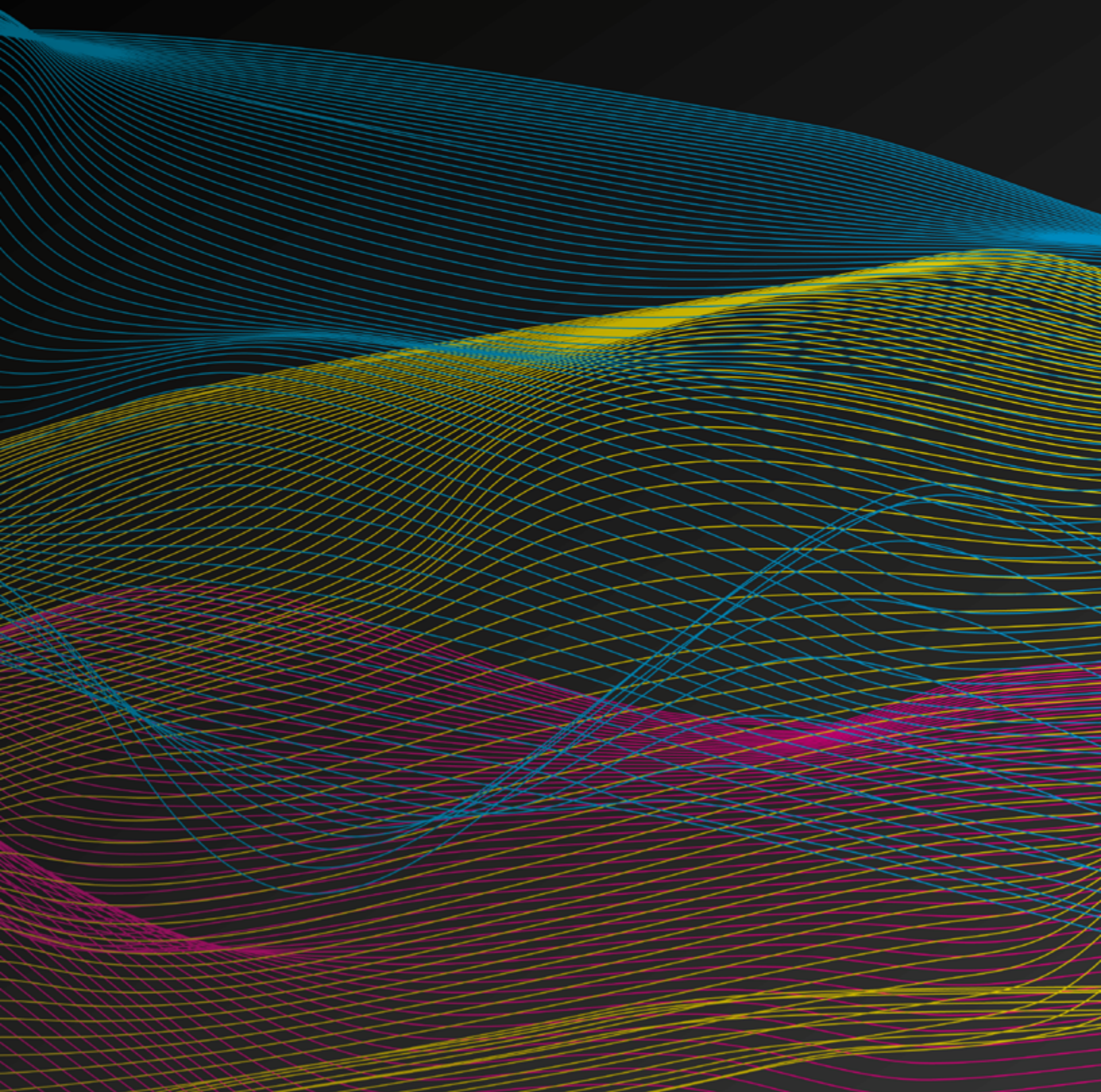
SMD AI WORKSHOP
12-14 MAY 2021

National Aeronautics and Space Administration

NASA SMD AI WORKSHOP REPORT

SEPT 2021



Part 1 The Workshop & Opportunities 4
Foreword 7
Executive Summary and Synopsis of Workshop 8
Introduction 12
Workshop Overview 16
Traceability Matrix 18

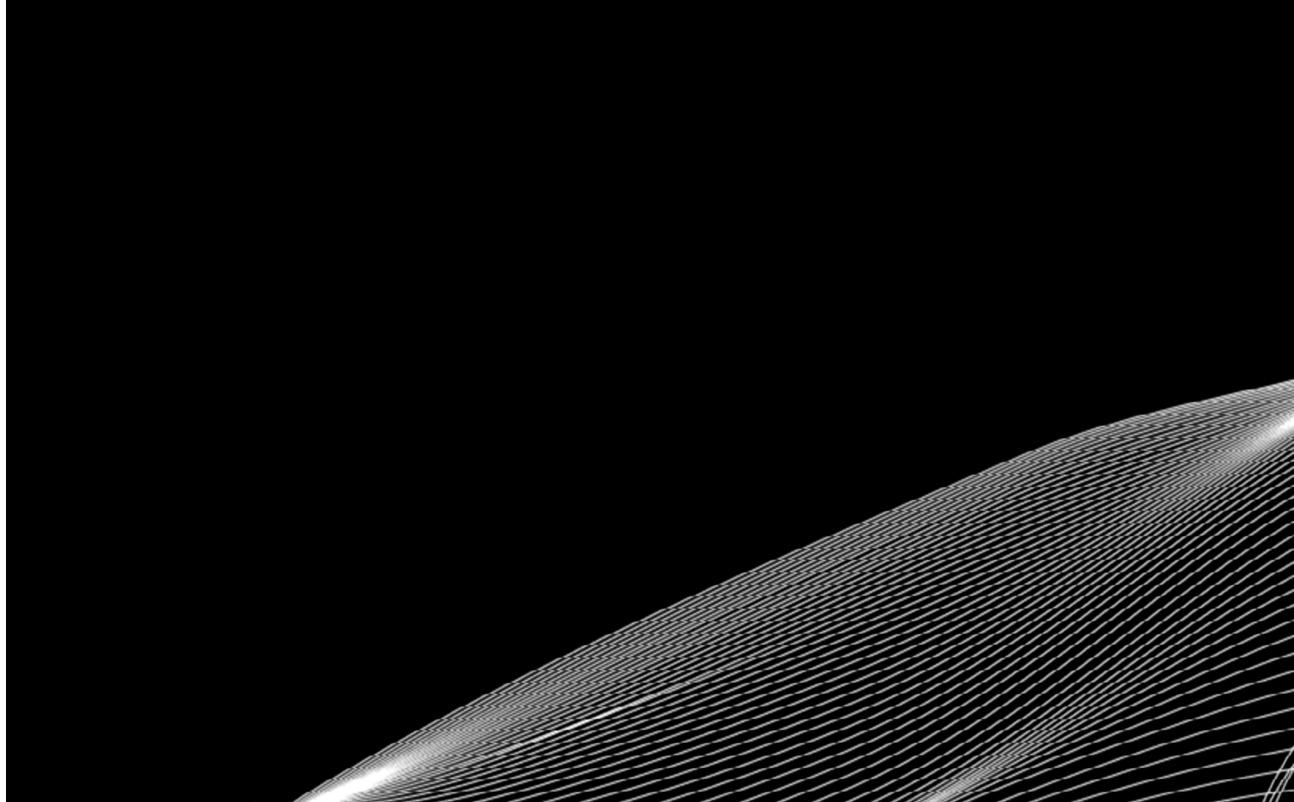
Part 2 Technical Memos 20
Standards for AI Readiness 22
Data Sparsity and Heterogeneity 30
Uncertainty and Bias 38
Reproducibility 54
Cataloging and Sharing AI Ready Data and Models 64
Computational Platforms 72
Cross Divisional Projects 88
Adapting Tools and Methods Across Domains 96
Practitioners Checklist and AI Ethics 106

