NIST Special Publication
NIST SP 1320

Driving U.S. Innovation in Materials and Manufacturing using AI and Autonomous Labs

Building a National Center and Testbed for Autonomous Materials Science

Howie Joress
Zachary Trautt
Austin McDannald
Brian DeCost
Aaron Gilad Kusne
Francesca Tavazza

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1320

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Special Publication
NIST SP 1320

Driving U.S. Innovation in Materials and Manufacturing using AI and Autonomous Labs

Building a National Center and Testbed for Autonomous Materials Science

Howie Joress
Zachary Trautt
Austin McDannald
Brian DeCost
Aaron Gilad Kusne
Francesca Tavazza
Data and Artificial Intelligence Driven Materials Science Group
Material Measurement Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1320

August 2024

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

Publication History
Approved by the NIST Editorial Review Board on 2024-07-10

How to cite this NIST Technical Series Publication:
Joress H, Trautt Z, McDannald A, DeCost B, Kusne AG, Tavazza F (2024) Driving U.S. Innovation in Materials and Manufacturing using AI and Autonomous Labs. (National Institute of Standards and Technology, Gaithersburg, MD), NIST SP 1320. https://doi.org/10.6028/NIST.SP.1320

Author ORCID iDs
Howie Joress: 0000-0002-6552-2972
Zachary Trautt: 0000-0001-5929-0354
Austin McDannald: 0000-0002-3767-926X
Brian DeCost: 0000-0002-3459-5888
Aaron Gilad Kusne: 0000-0001-8904-2087
Francesca Tavazza: 0000-0002-5602-180X

Contact Information
howie.joress@nist.gov, francesca.tavazza@nist.gov

Abstract

With the goal of advancing US competitiveness and excellence in the materials and manufacturing industries, we present our vision for the National Center for Autonomous Materials Science. The objective of this center is to enable and promote the use of autonomous methodologies for materials science in industrial applications. Virtually every industry is defined and limited by the materials and processing available to it. The foundation of material science is to elucidate and exploit the relationships between structure, processing, and properties within materials which affect its ultimate performance. Using traditional research approaches this is typically a lengthy, costly, and complex process. New paradigms of conducting research, leveraging cutting edge AI and automation technology, are being developed that can address grand challenges in materials research and development that are otherwise intractable. These new research paradigms additionally accelerate the acquisition of knowledge and understanding of critical materials problems. Unfortunately, most industrial R&D as well as some academic research does not take full advantage of these new research paradigms. The U.S. needs to foster the adoption of these new research paradigms as a national effort in order to accelerate and maintain global technological leadership. Our vision is for the US to support an initiative that would facilitate the industry-wide adoption of these new materials science research paradigms.

Keywords

Autonomous materials science; Self-driving lab; Testbed facility.

Table of Contents

List of Figures

Acknowledgments

The authors thank D Holbrook and J Warren for useful discussion and feedback.

Executive Summary

In response to the Senate Roadmap for Artificial Intelligence Policy in the U.S., we present a vision to empower and build the proposed *"testbed to identify, test, and synthesize new materials to support advanced manufacturing through the use of AI, autonomous laboratories, and AI integration with other emerging technologies, such as quantum computing and robotics."*

Materials Technology Gap: All modern technology is limited by the capabilities of current-generation materials. Currently, U.S. industries rely on methods spanning empirical trial and error to physics- and data-based methods to develop new advanced products. Unfortunately, some legacy R&D approaches may not be competitive in a world currently undergoing an AI revolution. This effort aims to enable U.S. industries to leverage the most powerful paradigms of R&D approaches to develop new advanced and sustainable materials and products. These new approaches expand the limits of autonomous laboratories and scientifically explainable AI to solve our nation's most urgent technical challenges.

Opportunity: Autonomous experimentation (AE), or self-driving laboratories, combines AI and automation while leveraging human intuition and creativity to guide campaigns of experiments. AE can reduce the time and resources needed to discover new materials by orders of magnitude and reduce the time-to-market for critical technologies based on them. Further, AE has the power to address problems that are otherwise intractably complex or ever mutating. Materials R&D involves combining and processing different kinds of raw materials in different ways, and there are more possible combinations than stars in the known universe. This near infinite solution space is the underlying basis of all grand challenges in materials and manufacturing. An initiative is needed to develop tools that enable industry to autonomously navigate this infinite solution space efficiently with automated laboratories. AI has the power to discern complex patterns in large datasets including multiple streams of input materials data. Automation enables vital materials data to be collected rapidly and dynamically. To enable US Industry to maintain technical leadership in developing cutting-edge materials and technologies, the US must become a leader in AE.

Autonomous Experimentation Challenges: While the underlying technology for AI and automation exists in many domains, their application to materials research is still in its infancy. Currently, there is no standardized AE ecosystem for materials R&D, which is a major obstacle to industrial adoption. Creating a standards based ecosystem will dramatically reduce the cost of engineering a platform, as well as reduce the risk of obsolescence by ensuring it is expandable and upgradeable. Relatedly, a defined standard will empower equipment and software vendors to design their products for autonomous integration, reducing their cost and liability associated with bespoke engineering and ensuring a customer base for their products. Beyond these standardization challenges, there are many basic and applied research challenges to increase efficacy of AE, including development of new explainable AI algorithms as well as synthesis methodology and materials metrology.

Goals: Promote US technological leadership by:

1. **Revolutionizing how U.S. industry does materials research and development:** This effort seeks to propel U.S. industries into more powerful paradigms of research, accelerating the discovery and commercialization of new technologies and devices that solve pressing societal needs.

2. **Eliminating the formidable barriers to autonomous technology adoption:** A critical outcome of this effort will be the availability of off-the-shelf, autonomous solutions that will drastically reduce the investment and risk associated with implementing autonomous R&D workflows, which is critical for small- and medium-sized businesses.

3. **Making previously impossible materials "Grand Challenges" tractable and achievable:** This effort seeks to develop tools to navigate near infinite solution spaces in a rapid, intelligent, and meaningful way.

