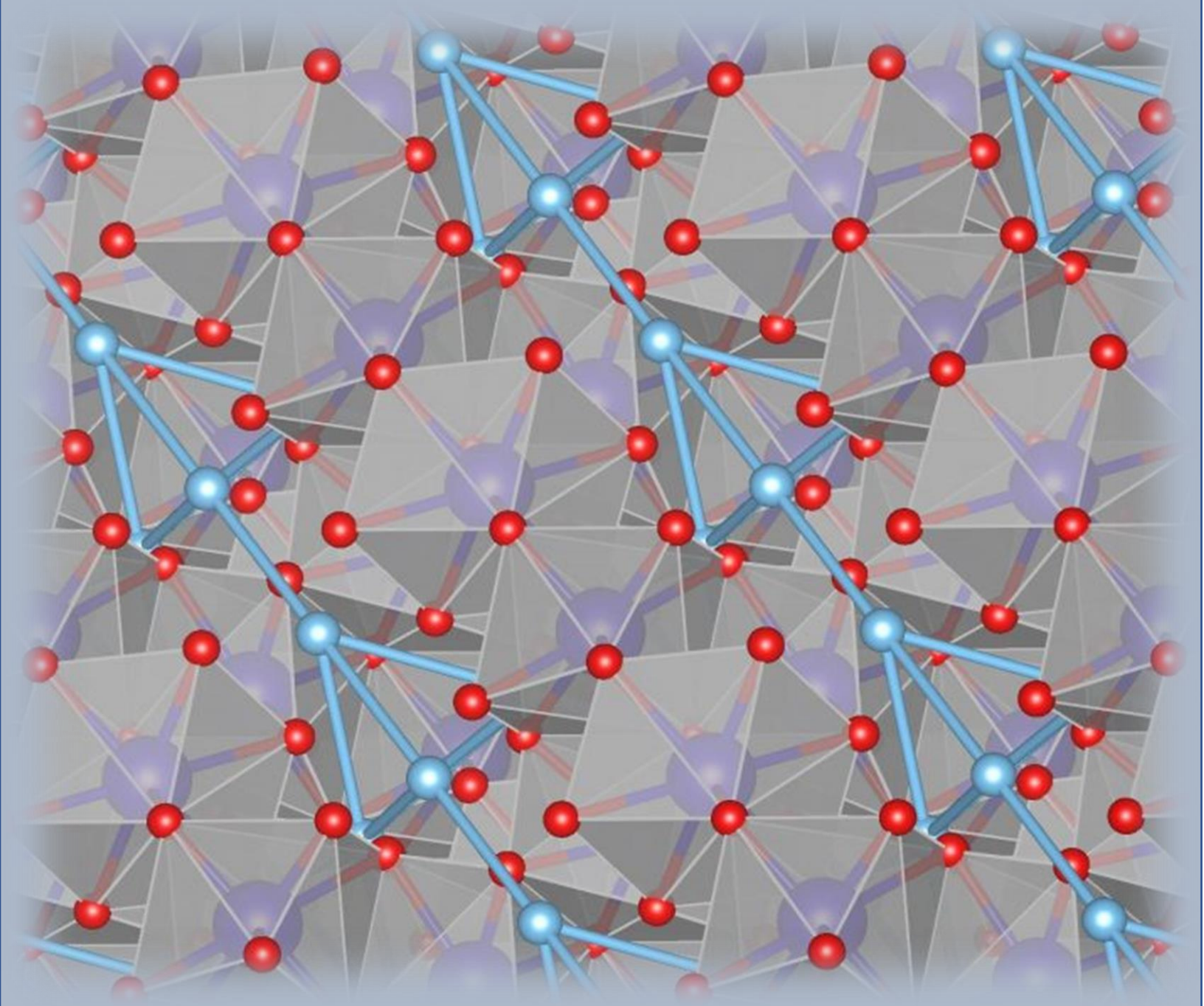
Opportunities and Challenges from Artificial Intelligence and Machine Learning for the Advancement of Science, Technology, and the Office of Science Missions



A report for the Advanced Scientific Computing Advisory Committee from the Subcommittee on Artificial Intelligence, Machine Learning, Data-intensive Science and High-Performance Computing

Chair: Tony Hey

September 2020

Caption:
The cover image shows the crystal structure of $Dy_2Ti_2O_7$ in which the magnetic ion Dy^{3+} of the rare earth element Dysprosium (shown in cyan) occupies sites on a geometrically frustrated corner-sharing tetrahedral network. AI/ML methods were used to solve for the couplings in the material and to identify the origin of its freezing into a glassy state. Neural networks were used to extract the phase and couplings in the system from diffuse neutron scattering data by solving the inverse scattering problem. The strong single-ionic anisotropy of Dy^{3+} ions defined by their molecular environment restricts the effective magnetic moment to align either inward or outward. The magnetic moments impose an ice-rule where low-energy spin states are restricted to a two-in and two-out configuration for each tetrahedron and the breaking of this ice-rule creates two fractionalized magnetic monopoles with opposite polarity.

Acknowledgement:
Anjana Samarakoon and Alan Tennant, Oak Ridge National Laboratory.

U.S. DEPARTMENT OF ENERGY | Office of Science

Table of Contents

Executive Summary 6
Introduction 6
Context 7
Key Findings 8
Recommendations for DOE's Office of Science 12
Report 16
1. Introduction and Background 16
2. Charge Letter to ASCR 17
3. Subcommittee Information Gathering Activities 17
4. DOE as the lead agency for AI/ML applied to Facilities Science 19
5. Opportunities and challenges from Artificial Intelligence and Machine Learning for the advancement of science, technology, and Office of Science missions 20
6. Strategies for the DOE Office of Science to address the challenges and deliver on the opportunities 21
6.1 Introduction 21
6.2 AI Applications 23
6.3 AI Algorithms and Foundations 31
6.4 AI Software Infrastructure 37
6.5 New Hardware Technologies for AI 40
6.6 Instrument to Edge Computing 41
6.7 AI/ML Workforce: Training, Focusing, and Retention 42
6.8 University Partnerships 44
6.9 Collaboration with Industry 45
6.10 Inter-Agency Collaboration 46
6.11 International Collaboration 47
6.12 Importance of ASCR's long-term Applied Mathematics and Computer Science Research Programs 49
7. Summary of Conclusions 49
Figures 51
Figure 1: AI, Machine Learning, Deep Learning in a Nutshell 51
Figure 2: What is a Data Scientist? 52

Figure 3: Structure of SC *AI for Science* 10-year Initiative 53
References and URLs..... 54
Appendix A: Charge Letter 56
Appendix B: Subcommittee Members..... 58
Appendix D: List of Acronyms 64
Acknowledgements..... 65

“AI won’t replace the scientist, but scientists who use AI will replace those who don’t.”

Adapted from a Microsoft report, “The Future Computed”

Executive Summary

Introduction

In February 2019, the President signed Executive Order 13859, *Maintaining American Leadership in Artificial Intelligence* [1]. This order launched the American Artificial Intelligence Initiative, a concerted effort to promote and protect AI technology and innovation in the United States. The Initiative implements a government-wide strategy in collaboration and engagement with the private sector, academia, the public, and like-minded international partners.

Among other actions, key directives in the Initiative called for Federal agencies to:

- Prioritize AI research and development investments,
- Enhance access to high-quality cyberinfrastructure and data,
- Ensure that the US maintains an international leadership role in the development of technical standards for AI, and
- Provide education and training opportunities to prepare the American workforce for the new era of AI.

The mission of the Department of Energy (DOE) is to ensure America's security and prosperity by addressing its energy, environmental, and nuclear challenges through transformative science and technology solutions. In terms of Science and Innovation, the DOE's mission is to maintain a vibrant US effort in science and engineering as a cornerstone of our economic prosperity with clear leadership in strategic areas.

From July to October in 2019, the Argonne, Oak Ridge, and Berkeley National Laboratories hosted a series of four ***AI for Science*** Town Hall meetings in Chicago, Oak Ridge, Berkeley, and Washington DC. The four meetings were attended by over 1300 scientists from the 17 DOE Labs, 39 companies, and over 90 universities. The goal of the Town Hall series was *'to examine scientific opportunities in the areas of artificial intelligence*, *Big Data, and high-performance computing (HPC) in the next decade, and to capture the big ideas, grand challenges, and next steps to realizing these.'* The discussions at the meetings were captured in the final report of the ***AI for Science*** Town Hall meetings [2].

In response to a charge letter from the DOE's Office of Science (SC), the Advanced Scientific Computing Research (ASCR) program asked its Advisory Committee (ASCAC) to establish a subcommittee to explore the scientific opportunities and challenges arising from the intersection of Artificial Intelligence (AI) and Machine Learning (ML) with data-intensive science and high performance computing. Specifically, this ***AI for Science*** subcommittee was asked to:

- *Assess the opportunities and challenges from Artificial Intelligence and Machine Learning for the advancement of science, technology, and the Office of Science missions.*
- *Identify strategies that ASCR can use, in coordination with the other SC programs, to address the challenges and deliver on the opportunities.*

This report is the result of the Subcommittee's investigation of these charge questions. To set the context a summary of AI, ML and Deep Learning is included here along with a characterization of different roles for data scientists. This executive summary reports the subcommittee's key findings and recommendations.

Context

The term Artificial Intelligence was coined by John McCarthy for a workshop at Dartmouth College in New Hampshire in 1956. At the workshop, McCarthy introduced the phrase 'Artificial Intelligence' which he later defined as [3]:

> *'The science and engineering of making intelligent machines, especially intelligent computer programs.'*

By contrast, the field of Machine Learning is less ambitious and can be regarded as a sub-domain of artificial intelligence [4]:

> *'Machine learning addresses the question of how to build computers that improve automatically through experience. It is one of today's most rapidly growing technical fields, lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science. Recent progress in machine learning has been driven both by the development of new learning algorithms and theory and by the ongoing explosion in the availability of online data and low-cost computation.'*

Finally, Deep Learning neural networks are a subset of Machine Learning methods that are based on artificial neural networks (ANNs) [5]:

> *'An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron that receives a signal then processes it and can signal neurons connected to it. The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds.'*

The artificial neurons in these networks are arranged in layers going from an input layer to an output layer with connections between the neurons in the different layers. Deep learning neural networks are merely a subset of such ANNs with very large numbers of hidden layers. On the ImageNet Image Recognition Challenge, the 2015 competition was won by a team from Microsoft Research using a very deep neural network of over 100 layers and achieved an error rate for object recognition comparable to human error rates [6]. Figure 1 tries to capture the essence of this AI, Machine Learning, and Deep Learning hierarchy [7]

Figure 2 attempts to define three different roles for a data scientist [8]. The first role is that of a data engineer who is expert at operating close to the computers, instruments, and sensors that generate the data. The second role is that of a data analyst who uses advanced statistics and AI/ML methods to explore the experimental data sets and assist the researcher to extract new science. Finally, in this classification, there is a third role of data curator who is expert in managing large data sets, curating the data with suitable metadata for re-use, and later archiving. All three of these aspects of data science are relevant for the proposed ***AI for Science*** initiative.

Key Findings

- **Finding A**

The growing convergence of AI, Data, and HPC provides a once in a generation opportunity to profoundly accelerate scientific discovery, create synergies across scientific areas, and improve international competitiveness.

Science and computing are now in an era of post-Moore's Law silicon technologies and there is an urgent need for a sea-change in the programmability and productive use of increasingly complex/heterogeneous systems and the seamless integration of data, algorithms, and computing resources. Doing so will help manage the challenges of Big Data, carrying out science at scale using DOE's most advanced facilities, leverage the workforce at the Labs, and set the stage for the emergence and development of robust and reliable AI systems with the ability to learn for themselves in domain-science specific areas.

- **Finding B**

Science can greatly benefit from AI methods and tools. However, commercial solutions and existing algorithms are not sufficient to address the needs of science automation and science knowledge extraction from current and future DOE facilities and data.

Current AI solutions can be successfully applied to conduct a variety of data analyses. However, new algorithms, foundations, and tools are essential to addressing unique science concerns in a broad spectrum of science applications. AI algorithms need to be able to deal with sparse, heterogeneous, and un-labeled data sets that are often expensive to collect and archive and be able to generate models that incorporate domain knowledge and physical constraints. AI-enabled experimental design and control are necessary for optimal use of DOE facilities. In the science context, AI methods need to have provable correctness and performance, be able to expose biases, and to quantify uncertainties, errors, and precision.

- **Finding C**

Adopting *AI for Science* technologies throughout the Office of Science will enable US scientists to take advantage of the tremendous new advances in the DOE's scientific user facilities.

The DOE's Office of Science provides US researchers with access to the largest and most diverse suite of scientific experimental facilities in the world – from X-ray synchrotrons and neutron sources to integrative genomics and atmospheric radiation facilities – as well as to the world's most capable high performance computing facilities. Upgrades to these user facilities and new nuclear physics facilities coming online now and over the next decade will dramatically increase the amount of new data produced across all of the scientific domains supported by the Office of Science, posing new challenges and new opportunities. Science-aware AI technologies will allow us to extract information and scientific understanding from these tremendous new data sources.

- **Finding D**

Realizing the potential for a generational shift in scientific experimentation at the DOE Laboratories due to science-driven AI/ML technologies requires far more than simply compute power and encompasses the full spectrum of computing infrastructures, ranging from ubiquitous sensors and interconnectivity across devices to real-time monitoring and data analytics, and will require a concerted and coordinated R&D effort on AI/ML algorithms, tools, and software infrastructure.

Across the SC programs, scientific applications of Artificial Intelligence (AI) and Machine Learning (ML) can build on the power of sensor networks, edge computing, and high performance computers to transform science and energy research in the future. Given the highly specialized nature of many DOE facilities and scientific research domains, it is not possible to rely solely on third-party AI/ML research and development (R&D) for this transformation. The DOE will need to build its own R&D programs that focus on the most challenging science-driven applications. Software infrastructure will be required that combines leadership in AI/ML tools and algorithms with the DOE's traditional strengths in simulation and modeling technologies and that can execute on new computing platforms capable of high performance on both types of applications. The anticipated returns will help ensure that the US continues to maintain and enhance leadership in both data-intensive science and high performance computing.

- **Finding E**

The DOE Labs are uniquely positioned to integrate AI/ML technologies across a host of scientific challenges thanks to the enviable culture of co-design teams consisting of scientific users, instrument providers, theoretical scientists, mathematicians and computer scientists that has proven so successful in the Exascale Computing Project.

The subcommittee, therefore, sees a compelling need for AI/ML technologies to be incorporated into all of the DOE's scientific research capabilities in order to effectively support the Office of Science's missions in energy, national security, fundamental sciences, and the environment. DOE's National Laboratories, together with US university and industry partners, have the necessary assets to initiate a large-scale program to accelerate the development of such capabilities and the necessary workforce to not only meet their SC mission needs but also benefit all of DOE's activities.

- **Finding F**

The impact of a DOE-driven AI/ML strategy for science will have national implications far beyond the Office of Science and will drive new industrial investments, including accelerating engineering designs, synthesizing materials, and optimizing energy devices, as well as advancing hardware and software computing capabilities.