Part 3 AI Auto Summaries 116
About the NLP Auto Summarization Project 120
NLP Auto Summaries 126

Part 4 Workshop Details 144
Organizing Team 146
Participants List 148

This document provides an overview of the inaugural **NASA SMD AI workshop** in May 2021.

We encourage interested readers to explore the memos herein in detail and look forward to the rich discussions that will undoubtedly follow.



PART 1
THE WORKSHOP & OPPORTUNITIES

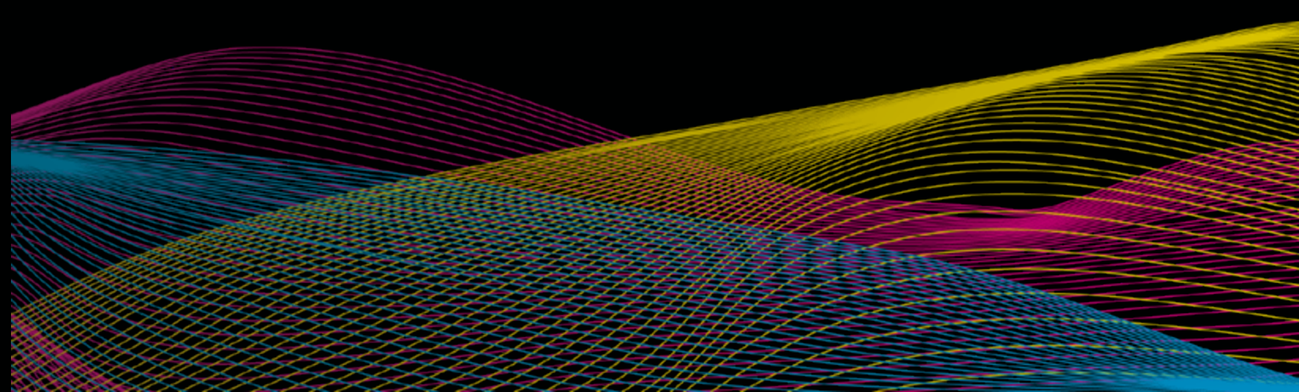
“Deep Learning excels at unlocking the creation of impressive early demos of new applications using very little development resources.

The part where it struggles is reaching the level of consistent usefulness and reliability required by production usage.”

“In general, there is very little research done on best practices for data curation / cleaning / annotation, even though these steps have far more impact on applications than incremental architecture improvements.

Preparing the data is an exercise left to the reader.”

François Chollet
Google Brain
Creator of Keras
Author of ‘Deep Learning with Python’



FOREWORD

This is the final report of a workshop carried out by the NASA Science Mission Directorate (SMD) Strategic Data Management Working Group (SDMWG) Artificial Intelligence (AI) team from May 12-14, 2021. The workshop was a direct response to the SDMWG Strategy for Data Management and Computing for Ground-breaking Science 2019-2024 recommendations.

The purpose of the workshop was to bring together NASA science community, academic partners, and industry experts to discuss ways to accelerate the adoption of AI across NASA science divisions and foster cross-disciplinary science enabled by AI utilizing large science data archives and computing platforms. This report is a summary of the three-day workshop, which consisted of keynote addresses, panel sessions, and breakout sessions that discussed various focus areas addressing the purpose of the workshop. This report has been reviewed by individuals with diverse perspectives and technical expertise.

In preparing the report, the aim has been to reflect the spirit of the discussions in a form from which participants and the stakeholders may gain insights and use the inputs to plan future activities. Some of the ideas discussed at the workshop have been further refined and additional background information added.

The workshop's success would not have been possible without the invaluable contributions by many speakers, panellists, moderators, and other participants who donated their time and expertise to inform these discussions. We wish to also extend a sincere thanks to each member of the NASA SMD SDMWG AI team for formulating the workshop, the Frontier Development Lab team for planning and hosting the workshop, and the reviewers and editors for assistance in developing the report.

The SDMWG will use this report to guide future SMD AI activities.

While we implement the AI activities, we will continue to receive additional input from the community through activities such as meetings, workshops, and webinars.

Manil Maskey
Lead, AI/ML Team, NASA SMD
Strategic Data Management Working Group
NASA HQ

EXECUTIVE SUMMARY AND SYNOPSIS OF WORKSHOP

EXECUTIVE SUMMARY

NASA Science Mission Directorate (SMD) Strategic Data Management Working Group (SDMWG) Artificial Intelligence (AI) team collaborated with the Frontier Development Lab (FDL) to host a virtual SMD AI workshop from May 12-14, 2021. The goals of the workshop were to explore ways to transform science data into AI-ready data, examine opportunities to utilize computational platforms for AI, foster applications of AI across multiple science domains, and develop next steps to realize the opportunities.