Approach:

1. **Build a new world-class center to host autonomous testbed facilities:** We envision the construction of a new NIST-led national center to work side-by-side with industry, government, academia to develop a standards-based, modular ecosystem of new hardware, software, and methods for AE in materials R&D. This center will host a testbed that will serve as a technology demonstrator, technology incubator, national user facility, and reference data generation platform. This testbed will consist of a set of autonomous-ready scientific instruments for on-demand materials synthesis and materials characterization. The ecosystem will be designed to be readily implementable in labs across the research sector including within industry, government, and academia. This testbed and ecosystem will empower human-AI teaming for materials research.

2. **Develop a modular autonomous laboratory ecosystem:** Just as the internet revolution was the outcome of low-level communication standards, we aim to initiate a laboratory revolution that will be powered by a standards-based modular laboratory ecosystem. In partnership with industry stakeholders, NIST will lead development of this next-generation laboratory ecosystem that is based on and enabled by voluntary consensus standards. These standards will enable components for the ecosystem, produced by industry partners, to be modular and plug-and-play and will reduce the risk for product vendors. NIST and partners will use the testbed as an enabling tool to coordinate and develop standards and protocols for industry-wide interoperability, including sample management, instrument communication, data management, and AI algorithms. Development of new marketable ecosystem components by industry will be incentivized by direct financial support through competitive development grants and other mechanisms. This effort seeks to initiate a revolution in how laboratory infrastructure is developed, commercialized, and used.

3. **Cultivate public-private partnerships and workforce development:** We envision NIST leading the coordination of multi-agency efforts and public-private partnerships to realize the full potential of the AE ecosystem within U.S. industry. This NIST-led center would work to provide organizational frameworks and incentive mechanisms for priority collaborations. These collaborations would revolve around the development of new hardware, software, and algorithms to expand the ecosystem of AE components. Educating the next-generation workforce will be critical to ensure the full and sustainable impact of this AE revolution. The testbed will also serve as a platform for hands-on practical training across a range of educational levels, from vocational to doctoral programs.

Time and Resources: We estimate that this vision will take 10 years to come to full fruition and require $1B for the center.

1. Introduction

1.1. Filling the Materials Technology Gap

As expressed best by Green *et al.* [2]:

> Materials are technology enablers. There would be no skyscrapers without steel girders. There would be no commercial aviation industry without high-strength aluminum alloys and advanced composites. There would be no information age without silicon. There would be no mobile phones without functional ceramics. There would be no solar electricity without photovoltaic materials. There would be no modern medicine without biocompatible soft materials. Virtually every industry is defined and limited by the materials and processing available to it.

The foundation of material science is to elucidate the relationships between structure, processing, and properties within materials and how they affect their ultimate performance. The work of materials design is to harness the knowledge of those relationships to make materials with the needed properties and performance to meet critical societal needs.

The space of possible material processing, the resultant structures, and their respective properties is too large to explore and optimize efficiently with traditional research methods. Most technologically relevant materials are too hierarchically intricate for effective theory-driven approaches, and brute-force trial-and-error approaches suffer from the “curse of dimensionality”, the exponential explosion of possibilities with increasing number of design parameters. Yet, these are still the default approaches to materials discovery and processing optimization problems in many industries. Autonomous laboratories implement hardware and software tools that help navigate the search space of these materials science challenges with more efficacy. The overarching objective of the effort described here is to demonstrate and facilitate the use of autonomous labs for materials science in industrial applications. Our vision is to create a national center to enable and promote autonomous methods for materials science.

1.2. Autonomous R&D: A New Paradigm

To meet the materials needs of the 21st century it is necessary to change the way materials R&D is done. Research is typically categorized into 4 major paradigms of scientific study: (i) empirical science, (ii) theory driven science, (iii) computational science (*e.g.,* simulations), and (iv) data driven science. U.S. industries are at various levels of embracing each paradigm. The federal government has been actively incentivizing advancement of the materials science community for some time through programs such as funding Integrated Computational Materials Engineering (ICME) [3] efforts and the Materials Genome Initiative (MGI) [4]. These initiatives have propelled industry forward in harnessing computational and data driven methods. The 4th paradigm leverages ”big data” and machine learning to make predictions and interpolations from what is already known, often in cases that are too complex to create theory based models. [5] However, many materials problems require traversing a parameter space that has not been traversed before, where even conventional (4^{th} paradigm) "big data" extrapolations will fail. Furthermore, many important materials problems (parameter spaces) are impossible to explore with any of the 4 existing paradigms of scientific discovery.

Currently, most R&D experiments are carried out using independent instruments with a large amount of manual interaction for operations such as sample synthesis, characterization, data analysis, and knowledge extraction (*e.g.,* finding patterns and trends). Autonomous experimentation (AE), colloquially referred to as self-driving labs, offers a new operational paradigm that systematically generates the most informative data toward the end goal of the particular problem at hand, greatly increasing the efficiency of research budgets. AE can leverage cutting-edge advances in automated experimentation and machine learning algorithms, along with theoretical computation, for rapidly acquiring new materials knowledge. This frees human scientists to focus on imbuing the system with intuition and leveraging their creativity to guide the system towards important work. Fig. 1 illustrates this conceptually. The core of an autonomous system is a tightly coupled feedback loop. To start the loop an experimenter frames the problem with a set of objectives and constraints. Prior knowledge including databases of known materials and their properties, physicochemical heuristics, and physical laws can also be provided to the platform. The AI-based agent will then take all of this information and use it to generate a computationally inexpensive representative model of the system, then use it to make predictions across the vast parameter space. Based on the objective and the model, the agent will decide what new data will provide the most critical piece of information to increase its knowledge and reach its objective, typically through some combination of exploration and optimization. The automated system will then perform an experiment or theory based simulation to produce that knowledge. The platform then analyzes, interprets, and stores that data. The AI agent will then look at the update information and make a decision on the next data point. This loop is repeated until sufficient information is gathered to achieve the objective.

1.3. Benefits of Autonomous Methods

Autonomous methods have several advantages over traditional research methods. The use of active learning, along with inclusion of theory-based modeling, can greatly increase the value of each experimental data point collected. Conversely, these methods reduce the amount of low-value data that needs to be collected in order to generate the same knowledge. Particularly in fields like materials science, generating data can be fiscally expensive and resource intensive. Reducing the number of experiments also improves the environmental sustainability of research by reducing the required experimental precursors and experimental waste. Furthermore, AI-based decision-making systems with well-characterized assumptions can reduce the unquantifiable effects of bias on the part of human experimenters, both on decision making and analysis.