The benefits to the nation in developing powerful and broad-based ***AI for Science*** capabilities in the DOE Laboratories will extend well beyond the DOE's programs. The development of comprehensive AI/ML capabilities will benefit other government agencies and a broad range of industries in this country, including energy, pharmaceutical, aircraft, automobile, entertainment, and others. More powerful AI capabilities will allow these diverse industries to more quickly engineer new products that can improve the nation's competitiveness. In addition, there will be considerable flow-down benefits that result from meeting both the hardware and software AI challenges. Initiating a major program focused on applying AI/ML technologies to the DOE scientific challenges would be likely to lead to significant gains in US competitiveness in several critical areas and technologies.

- **Finding G**

A workforce trained in advanced AI/ML technologies would play a pivotal role in enhancing US competitiveness.

The training, focusing, and retention of a cadre of young people, experts in both inventing and delivering the techniques and technologies of AI/ML for science and engineering applications, is critical to the success of the ***AI for Science*** agenda. The Office of Science DOE Laboratories can play a key role in cooperation with the National Science Foundation (NSF). Over the past 20 years, the Information Technology (IT) industry has expanded dramatically, driven by e-commerce, social media, cloud services, and smartphones. In recent years, the emergence of the Internet of Things (IoT), the widespread deployment of healthcare sensors, increasing industrial automation, and the development of autonomous vehicles have further expanded the domain of AI/ML data analytics and services. In response to these growing workforce demands, most students are now trained in software tools and techniques that target commercial opportunities. At present, commercial tools are rather generic and not well-targeted to scientific applications. An ***AI for Science*** initiative would deliver scientific AI/ML tools and environments appropriate for training a new generation of scientists and engineers.

- **Finding H**

Partnering with other Agencies and with international efforts will be important to deliver on the ambitious goals of an AI for Science initiative.

The NSF and NIH, the two other major science-focused funding agencies in the US, also have or are planning, major investments in AI/ML programs for their scientific domains. In several areas there are clear synergies of research interest and the DOE should explore possible mechanisms

for collaborative projects with other agencies such as NIST and DOD in any DOE ***AI for Science*** initiative.

Other countries have also recognized the potential benefits of applying AL/ML technologies to science. The subcommittee believes that there would be a benefit in the DOE collaborating with 'like-minded international partners' on aspects of an ***AI for Science*** research agenda that are likely to be of mutual benefit.

Recommendations for DOE's Office of Science

1. Creation of a 10-year *AI for Science* Initiative

In order to create the world-leading AI systems and applications needed to drive scientific productivity and discovery in science and technology dramatically beyond that achievable with traditional scientific supercomputing, we recommend that the DOE Office of Science start a ten-year program to develop an ambitious ***AI for Science*** initiative, as recommended in the recent PCAST report [9]. This program should encompass foundational research into new, science-aware AI methodologies, specifically designed for DOE mission-critical challenges, and AI solutions that can be deployed in operational settings at leading DOE research facilities. The initiative should provide a clear, guided roadmap from research to deployment. The DOE laboratories can play a key role here, offering leading-edge exascale supercomputers and large experimental facilities generating increasingly large scientific datasets, as well as providing critical expertise in mathematics, computer science, and experience with DOE mission-specific applications. No other agency has the breadth, critical mass, or recent large project management experience to undertake this cross-disciplinary ***AI for Science*** challenge. However, there is a clear case for the benefits of collaboration with other agencies and other countries, to leverage existing expertise to maximum advantage. Partnerships with other funding agencies and other countries are therefore strongly encouraged.

2. Structure of an SC *AI for Science* Initiative

It is recommended that this ***AI for Science*** initiative be structured around four major AI R&D themes:

- AI-enabled applications
- AI algorithms and foundational research
- AI software infrastructure
- New hardware technologies for AI

The subcommittee believes that this ten-year ***AI for Science*** initiative should be funded at the same scale as the successful Exascale Computing Initiative (ECI) and Exascale Computing Project (ECP). Essential for the success of such an initiative is that the work of these four themes must be closely-coupled in a manner similar to that used in the ECP, as the advances and improvements in one area can inform advances and improvements in other areas.

Figure 3 illustrates an overview of a possible roadmap for such an ***AI for Science*** initiative. As for the ECI and ECP, the roadmap for this proposed AI for Science initiative envisages an initial 'incubation' research phase of coordinated projects with co-design centers connecting the four major themes. Partnerships across all Office of Science domains, with participation from universities and private industry, would be initiated early in the program. The goal of this research phase is to specify the application grand challenges and AI/ML tools and services required as deliverables in the more focused project R&D and Deployment phases, where broad

engagement of the DOE research community becomes critical. Since these applied R&D and Deployment phases will inevitably generate new questions and challenges, having the research phase continuing and overlapping with the R&D and Deployment phases will significantly increase the chances of success for the ***AI for Science*** Project.

3. An Instrument-to-Edge Initiative

The subcommittee believes that ASCR, in close cooperation with BES and with the other science programs in the Office of Science, should work with scientists, users, and the broad academic community to define requirements, conduct research, competitive procurement and design a highly integrated end-to-end system and software stack that connects instruments at the edge to the needed AI computing resources. Integrating national and global data sources (large scale experimental facilities, observational networks terrestrial & space-based, etc.) poses unique opportunities and challenges that require addressing foundational research in the context of leading-edge scientific experiments. Integrated systems for acquiring, analyzing, transforming, storing, and maintaining scientific results, capturing provenance, and contributing broadly accessed analytical workflows within DOE supported computational infrastructure could be transformative. There are, however, severe challenges that will need to be confronted in terms of privacy, security, commercial licensing of data, and integrated data services.

Building on ASCR's co-design experience in ECP, application users, software infrastructure developers, AI/ML researchers, and Lab and industry hardware specialists should be encouraged to define, develop, and contribute to a common software stack for AI/ML Edge computing resources across the different facilities. The software infrastructure should support some generic services at the facilities but also allow the easy creation of specialized AI-based software pipelines specific to the facility and capable of supporting coupling to particular instruments in some cases.

4. Training, focusing, and retention of AI/ML workforce

Industry, national laboratories, government, and broad areas of academic research are making more use than ever before of AI, ML, and simulation-based decision-making. This trend is apparent across many domains such as energy, manufacturing, finance, and transportation. These are all areas in which AI is playing an increasingly significant role, with many more examples across science, engineering, business, and government. Research and innovation, both in academia and in the private sector, are increasingly driven by large-scale computational approaches using AI and ML technologies. With this significant and increased use comes a demand for a workforce versed in technologies necessary for effective and efficient AI/ML-based computational modeling and simulation and big data analytics, as well as the fundamentals of AI/ML algorithms. Graduates with the interdisciplinary expertise needed to develop and/or utilize AI techniques and methods in order to advance the understanding of physical phenomena in a particular scientific, engineering, or business field and also to support better decision-making are in high demand.

A strong research program will crucially rely on a complementary education and skills component, which is as important as providing adequate infrastructure support. As emphasized in the ASCR ECP Transition report [10], this is also a timely and important opportunity to focus SC efforts to create a more diverse and inclusive workforce. A continuing supply of high-quality computational and data scientists available for work at DOE laboratories is of vital importance. In high performance modeling and simulation, for example, the DOE Computational Science Graduate Fellowship (CSGF) program has successfully provided support and guidance to some of the nation's best scientific graduate students, and many of these students are now employed in DOE laboratories, private industry, and educational institutions. We need a similar fellowship program to meet the increasing requirement for computational and data scientists trained to tackle exascale and data-intensive computing challenges. In addition, the DOE SC should explore the possibilities for collaboration with the NSF about the provision of relevant training programs in AI/ML technologies and their application to science.

5. Inter-Agency collaboration

Although the NSF has long been regarded as the lead agency for fundamental AI research, DOE is clearly the lead agency for research involving the intersection of 'Big Science, Big Data, and Big Computing.' DOE has not only established national and international leadership in HPC and supercomputing but is also a leader in the application of AI/ML technologies to the very large scientific datasets generated by their large-scale experimental facilities.

With the NIH, the DOE SC has a successful collaboration with the National Cancer Institute (NCI) in the CANDLE project [11]. DOE is now developing an MOU with both the NSF and NIH on a program of collaborative research in Computational Neuroscience. The subcommittee, therefore, recommends that the SC explore new opportunities to work with both NSF and NIH in areas where there would be a clear benefit for scientific progress under a DOE-led ***AI for Science*** initiative. There may also be opportunities to work with other US funding agencies, such as NIST and DOD, in areas of mutual interest.

6. International collaboration

There is a need for broad-based, coordinated action by like-minded international partners to harness the global scientific software community to address the tremendous opportunities in data-intensive science stemming from the huge increase in scientific data collection rates. Computational and data analytical methods driven by AI/ML are now universally accepted as indispensable for future progress in science and engineering.

International leadership in ***AI for Science*** over the coming decade will hinge on the realization of an integrated set of programs spanning the four interdependent areas noted above – AI-enabled applications, AI algorithms and foundational research, AI software infrastructure, and new hardware technologies for AI. Scientists in nearly every research field in every country will now depend on the development of such software infrastructure for high-end computing and big data analytics to open up new research fields and to dramatically increase their research productivity.

Such AI/ML software infrastructure and algorithms capable of scaling up to exascale systems will underpin the work of global scientific communities working together on problems of global significance and enable them to leverage distributed resources in transnational configurations. In terms of feasibility, the dimensions of the task – totally re-thinking, re-imagining, and expanding, in the period of just a few years, the massive software foundation of computational and data science to meet the new realities of ***AI for science*** – are simply too many and too large for any one country to undertake on its own.

To realize this vision for an international cooperative effort, the Office of Science needs to:

- Provide a framework for organizing the software research community
- Create a thorough assessment of needs, issues, and strategies
- Initiate development of a coordinated software roadmap
- Encourage and facilitate collaboration in education and training
- Engage and coordinate the vendor community in cross-cutting efforts

In DOE's Office of Science, ASCR is well suited to lead on bringing the international community together to work on these challenges.

Report

1. Introduction and Background

Executive Order on AI

On February 11, 2019, the President signed Executive Order 13859, *Maintaining American Leadership in Artificial Intelligence* [1]. This order launched the American AI Initiative, a concerted effort to promote and protect AI technology and innovation in the United States. The Initiative implements a whole-of-government strategy in collaboration and engagement with the private sector, academia, the public, and like-minded international partners.

Among other actions, key directives in the Initiative called for Federal agencies to prioritize AI R&D investments, enhance access to high-quality cyberinfrastructure and data, ensure that the Nation leads in the development of technical standards for AI, and provide education and training opportunities to prepare the American workforce for the new era of AI.

DOE Office of Science Town Hall Meetings on *AI for Science*

From July to October in 2019, the Argonne, Oak Ridge, and Berkeley National Laboratories hosted a series of four ***AI for Science*** Town Hall meetings in Chicago, Oak Ridge, Berkeley, and Washington, DC. The four meetings were attended by over 1300 scientists from the 17 DOE Labs, 39 companies, and over 90 universities. The goal of the Town Hall series was *'to examine scientific opportunities in the areas of artificial intelligence (AI), Big Data, and high-performance computing (HPC) in the next decade, and to capture the big ideas, grand challenges, and next steps to realizing these* opportunities.'

The term ***AI for Science*** was used as a shorthand for the next generation of methods and scientific opportunities that will be enabled by the development and application of a variety of AI technologies including machine learning, deep learning neural networks, approximation and statistical methods, data analytics, and automated control as well as agent-based AI models. The Town Hall discussions focused on capturing the transformational uses of AI that employ HPC and/or data analysis, leveraging datasets generated by DOE instruments and user facilities as well as HPC simulations.

These discussions were captured in the 16 chapters of the ***AI for Science*** report of the Town Hall meetings [2]. The study of these chapters showed that the discussions contain *'common arcs revealing classes of opportunities to develop and exploit AI techniques and methods to improve not only the efficacy and efficiency of science but also the operation and optimization of scientific* infrastructure.'

2. Charge Letter to ASCR

The charge letter from the Director of the Office of Science sets out the context of the challenge to the subcommittee as follows:

- Artificial Intelligence and Machine Learning have the potential for providing new insights and even new discoveries from this data, including the correlation of experimental and computational data.

- However, the technical aspects of 'AI/ML for Science' may be more challenging than currently envisioned. Over the last few years, several workshops and subcommittee reports have identified and enumerated the scientific opportunities and some challenges from the intersection of AI/ML with data-intensive science and high performance computing.

The subcommittee is tasked to deliver a report that specifically:

- Assesses the opportunities and challenges from Artificial Intelligence and Machine Learning for the advancement of science, technology, and the Office of Science missions.

- Identifies strategies that ASCR can use, in coordination with the other SC programs, to address the challenges and deliver on the opportunities.

The letter also noted that, due to the cross-cutting nature of this effort, members from the other Office of Science Federal Advisory Committees should be included in the make-up of the subcommittee as well as input from Industry and other relevant Federal agencies.

3. Subcommittee Information Gathering Activities

The subcommittee has gathered information from a wide range of sources. At its first meeting in February 2020, there was a presentation and discussion session with the organizers of the ***AI for Science*** Town Hall meetings from Argonne, Oak Ridge, and Berkeley Laboratories. There were also presentations from the 'big five' IT software companies about their thoughts on ***AI for Science*** – Sanjay Padhi from Amazon, Larry Zitnik from Facebook, Peter Norvig from Google, Jed Pitera from IBM, and Sarah Bird from Microsoft.

This session was followed by presentations from both the NIH and NSF. Susan Gregurick, Associate Director for Data Science, highlighted the recommendations of an NIH AI Working Group, and Grace Peng from the National Institute of Biomedical Imaging and Bioengineering (NIBIB), talked about their programs in AI, Machine Learning and Deep Learning, Mathematical Modeling, Simulation and Analysis, and summarized the recommendations of several recent AI Workshops.

For the NSF, Manish Parashar, Director of the Office of Advanced Cyberinfrastructure, explained how AI cuts across all of NSF's Priorities and their '10 Big Ideas'. In the words of France Cordova: *'AI is the universal connector that interweaves all of our Big Ideas; data science is changing the very nature of scientific inquiry, and AI's use of data has the potential to revolutionize everything we do in science.'* In FY19, NSF's investments in AI totaled over $450M. Denise Caldwell, Physics Division Director, then summarized some science projects incorporating AI/ML and gave details of the recent call, in collaboration with several other federal funding agencies, to establish six National Artificial Intelligence Research Institutes.

In addition, the subcommittee heard presentations from James Sethian (LBNL) on the ASCR-BES funded CAMERA project [12], from Laura Freeman (VA Tech) on 'Statistical thinking and accelerating insights from Machine Learning and Artificial Intelligence,' and from Tanmoy Bhattacharya on 'Uncertainty Quantification.'