The three-day virtual workshop hosted a diverse group of researchers, AI practitioners, and stakeholders from academia, industry, government, and non-profit organizations.

The workshop was organized around nine focus areas:

1. standards for AI readiness;
2. data sparsity and heterogeneity;
3. uncertainty and bias;
4. reproducibility;
5. cataloging and sharing AI-ready data and models;
6. computational platforms;
7. cross divisional projects;
8. adapting tools and methods across domains; and
9. practitioners' checklist and AI ethics.

This final workshop report is based on expert opinion and information provided by the participants and curated by the focus area leads and the organizing committee.

SMD AI WORKSHOP
12-14 MAY 2021

SYNOPSIS OF WORKSHOP

The workshop identified common themes and limitations in the adoption of AI along with opportunities to accelerate its usage. These include increasing the trustworthiness and reproducibility of AI data and models, preparing AI-ready data and sharing training data and models, increasing the access to computing resources, and fostering an inclusive, collaborative community through training and engagement that are able to practice cutting edge ethical AI for science.

Increasing the trustworthiness of AI models and data is a critical step to increasing the usage and adoption in different communities. During the meeting, experts discussed several areas to improve the trustworthiness of AI models. These include:

- **development of best practices for overcoming sparse data that can cause overfitting of models;**
- **consideration of methods to measure appropriateness and the impact of the techniques used to overcome data sparsity or heterogeneity issues.**
- **review of uncertainty and bias at the SMD proposal stage and peer reviewed to increase the trust in AI training datasets and models; and**
- **inclusion of uncertainty quantification and known bias in science results.**

Reproducibility of AI experiments is difficult and methods to incentivizing reproducibility of science results based on AI need to be investigated to foster a culture of open reproducible science. Activities to help incentivize reproducibility include reproducibility challenges, identifying and documenting best practices and processes to address versioning and maintenance of AI-ready data and models, and open-source access to AI software, training data, and models.

With its long history of openly sharing data and funding large science teams, NASA SMD has an opportunity to enable greater sharing of AI-ready data and models. AI-ready datasets significantly lower the barrier to entry to using AI for science and allow for easier reproducibility of the results. The following activities can help to increase the sharing of AI-ready data and models:

- **development of subject matter expert (SME) informed AI-ready dataset standards to ensure that datasets and models are prepared consistently and ethically;**
- **development of reusable AI-relevant data management tools; and**
- **development and publication of labelled training data and trained models for AI applications and benchmarking.**

Access to computing resources is often a limitation to the adoption of AI for science. NASA SMD should lower the barriers to entry to access computing resources for AI to maximize its investment in high end computing capability (HECC). Additionally partnerships with commercial cloud providers could expand the availability of computing to users without NASA credentials. NASA should also expand current HECC to advance AI by fostering a collaborative hybrid open science environment where NASA and non-NASA researchers can work together.

AI techniques can be applied across different domains for analysis and knowledge discovery from large data archives. NASA SMD has an opportunity to accelerate the adoption and usage of AI across science disciplines through:

- **supporting cross domain collaboration and sharing of code, workflows, and best practices;**
- **training and education on possibilities of AI, skill development, and adaptation of existing models for different domains; and**
- **fostering an inclusive and collaborative culture in both data management and AI implementation within and across SMD programs.**

SMD AI WORKSHOP
12-14 MAY 2021

SYNOPSIS OF WORKSHOP

The expert group identified that NASA science needs practical guidance for ethically applying AI. The group encouraged NASA SMD to apply existing responsible research practices, develop a checklist for best practices, conduct ethics review of research, develop systems to examine unintended consequences of AI research, and include additional ethics experts into NASA discussions.

The level of enthusiasm and engagement of participants from all of NASA's science divisions and external partners indicates the importance of addressing the challenges identified to accelerate adoption of AI for NASA science. Identified opportunities to address these challenges include :

- **increase investments in interdisciplinary and collaborative AI projects using cross divisional competitive solicitations;**
- **invest in the generation and sharing of AI-ready data and models;**
- **incentivize reproducibility and open sourcing of AI artifacts;**
- **consider optimal and more open use of HECC and cloud computing for AI;**
- **develop best practices and guidelines for trustworthy AI for science;**
- **embed ethical considerations of AI into the science research process; and**
- **establish an AI center of excellence to provide leadership, planning, guidance, communication, and implementation of initiatives to accelerate AI across the SMD.**

INTRODUCTION

AI FOR EARTH AND SPACE SCIENCE

Artificial intelligence amplifies and extends our reach as scientists—expanding capabilities and creating new efficiencies. It's emerging at the same time as a rapid scaling in our data gathering capabilities, allowing us to observe our planet, star, and sky from multiple vantage points. Enormous quantities of streaming data can now be used to unveil dynamic, real-time insights in a way we have never been able to before.

In parallel, AI is changing how we think about the outputs of science, shining a forensic eye on the limitations of journal publication, how data is shared, and how results are reproduced. Its trans-discipline, iterative, and data hungry nature is also challenging classical definitions of how science is done. In AI, the journey starts with the data and the hypothesis often comes later.

Over the last few years, we have also seen excellent examples of the ability of neural networks to predict and simulate physical phenomena, from fluid mechanics, the wave equation, particle behavior, and climate dynamics. AI pipelines can now be developed which integrate prior scientific knowledge into workflows, both improving the accuracy of predictions, but also allowing results outside their initial training distribution. Once trained, physics-informed neural nets (PINNs) can predict phenomena with greater efficiency and confidence —a tantalizing new capability for applied science.

A DATASCOPE

In the same way that the telescope and microscope gave us the laws of gravitation and germ theory, the ability of AI to act as a 'datascope' promises an exciting new chapter of discovery. It is therefore worth emphasizing that this new chapter is different to the 'code and ship' mindset of the last generation of computing. In the same way that reliable air traffic control enabled jet travel, or allocation of radio spectrum facilitated the universal connectivity we now take for granted, there remains a substrate of challenges to AI adoption which require a new mindset at the ecosystem level, with data management at its very heart.

