPC, Government system, or Cloud).

GURU is capable of utilizing ACSR-funded software computational physics suites, such as those from the Exascale Computing Project (ECP), and fostering mass adoption within the DOE and the public to ensure greater operational effectiveness.

GURU is a framework to scale up users of ASCR-funded software by addressing two fundamental problems preventing commercial adoption:

- Learning one simulation package takes an engineer months to years — and there are more than 1000 such packages covering a broad range of applications.
- Setting up a single new simulation takes hours of manual work.

GURU has an entirely new, easy to use user interface that is able to play from a browser on computer, tablet, and phone; no competitor is able to do this. We have a containerized orchestration system based on the latest IoT standards and technology that enables our system on the backend to send compute jobs to any system, such as a cloud system or government supercomputer.

GURU leverages the experience of MSBAI's multidisciplinary team, with decades of first-hand experience working on complicated engineering technology development projects, HPC, both technical and business experience in engineering software, and capabilities and experience in Artificial Intelligence. MSBAI believes that GURU will be the solution that the DOE is searching for, and will lead to an increased value in ASCR-funded software for the public.

Technical Discussion

MSBAI's GURU will allow the DOE to capture the immense potential of simulations, saving the organization time and money and increasing the scalability of operations.

- GURU solves this challenge by providing a modern, AI-powered interface that allows users with little to no technical expertise.
- Allows users to harness the power of the HPCs to run and code simulations.
- GURU allows the DOE to finally take advantage of the immense computing power developed and run simulations at a scale that was previously unavailable in the past.
- The increase in simulations will increase DOE safety, operations, and efficiency, while reducing costs and time.

GURU is designed to work within multiple data packages to provide the content with an easy to use interface. This allows the most advanced computational physics to be presented to the user in a readily usable format. This user interface also allows the DOE to write new source code for new simulations. What normally takes hours to format and calculate, GURU can do in minutes. For the DOE, GURU will be scaled to the cloud to allow it to interface with multiple ASCR workflows.

Background

GURU is a highly modular agent-based, blackboard-driven system. At the platform level, the goal manager interprets inputs from the user ('intents' via the UI) and assembles workflows consisting of skills/capabilities trained in agents. We have chosen a common hybrid AI architecture for each agent. The hybrid of symbolic learning and geometric/machine learning enables the system to perform trusted explainable procedures, while sufficiently generalizing to real world requirements.

ARTIST

The following data are expensive, so practical systems must use them sparingly:

- monitored human expert sessions
- large sets of simulations

We therefore developed an approach to assemble and train agents from inexpensive data, and only require sparse datasets of the expensive samples. This enables each module or 'agent' in the system to learn a new specialized task in a practical timeframe. ARTIST utilizes a process that begins with collecting the cheapest 'primitives' data available to build a foundation of learned states/features/actions into sub-workflows in order to minimize the amount of required expensive human expert sessions to monitor, or simulations to run. Key elements of this process include: 'program exploration' — building state and feature graphs, training transformer based Q-value estimation, and a history-aware Reinforcement Learning approach shaped by Distance Partial Rewards + sampling to enable Hindsight Experience Replay.

Current DOE Environment

For decades, the DOE has invested in HPC software packages that operate on large, heterogeneous supercomputers; however, due to the complexity of the software packages, only a select few have had the technical expertise to fully utilize the immense power of the simulations they create. As a result, HPC has been underutilized and the DOE has not fully realized the results of their initial investment in this technology.

While many ASCR software packages are open-source, they are complicated to use. Most of the software packages distribute their source-code from targeting common HPC systems and, as a result, potential software adapters to modernize DOE operations are limited from pursuing current commercial options. Due to the lack of commercial overlap of computing systems, there is largely a lack of commercial interest due to the high barrier of expertise required to install and use these software packages. This severely limits DOE operations as they are unable to fully realize the HPC software packages for scientific discovery and research.

The ASCR software packages have cost taxpayers substantial resources to develop and maintain, and as a result of their complexity, have been underutilized. In order to realize the benefits of the software, the user base needs to grow. The most direct way to grow the user base is to make the software packages easier to use and more accessible to the DOE stakeholders.