Subsequent meetings of the subcommittee were held virtually as Zoom meetings. The nominated representatives from the five domain science directorates of the Office of Science presented reports on their community's thoughts about the potential for AI and ML technologies in their respective fields:

- Kerstin Kleese van Dam for BERAC on 'BER AI Strategy and Requirements'
- Phil Snyder for FESAC on 'Strategy and Potential of AI/ML in the Fusion Energy Sciences Program'
- Mike Hildreth for HEPAP on 'AI and HEP: An overview of opportunities'
- Tanja Horn for NSAC on 'Artificial Intelligence for Science – Nuclear Physics Overview'
- Abbas Ourmazd for BESAC on 'Impact of AI on Basic Energy Sciences'

There were three presentations on AI research from:

- David Womble for Oak Ridge National Laboratory on the 'ORNL AI Initiative'
- Becca Willett from the University of Chicago on 'AI: Challenges & Opportunities'
- Pedro Domingos from the University of Washington on the potential for 'AI for Science'

The subcommittee also heard two non-IT industry talks:

- Kim Branson, Global Head AI/ML for Medicinal Science and Technology R&D at GSK, gave a presentation on 'AI and ML at GSK'
- Rick Arthur, Senior Principal Engineer, Advanced Computational Methods Research, GE Research, gave a talk about more than 30 years of applying 'industrial AI' at GE [14]

The final two presentations were from IT hardware companies:

- Andy Hock, Head of Product at Cerebras Systems, on 'Wafer-Scale AI Computing for Science'

- Raja Koduri, Senior Vice President, Chief Architect, and General Manager of Architecture, Graphics, and Software at Intel, on 'Exascale for Everyone'

A collection of recent reports on AI and science were available in an online archive together with copies of most presentations to the subcommittee.

4. DOE as the lead agency for AI/ML applied to Facilities Science

Although the NSF is the lead agency for fundamental AI research, the DOE is the lead agency for AI research involving 'Big Science, Big Data, and Big Computing.' DOE has not only established national and international leadership in HPC and supercomputing but is also a leader in the application of AI/ML technologies to the very large scientific datasets generated by their large-scale experimental facilities. DOE's Office of Science provides US researchers with access to the largest and most diverse suite of scientific experimental facilities in the world – from X-ray synchrotrons and neutron sources to integrative genomics and atmospheric radiation facilities – as well as to the world's most capable high performance computing facilities. Upgrades to these user facilities and new nuclear physics facilities coming online now and over the next decade will dramatically increase the amount of new data being produced across all of the scientific domains supported by the Office of Science.

Across the SC programs, scientific applications of Artificial Intelligence and Machine Learning can build on the power of sensor networks, edge computing, and high performance computers to transform science and energy research in the future. Given the highly specialized nature of many DOE facilities and scientific research domains, it will not be possible to rely solely on third-party AI/ML R&D for this transformation. The DOE will need to build its own R&D programs that focus on the most challenging science-driven problems and computing capabilities. Software infrastructure will be created that combines leadership in AI/ML tools and algorithms with the DOE's traditional strengths in simulation and modeling technologies and that can execute on new computing platforms capable of high performance on both types of applications. The new ***AI for Science*** initiative can also build on the lessons learned in the successful Exascale Computing Project [10]. The anticipated returns will help ensure that the US continues to maintain and enhance leadership in both data-intensive science and high-performance computing.

Science and computing are now in an era of post-Moore's Law silicon technologies and there is an urgent need for a sea-change in the productive use of increasingly complex/heterogeneous systems, and in the seamless integration of data and computing resources. There are also major challenges in the management, reduction, visualization, provenance, and curation of the scientific Big Data generated at scale by DOE's most advanced facilities. With a major initiative in ***AI for Science***, a growing convergence of AI, Data, and HPC can be enabled across the DOE to accelerate scientific discovery and improve our international competitiveness. With such a coherent initiative and the right research programs, building blocks, and sustained effort, it should be possible to engineer the transformation of this once in a generation science/technology convergence into a profound new way of doing science within a decade.

Only the DOE's National Laboratories, together with US university and industry partners, have the necessary assets to initiate a large-scale ***AI for Science*** program that can accelerate the development of such capabilities not only to meet their SC mission needs but also to benefit all of DOE's activities. With their experience in the ECI/ECP, the DOE Labs have established a vibrant co-design culture involving teams of scientific users, instrument providers, mathematicians, and computer scientists. In the proposed ***AI for Science*** initiative, this unique experience can be leveraged to develop new capabilities and tools so that they can be readily applied across the agency's diversity of instruments, facilities, and infrastructure [10].

5. Opportunities and challenges from Artificial Intelligence and Machine Learning for the advancement of science, technology, and Office of Science missions

The application of AI has the potential to revolutionize how science is conducted – substantially advancing not only the efficiency of the scientific enterprise but also the operation of the scientific infrastructure. The combination of ML, high performance computing (HPC), and advanced data acquisition and handling will uncover a range of opportunities for breakthrough science – allowing the analysis of huge datasets, the exploration of enormously complex parameter spaces and the discovery of extremely subtle effects, leading to unforeseeable discoveries that will benefit the nation and, ultimately, the world.

The historical result of scientific discovery is innovation. Whether in the creation of new technologies that transform an industry or in the creation of entirely new industries, scientific discovery is the fuel for the engine of innovation that powers the US economy. Thanks both to a technology culture that embraces and rewards innovation together with world-class centers of higher education, the US has earned an enviable reputation as the innovation leader in the world. However, other countries have made great strides in closing the gap and are keen to erase it entirely. With much of the innovation that will be delivered to humanity in the coming decades likely to be fueled by AI, the US must lead in AI if it is to continue to lead the world in innovation and continue to reap the enormous societal and economic benefits that come with that position [1].

For AI to fulfill its enormous potential for scientific advancement (and for the nation to maximally benefit from this potential), we must see widespread adoption and utilization of AI across the entire scientific enterprise. Scientific competition will drive many researchers to rapidly embrace AI/ML technologies, but each scientific discipline comes with its own culture and sets of requirements and constraints. The resulting patchwork of AI advances will therefore almost certainly be sub-optimal. The subcommittee, therefore, recommends that a comprehensive strategy for advancing the ***AI for Science*** agenda be created to guide developments across the entire mission space of the Office of Science.

The expected benefits of such an initiative will result in a transformation of scientific research at the DOE Laboratories and other DOE supported facilities. For example, the use of AI methods in science will:

- accelerate the design, discovery, and evaluation of new materials
- advance the development of new hardware and software systems, instruments and simulation data streams
- identify new science and theories revealed as a result of increasingly high-bandwidth instrument data streams
- improve experiments by inserting inference capabilities in control and analysis loops
- enable the design, evaluation, autonomous operation, and optimization of complex systems from light sources and accelerators to instrumented detectors and HPC data centers
- advance the development of self-driving laboratories and scientific workflows
- dramatically increase the capabilities of exascale and future supercomputers by capitalizing on AI surrogates
- automate the large-scale creation of "FAIR" (findable, accessible, interoperable, reusable) data [15]

To achieve such advances, the DOE community will need to work closely with the NSF and with the university AI research community to pursue the research needed to create new and powerful AI methods and technologies. For example, new AI architectures capable of incorporating mathematical and physical constraints that require less data to train effectively will allow the design of a whole new range of applications optimized for specific scientific tasks.

6. Strategies for the DOE Office of Science to address the challenges and deliver on the opportunities

6.1 Introduction

The purpose of an AI strategy for SC is to guide the development of AI technologies and foster the utilization of AI capabilities in such a way that leads to widespread adoption of AI across the entire Office of Science. Successful execution of this strategy would see, after an appropriate period of time (5-10 years), a deep understanding and pervasive exploitation of AI in all areas of DOE science, leading to substantial advances and discoveries as well as to significant efficiency gains in the process of scientific experiments.

In order to achieve this vision, the entire ecosystem of AI must be advanced, in much the same way as the entire ecosystem of high performance computing needed to advance in order to deliver usable exascale systems. The AI ecosystem has similarities to that for high performance computing, with some interesting (and important) differences. Crudely speaking, the HPC 'stack' can be thought of as Applications built on top of Software that is built to run on Hardware. In a

process known as co-design, the DOE has pioneered an approach whereby all three are advanced together: hardware and software are designed together while informed by the application that drives the entire process. In the case of AI, substantial progress is required in foundational mathematics and AI algorithms if we are to move the science forward significantly. The four integral components needed for a successful ***AI for Science*** initiative are then:

- **Application-specific solutions based on hardware/software/algorithm co-design**

  The driver for the entire initiative is the acceleration of scientific discovery leading to innovation. It is here, through development and deployment of domain-specific, AI-based solutions, that a transformation of the scientific research process can be achieved.

- **Research in AI algorithms and foundations**

  The lack of a fundamental basis for understanding how ML methods such as Deep Learning neural networks arrive at their 'conclusions' is a cause for concern. For AI/ML technologies to gain the confidence of scientists and to be useful in contexts where there can be consequential decisions, substantial research is needed to develop the underlying theory and our understanding of the robustness, trustworthiness, and limitations of these technologies.

  The space of AI algorithms is changing very rapidly and includes some of the most impactful innovations such as Deep Learning, Reinforcement Learning, and Generative Adversarial Networks (GANs). A continued focus on the exploration and development of new AI algorithms must be an important component of an ***AI for Science*** initiative.

- **Development of AI software infrastructure**

  In terms of software infrastructure, in order to take advantage of changes in both AI hardware and algorithms, significant advances in areas such as programming environments, workflow and resource management, data management and engineering, and other system software will be required.

- **AI-specific computing architectures and hardware**

  While much progress has been made using general-purpose high performance computing architectures, the workflows of the future can be substantially accelerated by the development of AI-specific hardware. Although numerous start-up companies are exploring new hardware solutions, the DOE has unique computational requirements and needs to be engaged in this co-design process with industry. A specific DOE focus area will be the integration of such new AI-specific hardware into the harsh environment of scientific instruments.

Successful integration of all four components of an ***AI for Science*** initiative will require:

- A full partnership between all areas of the Office of Science
- Engagement of the National Laboratories and their user facilities
- Involvement of the university and private industry research community
- Mechanisms for collaborative projects with agencies such as the NSF and NIH
- Collaboration with expert organizations from similarly minded countries
- An organized process for dissemination to the scientific community

6.2 AI Applications

AI is expected to revolutionize scientific discovery. The transformative impact of AI on science has been carefully evaluated by the Office of Science in a series of ***AI for Science*** Town Hall meetings involving more than 1300 scientists [2] as well as in Roundtable and Workshop Reports from appropriate areas of the Office of Science (see Appendix C). These activities have identified the exceptional opportunities offered by AI for the Office of Science, crystallized the key 'grand challenges' associated with these opportunities, and highlighted the research priorities needed to meet these grand challenges. Note that the Priority Research Directions in the submissions from each of the five Office of Science programs have not been integrated so that several generic priorities common to several research areas are separated from those that are specific to a given area.

Priority Research Directions: Basic Energy Sciences (BES)

AI/ML methods for data analysis, control, and modeling hold promise for greatly accelerating experimental and computational discovery. The Office of Science has identified key Priority Research Opportunities to realize the vision that, in the next 10 years, AI/ML will be an integral part of the discovery and design workflow, just as experimental, theoretical, and computational tools are today. The BES scientific user facilities can work in close synergy with experts across the DOE national laboratories and universities to realize these opportunities and attain the vision of broadly incorporating AI/ML methods in facility operations and scientific experiments. These advances will result in new insights that will drive innovation and enable exploration of new scientific opportunities far beyond the current horizon.

1. **Efficiently extract critical & strategic information from large, complex datasets.**

   Key Question: How to extract robust, meaningful information from increasingly vast, complex data produced at BES' scientific user facilities?

   AI/ML approaches are based on training algorithms with appropriate data, in order to extract information from data not used in training. This approach gives rise to a number of important questions. These include: whether an algorithm trained on one dataset can be used to produce reliable answers about a different dataset; whether a particular algorithm is robust against noise or attempts to deceive it; what the basis is for the answers an algorithm provides and whether these answers are free of bias. Answers to

such critical questions are required in order to gain confidence in the veracity and reliability of AI/ML based algorithms.

2. **Address challenges of autonomous control of scientific systems.**

   Key Question: How to address challenges inherent in real-time operation of large, complex scientific user facilities?

   AI/ML is expected to enable breakthroughs in operating complex infrastructure. In order to realize this potential, new paradigms must be developed and tested to integrate real-time information from powerful sensors, combined, where appropriate, with edge computation, to enable real-time AI-based predictive analytics, control, and optimization. AI-driven, real-time intelligence will likely couple real-time data with infrastructure models (e.g., a digital twin). Similarly, AI/ML-enabled predictive models trained with infrastructure data will be indispensable for exploring the design spaces for smart energy and transportation infrastructure, HPC computing systems, data centers, and communications networks. In a similar fashion, they will also become indispensable for particle accelerators, light sources, and complex instruments, many of which involve interconnected subsystems of magnets, mechanical, vacuum, and cooling equipment, power supplies, and other components. Such instruments have many control points, and require high levels of stability, making their operation a complex optimization problem. The operation of these instruments will require AI/ML-based solutions but this remains a challenging problem due to lack of prior models for reliable and safe control. Initial results have already highlighted the potential of AI/ML-based solutions, but key challenges remain. Finally, critical infrastructures increasingly rely on information systems. Removing the human-in-the-loop is increasingly necessary for defensive responses on the same millisecond timescales as digital attacks. The development of robust and transparent AI/ML solutions is expected to offer the best approach to detecting and diagnosing cyber and physical attacks and threats in real-time.

3. **Enable offline design & optimization of facilities & experiments.**

   Key Question: How can we catalyze scientific discovery by leveraging the wealth of diverse and complementary data recorded across the BES scientific user facilities?

   Advanced user facilities generate terabytes of data per experimental run. The unabated advance of high repetition-rate instruments will exacerbate the data torrent. Techniques are needed to rapidly pre-process data, and extract meaningful information from very large datasets, to guide experiments in real-time. AI/ML approaches promise to play a prominent role in this endeavor. Science, however, is often driven by exceptions to the ordinary. New AI/ML approaches are needed to identify and help interpret previously unseen rare ('transition') states in very large datasets. New paradigms are also needed to integrate multisensory data, combining, for example, structural and spectroscopic snapshots from complex systems.

Taken together, the priority research directions outlined above show that ***AI for Science*** is poised to fundamentally transform the way scientific experiments are designed and conducted, and the nature and extent of the information extracted from the resulting data. Continued US leadership in basic energy sciences depends critically on integrating AI into the fabric of BES activities.

Priority Research Directions in Biological and Environmental Research (BER)

Climate and Environmental Sciences

The BER Climate and Environmental Science research community have been at the forefront of applying novel AI/ML methods, particularly in its computational efforts. Core methods such as clustering, regression, and feature and anomaly detection have been successfully applied in a range of projects. One such project won the Gordon Bell Prize in 2018. Furthermore, a number of projects are now underway that are exploring the use of surrogate models in climate modeling with excellent early results. From this basis, the community has set its sights on tackling more advanced challenges in the future to provide novel, robust and meaningful scientific insights through the use of AI such as:

- **Create new AI enabled observational instruments that can adaptively capture dynamic environmental processes with greater precision**
    - **Key Question:** How can we integrate AI hardware and software into observational capabilities at the edge and link these seamlessly to other instruments and computing capabilities.
- **Develop hybrid process-based/AI modeling frameworks to gain a predictive understanding of the Earth system at global, regional and local scales under a changing environment.**
    - **Key Question:** How can we extrapolate sparse measurements across space and through time to improve our understanding of the functional traits of biological and hydrological systems, as well as dynamic processes important for closing the carbon cycle or secure water resources, and to develop models that improve climate predictability and reduce uncertainty in future projections.
- **AI enabled multi-scale, real-time data-model assimilation to predict environmental risk and develop resiliency in a changing environment - for energy infrastructures and subsurface applications.**
    - **Key Question:** How can we integrate smart sensing systems, built-for-purpose models, large ensemble forecasts to quantify uncertainty, and dynamic decision support systems for critical infrastructure - in real-time and across scales?