SMD AI WORKSHOP
12-14 MAY 2021

AN AI-READY MINDSET

Like humans, machines that learn are often fallible. AIs need to be taught and the data inputs constantly maintained before they can be trusted. This mindset is useful. It's important to frame AI as endlessly self-learning systems, where humans are very much in the loop.

As our learning machines get more capable, so must our human systems that maintain them. In other words, to successfully grow and benefit from this next chapter of AI, we need to establish management systems that work in harmony with AI and data, in a cycle of continuous improvement.

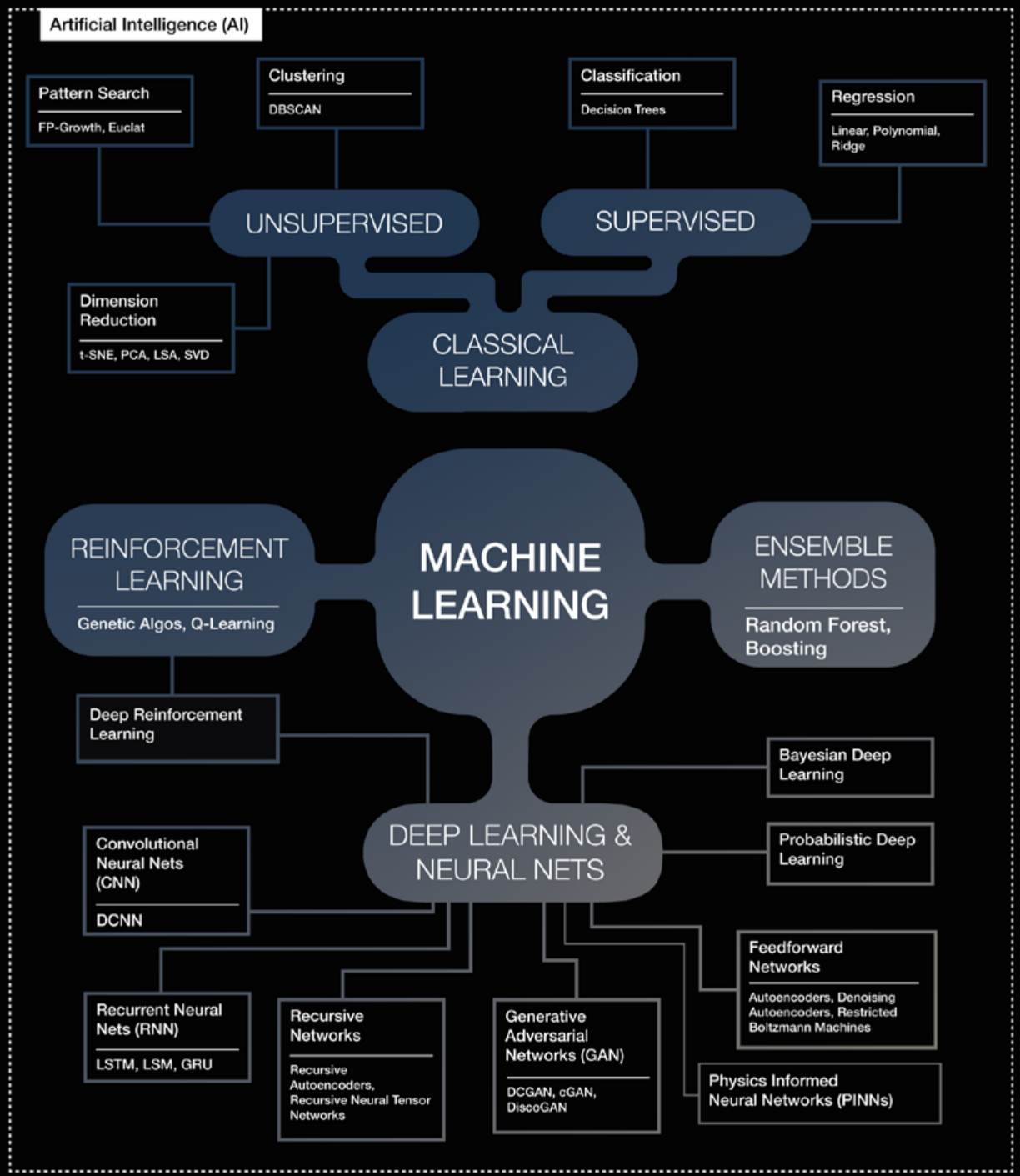
The broad outlines of these ecosystem level requirements are already well understood by NASA's SMD: AI-ready data, reproducibility, bias and uncertainty, and AI ethics, to name a few. Examining these core components became the goal of the inaugural AI SMD workshop in May 2021, to dive deeper into each component's aspects, understand the central and interconnected tensions, and surface opportunities.

The outcomes are captured in detail in the technical memos that follow. We'd like to thank everyone who lent their time, passion, and wisdom to these fascinating papers and invite readers to contribute to the discussion as our understanding evolves.

James Parr,
Director
Frontier Development Lab,
September 2021

In this report for the purposes of brevity, we are using the term artificial intelligence (AI) as a catch-all for machine learning / deep learning, artificial neural networks (ANNs), and traditional data science.

A non-exhaustive overview of the AI toolbox can be viewed here:



WORKSHOP OVERVIEW

Supporting NASA Science Mission Directorate research teams to build trusted and reproducible artificial intelligence/machine learning (AI/ML) pipelines for science will require a universal understanding of AI best practice to be shared across the community.

During May 12-14, 2021, NASA SMD Strategic Data Management Working Group (SDMWG) organized a workshop that brought together over 100 AI/ML experts from academia, government, and industry. The participants discussed nine focus areas on how to best utilize and advance AI/ML techniques for NASA science. This report captures individual memos from the nine focus area discussions that contain opportunities, challenges, and next steps for exploiting AI/ML using NASA science data and computational capabilities. The diverse community's experience with AI/ML, high performance computing (HPC), data systems, data curation and management, and policy development revealed a rapidly growing set of unique opportunities for groundbreaking science, novel discoveries, interdisciplinary applications, and stronger collaborations.

The nine focus areas discussed are shown in the diagram below. A technical memo was produced for each focus area, and the high level implications aggregated into the executive summary.