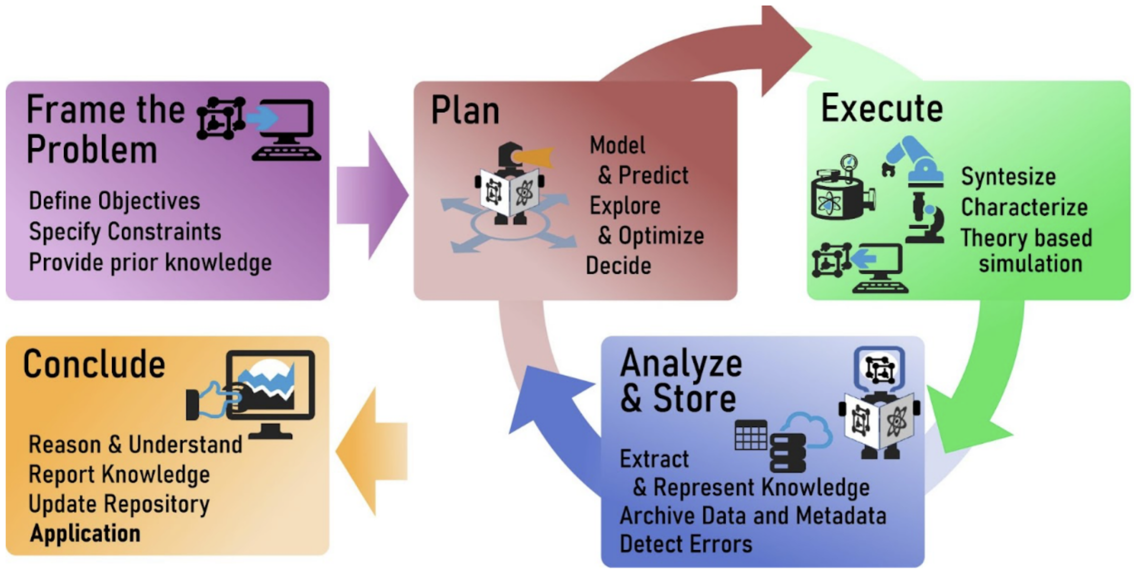

Fig. 1. The major functional components of an autonomous laboratory. Adopted from Ref [1]

Collecting data at scales sufficient to meet societal needs requires agile experimental capabilities to dynamically explore disparate hypotheses. Robotic, automated platforms achieve this by quickly and accurately executing experiments on demand. These automated platforms bring several distinct advantages: First, the data is generated with robotic control, which has the advantage of being more precise, traceable, repeatable, and systematic. Second, the digitally native data produced by automated labs streamlines collection and recording of crucial metadata about the experiment. Third, these laboratories can also run without the need for downtime required by human lab workers. The time and cost savings associated with these platforms is not simply a difference in degree, but in many cases is a difference in kind; making problems which were previously intractably complex tractable.

1.4. Autonomous Methods can Solve Grand Challenges in Materials and Manufacturing

Materials R&D involves combining and processing different kinds of raw materials in different ways, and there are more possible combinations than stars in the known universe. This near infinite solution space is the underlying basis of all grand challenges in materials and manufacturing. There are some technical areas where AE platforms may be the only way to make materials design tractable. Some areas such as high entropy alloys—alloys with multiple elements in large proportions which promise extraordinary properties for applications as varied as hypersonic coatings and water splitting catalysts — exist in such a large compositional and processing design space that collecting informative data on these alloys through standard experimental approaches is intractable. Another place autonomous platforms are crucial is in domains where the problem is constantly mutating. Two examples of this are metals-based additive manufacturing (AM) and circular economy. In the former example, optimization of AM processing parameters is a complex problem. Each unique combination of part geometry, materials, and printing platform requires a new optimization. Efficiently mapping this optimization surface is the only path for being able to design parts without *de novo* experimentation for each part and process. Similarly, recycling of materials as part of a circular economy requires continually optimizing the processing to account for the ever mutating input material stream, which can have strong impacts on the end-product properties and performance. Understanding what effects are caused by the changing input stream, what materials can be made, and what adjustments to the processing are needed is an ongoing design challenge. In all of these cases, the materials physics is too complex to be modeled using currently available, first-principles based approaches, so experimentation is a necessary driving force for innovation.

1.5. Current Challenges to Autonomous Implementation

1.5.1. Lack of standardized protocols

R&D labs that would like to take advantage of the benefits that AE can provide face a variety of challenges, and their associated costs, when it comes to implementing an autonomous platform. Many of these challenges have been highlighted in a recent government wide report: Ref. [6]. There are currently no agreed protocols for how samples should be transferred between instruments, nor standardized protocols for interoperability for data sharing or instrument control. This leaves researchers with two options: seek an outside equipment *vendor* to put together an instrument, typically at great cost, or take the time to cobble together instruments in-house. Because there is no standardization, these platforms are typically designed for a single purpose, meaning the engineering work put into them cannot typically be leveraged for future platforms. Further the platforms are typically built for one purpose and when the scope of inquiry shifts the platform becomes obsolete, necessitating a major redesign.

The current rarity of these systems has led to and is caused by a “chicken and egg problem” between *vendors* and *end-users*. Because of the large investment currently needed to procure or develop an autonomous system, currently few *end-users* are attempting to use one as part of their R&D efforts. As a result, *vendors* do not believe there is a market to create hardware and software to support these types of systems and therefore are reluctant to make the investment to develop these products. In turn, this makes building an autonomous system more challenging and expensive.

A clear example of this was discussed in detail at the October 2023 Autonomous Methodologies for Accelerating X-ray Measurements workshop hosted by NIST and the International Center for Diffraction Data. Several manufacturers of common x-ray diffractometers were present and many described their reluctance to create interfaces necessary to enable autonomous control of their instruments. The clear message was that, at present, there was not significant customer request for these features (but not none). The further concern was that the lack of a standard interface meant different customers would request different interfaces, requiring repetitive custom engineering. A related risk that was discussed is the liability for *vendors* that changes to software that interacts with their interfaces but is out of their control would inversely affect the functionality of their tool; a problem for which customers would demand fixes. Given these challenges, these equipment manufacturers decided that, for the moment, they would avoid offering this feature set.