In addressing the above challenges it is expected that researchers will gain a deeper, predictive understanding of the processes that govern our climate and environmental systems, helping

them to identify future risks and develop effective mitigation strategies with reduced uncertainty.

Biological Systems Science

Biological System Science in BER is focused on improving the understanding of the fundamental processes that govern genomic and metabolic processes that influence the secure production of biofuels and bioproducts. Furthermore, the community studies the role of microbes in the environment, and earth system and subsystems. As such, AI/ML efforts have focused on data analysis and subsequent inference of process knowledge. Moving forward, the community is now looking to solve:

- **AI enabled systematic mapping of the small-molecule chemical space to find new applications and biological knowledge by eliminating biological 'Dark Matter'.**
    - **Key Question:** How can the multitude of experimental, computational, and literature-based knowledge be harnessed to infer new biological compounds, functions, and behavior?
- **Build AI based higher order integrated biological models that fully capture the complexity of interactions and lead to a predictive understanding of biological systems.**
    - **Key Question:** How can we learn and represent the complexity of biological processes and their interactions in an efficient, yet accurate manner?

In solving the challenges above, researchers will not only gain a predictive understanding of biological and environmental systems as they pertain to biofuels and bioproducts, but also be able to develop effective strategies to maximize bio-production.

Priority Research Directions: Fusion Energy Sciences (FES)

As fusion energy research advances toward reactor-scale experiments, exascale computation, and sophisticated next-generation diagnostics, new and exciting opportunities are emerging for application of AI/ML techniques. AI/ML has been successfully applied across a wide range of problems in fusion, including development of reduced models, disruption prediction, plasma control, and physics discovery. As the state-of-the-art advances, and experimental and computational fusion datasets grow dramatically in size, AI/ML is expected to play an ever more important role in scientific discovery, experiment planning and analysis, and device control and operation. Grand challenges identified in the ***AI for Science*** Town Hall Report include:

- **Enable real-time understanding in long pulse tokamak experiments**: Next generation experiments such as ITER will involve long pulse lengths and 'burning plasma' conditions in which the plasma is self-heated by alpha particles. This strongly-coupled system will introduce real-time data streaming and analysis challenges similar to those posed by an operational fusion power plant.

- **Develop models that bridge gaps in fusion plasma confinement and stability prediction**: Capability to predict plasma confinement and stability has improved dramatically across the past decade. Remaining gaps in understanding, such as physics of the boundary plasma and tearing modes, often involve bridging a very wide range of spatio-temporal scales, and the interaction of plasmas with neutrals and material surfaces. The development of interpretable AI/ML methods and model extraction and reduction techniques can facilitate both understanding and performance optimization.

- **Establish the plasma prediction and control solutions for sustained fusion power plant operation**: Power plants must have very high reliability and availability. Control physics and control algorithm mathematics requirements for fusion are challenging due to extreme nonlinearity, and multi-physics overlaps. Data-driven methods can contribute to control level modeling, management, and interpretation of real-time data for control and determination of optimal parametric trajectories.

In addition to these grand challenges, numerous other problems in the FES domain would benefit greatly from application of AI/ML technologies. Examples include high repetition rate lasers, which produce enormous data sets, and fusion material simulations and experiments, whose multi-scale nature challenges the capabilities of existing methods.

To address these and other key challenges in fusion and plasma science, the FES and ASCR communities have developed a set of Priority Research Opportunities (PROs) for application of AI/ML methods to enable accelerated solution of fusion problems. Seven PROs were identified, including three in each of two broad categories, Accelerating Science (1-3) and Enabling Fusion (4-6), and a cross-cutting opportunity (7):

1. **Science Discovery with Machine Learning**:

   Key Question: How can AI/ML approaches bridge and close gaps in understanding, accelerate hypothesis generation and testing, and optimize experimental planning?

2. **Machine Learning Boosted Diagnostics**:

   Key Question: How can AI/ML maximize the information extracted from complex measurements, enhance interpretability, combine multiple data sources, and generate synthetic diagnostics that enable inference of quantities that are not directly measured?

3. **Model Extraction and Reduction**:

   Key Question: How can AI/ML enable extraction of models of fusion systems and plasmas to enhance understanding of complex processes? How well can sophisticated AI/ML approaches accelerate computational algorithms while maintaining accuracy?

4. **Control Augmentation with Machine Learning**:

   Key Question: How can AI/ML improve control-level models and cope with massive data streams in real time to optimize control and analysis algorithms?

5. **Extreme Data Algorithms**:

   Key Question: How to develop methods for in-situ, in-memory analysis and reduction of extreme scale simulation data, and methods for efficient ingestion and analysis of fusion experimental data.

6. **Data-Enhanced Prediction**:

   Key Question: How to use AI/ML to develop reliable algorithms for prediction of key plasma phenomena, including disruption avoidance and mitigation.

7. **Fusion Data Machine Learning Platform**:

   Key Question: How to develop a novel system for managing, formatting, curating, and enabling access to fusion experimental and simulation data for optimal usability in applying AI/ML algorithms.

Priority Research Directions: High Energy Physics (HEP)

AI/ML techniques have been eagerly embraced by investigators in HEP. The application of AI/ML algorithms to particle physics problems has enabled marked advances in sensitivity or scientific output, in many cases leading to state-of-the-art results. Many other problems remain intractable, however, because current ML techniques are not suited to these challenges. Advancing the discovery potential and reaping the rewards of the investment in the large facilities that support HEP requires a new era of AI development and application. The ***AI for Science*** Town Hall events identified some major scientific challenges arising in the next decade that can be met with the development of suitable AI technology to acquire, analyze, and simulate datasets of unprecedented size and accuracy:

1. **Create usable tools for large-scale distributed training and optimization of ML models to enable physicists to scale up the complexity of their models by several orders of magnitude above the current "laptop-size."**

   The larger AI ecosystem provided by DOE Leadership Class Facilities will be a critical element in the development and training of the sophisticated AI models needed to accomplish the scientific goals. Enabling rapid training and optimization will accelerate discovery.

   Key Question: Can we reconstruct the History of the Universe? Rich datasets from a host of new telescopes plus exascale cosmological simulations will generate a wealth of astrophysical data with unparalleled resolution and depth. To optimally extract

information from this data while maintaining robustness, new AI techniques combined with statistical methods and HPC simulations are needed. This combination will enable predictions deep into the nonlinear regime of structure formation, spanning a large mass and spatiotemporal dynamic range. This will enable us to enhance our understanding of dark matter, dark energy, and the history of the early universe back to the era of Inflation and beyond.

2. **Develop training methodologies that can detect rare features in high-dimensional spaces while being robust against systematic effects.**

   Unsupervised learning, AI-based system controls, and many other aspects of the future HEP science program will rely on AI algorithms to make decisions and react to rare events in highly complex environments. This could include particle physics triggers, transient detection in astronomy, accelerator control, and new particle searches. Robust, trustworthy AI will be required.

   Key Question: Can we discover New Physics with zettascale datasets? New instruments capable of generating zettabytes of data will be deployed, challenging all aspects of the scientific process. In order to maximize the chances of discovery, these instruments will need to be AI-controlled and have AI-enabled filters to examine incoming data with near-zero latency. AI-based simulation and analysis will be required to process rapidly the vast quantities of data with finite computing resources. Unsupervised AI searches for "new physics" will be deployed to exploit the data fully.

3. **Design tools to quantify the impact of systematic effects of the accuracy and stability of complex ML models.**

   Uncertainty Quantification (UQ) is imperative for physical measurements. While AI algorithms have proliferated, general techniques for understanding their accuracy and stability in complex environments have not. If these techniques are to become fully useful in deriving measured quantities, work is needed in this area to develop models whose uncertainties can be described in a manner appropriate to the measurement.

   Key Question: Can we understand Large Scale Cosmic Structure Formation? Advances in AI could allow an automated cosmology experiment that would be able to combine new survey data and detailed simulations to generate optimized observing strategies. Such an endeavor could benefit from the increased processing speed afforded by AI algorithms to oversee a cluster of instruments and to gather data in such a way as to minimize systematic errors, to optimize calibrations, and to maximize sensitivity to cosmological parameters that could shed light on the dark universe.

These are not the only scientific areas that can be dramatically enhanced with developments in ML/AI. AI techniques are heavily used in cosmological studies, for example in red-shift estimation from photometric data, galaxy image processing, and feature extraction. AI Applications in theoretical calculations include kernel estimation, model optimization in large parameter spaces,

searches for new models, and the estimation of parton distribution functions. As a specific example of a massive theoretical computational problem, Lattice QCD calculations could realize exponential gains in speed if AI can be applied to path integral estimation. Advances in the foundations of AI, especially including constraints from physical systems in ML, will have dramatic impacts on these and many other physical problems.

Priority Research Directions: Nuclear Physics (NP)

AI has tremendous potential within NP Research. It can provide new insights and discoveries from both experimental and computational data produced at user facilities. As identified in the January 2020 Roundtable Meeting on AI/ML in NP Facilities [13], all the top priorities of the 2015 NP Long-Range Plan on Research Opportunities and Directions can benefit from AI. At the same time, a number of activities and technologies in the diverse NP research portfolio have the potential to contribute to the emerging AI programs. For example, NP presents data on short time scales and with many different configurations that expose the limitations of current methods and could contribute to making AI more interpretable for the long term. Some of the most notable grand challenges for NP identified in the ***AI for Science*** Town Hall report include:

- **Automate and/or optimize the operation of accelerators and detector systems**

  Key Questions: Development and validation of virtual diagnostics; improvement to beam sources and injector performance; data-driven system maintenance; automated learning for operator support; anomaly detection and mitigation?

- **Improve experimental design and real-time tuning**

  Key Questions: improving experiments by intelligently combining disparate data sources such as accelerator parameters, experimental controls, and detector data. AI enables intelligent decisions about data reduction and storage and can improve the physics content through data compactification, sophisticated triggers (both software and hardware-based), and fast-online analysis.

- **Improving simulation and analysis**

  Key Questions: Improving sensitivity to allow more information to be extracted from datasets, which decreases uncertainty in results and increases discovery potential. Decreasing simulation and analysis time to save costs and ultimately allow for a higher volume of scientific output by accelerating the feedback loop between experiment, analysis, and theory.

- **Game changers in nuclear theory**

  Key Questions: Several case studies were identified by the community including identifying rare events, quantified computations of heavy nuclei using realistic inter-nucleon force, and dense matter equation of state.

In the ***AI for Science*** Town Hall report, it was identified that the NP community could benefit from using existing AI/ML solutions. For example, the Exascale Computing Project (ECP) has created tools that can accelerate computationally expensive tasks. Similarly, the ECP ExaLearn project is now developing scalable tools to address common AI/ML challenges such as developing surrogate models, inverse problems, and automated design and controls challenges. These tools could save a significant amount of development time and allow the NP community to focus on solving domain specific challenges. However, as the NP community expands its use of AI/ML technologies, it will require access to greater computing resources. In addition, the current AI tools and methodologies have limitations that have to be addressed in the long term. The following needs have been identified by the NP community for successful collaboration with AI/ML and increasing scientific output:

- **Need for problem-specific tools**

  NP applications are unique in that they are often aimed at accelerating calculation, whether in the evaluation of models where one can use AI techniques to identify the most promising calculative pathways to simulations where AI-determined parametrizations can be used to circumvent performance-limiting elements. While traditional ML tools may be applied to these problems, significant effort is required in the careful tuning of ML tools (hyperparameter determination) to optimize performance in each application domain.

- **Enabling Infrastructure for AI in NP**

  To maximize the usefulness of the data, it will be important to have standards on the processing of data, the application of theoretical assumptions, and the treatment of systematic uncertainties that will be used as training samples or as part of the combined analysis. AI techniques are computationally intensive and success in using these techniques will require access to GPU computing and disk storage at appropriate scales.

- **Need for uncertainty quantification**

  A common theme is to investigate and apply AI methods with well-understood UQ, both systematic and statistical, to accelerator science, NP experimentation, and NP theory. The commonly used ML algorithms do not provide error estimations with model predictions, which are essential to understand outcomes. In addition, an assessment of metrics for the evaluation and comparison of uncertainty predictions using different modalities is required for the widespread use of AI in NP.

6.3 AI Algorithms and Foundations

Advances in algorithms have given scientists the tools to model and simulate nature at an unprecedented range of scales: from computing the history and fate of the cosmos and the explosion of supernovae to the evolution of the climate system and the properties of materials to the smallest of subatomic particles. These efforts have traditionally relied on mathematical, modeling, and computational building blocks whose properties are well established. Despite

having access to tremendous computational resources, the fact remains that scientists cannot explore all possible theories or easily simulate phenomena across multiple space and time scales.

AI presents a unique opportunity for bridging this gap. Modern methods in AI, defined broadly to include ML, optimization, statistical inference, and supporting systems, are yielding unprecedented results in many application areas. Recent AI successes, unimaginable even a short time ago, can be attributed to colossal datasets, enormous computing power, and innovations in the underlying mathematics, statistics, formulations, and algorithms. Meanwhile, in the physical sciences, increasingly powerful sensors and facilities are generating overwhelming quantities of data.

However, commercial or off-the-shelf AI tools are often unable to fully extract the knowledge contained in these datasets, for several reasons:

- First, scientific datasets are often limited in size and seldom annotated with the sort of information required for many algorithms designed for large-scale data.
- Second, scientific data can often be very expensive to obtain: experiments carried out at facilities can require tremendous resources.
- Third, extracting correct information often requires enforcing additional constraints and physical principles.
- Fourth, in many scientific and engineering applications, the price of being wrong is high, and guarantees on applicability and errors are required.

Developing new algorithms that meet these challenges is a core effort that will profoundly advance the impact of AI/ML for science.

Additionally, many AI/ML algorithms can be 'brittle,' failing to perform when applied to situations or problems different from those for which they were designed. In order to build and apply these new algorithms with confidence, a parallel effort is needed to understand some of the key foundational issues underlying the applicability and reliability of these new methods, particularly in their application to scientific and engineering challenges. Common challenges arising in the physical sciences include AI-informed adaptive design of experiments, statistical inference, and quantification of uncertainties, understanding of dynamical systems in far-from-equilibrium regimes, integration of physical models and simulations into ML methodology, anomaly or rare-event detection, inference based on heterogeneous data, and the design interpretable, robust AI methods. A central challenge will be constructing ML algorithms that naturally comprehend what is already understood, and then build on this to better understand what is observed.

Altogether, we have a unique opportunity to develop algorithms, new foundations, and new tools that will dramatically advance the frontiers of both AI and physical sciences. Key areas of algorithmic and foundational AI emphasis for discoveries in physics, material science, and other scientific DOE domains include:

- **Effective ML training for small or sparse datasets.** Outside the realm of computer vision and natural language processing, many scientific areas have very small numbers of

datasets with *labeled* training data, either because they are difficult to generate or the parameters of the systems are such that they cannot be produced. This applies to fields ranging from material sciences and astronomy to various engineering disciplines and biological applications. Many current AI algorithms, for example, deep neural networks, cannot perform well when trained on these limited samples. New developments in foundational AI, such as few-shot learning, self-supervised learning, mixed-scale dense networks, and meta-learning techniques, need to be developed to enable broad AI applications to many new fields.