SMD AI WORKSHOP
12-14 MAY 2021

DAY01
SCIENCE DATA: OPEN, AI-READY, AND ETHICAL USE

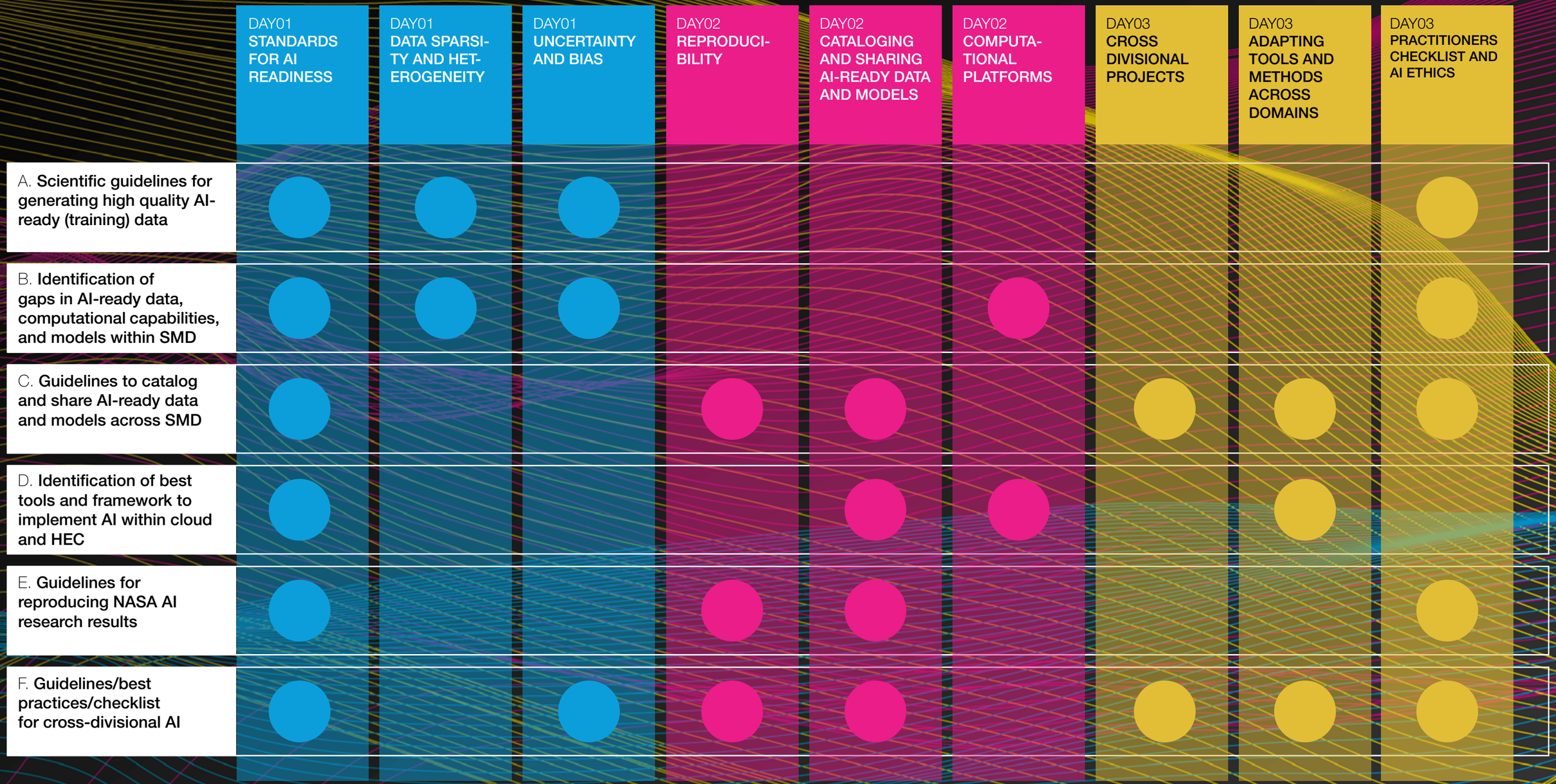
DAY02
TOOLS, SERVICES, WORKFLOWS, AND PLATFORMS TO CATALOG AND SHARE ML DATA AND MODELS

DAY03
APPLIED AI ACROSS DIVISIONS

TRACEABILITY MATRIX

SMD AI WORKSHOP: GOALS AND TRACEABILITY

Specific outcomes: while the case for AI and science is now well established within the scientific community, there is much opportunity to optimize. The following traceability matrix captures the key areas of emphasis and breakout session goals



PART 2
TECHNICAL MEMOS

FOCUS AREA 1

STANDARDS FOR AI READINESS

Exactly what do different divisions mean by AI ready and what are the guidelines for creating AI-ready data?

KEY CONCEPTS

AI readiness is the process in which raw data is converted into a dataset that can be ingested by a machine learning (ML) workflow. This process involves calibration, (such as the correction of an offset introduced by a measuring instrument), interpolation, framework adaptation, and other preprocessing steps. Additionally, AI readiness could involve ensuring some global properties of the dataset, such as uniformity across the dataset or statistical properties (e.g. presence of outliers or population imbalance).

WHY DOES IT MATTER?

AI-ready datasets allow for easier reproducibility (defined as the ability to reproduce an ML workflow to reach the same conclusions as the original work) of the results and help researchers and scientists create benchmarks for new ML models. However, there are no generalized guidelines for this process. The consequence: large amounts of duplicated implementations of code, hidden costs, and unvalidated scientific research (to name a few).

IMPLICATIONS

Should NASA SMD align efforts to develop a novel method to standardize the process of AI readiness?

Currently, a significant amount of time and resources of any ML project are invested in data labeling and preparation for ML applications. Standardization of the AI readiness of the dataset could help reduce the duplication of datasets in different divisions, avoid custom pre-processing, and therefore reduce the time employed in obtaining an AI-ready dataset. We can envision an automated and reproducible data prepossessing pipeline that could convert raw data into ML-ready data based on different specifications. Also, we should discuss the standardization of data augmentation, uniformization, and ways to smoothly tie the datasets to the labeling process.