Another challenge that is slowing the widespread adoption of these technologies is the technology transfer of tools, both hardware as well as software, from the many academic research labs working in the AE space to industry. This is a particular issue for software tools; many tools written by students are often not maintained at a professional level necessary for adoption by equipment manufacturers and industrial users.

1.5.2. Advanced domain specific algorithms

Autonomous physical science is still in its early stages. In addition to the engineering challenges described above, there are several opportunities for further innovation to improve these autonomous workflows. Some of these innovations can boost the scope and speed of what can be accomplished within the ecosystem. These include development of fully automated analysis algorithms for a variety of measurement techniques.

Often one measurement technique is insufficient for fully characterizing a material. If there was a common ecosystem of autonomous tools - as this center aims to foster - then algorithms and workflows could be designed that choose between types of measurement instrumentation, and combine these data-streams to increase the scope and accuracy of the information about that material. Additionally, inclusion of sample fabrication and material synthesis instruments allows for larger scope questions to be asked. Such an ecosystem could autonomously address the open ended tasks of: “find a material composition and processing conditions that maximize a particular property (or suite of properties, *etc.*)” - with the important caveat that the scope of the search space is limited by the regions the instrumentation can access as well as any user defined boundaries.

Tying together parts of the ecosystems that can handle different sample morphologies (*e.g.,* bulk samples, powders, thin films) opens up the possibility of autonomously choosing between different architectures. For the next generation energy storage, do we need materials as nanoscale thin-films or as bulk ribbons? Is it best to study superconductivity using sintered blocks, or as a large single crystals? These questions could be addressed if there were well developed ecosystems of autonomous tools that could handle different sample morphologies.

Eventually, a mature ecosystem of autonomous tools could begin to tackle truly open-ended engineering challenges and scientific questions. An ecosystem that includes not only synthesis and characterization instruments but also theoretical simulations could begin to test boundaries of current scientific understanding - searching for new operant mechanisms to explain observed behavior. What causes batteries to degrade, and how can we make them safer? Do room-temperature superconductors exist, and what is their underlying mechanism of operation? A mature ecosystem of autonomous tools could be leveraged to address grand challenges in engineering: What is the best way to capture energy from the sun? How best to enable resilient infrastructure given increasingly frequent and severe natural disasters? How best to mitigate the climate impacts of atmospheric CO_2?

The realization of such a mature ecosystem of autonomous tools is predicated on the ability to not only sample interchange and software communication protocol standards, but also standards on the trustworthiness and explainability of the algorithms used, and the ability to store and handle the required interconnected and complex data structures.

1.6. The Path Forward

In order to radically transform the way materials R&D is done, we propose a new national center and testbed, which will *"identify, test, and synthesize new materials to support advanced manufacturing through the use of AI, autonomous laboratories, and AI integration with other emerging technologies, such as quantum computing and robotics."* [7] The objective of this center is the development and demonstration of an ecosystem of materials R&D infrastructure that enable autonomous materials workflows. This ecosystem will drastically reduce the cost and risk of companies developing and acquiring autonomous systems by increasing system modularity and interoperability. In turn this will enable a reduction of duplicate engineering through assembly of off-the-shelf, ecosystem-enabled components (both hardware and software) available in the open market. To achieve this we propose the establishment of a new consortium driven center. This center, revolving around an autonomous testbed facility, will establish new documentary standards that will define the ecosystem and fund, through grants, the development of new components for the ecosystem. The testbed facility will serve as a technology demonstrator, test-platform, and user facility.

Similar efforts exist in other countries. For example, Canada has invested in the Acceleration Consortium [8] which is a collection of primarily academic research efforts in autonomous systems. In the United States, the autonomous research programs are generally isolated, either independent academic research efforts or one-off user facility systems such as the Molecular Foundry [9], typically focusing on liquid-based chemistry platforms. In order to maintain technical leadership, the U.S. needs to establish a centrally coordinated effort around facilitating industrial adoption of autonomous systems for materials research. This central coordination is necessary to generate critical mass sufficient to overcome the problems described above, enducing commercial interest in developing autonomous hardware and software, as well as reducing duplication of efforts, ensuring efficient use of research funding. This coordination additionally facilitates the sharing of methods, tools, and best practices.

A U.S. initiative is needed to develop tools that enable industry to autonomously navigate the infinite solution spaces in materials science with efficacy using autonomous laboratories. To enable US Industry to maintain technical leadership in developing cutting-edge materials and technologies, the US must become a leader in AE.

2. Major Goals

The overarching goal of this initiative is to ensure U.S. leadership in the emerging technologies of autonomous "self-driving" laboratories. Autonomous laboratories represent a major frontier that will determine the future economic competitiveness of global industries through the development and discovery of new materials. The goals of this initiative are to promote US technological competitiveness by:

2.1. Revolutionizing How U.S. Industry Does Materials Research and Development

U.S. industries are variously distributed across the previous paradigms of research [5]. This effort seeks to propel U.S. industries into more powerful paradigms of research, accelerating the discovery and commercialization of new technologies and devices that solve pressing societal needs. This will ensure U.S. technological leadership in autonomous laboratories - and the result of that will be leveraged to help ensure technological leadership across many industrial sectors.

2.2. Eliminating the Formidable Barriers to Autonomous Technology Adoption

Poor interoperability and modularity of cyber-physical laboratory infrastructure have plagued the industry for more than a decade, meaning acquiring these platforms is a major investment for companies. A key goal of this effort is to overcome this barrier, resulting in the availability of off-the-shelf, autonomous-enabled hardware and software solutions that will drastically reduce the investment and risk associated with implementing autonomous R&D workflows in industry and beyond, critical for small- and medium-sized businesses.

2.3. Making Previously Impossible Materials "Grand Challenges" Tractable and Achievable

Many technologically relevant materials research problems require innovations in algorithms, deployed in autonomous laboratories, to traverse highly multidimensional input spaces and to address problems that are continually mutating. The third goal of this effort is to develop tools to navigate these high-dimensional spaces in an intelligent and meaningful way. Through these efforts we will enable an increase in exploration and optimization speed of 2 orders of magnitude.