- **Incorporation of physical models into AI structure.** The current theory and practice of AI struggles to best incorporate mathematical models into the learning process. Here, prior physics domain knowledge is expressed in terms of equations, including differential equations, conservation laws, invariances, symmetries, and distributions, that can be computationally encoded. At the same time, AI frameworks that incorporate known physics will perform better, require less data to train effectively, and will be more robust. Formulations of new AI algorithms structured to include physical models, rather than having to "learn" the physics, will allow the designs of new classes of AI applications optimized exactly for the scientific tasks that require them. To do so will require advances in such formalisms as projection operators that enforce physical principles, data structures that encode symmetries and constraints, and the construction of physical priors and their injection into mathematical models.

- **Creation of surrogates.** AI presents a unique opportunity for creating data-driven surrogate models that are potentially orders of magnitude faster to run than first-principles simulation codes and can be particularly effective in the ability to simulate physical processes that span many spatial and temporal scales. Some of the unique challenges for AI systems revolve around a careful characterization of the generalization limits, proofs of interpolation/extrapolation, robustness, assessment of confidence associated with predictions, and effects of the input data.

- **Dimensional reduction, data synthesis and compression, and reduced order models.** Extracting smaller models from large datasets provides a powerful way to rapidly evaluate and test theories, examine the range of viable parameters, and point to under-resolved areas. Developing workable and robust algorithms to do so requires linking together aspects of approximation theory, statistical and probabilistic methods, and compression methods, as well as coupling coarse-fine grain representations of equations of motion and state into the development of new AI/ML methods.

- **Control and AI-enabled experimental design.** Facilities for high-throughput screening and automated control of experiments have the potential to transform how the DOE collects data. The scientific method of conducting experiments, analyzing data, forming hypotheses, and designing future experiments may be revolutionized by AI methods that can identify patterns in data with potential scientific significance. However, existing AI tools generally do not account for the interactive nature of this process and can result in

biased or misleading analyses. Control, active learning, reinforcement learning, derivative-free optimization, Gaussian processes, surrogate models, and optimization methods incorporating scientific priors incorporating known constraints and principles, and experimental design are all critical elements of an AI-enabled scientific method.

- **Operation of experimental facilities.** AI/ML technologies are rapidly permeating the design and operation of scientific instruments as well as the experimental facilities. This trend is expected to grow in extent and intensity, strongly eroding the partition between hardware and software. Optimal design, construction, and operation of major user facilities require a deep co-design approach integrating hardware, software, and edge computing on an equal footing. It is therefore important that ASCR further enhance its strong and concerted support for AI/ML based co-design throughout the design, construction, and operation cycles of major facilities.

- **Inference and calibration.** Inference refers to providing confidence intervals or quantifying the uncertainty about the output of an AI system. Calibration refers to avoiding or correcting for overfitting in AI systems to ensure the validity of inference claims. DOE-relevant AI challenges often require precise measures of uncertainty, particularly when a guess-and-check method development is not feasible due to safety concerns or expense. The DOE needs a comprehensive framework for assessing the uncertainty associated with AI predictions and leveraging this knowledge to develop better predictors.

- **Real-time processing.** Many DOE experiments and facilities produce large quantities of data, often so large that the data cannot all be stored even temporarily and therefore must be filtered in real-time to identify experiment-relevant components for future analysis. Many existing AI systems are not amenable to large-scale streaming data and integrating systems constraints into state-of-the-art learning systems is a pervasive challenge. Additional applications including the steering of experiments by using results of past experiments to construct surrogate models which then suggest new experiments targeted at under-resolved/underexplored areas.

- **Self-learning algorithms.** The availability of precise mathematical models and precise microsimulations of various physical processes, combined with advances in the understanding of inverse problems, can enable the creation of AI frameworks that can effectively do self-learning. Techniques such as GANs (Generative Adversarial Networks), ANNs (artificial neural networks), and Auto-encoders are being explored to 'learn' how to express the results of complicated multi-scale simulations in an AI algorithm. This work, as applied to physics processes, is still largely in its infancy. Dramatic improvements in the understanding of these techniques would lead to, for example, AI versions of heavily computational simulations codes that could run exponentially faster than their stochastic counterparts.

- **Time-series data and real-time decision making.** Many facilities in the DOE portfolio and beyond generate vast quantities of potentially heterogeneous data that must be analyzed

in real-time. While advances in the processing of such data have been made in the case of autonomous vehicles, for example, further work is required to achieve sub-microsecond decision times for applications such as particle physics, real-time particle accelerator or fusion reactor control, or the kinds of processing required for applications such as radio astronomy.

- **Predictive maintenance and event prediction.** Predicting the occurrence of events such as material failures, incipient fault detection, condition assessment, and failure progression prediction in complex scientific experimental facilities requires the development of new AI algorithms and new training strategies. Predictive Maintenance methods are increasingly used to predict the failure of assets and have been used in IoT industrial applications and predictive maintenance where likely equipment faults are foreseen and proactively handled. There has been some success in accomplishing analogous tasks in biomedicine where methods are being developed to predict which patients will develop a given disease and to integrate multi-scale information over time to predict which patients will respond to a particular treatment. There are many different ways prediction problems can be formulated and solved often requiring the solution of a myriad of computational and mathematical challenges associated with the coupled analyses of highly interrelated but structurally disparate information sources.

- **Anomaly detection.** Many scientific endeavors are searching for the "needle in the haystack" of tiny signals or transients hidden among enormous backgrounds. As in many other situations, being *physically anomalous* is potentially different from being *statistically anomalous*. The development of AI anomaly detection algorithms that can incorporate physical (or other) models in order to define anomalies would lead to powerful new methods of discerning small signals of interest.

- **AI methods for management of computational resources and workflows**. Computer systems employed in high-end scientific applications have become extremely complex. Computer architectures have deep hierarchies with multi-core CPU and GPU components, are often heterogeneous, and are increasingly integrated into distributed systems that incorporate edge computing. AI methods are being developed to schedule and coordinate execution of workflows and to choreograph applications that include capture, reduction, assimilation, and analysis of streaming data.

- **AI Models and Security.** It is generally advantageous to train AI models with data from a variety of sources. Some attributes of training data can be reconstructed from models. Development of methods that create models that cannot be used to reconstruct secure data would make it possible to use data from a variety of secure and open applications to train models. It is of equal importance to develop methods to determine when the release of a model, trained on secure data, would constitute a security breach.

- **Training data variability and noise**. Variability in sensor characteristics and varying statistical characteristics of noise create challenges in quantifying the accuracy of models trained using sensor data. These impact models created from photon and neutron

sources as well as weather and climate models. Further challenges arise in models that are trained using combinations of experimental and simulated data.

- **Interpretable models and algorithms.** There is a dichotomy between two views of what it means to "understand" a phenomenon: on one hand is a model that explains the result based on known or discovered laws, equations, etc. and on the other hand is a model that predicts results with accuracy, but provides no insight as to its inner workings. The quest to build interpretable models and algorithms bridges this gap, and aims to provide extractable insight in to why and how a model is producing a particular output. This is important, both to provide understanding, as well as to shed light as to the applicability of the model.

- **Robustness and applicability.** It is critical to understand when and where an AI/ML algorithm can be applied. While an AI/ML algorithm may be well-suited for interpolation, in which the algorithm is applied to an input that lies within the space of its training, when is it possible to use it for extrapolation when the algorithm predicts results outside the arena of what it has previously seen? Typically, the parameter space of possibilities is very large—how can one faithfully know the limits of an AI/ML algorithm?

- **Coupling simulations and experiment.** A tantalizing opportunity comes in using AI/ML to couple experiment to simulation, iterating between using data to refine parameters and terms in equations, and using simulations to solve these equations to further interpret the data. Doing so will require marrying a host of mathematical methods across such areas as partial differential equations, stochastic modeling, dimensional reduction and reduced order models, to new AI/ML algorithms.

- **Graph-based ML and AI.** Graphs arise naturally in many scientific domains (e.g., molecules, protein interaction networks, community networks). Structuring data and knowledge representations in terms of graphs and exploiting the topology information available from a graph representation can be critical to realizing tractable algorithms and obtaining better outcomes in tasks such as classification, clustering, and prediction of missing data.

This list is **not** comprehensive. Continuing use-inspired research constantly reveals the limitations of current methodology and opportunities for new foundational advances. Tackling these areas of emphasis will require a combination of new algorithm and methods development along with the need to understand, adapt, and generalize AI methods whose creation was motivated by application scenarios that arise in many areas of science, engineering, and medicine. AI challenges arising from several of the areas of DOE AI emphasis listed above have a great deal of overlap with AI challenges in the NCI domain, i.e. prediction of treatment response in cancer patients requiring analysis and integration of temporal sequences of combined clinical observations, clinical notes, radiological imaging, pathology images, molecular data, and treatment data.

6.4 AI Software Infrastructure

There are a few notable gaps between state-of-the-art and DOE scientific requirements when it comes to software for AI. First, DOE researchers produce massive amounts of data from simulations and models that can benefit from the integration of AI capabilities. These are often challenging datasets with multidimensional data and can also include non-image-based data. Second, DOE's unique user facilities that produce petabytes of data, have no counterpart in industry and require new AI software and capabilities. Third, many of the DOE scientific datasets need the scale of HPC systems for analysis, and those systems can have unique architectural features that require software attention and investment, such as large-scale I/O subsystems and heterogeneous compute elements. With DOE's challenging datasets and deep expertise in data analytics, simulation, and modeling, DOE researchers are well-positioned to contribute unique enhancements to the AI software stack.

As ***AI for Science*** matures, DOE computing workloads will primarily consist of AI-driven scientific computing. These computations will be deployed, in part, on evolved versions of the specialized accelerators that are emerging to execute AI applications such as artificial neural networks, machine vision, and machine learning. These new AI accelerators will, in turn, need a specialized software stack if they are to be able to optimize the performance of scientific applications within a node and across nodes. This will require a rethinking of all elements of the software stack to take the new hardware as well as these fundamentally new requirements into account. There is thus an urgent need to design and develop a software stack to meet DOE needs in this space, consisting of specialized programming models, compilers, and runtime (including communication software) support for AI-based scientific computing.

The coupling of scientific codes with AI workloads is not a straightforward process. The framework, software, and libraries are distinct and currently there do not exist any APIs (application programming interface) that might help a domain scientist couple simulation codes and ML codes seamlessly. Moreover, AI technologies being deployed are continuously evolving. For example, graph-based machine learning is a new learning technique that demands new specialized APIs for effective deployment on AI accelerators. The ML programming environment should be rich enough to handle all these challenges. Because scientific datasets are very large, distributed machine learning is critical for DOE ML workloads. The Horovod framework from Uber has become the lingua franca for distributed machine learning. In particular, Horovod has been shown to scale to the full scale of current DOE HPC facilities. However, although the Tensorflow and Pytorch frameworks both support distributed machine learning they do not scale so well. Significant investment is needed to further develop distributed parallel programming models to overcome these scaling and performance issues.

We highlight several specific areas that need R&D and deployment in the subsections below:

- **Data Management and Engineering.** There is a distinct mismatch between existing data management tools and the needs of AI solutions in science. Designed for large, static dataset management (Spark, Hadoop, iRODS), these systems do not cope well with the

flexible streaming nature of many AI training frameworks. Furthermore, while the overall training data volume for AI applications is very large, the actual data points are small and require a much more fine-grained data management and movement approach than these tools can provide. To fully support AI applications during training and operation will require new customized solutions. These also need to take into account topics such as metadata descriptions appropriate for selecting data samples suitable for AI model training.

- **User environments.** Software areas highly relevant to ***AI for Science*** include core infrastructure and software development and information-technology operations, parallel computing (MPI), big data machine/deep learning kernels, data management and engineering, workflow management, resource management, edge computing, and simulation. The deep learning kernels have the world's attention today with their implementations of the latest AI algorithms and optimization of the compute-intensive deep learning that has revolutionized AI. Here, there is a natural collaboration with Industry with DOE aiming at scalable supercomputer solutions exploiting the rich available parallelism (data, model, pipeline, hyper-parameter). However, all the software areas are important and need to be combined into a system to give the required user-friendly AI environments for science. The deep learning kernels are probably less than 10% of the needed software.

- **Workflow and resource management.** Scientific workflows are already transforming modern science. They enable the composition and automated execution of complex computational and data management tasks. The workflows today are somewhat disconnected in that data and initial data processing are done close to the instruments while other workflows retrieve the calibrated or pre-processed data from storage and perform additional analysis. There is a clear need to merge the two workflow phases and enable semi- and fully automated experiment steering. The challenges for achieving this include robustness of workflow management systems, efficient workflow execution, efficient resource management, and meaningful communications between the software layers working together to support the workflow execution, among others. AI has the potential to make workflow management systems better performing, more adaptive, and more robust. AI can also push the boundaries of automation, suggesting or building appropriate workflows based on the data a scientist collects and their research goals.

  At the same time, AI workflows have their challenges as well, requiring a diverse set of resources, access to large amounts of data, exploration of potentially large hyperparameter spaces, and coordination of distributed learning processes. In some cases, AI workflows need to run at the edge, imposing additional resource constraints and the need to operate in potentially volatile environments, where power and network demands pose challenges that are not well addressed by current workflow management systems. New solutions are needed to efficiently and reliably address and perform resource and workflow selection, management, checkpointing, and coordination, as well as handling many other aspects of reliable and efficient management of AI workflows.

- **Simulation.** In the simulation area, DOE has clear leadership in developing simulations at all scales, but there is a software gap in the support for incorporating AI, with its multiple opportunities for enhancement, into simulations. One such gap is in the provision of infrastructure to enable real-time machine learning needed for the time-critical response for instrument control. Recent DOE progress has been seen in real-time identification of tokamak instabilities and real-time analysis of light-source data, but training datasets are difficult to acquire, and validation of performance is difficult. Similarly, virtually every area of science would like to advance their capability to simulate processes (whether physical, chemical, or biological) across a broader range of spatial and temporal scales. Through the application of AI techniques, researchers are beginning to extend the reach of multiscale simulations. However, the workflows required are complex, dynamic, and not well supported in existing software infrastructure. In many cases, surrogate AI models can be used to provide substantial acceleration for simulations. Here, too, the software infrastructure to support multiple execution modalities is lacking.

  As we move to the new applications opened up by ***AI for Science*** it is important to look at the whole software stack that supports simulation methods and derive updated requirements that can be compared to the capabilities of existing software. Such an analysis should be coordinated across DOE, academia, and industry; there will be important similarities and differences. This analysis will pave the way for the development of support cyberinfrastructure that will enable and accelerate the creation of AI-based simulations.

- **Compilers.** Domain-specific machine learning compilers are emerging that offer the promise of more flexible ML frameworks and better exploitation of hardware. Today, these compilers struggle to effectively optimize the low-level computation within ML applications. Moreover, these optimizations are focused on goals that will not lead to meeting the requirements of DOE ML workloads. There is a great need to further develop machine learning compiler technology to improve the performance of DOE applications and to enhance the expressivity of machine learning frameworks. This may include rethinking the ML compiler design and scope to meet DOE-specific needs, from tightly integrated AI-driven scientific computations through edge ML scenarios.