WHAT CONTRIBUTORS SAY

“When we are trying to set up, say sensor classification with some of the Mars imaging data, there are things that we might need to go back and expand the metadata requirements that we had to add-on. Often there is a lot of cleanup to really make the data usable and searchable and analyzable, in that way, there are considerations that could have been brought in sooner.”

“We have a lot of image data, returned by the Stardust mission, we’re trying to recognize the tracks of interstellar dust particles, which we have had humans doing previously. Now we want to use machine learning to recognize these tracks. Our main challenge is not the data, the data is ready. It’s the training and generating a training set that works effectively”

“There is a major upfront investment to get going.”

TECHNICAL MEMORANDUM: FOCUS AREA 1

STANDARDS FOR AI READINESS

Authors: Lauren M. Sanders[1], Sylvain V. Costes[1]

[1]Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

As artificial intelligence (AI) and machine learning (ML) implementation picks up speed across NASA domains including the Science Mission Directorate (SMD), the need has grown for standardized guidelines surrounding AI readiness.

AI readiness encompasses many data preprocessing, curation and labeling steps between the generation of raw data and the implementation of AI methods. Standards for AI readiness become particularly important in light of recent findings by the NASA Advisory Council's Ad-Hoc Big Data Task Force that at least half of all scientific papers report results using archived rather than newly generated data[1]. SMD alone is projected to create over 100 petabytes of data per year, much of which is made widely available long-term to NASA and external researchers for analysis and modeling[2].

However, as Barbara Thompson noted in the focus area 1 main meeting, AI readiness is more than data processing and preparation. A process of AI readiness involves designing future missions and experiments with the intent to generate machine-interpretable, AI-relevant, useful, and ready data. This process will be aided by leveraging existing data standards such as the FAIR data principles for open science (findability, accessibility, interoperability, and reusability)[3], as noted by Sylvain Costes, focus area host. Moreover, SMD focus on interdisciplinary scientific discovery will necessitate the development of AI-readiness guidelines to allow for integration of datasets across scientific boundaries.

Here we summarize the topic identified from the 2021 NASA SMD AI Workshop on the current challenges, opportunities, and suggestions for developing SMD standards for AI readiness.

DIGGING DEEPER

Overview
The SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024 recommended SMD investment in incentives and education for AI/ML use for scientific discovery[2]. The same report noted that NASA's focus has traditionally been mission-oriented rather than strategically facilitating the use of cutting-edge applications. Thus, to ensure that machine learning data are prepared and analyzed consistently, ethically, and according to best practices, there is a need for a hierarchical, iterative, domain-sensitive set of standards and guidelines for AI-ready data generation within SMD.

Leveraging Diverse Data Types
AI-readiness guidelines must encompass the diverse types of data generated by the five science divisions within SMD. For example, wind measurement data from climate scientists within the Planetary Science division may be so large that the datasets must be subset prior to applying AI/ML methods. Subsetting guidelines must be designed in a statistically robust way so as not to introduce unnecessary bias into the downstream analysis. However, scientists working with biological datasets from the ISS face the opposite problem. Due to time and space constraints on the ISS, the sample numbers for space flown biological experiments are necessarily very small. Merging datasets together prior to AI/ML analysis increases statistical power but can introduce technical bias or batch effect. To address this, SMD AI-readiness standards could provide guidelines for overcoming the statistical limitations of diverse data types.

Even within a specific domain, similar datasets may have been generated from different platforms or using different parameters. For example, datasets from different types of X-ray space-based telescopes generate spectra with different properties, while gene expression datasets generated with different sequencing parameters display global shifts in gene abundance composition. Combining or comparing such datasets requires robust interconversion measures, without risking data degradation by imposing stringent data formatting requirements. Consultation with a subject matter expert (SME) is often required for appropriate data management, but the development of standardized ontologies by SMEs has in some cases alleviated this requirement. For example, different resources can have different versions of chemical names for the same compound, making it difficult to synthesize chemical datasets. SMD AI-readiness guidelines could include efforts to develop SME-approved standardized ontologies across domains to facilitate broad usage of diverse datasets in AI/ML applications.

In the development of metadata standardization and ontologies, SMD may benefit from the application of reinforcement learning algorithms to allow some datasets to label themselves based on small examples of labeling from SMEs. Similarly, natural language processing could be used as a tool to curate metadata in real-time. Alternatively, some groups have had success soliciting citizen scientists to label datasets, with secondary SME review. In all cases, robust measures are necessary to mitigate reinforcing any bias present in the existing labels.

SMD AI WORKSHOP
12-14 MAY 2021

Data Sharing and Reuse
Because many AI methods greatly improve outcomes with more observations, reusing and converting previously generated datasets to AI readiness will take advantage of the abundance of data both within SMD and externally. However, a scientific culture of data sharing is important to ensure that existing datasets are fully leveraged. In many cases, this can be communicated top-down from leadership and motivated effectively through funding incentives. For example, NIH-funded researchers are required to submit a data management and sharing plan to make data accessible and reusable. Subsequently, costs associated with data management and sharing are often covered by NIH funding. When possible, similar incentives at the SMD or NASA level may help foster a culture of data sharing throughout our research institutions.

Similarly, transparency and reproduction will be encouraged through appropriate incentives for storing and maintaining raw and processed data and generating detailed documentation on processing steps. This should include documentation on any domain-specific knowledge required for preprocessing, as well as contextual scientific information. Detailed documentation of the preprocessing steps is essential, as using preprocessed data without understanding can lead to bias and misuse in downstream results. It is important to note that even using raw data can lead to bias; for example, in the biological sciences, the available data may be limited and biased for technical or scientific reasons.

Beyond incentives for data maintenance and sharing, initiatives for educating science experts on the potential power of their data may facilitate a culture of AI-ready awareness. In some cases, life scientists with no statistical background may be unaware that their experimental data could be reused for an AI/ML application. Education initiatives which demonstrate the value of their datasets in an AI/ML context could encourage compliance with AI-readiness standards.

Considerations and Implications
Standards for AI readiness may impact research feasibility and continuity in multiple ways. Awareness of these impacts while designing and implementing standards will help avoid negative downstream effects and noncompliance.