3. A Vision for a National Center for Autonomous Materials Science

To achieve these goals, we envision the creation of a NIST-led center. This center, guided by stakeholders, would provide several integrated functions to create and demonstrate an ecosystem of autonomous tools, which industry could leverage to accelerate their materials R&D cycles. The functions of the center include: testbed facility, standards development, workforce development, and funding body.

Before we describe the center, we define two types of stakeholder roles: *end-users* and *vendors*. *end-users* are entities interested in discovering and developing new materials that can be used to meet societal needs. *Vendors* are organizations that make R&D products to support the work of the *end-users*. In both cases stakeholders will primarily be companies who fall into one category or the other, but may also be academic groups, government labs, or non-profit labs (and in rare cases companies) who may act as both *vendors* and *end-users*.

Approach 1: Build New World-Class Autonomous Testbed Facilities

The core of the center will be a multifunctional autonomous laboratory. This laboratory will comprise various instruments, both for materials synthesis as well as materials characterization, of interest to industrial and governmental partners. These instruments will all be interconnected at both the hardware and software levels, allowing seamless movement of samples and data between instruments with minimal human interaction. The laboratory will serve several purposes and have interaction modes related to each of them, as described below. In addition to the centrally located facility, pipelines for samples and data could be made to other critical materials R&D infrastructure, including x-ray and neutron beamlines and nanofabrication facilities.

Approach 1.1: Development Sandbox

The first role of the testbed facility will be to serve as a sandbox and testbench for developing new hardware, software, and methods. Critically, the building of the test-bed will require a set of protocols for integration of the various instruments and software from a variety of *vendors*. The development of these protocols, described below, will form the basis of the future ecosystem. As the testbed is built up over time, it will continue to serve as a substrate for refining or expanding these standards. Once established, the testbed will allow for testing and development of the interoperability of new products and features within the ecosystem. Also importantly, it can be used to develop and test new methodologies, both for new characterization methods as well as new AI algorithms and new data analysis methods. For instance, the platform can be used to benchmark the efficiency of new recommendation engines for solving cutting edge materials science problems.

Approach 1.2: Exemplar System

The second purpose is to act as an exemplar system. The system will showcase the capabilities and features of the autonomous ecosystem. *Vendors* will be able to gain an appreciation for how the various components interact and how their instruments/software can interact with the ecosystem. The center will also use the platform to educate *end-users* on its capabilities as well as how to frame research questions for autonomous laboratories. If they are interested in these types of AE systems, they can then use the platform as a user (see below) or procure the necessary components for their own lab.

Approach 1.3: User Facility

Third, the physical laboratory will act as a user facility, similar to a synchrotron light source or a nanofab. In this mode, *end-users* would submit proposals to perform experiments on the platform. These proposals may be from, for instance, be small companies or academic groups looking to perform a simple optimization that is beyond the capability of their in-house resources, or they may be from larger companies looking to explore in a deeper way what autonomous R&D can do for their company, to de-risk investment in their own autonomous platform. By default the data generated by these user-initiated experiments can be used to populate a publicly available database of experimental results. Alternatively companies can opt to perform their experiments in a proprietary mode, maintaining their own IP. In concept, this would be a proof of concept for materials science focused "cloud lab" (current instantiations of cloud labs are chemistry based due to the simple interchange of samples), often discussed as a potential future for research. Because facility users do not need to purchase their own experimental equipment, these types of facilities can serve to democratize science and innovation. The typical interaction will be in the form of the user querying the platform to optimize or explore one or more properties in a given region of materials composition and processing space. The algorithms will then select specific materials to synthesize and measure to generate knowledge of that space (to a defined uncertainty) in the most efficient way.

Approach 1.4: Data Generator

One of the advantages of an autonomous laboratory is that the high degree of automation means that the required down-time of the equipment is negligible. The fourth mode of operation of the facility would be as a data-generation platform, adding to the data generated by user-initiated experiments. The platform's planning algorithm can identify knowledge gaps in the testbed facilities database, as well as other publicly available experimental databases, and autonomously perform experiments to fill those gaps. In this mode, the autonomous laboratory will generate high-quality, highly consistent, informative, and publicly accessible data sets. This data can be used to expand the predictive power of models trained on the database, accelerating future efforts.

Approach 2: Develop a Standards-Based Modular Laboratory Ecosystem

A critical step will be to create, in concert with relevant stakeholders, a set of protocols, which can be transitioned to formal documentary standards, upon which the autonomous ecosystem will be based. This fits squarely into NIST's traditional role as a neutral convening body, which can help promote the development of new standards. The goal of these standards is to enable modularity. Modular hardware and digital components will greatly reduce integration costs and will allow autonomous systems to seamlessly evolve with changing research scope, goals, and technologies. Protocols will need to be developed in a number areas, all related to the interaction between various instruments: sample transfer, instrument control, data transfer, and algorithm design. As described above, the testbed facility will be used as a substrate for development of these protocols. Any hardware or software developed for use with the testbed facility will be designed to interact through these protocols.

Approach 2.1: Develop Sample Management Standards

Compared to the many existing AE platforms in the chemistry and bio-pharmaceutical fields, a particular challenge in automated systems for materials science is the need to move around solid samples. Solid materials almost entirely fall into three categories: thin films on substrates, bulk samples, and powders. Each of these has unique sample handling challenges and requirements. The challenge here would be to create a series of universal sample holders that could handle one or more of these sample form factors. Consider the USB-C standard; it defines a variety of aspects (*e.g.,* physical form factor, data transfer protocols and rate, power transfer rate, alternate use modes) which can be supported (or not-supported) at a variety of levels. Similarly, the sample holder interchange standard can describe a variety of aspects, such as sample form factor, size, number of samples, temperature, sample atmosphere. Each instrument and various sample holders can be designed to handle each of these aspects to a different degree. For example, one instrument may be designed to only support thin films while another can support thin films, bulk samples, and powders. Another example is that some sample holders may be designed to be heated and others can only be used at room temperature. Depending on the nature of a particular instrument, the sample holder may be used natively in that instrument (x-ray diffractometer being a likely instance), while in other cases the sample holder will not be conducive to the tools operation (a likely instance being a device to test a material as part of a device like a battery) and the sample will need to be manipulated by the instrument.