- **Runtime.** Future node architectures may be highly heterogeneous systems, with a variety of accelerators configured. As with scientific computations, the behavior of AI applications on these platforms may be complex and difficult to characterize and predict. New mechanisms for deploying ML computations may include the creation of smarter runtimes that can respond to system inefficiencies and changes in workload by adapting execution details to provide the best possible performance. Such a runtime might facilitate the dynamic coupling of simulations and machine learning in AI-driven scientific computations and contribute to their further integration. DOE must invest in adaptive runtime systems that can take on the challenge of managing such workloads on future, complex node architectures to effectively exploit their heterogeneous resources.

- **Reproducibility/reuse**: A fundamental question that can be asked of any AI algorithm is what are the limits of its applicability. An answer to this should include information about its structure, training methods, training data sets, and, potentially, the hardware on which it was trained. Infrastructure capable of automatically capturing such provenance information needs to be created as part of the large-scale training and validation workflows that will be necessary for complicated AI applications. Without this, reuse will be difficult and the sharing of resources non-existent. A related topic is the creation of large, annotated datasets for training. Datasets that are created to be FAIR (Findable, Accessible, Interoperable, and Reusable) can form the cornerstone of AI research [15]. Significant effort is required to provide tools that will enable the creation of massive FAIR datasets from target scientific applications. Only these will contain the appropriate physical constraints, salient features, and underlying complexity that can enable training and validation of new AI algorithms targeted at the science.

- **Workload analysis and benchmarking**. These are critical areas of work for all four integral components described in this subsection. In addition to co-design, a standardized and rigorous process for characterizing and benchmarking existing AI application workloads is needed to understand computational requirements and evaluate system performance. A collection of AI science benchmarks with FAIR datasets and reference implementations would have multiple benefits. The benchmarks could be used for the comparative measurements of the performance of different systems; they would provide exemplars for users having similar requirements and allow Kaggle style challenges to develop improved algorithms and systems.

6.5 New Hardware Technologies for AI

The DOE computing systems such as Summit, Perlmutter, Aurora, and Frontier will simultaneously support the use of existing large-scale simulations, development of new hybrid HPC models with AI surrogates, and the exploration of new types of generative models emerging from multi-model data streams and sources. Future systems envisioned over the next decade may need to support even richer workloads of traditional HPC and next-generation AI-driven scientific models.

The types of hardware that exist in DOE leadership computing (LCF) facilities to support training may differ from the hardware needed to support inference at the edge or in the real-time use cases. Edge devices typically require very low power, while devices for training may be among the most power-hungry chips ever produced. In addition, many edge devices need to be able to operate in harsh conditions such as low temperatures or with radiation exposure. The industry is developing a wide variety of solutions that may fit multiple use cases, ranging from purpose-built devices designed for low-precision tensor operations to more traditional GPU and CPU architectures optimized for machine learning. However, some deployment situations are so specialized that the DOE may need to provide the required solutions, as it does today for more traditional computing support at the edge. Both the hardware demands of AI software (as noted

in the previous section) and the marketplace of devices are likely to continue to dramatically evolve in the coming years.

DOE should create a focused strategy to shape AI hardware to serve its science mission. Key to success is a strategy that leverages DOE, community, and industry investments in technology and scalable solutions. The industry will continue its dramatic pace of advancement over the next decade, but those advances are focused on goals that will not lead to meeting the requirements of DOE computational science and experimental data applications. In particular, the AI use cases for scientific applications will differ significantly, requiring extreme data rates, low-latency response, and extensive exploitation of explicit knowledge. Second, the rapid growth of AI training costs will create sustainability challenges to the growing burden of AI computing, forcing new approaches.

Looking further ahead, the ASCR facilities will continue to design complex, technically advanced networking and computing systems for future science generations where the needs of the AI ecosystem will be an integral part of any initial design. Given the pace of change in AI technology and techniques, these future facilities will also need to be designed with flexibility in mind to take advantage of the advances that will inevitably come from application work over the next decade.

6.6 Instrument to Edge Computing

Edge Computing

At DOE's modern user facilities, it is well known that it can take months for users to analyze data collected in only a few experimental shifts. The shortage of widely available, science-driven, AI/ML tools for rapid data-analysis and inference, together with appropriate computing, data storage, and networking support, is now a critical bottleneck hindering the extraction of new scientific knowledge from major DOE facilities. With new facilities coming on-line, and the planned upgrades of existing experimental facilities, this data-analysis bottleneck will become more severe and scientists will genuinely be overwhelmed with data. Some have referred to this data overload as a 'looming crisis' while Turing Award winner, Jim Gray, preferred to work with scientists who were 'drowning in data' to help them exploit modern computer science technologies. Thus, an important component of any ***AI for Science*** initiative must be the mitigation of this data analysis bottleneck with the development of new, domain-aware AI/ML methods and provision of appropriate software and hardware infrastructure.

The number of network-connected devices (instruments, computers, data stores) is growing at an exponential rate. New AI/ML technologies, including deep learning, will be critically important to fully exploit complex instruments and facilities, replacing pre-programmed hardware event triggers with algorithms that can learn and adapt, as well as discover unforeseen or rare, rate-limiting events that would otherwise be lost in compression. Sensors and other real-time hardware and software together with other significant edge computing resources located at the facilities are needed both to detect events and anomalies more efficiently and to reduce the raw instrument data rates to manageable levels.

The goal of DOE's 'Instrument-to-Edge' activity is to chart a course for deploying data analysis methodologies, tools, and services for SC experimental user facilities that exploit new AI/ML technologies, based on some commonality of hardware and software infrastructure. These edge resources at the facilities need also close linkage to the major SC computing facilities (LCF, NERSC) that can support the advanced use of simulation, and AI to optimize and steer experiments as needed.

Hardware-Software-Algorithm Co-design

The use of AI/ML technologies is rapidly permeating the design and operation of scientific instruments. This trend is expected to grow in extent and intensity, eroding the traditional partition between hardware and software. Optimal design, construction, and operation of major user facilities, therefore, require a deep co-design approach that integrates user experiments with developments in hardware, software, and edge computing. The DOE Labs' co-design culture has matured during the Exascale Computing Project (ECP) and involves teams of scientific users, instrument providers, physicists, mathematicians, and computer scientists as well as hardware designers. As recommended by the ASCAC ECP Transition Report [10], ASCR should take advantage of this legacy of co-design expertise to develop a major 'Instrument-to-Edge initiative with BES and the facilities. For example, two of the six ECP co-design centers are of immediate relevance to an ***AI for Science*** initiative:

- The ExaLearn Co-design Center is identifying the fundamental machine learning challenges associated with ECP applications and developing scalable AI/ML technologies for the analysis of data generated not only by exascale applications but also by the DOE user facilities.
- The goal of the Center for Online Data Analysis and Reduction at the Exascale (CODAR) is to produce a coherent software infrastructure and to release appropriate software tools and libraries.

Much of the ECP experience in co-design and software infrastructure development can, therefore, be leveraged to assist in the creation of new AI/ML capabilities and tools that can be readily applied across the agency's diversity of instruments, facilities, and infrastructure. ASCR therefore needs to provide strong and concerted support for AI/ML based co-design throughout the design, construction, and operation cycles of major facilities.

6.7 AI/ML Workforce: Training, Focusing, and Retention

Industry, national laboratories, government, and broad areas of academic research are making more use than ever before of AI, ML, and simulation-based decision-making. This trend is broadly apparent across many domains such as energy, manufacturing, finance, and transportation. These are all areas in which AI is playing an increasingly significant role, with many more examples across science, engineering, business, and government. Research and innovation, both in academia and in the private sector, are increasingly driven by large-scale computational approaches using AL and ML technologies. With this significant and increased use comes a

demand for a workforce versed in technologies necessary for effective and efficient AI/ML-based computational modeling and simulation as well as in data science. There is a high demand for graduates with the interdisciplinary expertise needed to develop and/or utilize AI techniques and methods to advance the understanding of physical phenomena in a particular scientific, engineering, or business field and to support better decision-making.

Training and retention of a cadre of young people for DOE SC in the area of AI is vital. Over the past 20 years, the IT industry has expanded dramatically, driven by e-commerce, social media, cloud services, and smartphones, with the IoT, healthcare sensors, industrial automation, and autonomous vehicles further expanding the domain of big data analytics and services. In response to seemingly insatiable workforce demands, most students are now trained in software tools and techniques that target these commercial opportunities rather than scientific computing and HPC. Few students outside of scientific domains learn C, Fortran, or numerical methods, which could be considered the traditional 'tools of the trade' in computational sciences, and engineering. This trend is an extension of one that began in the 1990s and is irreversible. Consequently, the scientific community is beginning to embrace new tools and approaches for artificial intelligence, and machine learning for science while also encouraging students to learn both HPC and data analytics tools.

A strong research program will crucially rely on a complementary education component, which is as important as adequate infrastructure support. A continuing supply of high-quality computational and data scientists available for work at DOE laboratories is critical. For example, the DOE Computational Science Graduate Fellowship (CSGF) program has successfully provided support and guidance to some of the nation's best scientific graduate students, and many of these students are now employed in DOE laboratories, private industry, and educational institutions. To meet the increasing need for computational and data scientists trained to tackle exascale and data-intensive computing challenges, there is now a significant requirement for a similar fellowship program supporting training in exascale and data intensive computing and related areas, as outlined previously in this report.

Other examples that could be replicated to help train both new scientists, as well as focus and retrain existing scientists to utilize exascale computing and data analytics are the Argonne School on Extreme-Scale Computing and the Berkeley Summer School on Deep Learning for Science. Expanding or replicating such summer schools on an ***AI for Science*** theme and introducing similar courses on data-intensive computing could help train a new generation of scientists capable of tackling the challenges of ***AI for Science*** and data intensive computing. Such schools would also serve to update and upgrade the skills of existing experimental scientists and computational scientists in these important areas as well. The schools would also provide a unique opportunity for the Office of Science to create a skilled diverse and inclusive workforce that is valued both by the DOE Lab community and by US industries [9, 10].

In addition to courses and schools, training the workforce for ***AI for Science*** will require the availability of well-curated, relevant datasets. Much of AI instruction today relies on commonly available public datasets (e.g. Youtube videos) that are not appropriate for teaching and training

the appropriate AI techniques relevant to the DOE mission. The DOE could produce and provide open datasets relevant to various areas of the DOE mission and share them broadly both for in-house programs and for universities.

6.8 University Partnerships

University-based investigators have driven conceptual advances in fundamental computer science that have furthered the frontiers in AI from the start. Many important breakthroughs in AI have been accomplished at universities, often in collaboration with technology companies and national laboratories. During the past two decades, US universities have been growing their research activities in AI to varying degrees. These have typically been cross-disciplinary activities involving a subset of computer science, mathematics, statistics, and electrical engineering departments as well as traditional science departments. At several institutions, joint institutes or centers and technology hubs have been successfully established. These university-based activities have played an essential role in establishing the currently existing AI workforce, including scientists, engineers, and developers at national laboratories, technology companies, and startups. They have been instrumental, along with the national laboratories, in training engineers and technical support for AI infrastructure. University engagement is typically through sponsored research agreements and is particularly effective via stable, long-term funding programs that allow sustained partnerships to flourish.

National laboratories, universities, and technology companies and startups have begun to convene consortia in various forms to extend capacity in AI. National laboratories and universities have established agreements with technology companies and startups to provide them with mutually beneficial collaborative access to computational resources and data science experts. The scope of research at the national laboratories typically encompasses elements integral to the AI programs supported by the DOE Advanced Scientific Computing Research (ASCR), Basic Energy Sciences (BES), Biological and Environmental Research (BER), Fusion Energy Sciences (FES), Nuclear Physics (NP), and High Energy Physics (HEP) Offices. Participants in these partnerships benefit from sharing expertise and resources, which accelerates the development of new AI-related ideas and empowers them to explore technologies built from these advances.

The stakeholders of ASCR, BES, BER, FES, NP, and HEP together should consider the following:

- New funding mechanisms to encourage universities and laboratories to hire scientists, engineers, and developers with expertise at the interface of AI and BES, BER, FES, NP, and HEP into early-career positions that have long-term opportunities.
- Development of flexible arrangements that support the engagement of DOE laboratory staff with technology companies and startups.
- Increasing the availability of fellowships and visiting scholar positions at the national labs to foster the growth of collaboration with the ASCR, BES, BER, FES, NP, and HEP communities needed to build this multi-disciplinary community.

AI for Science will be defined in many ways by new thinking. It is difficult to overstate the importance of bringing young people into this effort to create and explore the new ideas that will

be required. At the same time, one must ensure that sufficient funding is in place to ensure that the workforce development as a whole – from educating undergraduates, to nurturing young professionals and researchers, to establishing fulfilling permanent positions – is realistically supported. Collaboration with the NSF in relevant research and training programs will be important, as will funding to encourage the development of university partnerships so that universities can play a major role in this ***AI for Science*** agenda.

6.9 Collaboration with Industry

AI for Science has many opportunities for interactions between DOE and industry that have major benefits for both parties. These are naturally divided into two important classes: corresponding to the 'AI' and 'Science' parts of the proposed initiative. DOE is already interacting with industry on three aspects of AI technology – software, hardware, and algorithms. Software activities include optimizing DOE supercomputers and their AI/Exascale follow-ons to run the complex compute-intensive deep learning algorithms at the heart of AI that are often coming from technology companies. DOE is already looking at the products of the many startups developing novel AI hardware. DOE's research in algorithms (applied mathematics) and in computer science will be a key part of their interaction with the AI-oriented industry.

AI will certainly transform science over the next 10 years but it will also transform large parts of the industry. AI/ML methods developed for science will have both direct and indirect importance for Industry.

- DOE's work on the AI design of novel materials will be the basis of interactions with the manufacturing industry and illustrates the importance of AI designed surrogates to speed up complex simulations of material properties. Surrogates give a direct and fast map from structure to properties and are also seen for chemicals and the drug industry (the new QSAR). Such surrogates are also allowing engineering to be model-based, by using digital twins for designs and decision making. They are seen in General Electric's use of surrogates to provide their engineers with immediate response on the aerodynamic implications of design decisions for their engine industry [14].

- Fusion research is conducted as a partnership between national labs, universities, and a growing number of private companies engaging in both research and development of fusion concepts. DOE's AI for control of fusion devices such as the tokamak becomes the basis of an interaction with industry which is just one illustration of the opportunities for AI to assist in monitoring and control across industries.

- Naturally, AI is having and will have an important impact on the broader energy industry. AI opportunities here include optimization of energy production including both oil and renewable sources such as wind farms. Control of electrical power grids will increase the opportunities for AI as heterogeneous microgrids, virtual power plants, and smart energy-saving homes become more common. Optimization of distributed batteries including those in e-cars is a new possibility to support sporadic power sources such as solar. More

traditionally AI will monitor global power grids and help avoid instabilities leading to unplanned blackouts. It is also thought that AI will give better algorithms for the trading of power between regions and providers.

- Other major industries impacted dramatically by AI-enhanced science include automobile/transportation (reinventing itself as the mobility industry through the use of AI), space, and environmental engineering. AI for Space is a natural collaboration with NASA, with AI for remote system control, AI for remote robots, and AI for analysis of satellite imagery. Environmental engineering will need to combine smart sensors with novel hardware devices and real-time edge-based AI; it will contribute to a huge increase of new AI (today, deep learning) to interpret an explosion of geospatial time-series data.