Cutting-edge scientific data is often collected and formatted according to specific domain and data type constraints or economic considerations, rather than algorithmic input specifications. In some cases, imposing data standards may motivate removal of messy but valuable data. Care must be taken to ensure that standards for AI readiness do not become a lowest-common-denominator pressure that limits innovation and creativity. Engagement and involvement with each research community can help inform domain-specific constraints in this area. Further, buy-in from participants at all steps of data generation, processing, and maintenance is necessary to avoid blind rule following or even malicious compliance.

Coercing data into an AI-ready format can involve several preprocessing and statistical normalization steps. Documentation of these steps is clearly important,

but it is also important to make accessible the quantification of statistical error or uncertainty introduced in the process. As data uncertainty impacts model uncertainty, making it available or calculable when relevant is essential for responsible and ethical AI. For example, it is not unusual for SMEs to disagree or simply have different syntax for ontology or data labels. In some settings, saving the disparate SME assignments as a type of metadata can be useful in downstream applications. Certain analysts may want to keep only data points with full SME labeling agreement or use the level of certainty as metadata in their analysis.

IMPLICATIONS

An SMD AI-readiness framework should synthesize standards for newly generated AI-ready data, as well as harmonization guidelines for reusing existing datasets. Due to the varying degrees of AI readiness across experiments and datasets, designating an AI technology readiness level (TRL) could aid users in understanding the usability of a certain dataset for an application. A machine learning TRL framework has already been developed to rank technologies and could be used as a blueprint to design a similar framework for SMD scientific datasets[4]. An SMD AI-TRL platform could ask the user a few questions based on the needs of their application (e.g. “Are missing values acceptable?”) and based on the answers, could assign an AI-TRL value to all relevant datasets.

With respect to newly generated data, SMD scientists will benefit from top-down, basic AI-readiness standards for data and associated metadata. Due to the application specificity of many data preprocessing steps, it will likely be most useful to simply provide guidelines for making data generally usable for statistical analysis. More specific preprocessing steps can then be performed by the user. Similarly, rigid standards for all metadata may be counterproductive, but certain universal metadata could be required for a dataset to be considered AI-ready. For example, providing “time zero” and the time interval between data point collections would allow for comparison between experiments with different periodicity of collection. In some cases, it may be practical to implement these standards by simply updating existing data structure standards. Archives such as the Planetary Data System archive invest heavily in structuring data for long term preservation, and inclusion of AI-readiness guidelines would ensure that archived data are reusable for future AI applications.

In order to make best use of existing datasets, the SMD AI-readiness framework could be accompanied by a harmonization layer platform that maps a dataset into a standard space. This solution works as long as the dataset has sufficient metadata. Therefore, particular attention must be paid to designing comprehensive and widely applicable ontologies across all relevant domains. These ontologies can then serve as structure for a harmonization platform. Importantly, multiple versions of such a harmonization platform already exist. AWS has a proprietary harmonization method, while various highly specific

SMD AI WORKSHOP
12-14 MAY 2021

versions have been developed in academic groups (e.g. for image data[5], soil data[6], MRI data[7]). Rather than developing a novel method, SMD will likely benefit from evaluating and leveraging existing platforms from groups who have heavily invested in harmonization development. Leadership from SMD in this regard will generate buy-in across domains.

Overall, the outcome of an SMD AI-readiness framework must be informative enough to ensure compliance and reliability, while flexible enough to allow scientific creativity and maintain utility across domains.

CONCLUSION

As NASA SMD encourages widespread training and adoption of AI methods, stakeholders at all levels will benefit from a set of global, SME-informed AI data readiness standards. Foundationally, these standards can spring from broad guidelines for creating statistically useful datasets, which will allow flexibility for the wide diversity of data types generated within SMD. Accompanying these standards can be a series of standardized metadata ontologies and a harmonization platform for existing datasets which currently do not conform to AI readiness. Assignment of a technology readiness level value for AI to relevant datasets based on user needs will facilitate data reuse, and a culture of data sharing and reuse will encourage maximal return-on-investment from past and future SMD-generated datasets.

FOCUS AREA 2

DATA SPARSITY AND HETEROGENEITY

How are concepts such as data sparsity, imbalance, robustness, and heterogeneity breaking forces for NASA's SMD?

KEY CONCEPTS

Due to the broad spectrum of applications developed by the different divisions different issues arise when dealing with datasets. For example, medical data is usually sparse and Earth science data usually involves data fusion from different sources. This session will focus on the issues of sparsity and data heterogeneity. Data heterogeneity refers to data with great variability of types and formats. Heterogeneity occurs when a study requires data from different sensors or data is stored in different formats or coordinate systems. Sparsity refers to a lack of data in a certain dimension or paucity of a certain type of training data.

WHY DOES IT MATTER?

A large amount of time and effort are spent in harmonizing and unifying inconsistencies or synthesizing larger datasets from a sparse initial set (using autoencoders or GANS) so this issue could be viewed mainly as an economic problem. However, the current direction of AI-enabled research is moving towards leveraging data from multiple sources (e.g. different missions and historical data). Effective fusion of data is a powerful proposition for AI, improving the relevance of any obtained scientific results. In this context, managing data heterogeneity and sparsity becomes an eminent challenge and key amplifier of success.

IMPLICATIONS

Could heterogeneity and sparsity issues be prevented by introducing a set of universal guidelines across missions?

Solving the data heterogeneity problem would allow combining multi-domain and mission data and unlock the real power of data analytics. Standardizing the way effective homogeneity is achieved would reduce development time and allow for easier reproducibility of the results.

WHAT CONTRIBUTORS SAY

"One researcher's noise is another researcher's data. I have worked with missions where they do pre-processing on the spacecraft and by the time it gets down, it's conditioned and selected only for a particular type of research. I've run into situations where those who develop the onboard processing made it inconvenient for those to then use that data for other purposes. It would be wonderful if in future versions where there were liaisons to other fields of research that could at least provide some input into those who are developing that onboard processing."

"Data fusion often requires significant subject matter expertise just to manage and harmonize inconsistent data types."

"If you have a large mission, where the only source of data that you care about is from that mission, then you don't have a heterogeneity problem. Within minutes, you cross those domain and mission boundaries and seek to combine any other data, you immediately have a heterogeneity problem and quite frankly, given how science research is going, that is the future, as we find that single mission science is no longer the best science we can do."