Approach 2.2: Develop Instrument Control and Communication Standards

Digital connectivity between AI infrastructure, data infrastructure, and physical laboratory infrastructure is critical for the operation of all autonomous systems. For most of the past century, scientific instrumentation has been designed around human operators. The transition from operation via human intelligence to artificial intelligence has been mired by fragile hacks rather than robust interfaces. Efforts to address these issues have been made in internet of things (IoT) communication protocols (e.g., MQTT [10]) and within some sectors of scientific instrumentation (e.g., SiLA [11], EPICS[12]) as it pertains to controlling and monitoring instruments. This new body will evaluate existing protocols and identify paths forward that address the unique challenges faced by experimental materials hardware and related connected systems.

Approach 2.3: Develop Data and Knowledge Management Standards

In autonomous systems machine-actionable and AI-Ready data are required to inform advanced algorithms and models. Significant progress has been made in elevating the FAIRness [13] of computational data through the establishment of OPTIMADE [14] and related community efforts. Unfortunately, experimental equivalents have lagged behind. This new body will place an increased emphasis on building consensus on the data interchange formats including priority instrument types and knowledge graphs.

Approach 2.4 Develop Algorithm and Model Integration Standards

Currently there is a lack of portability that would enable an algorithm to engage multiple autonomous systems. Some of this challenge is addressed in Approach 2.2 with communication standards, but an additional layer of abstraction may provide significant benefit for future portability. The power of uniform high-level abstractions has been clearly demonstrated by scientific modeling software that enables generalized scientific problem statements to be addressed by diverse underlying computational codes. One example is the Atomic Simulation Environment [15], which provides a generic language for describing problem statements that can be seamlessly solved with a variety of different atomistic codes and methodologies. This new body will develop a comparable high-level interface for experimental materials science, such as an autonomous experimentation environment, which builds upon precursors such as BlueSky [16], ChemOS [17], Hermes [18], HELAO [19], and others. Once an open-source environment is established, outside *vendors* may produce their own proprietary innovation connected via the open standards.

Furthermore, such an autonomous experimentation environment allows for a more rigorous comparison of AI/ML algorithms for scientific tasks, which, in turn, fosters innovation. Some of the most impactful developments in generalist AI/ML have come from open-source libraries, such as PyTorch [20], Keras [21], TensorFlow [22], scikit-learn [23], and many others. Despite their success, they are not specialized to the specific needs of materials and manufacturing, nor are they informed by known physics. Unlike the unstructured data from social networks that the generalist AI/ML tools were developed to handle, scientific data is highly contextualized and imbued with specific physical meaning. Because of this, generalist AI/ML will fail to make meaningful insights on many scientific or engineering problems. Physics-aware scientific AI/ML is a nascent research topic that seeks to leverage the established physical knowledge and structure of the scientific data. However, the nascent scientific AI field lacks any standardization in terms of how it is implemented, documented, or evaluated. This effort will develop an ecosystem of open-source software packages for scientific AI. These packages will be empowered with known physics and have the capability to discover new physics. These packages, together with the autonomous experimentation environment, will provide broad benefits as well as stimulate the development of proprietary AI within US industries.

Approach 3: Empower Public-Private Cooperation via New Consortia

A set of consortia will be established to provide critical guidance and input to the center. These consortia will encompass a range of stakeholders from industry and the federal government as well as academia and other non-governmental organizations, while emphasizing the distinct needs and viewpoints of *vendors* and *end-users*. In particular, other government agencies will play a large role in selecting priority materials domains. We believe this approach will increase adoption of autonomous systems by industry and other end users by reducing risk and the time and financial investment needed to stand-up an autonomous platform. Simultaneously, it will increase the number of *vendors* supporting these platforms by producing a standards-based road map for them to develop products around and directly decrease their cost of product development, and increase their market for these products.

Approach 4: Incentivize Ecosystem Development and High-Risk Activities with Directed Funding Opportunities

While the testbed facility (Approach 1) will be useful to *end-users*, likely the biggest impact to the field will be in the expansion of the autonomous ecosystem. To accelerate the adoption of an AE ecosystem, the *vendors* may need to be enticed to support and develop products using the integration standards developed by the center. The first step of this would be to commission a techno-economic assessment of the gap between currently available instrumentation and software tools in this space and the needs of *end-users*. Once completed, the consortium could decide on the best path forward. As an example, one mechanism that might be employed is the use of grants such as SBIRs. These competitive grants would be guided by the consortium's needs and can be used by companies to modify existing products to meet the standards for use in the autonomous ecosystem or to develop new products to fill gaps in current market offerings. Some grant solicitations will be targeted while others will be open-ended, allowing for industry-driven innovations to prosper. The output of these grants being a saleable product, many of which will be acquired and integrated into the testbed facilities exemplar platform. The bulk of these grants will be to design new hardware, though many will also be used to develop new software solutions including data analysis tools, new modeling methods, or new active learning algorithms. Grant recipients may span sectors, from established companies to startups. In some cases grants may be directed to academic or national lab partners to produce cutting-edge open-source solutions, leveraging their talent to produce new IP and extending the power of the ecosystem and the testbed facility. Some grants aimed at software will take the latest research-level AI algorithms tested in Approach 2.4, and develop them into production-level software packages. While there will be global interest in the autonomous ecosystem, these grants would be preferentially awarded to domestic companies, giving them a competitive advantage in the global market and boosting the American economy.

Approach 5: Next Generation Workforce Education

As with any technological evolution, the advancement of AE technologies will require a workforce with a new set of skills. As such, educating the next-generation workforce will be critical to ensure the full and sustainable impact of this AE revolution. The testbed will also serve as a platform for hands-on practical training across a range of educational levels, from vocational to doctoral programs. NIST has helped support a machine learning for materials research bootcamp for nearly a decade[1], and this testbed will serve to expand this engagement mechanism.

Approach 6: *End-User* support

Many *end-users* looking to adopt the ecosystem in their own labs are likely to seek advice. Center staff will be able to serve as consultants, helping companies to frame their research projects and identifying ecosystem components that would work best to serve their particular needs. Simillary, many *vendors*, may seek help implementing standards into their prodcuts.