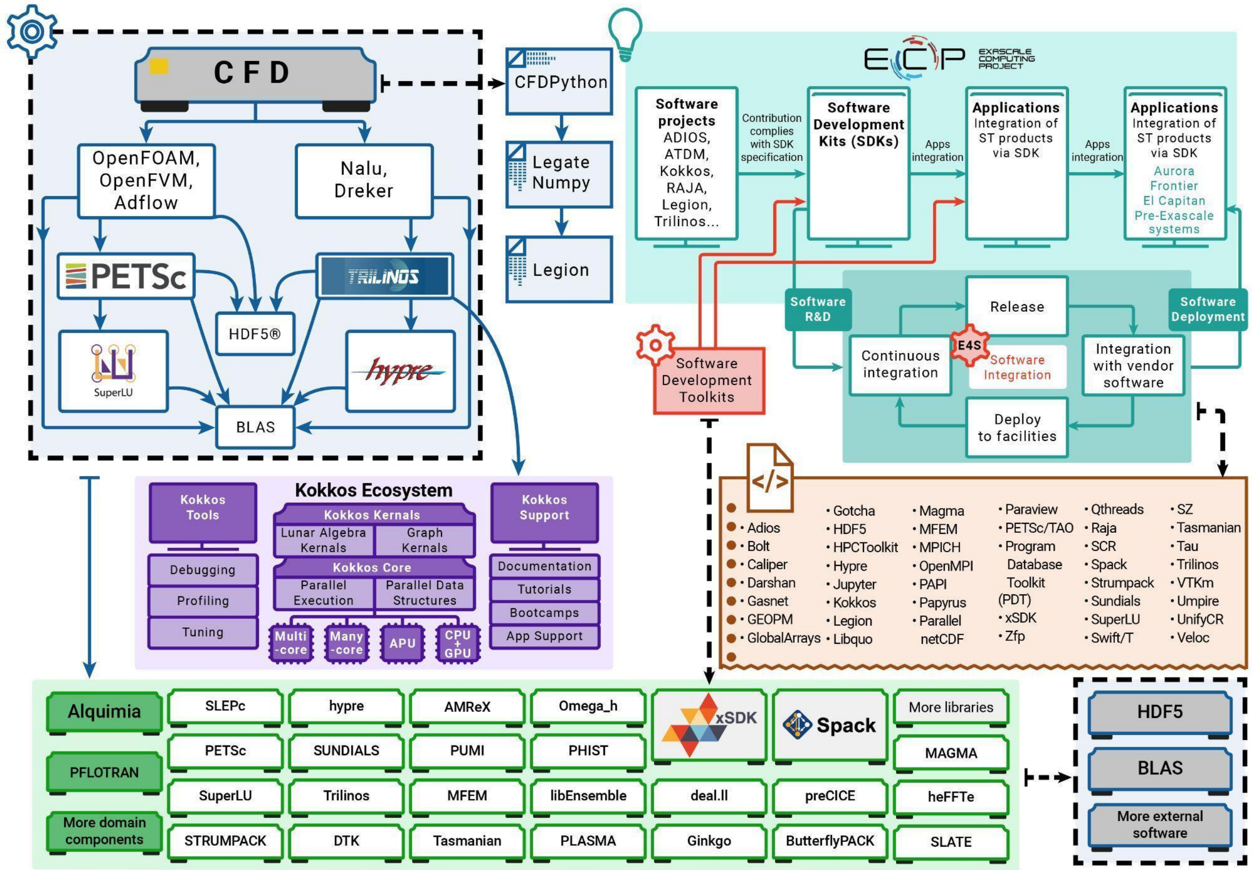

Figure 2: *Super Flowchart of ASCR-Funded Suites that will be Integrated by GURU*

Modern Computing Hardware

Super Computing Hardware has permeated society at large and is accessible, but many of the software packages and libraries that can take advantage of this heterogeneity have remained solely within the HPC ecosystem. As a result, despite having accessible supercomputing, these packages cannot be accessed except by highly-trained DOE individuals.

Prior Work

MSBAI proposed initial R&D focused on CFD applications in our DOE 2022 SBIR Release 1 Phase I proposal. We have built the fundamental capabilities in GURU under DOD contracts, commercial contracts, and investor funding. Our response to this RFI represents a natural expansion on top of that foundation.

E-Visionaries Bring Digital Engineering Revolution:

https://markets.businessinsider.com/news/stocks/evisionaries-bring-digital-engineering-revolution-1030598213

USAF selects GURU to combat hypersonic threats:

https://www.airforce-technology.com/news/usaf-selects-msbais-cognitive-ai-assistant-for-avatar-programme/

MSBAI Team

Allan Grosvenor, Founder & CEO:
20+ years of R&D in aerospace, HPC, engineering enterprise software, and AI
Dustin Sysko, Senior Architect:
15 years in software architecture, dev and leadership
Himanshu Pillai, HPC & Backend Dev:
10 years R&D HPC, Parallel computing, GPUs, Performance Optimization
Dwyer Deighan, Senior AI Engineer:
Graduate level experience in state-of-the art AI methodologies
Joshua Fredberg, Advisor:
30+ years VP bizdev, strategy, marketing leading CAE software co's PTC, Ansys, Aspen

Contact

Allan D. Grosvenor, founder and CEO
allan@msb.ai
2355 Westwood Blvd., Suite 961
Los Angeles, CA 90064
310-954-2049