- During the Covid-19 crisis, DOE's support for large-scale AI-driven simulations interacting with the drug industry is just one area where DOE research will accelerate the use of AI in the medical industry. Other examples are nanoscale sensors and complex systems simulations from the small (cells) to large scale (global pandemics).

One existing model for industry partners wishing to enhance its own AI/ML capabilities that could be followed is for SC to establish an industry collaboration for ***AI for Science*** similar to the High-Performance Computing for Manufacturing program (HPC4Mfg) supported by the Office of Energy Efficiency and Renewable Energy (EERE). This program connects industry to DOE Lab researchers in simulation and modeling and, in the first year, pays for a DOE employee or scientist to work on building and/or applying simulation packages to meet the specific industrial need. A similar ***AI for Science*** program to this HPC4Mfg program could direct industrial attention to the potential of DOE developed AI/ML technologies for their company. Another opportunity is to increase DOE's interaction with the largely industry-driven MLPerf organization that is developing open-source datasets and algorithm implementations for their large-scale performance-sensitive applications. DOE can collaborate with the IT Industry sponsors of the MLPerf consortium and expand the scope of the program to cover the performance of AI relevant to science.

We have deliberately discussed the science and engineering-related industries; however, algorithms and software will be generic to many fields and DOE will have the potential to interact with other industries transformed by AI, including commerce, entertainment, and sports.

6.10 Inter-Agency Collaboration

The DOE is clearly the lead agency for applications in big science, big data, and big computing at the DOE experimental facilities. It has also established national and international leadership in both Supercomputing and in the application of AI/ML technologies to the very large scientific datasets from their large-scale experimental Facilities. DOE has an existing successful collaboration with the NIH National Cancer Institute (NCI) through the ECP CANDLE project. In addition, the DOE is developing an MOU with both the NSF and NIH on collaborative research in a Computational Neuroscience program.

In the context of ***AI for Science***, SC should explore the creation of one or more Joint Research Institutes with NSF as well as the possibility of joint research calls. The subcommittee also believes it would be worthwhile for SC, working in coordination with DOE's Artificial Intelligence and Technology Office, NNSA, and other DOE programs as appropriate, to explore synergistic opportunities to work with other funding agencies such as NIST and DOD.

6.11 International Collaboration

There is a need for broad-based, coordinated action by like-minded international partners to harness the global scientific software community to address the research opportunities provided by data-intensive science. This reflects the fact that computational and data analytical methods driven by AI/ML are now universally accepted as indispensable to future progress in science and engineering. The last time a disruption of comparable dimensions occurred – during the transition from Petascale to Exascale supercomputers more than a decade ago – only a relatively small part of the scientific community felt the consequences of the struggle to manage the wholesale replacement of programming models, numerical and communication libraries, and all the other software components and tools on which application scientists were already building. AI/ML applied to computational and data science is still relatively young, and methods are still largely the province of relatively few scientific elites in a small number of physical sciences.

Today, aided by the success of the scientific software research and development community, researchers in nearly every field of science and engineering are beginning to turn to AI/ML approaches to assist in opening new areas of inquiry (e.g. the very small, very large, very hazardous, very complex). The goal is to dramatically increase research productivity and to amplify the social and economic impact of their work. Recent reports – such as the ***AI for Science*** Town Hall report and others (see Appendix C) – make a compelling case, in terms of both scope and importance, for the profound expansion of our research horizons that will occur if we can rise to the challenge of exploiting AI/ML technologies in computational and data science. However, in light of the radical changes in computing that are currently occurring, it is clear that the software infrastructure necessary to make these opportunities a reality does not yet exist and that we are a long way from being in a position to create it.

International leadership in ***AI for Science*** over the coming decade will hinge on the realization of an integrated set of programs spanning the four interdependent areas noted above – new applications, AI algorithms and foundations, software infrastructure, and hardware tools and technologies. In terms of an international effort rationale, scientists in nearly every research field and most countries will now depend on the development of software infrastructure for high-end computing and data analysis in order to open up new research fields and to dramatically increase their research productivity. Such an AI/ML exascale software infrastructure will underpin the global scientific communities who need to work together on problems of global significance and be able to leverage distributed resources in trans-national configurations. In terms of feasibility, the dimensions of the task – totally redesigning and recreating, in the period of just a few years, the massive software foundation of computational and data science to meet the new realities of ***AI for Science*** – are simply too large for any one country to undertake on its own.

To realize this vision, the Office of Science needs to:

- **Provide a framework for organizing the software research community**

  An organizational framework needs to be designed to enable the international software research community to work together to deliver a more capable and productive ***AI for Science*** environment. The framework should include elements such as initial working groups, outlines of a system of governance, alternative models for shared software development with common code repositories, feasible schemes for selecting valuable software research, and incentivizing its translation into usable, production-quality software for application developers, etc. This organization must also foster and help coordinate R&D efforts to address the emerging needs of users and application communities on new platforms.

- **Create a thorough assessment of needs, issues, and strategies**

  As part of its planning process, the international effort should assess the short-term, medium-term, and long-term needs of applications in an AI-enabled future. Participation in the effort from representative application communities and vendors will help ensure the adequacy of these assessments. The work of the organization that emerges from the effort must be prepared to provide DOE and other domestic and like-minded foreign research-oriented agencies with a series of well-crafted reports on the critical technical issues in the development of an ***AI for Science*** software infrastructure and with alternative strategies, both technical and programmatic, for solving these problems.

- **Initiate development of a coordinated software roadmap**

  Working with the results of its application needs assessment, the international effort will initiate the development of a coordinated roadmap to guide open-source ***AI for Science*** software development with better coordination and fewer missing components. This roadmap will help to guide both cooperative development and joint research efforts.

- **Encourage and facilitate collaboration in education and training**

  The magnitude of the changes in programming models, software infrastructure, and tools brought about by the transition to ***AI for Science*** will produce tremendous challenges in the area of education and training. The international effort will, therefore, provide a plan for cooperation in the production of education and training materials to be used in curricula, at workshops, and online.

- **Engage and coordinate vendor community in crosscutting efforts**

  To leverage resources and create a more capable software infrastructure for supporting ***AI for Science***, the international effort should engage and coordinate with vendors across all of its other objectives. Vendor participation in, and contributions to, all of these

objectives – the comprehensive application needs assessment, well-ordered but adaptive software roadmap, an organized framework for cooperation, coordinated R&D programs for new exascale software technologies – will be encouraged and facilitated.

Within the DOE's Office of Science, ASCR is well suited to take on these challenges.

6.12 Importance of ASCR's long-term Applied Mathematics and Computer Science Research Programs

Advancing the mathematical, statistical, and information-theoretic foundations of artificial intelligence are vital to realizing the full potential of ***AI for Science***. These foundations are now a bottleneck for scientific discovery and the practical application of AI/ML remains predominantly an art. As discussed in section 6.3 on AI algorithms and foundations, significant progress is required on multiple fronts. These efforts must be complemented by advances in the computer science and mathematical foundations of AI that will be required to complement capabilities in hardware and software and realize the full potential of AI in DOE's science and engineering missions.

Existing efforts within the ASCR Applied Mathematics and Computer Science program are complementary to key directions for AI foundations research outlined in Section 6.4. For instance, optimization algorithms, differentiation techniques, and models form the foundation of training in AI. Opportunities exist for fundamental advances in this area to impact ***AI for Science*** from the HPC facility to the edge. Similarly, graphs arise naturally in many scientific domains (e.g., molecules, protein interaction networks, community networks), and modern ML research focused on learning with graph structures complements ASCR investments in graph theory. Finally, given the data explosion at DOE facilities, there is a need for smart detectors and associated high performance embedded computing at the edge to complement work on AI software and hardware described in Sections 6.3 and 6.5. These efforts complement ASCR work on data management and advanced networking, e.g., real-time distributed computing from detector directly to a facility for processing.

Traditional applied mathematics and computer science research are essential to progress in the areas of AI and ML. Applied mathematics and computer science research are vital for scientific computing and its central role in US economic vitality, energy security, the environment, and national security. For the US to maintain an international edge in AI and ML, the subcommittee supports the recommendation of the ECP Transition report that ASCR should substantially reinvest in this research program and renew a stable environment for basic research [10].

7. Summary of Conclusions

The subcommittee makes six major recommendations to the Office of Science:

1. Creation of a 10-year ***AI for Science*** Initiative
2. Structure of an SC ***AI for Science*** Initiative
3. An Instrument to Edge Initiative

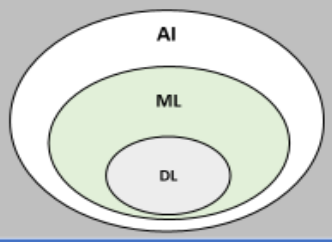
4. Training, focusing, and retention of AI/ML workforce
5. Inter-Agency collaboration
6. International collaboration

In addition, the subcommittee stresses the importance for all six of the Office of Science programs to work together on the issue of hardware-software-algorithm co-design and data analysis at their major user facilities. Finally, the subcommittee supports the recommendation of the ECP Transition report [10] that stresses the importance of ASCR's long-term Applied Mathematics and Computer Science research programs.

Figures

Figure 1: AI, Machine Learning, Deep Learning in a Nutshell

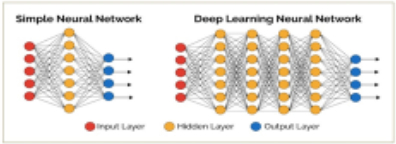
Artificial Intelligence (AI)
How can computational technologies be developed & used to assist, augment or automate human skills?



Machine Learning (ML) - Foundational basis for AI.

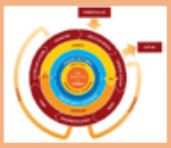
ML algorithms make predictions, decisions, & estimates from building a mathematical model or "learning" based on probabilities, samples, or training data.

Deep Learning (DL) - Includes neural network-trained approaches for tasks such as spam filtering, fraud/anomaly detection, image analysis.

Cognitive Skills (AI)	Specific Tasks (ML)	Neural Networks (DL)
Vision & Perception Natural Language Processing Search & Planning Problem solving Knowledge reasoning	Classification Clustering & regression Simplified or surrogate models Feature extraction Pattern recognition	

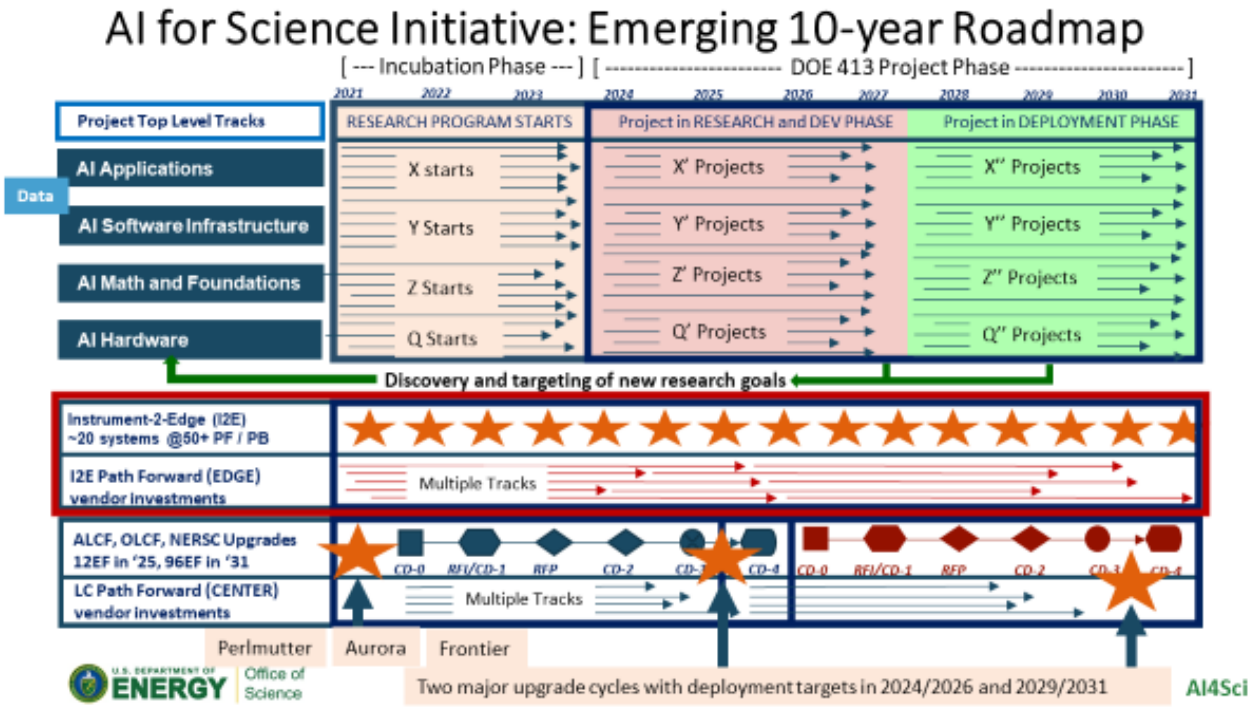
Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence (DOE/ASCR, 2019)

Figure 2: What is a Data Scientist?

Data Engineer	**People who are expert at** • Operating at low levels close to the data, write code that manipulates • They may have some machine learning background. • Large companies may have teams of them in-house or they may look to third party specialists to do the work.
Data Analyst	**People who explore data through statistical and analytical methods** • They may know programming; May be an spreadsheet wizard. • Either way, they can build models based on low-level data. • They eat and drink numbers; They know which questions to ask of the data. Every company will have lots of these.
Data Steward 	**People who think to managing, curating, and preserving data.** • They are information specialists, archivists, librarians and compliance officers. • This is an important role: if data has value, you want someone to manage it, make it discoverable, look after it and make sure it remains usable.

What is a data scientist? Microsoft UK Enterprise Insights Blog, Kenji Takeda
http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/01/31/what-is-a-data-scientist.aspx

Figure 3: Structure of SC *AI for Science* 10-year Initiative



References and URLs

[1] Maintaining American Leadership in Artificial Intelligence (February 2019). https://www.whitehouse.gov/ai/executive-order-ai/

[2] Report on the DOE Town Hall Meetings on AI for Science; Rick Stevens, Valerie Taylor, Jeff Nichols, Arthur Barney Maccabe, Kathy Yelick and David Brown (January 2020). https://www.anl.gov/ai-for-science-report

[3] J. McCarthy: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

[4] M. I. Jordan and T. M. Mitchell. 'Machine learning: trends, perspectives, and prospects', Science 349.6245 (July 2015), 255–260.

[5] Artificial neural network: https://en.wikipedia.org/wiki/Artificial_neural_network

[6] K. He, X. Zhang, S. Ren, J. Sun. 'Deep Residual Learning for Image Recognition' in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770-778; (arXiv:1512.03385; doi:10.1109/CVPR.2016.90).