4. Impact

This new center will provide a wide and long-lasting impact across the entirety of the materials R&D sector through a sea change in the way research is performed. The greatest impact will be the creation of an ecosystem for autonomous enabled hardware and software that will enable industrial researchers to apply the latest, cutting edge advancements in automation and AI to their research. The interoperability of this ecosystem will be enabled by standards and methods produced by this center. The market availability of components for this ecosystem will greatly reduce the investment in time and money that research laboratories in industry, academia, and government need to stand up these autonomous facilities and its modularity will minimize the risk of obsolescence. The physical platform will also reduce the (apparent) risk for companies, allowing them to try these platforms prior to investing. The grants for development of new autonomous components will greatly reduce the barrier for instrument *vendors* to develop these products and boost their availability on the market. This will also create a competitive advantage for American companies competing to produce autonomous-capable hardware. As these types of platforms become

[1] see https://www.nanocenter.umd.edu/events/mlmr/

more prevalent, companies will begin to compete to develop new products to add to these ecosystems, spurring new innovations in experimental hardware, software, and algorithms. All together, increasing the availability of this technology will decrease the time-to-market of new materials, many for critical end-products, advancing the safety, comfort, and economic well being of the American people and the global population.

5. Whole of Government Approach

Given NIST expertise and mission, we envision NIST having a central role in this new center, but success requires a whole of government approach. NIST's mission is "To promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life." This mission aligns perfectly with the goals and functions of this center including the development of documentary standards that are the core enabler of the autonomous ecosystem. Further, the core mission of the center and NIST is to act as an enabler for US innovation and industrial competitiveness through development of a new way of doing materials R&D. Much of NIST's expertise lies in methods development, as opposed to specific application spaces, conforming with the needs of the center. Perhaps most critically, NIST has a long reputation among industry as a neutral convening body capable of standing-up consortia and other bodies of disparate industrial stakeholders.

Beyond NIST, we envision that the Department of Defense (DOD), Department of Energy (DOE), National Aeronautic and Space Administration (NASA), and National Science Foundation (NSF) will all play a crucial role in the center. DOD, DOE, and NASA all have specific, mission driven materials needs. These needs will shape the priority research areas for the testbed facility. These agencies would coordinate with this new center for the development of tools specific to these applications. Examples might include the Air Force Research Lab (AFRL) or NASA funding for *in situ* extreme environment testing platforms or DOE funding of battery materials testing platforms. Additionally a portion of the overall facility output would be devoted to these priority areas. As mentioned above, there are many opportunities for new technology to support and improve current autonomous methods. NSF is ideally situated to fund research to develop these forward looking technologies that can be later integrated into the ecosystem.

6. Resources

In order to fully realize the impact of this center, we estimate a total budget of $1B over 10 years. As the center matures, government funding may be supplemented by user fees and consortia membership fees. We estimate costs to be:

6.1. Building Acquisition, Construction, and Outfitting: $200M

We propose the construction of a highly specialized new building to host the center and testbed facilities. As demonstrated by facilities like the NCCoE, concentrating necessary features (e.g., laboratories, meeting rooms, high-performance computing, etc.) in a single building will maximize the ability of NIST staff to work side-by-side with industry and other stakeholders. The center will also require construction of specialized lab spaces with necessary utilities (e.g., ventillation, compressed air, cooling water).

6.2. Experimental Hardware Acquisition: $120M

The core of the testbed facility will be a number of newly designed and purchased tools for materials synthesis, materials characterization, robotics, and automation. We envision starting by procuring instruments for 2 high-priority application spaces and scaling up to 4 over time (*e.g.,* advanced additive manufacturing, high entropy alloy design, or design for circular economy – as guided by the consortia). Each application space requires roughly two synthesis tool and five characterization tools; each tool costing on average $4M each, including maintenance, other equipment (e.g., hardware to interconnect the system and spare parts), and duplicates to remove throughput bottlenecks, leading to a total cost of $100M for addressing 4 application spaces.

6.3. Computational Hardware Acquisition: $100M

Complementing the cutting-edge experimental hardware will advanced cyber-infrastructure. We envision procuring high performance computing, storage, and networking systems to integrate directly with the experimental hardware at the facility. Computational infrastructure will include specialized systems, such as infrastructure for training ML models and edge computing. In addition to serving internal need for high-performance computing and data management, we envision procuring dedicated infrastructure for publication of open data.

6.4. Personnel: $180M ($18M per year)

We identify several full time roles to enable full operation of the facility:

- Scientific Staff: Materials domain experts, instrument experts, AI experts including postdoctoral fellows and graduate students
- Technical Support Staff: Technicians, software developers, data engineer, mechanical engineer
- Administrative Support Staff: office managers, conference support, system administrators, user support, program managers

We estimate in total we will need 35 full-time staff members, with a range of educational backgrounds.

6.5. Incentives for Industry Stakeholders: $200M

We anticipate one hardware grant and 3 software grants associated with each piece of hardware acquired for the testbedplus several software grants for layers of software that interconnect the various tools. We anticipate that hardware grants will average $1M and software grants will average $500k.

6.6. Other operating costs: $200M ($20M per year)

These funds will cover operation of the testbed facility including utilities, maintenance, and consumables. This will also cover resources for outreach and education.

7. Timeline

We propose an initial timeline of 10 years. Year 1 and 2 will focus on the initial build-up, including hiring staff and beginning to design the facility. More critically, work would begin to identify consortia members, both *end-users* and *vendors* to determine initial scope of the testbed facility (i.e., two initial application spaces) as well as beginning to gather consensus on the interchange protocols, culminating in a draft standard. Year 3 will kick off the construction of the facility and initial set of grants and acquisitions based around the draft standards. In year 4 and 5 we would begin to receive various physical and software components and begin to commission and integrate them into a unified platform, with the goal of having a working platform by the end of year 5. Years 6 and beyond will focus on using and expanding the platform as a demonstration and user facility while continuing to development and implement new technologies and assist *end-users* on implementing these systems.

References

[1] Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, Schrier J, Billinge S, Buonassisi T, Foster I, et al. (2021) Autonomous experimentation systems for materials development: A community perspective. *Matter* 4(9):2702–2726.

[2] Green ML, Espinal L, Traversa E, Amis EJ (2012) Materials for sustainable development. *MRS Bulletin* 37(4):303–309. https://doi.org/10.1557/mrs.2012.51

[3] Council NR, on Engineering D, Sciences P, Board NMA, on Integrated Computational Materials Engineering C (2008) *Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security* (National Academies Press).

[4] (2011) Materials genome initiative for global competitiveness (National Science and Technology Council), Available at https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf.

[5] Hey T, Tansley S, Tolle K, Gray J (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research). Available at https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/.

[6] Boswell-Koller C, Maruyama B, Warren J, Yang C (2024) Accelerated materials experimentation enabled by the autonomou s materials innovation infrastructure (AMII): A workshop report (Materials Genome Initiative Autonomous Materials Innovation In frastructure Interagency Working Group), Available at https://www.mgi.gov/sites/default/files/documents/MGI_Autonomous_Materials_Innovation_Infrastructure_Workshop_Report.pdf.

[7] Driving U.S. innovation in artificial intelligence: A roadmap for artificial intelligence policy in the united states senate, https://www.young.senate.gov/wp-content/uploads/Roadmap_Electronic1.32pm.pdf. Accessed: 2024-06-14.

[8] U of T receives $200-million grant to support acceleration consortium's 'self-driving labs' research, https://acceleration.utoronto.ca/news/u-of-t-receives-200-million-grant-to-support-acceleration-consortiums-self-driving-labs-research. Accessed: 2024-06-17.

[9] One of the world's premier nanoscience research institutions, https://foundry.lbl.gov/about/. Accessed: 2024-06-17.

[10] Mqtt: The standard for iot messaging, https://mqtt.org/. Accessed: 2024-06-17.

[11] Juchli D (2022) *SiLA 2: The Next Generation Lab Automation Standard* (Springer International Publishing, Cham), , pp 147–174. https://doi.org/10.1007/10_2022_204. Available at https://doi.org/10.1007/10_2022_204

[12] Thuot M, Dalesio L, Clausen M, Katoh T (1996) The success and the future of epics (Los Alamos National Lab.),

[13] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher

SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The fair guiding principles for scientific data management and stewardship. *Scientific Data* 3(1):160018. https://doi.org/10.1038/sdata.2016.18. Available at https://doi.org/10.1038/sdata.2016.18

[14] Andersen CW, Armiento R, Blokhin E, Conduit GJ, Dwaraknath S, Evans ML, Fekete Á, Gopakumar A, Gražulis S, Merkys A, Mohamed F, Oses C, Pizzi G, Rignanese GM, Scheidgen M, Talirz L, Toher C, Winston D, Aversa R, Choudhary K, Colinet P, Curtarolo S, Di Stefano D, Draxl C, Er S, Esters M, Fornari M, Giantomassi M, Govoni M, Hautier G, Hegde V, Horton MK, Huck P, Huhs G, Hummelshøj J, Kariryaa A, Kozinsky B, Kumbhar S, Liu M, Marzari N, Morris AJ, Mostofi AA, Persson KA, Petretto G, Purcell T, Ricci F, Rose F, Scheffler M, Speckhard D, Uhrin M, Vaitkus A, Villars P, Waroquiers D, Wolverton C, Wu M, Yang X (2021) Optimade, an api for exchanging materials data. *Scientific Data* 8(1):217. https://doi.org/10.1038/s41597-021-00974-z. Available at https://doi.org/10.1038/s41597-021-00974-z

[15] Larsen AH, Mortensen JJ, Blomqvist J, Castelli IE, Christensen R, Dułak M, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Jensen PB, Kermode J, Kitchin JR, Kolsbjerg EL, Kubal J, Kaasbjerg K, Lysgaard S, Maronsson JB, Maxson T, Olsen T, Pastewka L, Peterson A, Rostgaard C, Schiøtz J, Schütt O, Strange M, Thygesen KS, Vegge T, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW (2017) The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter* 29(27):273002. https://doi.org/10.1088/1361-648X/aa680e. Available at https://dx.doi.org/10.1088/1361-648X/aa680e

[16] Daniel Allan SC Thomas Caswell, Rakitin M (2019) Bluesky's ahead: A multi-facility collaboration for an a la carte software project for data acquisition and management. *Synchrotron Radiation News* 32(3):19–22. https://doi.org/10.1080/08940886.2019.1608121

[17] Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LPE, Hein JE, Aspuru-Guzik A (2018) Chemos: Orchestrating autonomous experimentation. *Science Robotics* 3(19):eaat5559. https://doi.org/10.1126/scirobotics.aat5559. https://www.science.org/doi/pdf/10.1126/scirobotics.aat5559 Available at https://www.science.org/doi/abs/10.1126/scirobotics.aat5559

[18] hermes, https://github.com/usnistgov/hermes. Accessed: 2024-06-17.

[19] Rahmanian F, Flowers J, Guevarra D, Richter M, Fichtner M, Donnely P, Gregoire JM, Stein HS (2022) Enabling modular autonomous feedback-loops in materials science through hierarchical experimental laboratory automation and orchestration. *Advanced Materials Interfaces* 9(8):2101987. https://doi.org/https://doi.org/10.1002/admi.202101987. https://onlinelibrary.wiley.com/doi/pdf/10.1002/admi.202101987 Available at https://onlinelibrary.wiley.com/doi/abs/10.1002/admi.202101987

[20] Ansel J, Yang E, He H, Gimelshein N, Jain A, Voznesensky M, Bao B, Bell P, Berard D, Burovski E, Chauhan G, Chourdia A, Constable W, Desmaison A, DeVito Z, Ellison E, Feng

W, Gong J, Gschwind M, Hirsh B, Huang S, Kalambarkar K, Kirsch L, Lazos M, Lezcano M, Liang Y, Liang J, Lu Y, Luk C, Maher B, Pan Y, Puhrsch C, Reso M, Saroufim M, Siraichi MY, Suk H, Suo M, Tillet P, Wang E, Wang X, Wen W, Zhang S, Zhao X, Zhou K, Zou R, Mathews A, Chanan G, Wu P, Chintala S (2024) PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)* (ACM). https://doi.org/10.1145/3620665.3640366. Available at https://pytorch.org/assets/pytorch2-2.pdf

[21] Keras 3: Deep learning for humans, https://github.com/keras-team/keras. Accessed: 2024-06-17.

[22] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org Available at https://www.tensorflow.org/.

[23] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.