[7] S. Lee. Scientific Artificial Intelligence and Machine Learning: Workshop on Predictive Models and High Performance Computing as Tools to Accelerate the Scale up of New Bio Based Fuels, June 9-11, 2020, Slide 3, https://www.energy.gov/sites/prod/files/2020/06/f76/beto-02-modeling-wksp-june-2020-lee.pdf#page=3

[8] Microsoft UK Enterprise Insights Blog, Kenji Takeda: http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/01/31/what-is-a-data-scientist.aspx

[9] PCAST report Recommendations for Strengthening America Leadership in Industries of the Future: https://science.osti.gov/About/PCAST/Meetings

[10] ASCAC ECP Transition Report: https://science.osti.gov/ascr/ascac/Reports

[11] DOE-NIH CANDLE project https://datascience.cancer.gov/collaborations/joint-design-advanced-computing/candle

[12] ASCR-BES CAMERA project https://www.camera.lbl.gov/

[13] Report of Roundtable Meeting on AI/ML in NP Facilities (January 2020) https://arxiv.org/abs/2006.05422

[14] Presentation to subcommittee from GE

15] D. Wilkinson, et. al, The FAIR Guiding Principles for scientific data management and stewardship; *Sci. Data*3:160018, 2016; (doi: 10.1038/sdata.2016.18).

Appendix A: Charge Letter

Department of Energy
Office of Science
Washington, DC 20585

Office of the Director

October 25, 2019

Professor Daniel Reed
Chair, Advanced Scientific Computing Research Advisory Committee
University of Utah
Salt Lake City, Utah 84112

Dear Professor Reed:

The Office of Science provides the largest and most diverse suite of scientific user facilities in the world – including the world's most capable high performance computing facilities. Planned upgrades will dramatically increase the amount of data produced across the SC scientific user facilities. Artificial Intelligence and Machine Learning (AI/ML) have the potential for providing new insights and even new discoveries from this data, including the correlation of experimental and computational data. However, the technical aspects of "AI/ML for Science" may be more challenging than currently envisioned. Over the last few years, several workshops and subcommittee reports have identified and enumerated the scientific opportunities and some challenges from the intersection of AI/ML with data-intensive science and high performance computing.

By this letter I am charging the Advanced Scientific Computing Advisory Committee (ASCAC) to assemble a sub-committee to look at the outputs from these activities and to analyze the opportunities and challenges for the Office of Advanced Scientific Computing Research (ASCR) and the Office of Science associated with Artificial Intelligence and Machine Learning. Specifically, I would like the sub-committee to deliver a report that:

- Assesses the opportunities and challenges from Artificial Intelligence and Machine Learning for the advancement of science, technology, and Office of Science missions.
- Identifies strategies that ASCR can use, in coordination with the other SC programs, to address the challenges and deliver on the opportunities.

Due to the cross-cutting nature of this effort, in assembling this subcommittee, please include members of, and recommendations from the other Office of Science Federal Advisory Committees, as well as Industry and other Federal experts.

We would appreciate the committee's preliminary comments by May 2020 and a final report by August 2020. I appreciate ASCAC's willingness to undertake this important activity. Your consideration on these issues will be an essential input to planning at the Department.

Printed with soy ink on recycled paper

If you have any questions regarding this matter, please contact either Barbara Helland, the Associate Director of the Office of Science for ASCR or Christine Chalk, the Designated Federal Official for the ASCAC.

Sincerely,

Chris Fall
Director
Office of Science

Appendix B: Subcommittee Members

ASCAC Members

Chair:
Professor Tony Hey
Chief Data Scientist
Rutherford Appleton Laboratory
Science and Technology Facilities Council
Harwell Campus
Didcot, OX11 0QX, UK

Professor Jack Dongarra
Innovative Computing Laboratory
EECS Department
1122 Volunteer Blvd
University of Tennessee
Knoxville TN 37996-3450

AITO Member

Dr. Fred Streitz
Chief Scientist
Artificial Intelligence and Technology Office
Department of Energy
1000 Independence Ave. SW
Washington DC 20585

University Research Members

Professor Ewa Deelman
Research Professor of Computer Science and Principal Scientist
Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marine Del Rey, CA 90292

Professor Anneila Sargent
Member, National Science Board
Cahill Center for Astronomy and Astrophysics
California Institute of Technology
1200 East California Blvd
Pasadena, CA 91125

Professor Geoffrey Fox
School of Informatics, Computing and Engineering
Indiana University
MESH Building,
2425 N. Milo B. Sampson Lane,
Bloomington, IN 47408

Dr. Dan Stanzione
Executive Director, Texas Advanced Computing Center
Associate VP For Research, The University of Texas at Austin
10100 Burnet Road (R8700)
Austin, Texas 78758-4497

Professor Joel Saltz
Associate Director, Stony Brook Cancer Center
Department of Biomedical Informatics
Stony Brook University
Stony Brook, NY 11794

Professor Rebecca Willett
Department of Computer Science
University of Chicago
5730 S. Ellis Avenue
John Crerar Library
Chicago, IL 60637

Office of Science Advisory Committee Members

BERAC
Kerstin Kleese van Dam
Director, Computational Science Initiative
Brookhaven National Laboratory
Upton, NY 11973

BESAC
Professor Abbas Ourmazd
Department of Physics
University of Wisconsin-Milwaukee
P.O. Box 413
Milwaukee, WI 53201

FESAC
Dr. Phil Snyder
Director of Theory and Computational Science
Energy and Advanced Concepts Group
General Atomics
P.O. Box 85608
San Diego, CA 92186

HEPAP
Professor Mike Hildreth
Department of Physics
University of Notre Dame
414 Nieuwland Science Hall
Notre Dame, IN 64556

NSAC
Professor Tanja Horn
Department of Physics
Catholic University of America
620 Michigan Avenue N.E.
Washington, DC 20064

Appendix C: Reports and Presentations

General AI Reports

- *Recommendations for Strengthening American Leadership in Industries of the Future*, Executive Office of the President of the United States, June 2020.
  https://science.osti.gov/-/media/_/pdf/about/pcast/202006/PCAST_June_2020_Report.pdf?la=en&hash=019A4F17C79FDEE
  Accessed 20 October 2020.

- *Preliminary Findings,* Secretary of Energy Advisory Board, AIML Working Group, March 2020.
  https://www.energy.gov/sites/prod/files/2020/04/f73/SEAB%20AI%20WG%20PRELIMINARY%20FINDINGS_0.pdf
  Accessed 20 October 2020.

- *Executive Order on Maintaining American Leadership in Artificial Intelligence,* The White House, February 2019.
  https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/
  Accessed 20 October 2020.

- *American Artificial Intelligence Initiative: Year One Annual Report*, The White House, Office of Science and Technology Policy, Feb. 2020.
  https://iapp.org/media/pdf/resource_center/US_AI_Initiative_year_one_annual_report_white_house.pdf
  Accessed 20 October 2020.

- *Summary of the 2018 White House Summit on Artificial Intelligence for American Industry*, The White House, Office of Science and Technology Policy, 10 May, 2018.
  https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf
  Accessed 20 October 2020.

- *Summary of the 2019 White House Summit on Artificial Intelligence in Government,* The White House, Office of Science and Technology Policy, September 2019.
  https://www.whitehouse.gov/wp-content/uploads/2019/09/Summary-of-White-House-Summit-on-AI-in-Government-September-2019.pdf
  Accessed 20 October 2020.

- *The National Artificial Intelligence Research and Development Strategic Plan,* National Science and Technology Council, NITRD Subcommittee, October 2016.
  https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf
  Accessed 20 October 2020.

- *Interim Report*, National Security Commission on Artificial Intelligence, November 2019.
  https://www.nscai.gov/reports
  Accessed 20 October 2020.

- *2016 – 2019 Progress Report: Advancing Artificial Intelligence R&D,* Artificial Intelligence Research & Development Interagency Working Group, National Science & Technology Council, November 2019.
  https://www.nitrd.gov/pubs/AI-Research-and-Development-Progress-Report-2016-2019.pdf
  Accessed 20 October 2020.

Office of Science Reports

- *AI for Science Town Hall Meetings*, DOE Technical Report, January 2020.
  doi:10.2172/1604756
  Accessed 20 October 2020.

- *ASCR/ECP Transition Report*, ASCAC Report, September 2020.
  https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202009/ExascaleTransition-ASCR-202009.pdf
  Accessed 20 October 2020.

- *Producing and Managing Large Scientific Data with Artificial Intelligence and Machine Learning*, BES Roundtable Report, October 2019.
  https://science.osti.gov/-/media/bes/pdf/reports/2020/AI-ML_Report.pdf
  Accessed 20 October 2020.

- *Advancing Fusion with Machine Learning,* Office of Science Report, September 2019.
  https://www.osti.gov/servlets/purl/1615241
  Accessed 20 October 2020.

- *Application of Artificial Intelligence and Machine Learning for the Operation of NP Accelerator Facilities*, NP Roundtable Meeting, January 2020.
  https://science.osti.gov/np/Research/ai/AI-and-ML-Roundtable
  Accessed 20 October 2020.

- *Report from the AI for Nuclear Physics Workshop,* Office of Science Report, June 2020.
  https://arxiv.org/abs/2006.05422
  Accessed 20 October 2020.

- *Machine Learning in High Energy Physics,* Community White Paper, July 2018.
  https://arxiv.org/abs/1807.02876
  Accessed 20 October 2020.

- *Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence*, ASCR Report, February 2019.

https://www.osti.gov/servlets/purl/1478744
Accessed 20 October 2020.

- *Data and Models: A Framework for Advancing AI in Science*, Report of Office of Science Roundtable on Data for AI, June 2019.
https://www.osti.gov/servlets/purl/1579323
Accessed 20 October 2020.

Other Agency Reports

- *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity*, DOD Report, February 2018.
https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.pdf
Accessed 20 October 2020.

- *NOAA Artificial Intelligence Strategy: Analytics for Next-Generation Earth Science,* National Oceanic and Atmospheric Administration Report, February 2020.
https://nrc.noaa.gov/Portals/0/Draft%20NOAA%20AI%20Strategy_v1-6.1_11.13.2019_update.pdf
Accessed 20 October 2020.

- *Strategic Plan for Data Science*, National Institutes of Health, June 2018.
https://datascience.nih.gov/nih-strategic-plan-data-science
Accessed 20 October 2020.

- *The Skilled Technical Workforce: Crafting America's Science & Engineering Enterprise,* National Science Board Report NSB-2019-23, September 2019
https://www.nsf.gov/nsb/publications/2019/nsb201923.pdf
Accessed 20 October 2020

Presentations

Office of Science Research Programs

- Kerstin Kleese van Dam for BERAC on 'BER AI Strategy and Requirements'.
- Phil Snyder for FESAC on 'Strategy and Potential of AI/ML in the Fusion Energy Sciences Program'
- Mike Hildreth for HEPAP on 'AI and HEP: An overview of opportunities'
- Tanja Horn for NSAC on 'Artificial Intelligence for Science – Nuclear Physics Overview'
- Abbas Ourmazd for BES on 'Impact of AI on Basic Energy Sciences'

AI Researchers

- David Womble on the 'ORNL AI Initiative'
- Becca Willett on 'AI: Challenges & Opportunities'

IT Software Companies

- Sanjay Padhi from Amazon
- Larry Zitnik from Facebook
- Peter Norvig from Google
- Ed Pitera from IBM
- Sarah Bird from Microsoft.

IT Hardware Companies

- Andy Hock, Head of Product at Cerebras Systems, on 'Wafer-Scale AI Computing for Science'
- Raja Koduri, Senior Vice President, Chief Architect, and General Manager of Architecture, Graphics, and Software at Intel, on 'Exascale for Everyone'

Non-IT Industry

- Kim Branson, Global Head AI/ML for Medicinal Science and Technology R&D in GSK, gave a presentation on 'AI and ML at GSK'
- Rick Arthur, Senior Principal Engineer, Advanced Computational Methods Research, GE Research, gave a talk about more than 30 years of applying 'industrial AI' at GE

Other Agency Presentations

- Manish Parashar, Director of the Office of Advanced Cyberinfrastructure, gave a presentation on 'AI/ML Investments at NSF'
- Denise Caldwell, Physics Division Director, gave a talk on 'AI and Science at NSF'

Other Presentations

- James Sethian (LBNL) gave a talk on the ASCR-BES funded CAMERA project
- Laura Freeman (VA Tech) gave a talk on 'Statistical thinking and accelerating insights from Machine Learning and Artificial Intelligence'
- Tanmoy Bhattacharya gave a talk on 'Uncertainty Quantification'

Appendix D: List of Acronyms

AI:	Artificial Intelligence
AITO:	Artificial Intelligence and Technology Office
ANL:	Argonne National Laboratory
ANN:	Artificial Neural Network
ASCAC:	Advanced Scientific Computing Advisory Committee
ASCR:	Advanced Scientific Computing Research program
BER:	Biological and Environmental Research program
BERAC:	Biological and Environmental Research Advisory Committee
BES:	Basic Energy Science program
BESAC:	Basic Energy Sciences Advisory Committee
CANDLE:	CANcer Distributed Learning Environment
CAMERA:	Center for Advanced Mathematics for Energy Research Applications
CPU:	Central Processing Unit
CS&E:	Computational Science and Engineering
CSGF:	Computational Science Graduate Fellowship
DOD:	Department of Defense
DOE:	Department of Energy
ECI:	Exascale Computing Initiative
ECP:	Exascale Computing Project
EERE:	Office of Energy Efficiency and Renewable Energy
FAIR:	Findability, Accessibility, Interoperability, and Reusability
FES:	Fusion Energy Sciences program
FESAC:	Fusion Energy Science Advisory Committee
FOA:	Funding Opportunity Announcement
GAN:	Generative Adversarial Network
GE:	General Electric Company
GPU:	Graphics Processing Unit
GSK:	GlaxoSmithKline plc
HEP:	High Energy Physics program
HEPAP:	High Energy Physics Advisory Panel
HPC:	High Performance Computing
HPC4Mfg:	High Performance Computing for Manufacturing
ITER:	International Thermonuclear Experimental Reactor
IoT:	Internet of Things
LBNL:	Lawrence Berkeley National Laboratory
LCF:	Leadership Computing Facility
ML:	Machine Learning
MPI:	Message Passing Interface
MOU:	Memorandum of Understanding
NBIB:	National Institute for Biomedical Imaging and Bioengineering
NCI:	National Cancer Institute
NASA:	National Aeronautics and Space Administration

NERSC: National Energy Research Scientific Computing Center
NIH: National Institutes of Health
NOAA: National Oceanic and Atmospheric Administration
NP: Nuclear Physics program
NSAC: Nuclear Science Advisory Committee
NSB: National Science Board
NSF: National Science Foundation
ORNL: Oak Ridge National Laboratory
PCAST: President's Council of Advisors on Science and Technology
QSAR: Quantitative Structure-Activity Relationship
SC: Office of Science
SEAB: Secretary of Energy Advisory Board
UQ: Uncertainty Quantification

Acknowledgements

The Chair would like to thank Vint Cerf, John Negele, and Dan Reed from ASCAC, Steve Lee from ASCR, Pushmeet Kohli and John Platt from Google, James Sethian from LBNL, Rick Stevens from ANL, and David Womble from ORNL for helpful discussions. The subcommittee also wishes to thank Christine Chalk from ASCR and Deneise Terry and Tiffani Conner from ORAU for their invaluable support of the work of the subcommittee. Finally, Tony Hey thanks Yvette Wallus and OSTI for assistance with report citations.