TECHNICAL MEMORANDUM: FOCUS AREA 2
DATA SPARSITY AND HETEROGENEITY

Authors: Megan Ansdell, NASA Headquarters, Washington, DC
Hannah Kerner, University of Maryland, College Park, MD

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

The topic of focus area 2 was data sparsity and data heterogeneity. Data sparsity and heterogeneity are truly cross-divisional issues—many areas of science across SMD are faced with data sparsity and/or the need to harmonize data before usage with AI/ML models. For the purposes of the workshop, data sparsity refers to the paucity of useful data in a given dimension for a given problem. For example, there may be few-to-no labeled examples for land cover classification in Kenya resulting in a sparse dataset when pairing these labels with Earth observations, in comparison to other regions like the United States with more publicly available labeled datasets; data sparsity in land classification could also arise when Earth observations are affected by significant and persistent cloud cover, resulting in a sparse dataset for a given region. Data heterogeneity then refers to data with large variability in one or more aspects, for example in terms of data types and formats, which could occur when a given problem requires data from multiple sensors on different spacecraft.

The guest speaker, David Donoho, professor of statistics at Stanford University, aptly set the stage by asking participants to think deeply about what parts of the recent progress touted by industry, in particular AI scaling, are truly useful and applicable to science research. Often large, complex neural networks that produce the highest accuracy results, but at the cost of requiring huge amounts of labeled data and compute, are not necessarily needed to answer science questions or are infeasible to deploy in scientific contexts, for example on rovers at distant locations such as Mars or Europa. The focus area hosts, Megan Ansdell and Hannah Kerner, urged participants to focus on practical discussions in the breakout sessions, for example by sharing useful tools and methods that have worked for their own research and could help to solve common problems faced by scientists applying AI/ML to NASA research areas where sparse and/or heterogeneous datasets are common. The result was a rich collection of common problems, practical solutions, and ideas for the future.

DIGGING DEEPER

Challenges of Data Sparsity in NASA Science
Data sparsity finds its way into many areas of NASA science research. For example, in Earth science, satellites orbiting our planet and monitoring its surface can be limited by persistent cloud cover that renders observations unusable for a given application and therefore can result in sparse datasets over certain parts of the world. Additionally, in-situ ground truth data is often important for interpreting Earth observation data but can be much sparser relative to the larger regions and/or more dense sampling obtained by the satellites, thereby requiring interpolations and other assumptions when combining the two. Identifying an appropriate near truth point on the ground in both space and time to a given satellite track is another dimension of data sparsity that requires a priori assumptions of where, when, and what is acceptable to sample as ground truth. These assumptions are not always obvious and can have significant implications for the ability of an AI/ML model to make accurate predictions.

Spacecraft telemetry is another area that faces data sparsity issues and is common to all of the NASA SMD divisions, as they all use space-based assets such as satellites, rovers, and/or landers. Telemetry is gathered by various onboard sensors that return information on the health and state of the spacecraft, but these sensors can report at very different cadences. Some sensors will only report a value when there is a change, others report a value at a regular cadence, and some sensors are not always turned on or are prone to erroring out. This presents a data sparsity (and data heterogeneity) problem that requires detailed understanding of what each reported sensor value actually means, and therefore what assumptions can and should be made to fill in values to complete the dataset in a manner that makes it useful for AI/ML models. This highlights an important aspect when tackling data sparsity issues: when there is missing data, it is critical to understand why the data is missing. This will then enable you to build accurate assumptions for filling in that data and whether doing so is valid. In addition, the “missingness” of data can even be used to learn patterns useful for science; for example, the frequency and location of missing information due to clouds could be used to infer information about weather patterns in the region.

Solutions to Data Sparsity in NASA Science
One solution to data sparsity when applying AI/ML models to scientific research problems involves the incorporation of detailed scientific models that encapsulate our current understanding of the underlying physics of the phenomenon being studied. The approach is to create a suite of physical models and then use them as an input into the AI/ML model, which is then trained on the delta between the physical model and the data. This delta can have a much simpler structure to learn and is therefore capable of being learned by lower-order AI/ML models on sparse datasets. This approach of learning the delta over the physical model essentially removes the burden from the AI/ML model to learn a potentially complex set of physical laws that we already understand, which would otherwise require large AI/ML models and large input datasets to re-learn this in a data-driven way.

SMD AI WORKSHOP
12-14 MAY 2021

However, there are caveats to the physical model approach. One is that this can limit real-time applications since running complex physical simulations for training AI/ML models on deltas requires significant time and compute investments, which may not be available in real-time, rapid-response situations such as natural disasters. A way around this is to train another AI/ML model to approximate the physical model; in this case, a large grid of the physical models just needs to be created once, and then the outputs for new samples can be quickly approximated by the trained AI/ML model in real time. However, an important caveat is that this approach requires that the physical model be an accurate representation of reality: you must have a strong physical understanding of the problem and a clear understanding of how any data sparsity correlates with the signal that you are trying to find. Indeed, the exploration work that NASA performs often implies a lack of clear physical understanding of the system, as NASA strives to explore new things for the exact reason that we do not yet understand them. However, science-driven tests that look for compliance with basic conservation laws and symmetries based on known physics can provide robust checks, while AI/ML approaches with GANs and autoencoders can also force out any noticeable differences between the models and actual data. In any case, when physics is included in an AI/ML approach, it is important to consider and understand what the physical models don’t know.

Another approach to the data sparsity problem is to employ simple models (e.g. random forests) to complex (e.g. high dimensional) data, rather than higher-order models (e.g. neural networks) to large/dense but simply labeled (e.g. one-dimensionally categorized) datasets. In healthcare, rare disease models are an example of this, where the sample is small but the datasets are rich due to focused experts collecting highly complex datasets on their human subjects. This has applications for astronaut studies where the numbers are small but the datasets are rich with complex situations of constantly changing physiology responding to a constantly changing environment. The caveat here is to avoid extrapolating from the small ends and to create not just predictions but also uncertainties so that you know the bounds of your predictions.

In general, data augmentation is not recommended for sparse datasets, as the effect is for the model to learn the details of your augmentation rather than to reduce the sparsity of your dataset.

Implications for Data Sparsity in NASA Science Research
Looking forward in the area of data sparsity, an exciting topic is the sparsity of explanation: eliminating as many features as possible (e.g. with ablation studies) to reduce the required number of feature types needed to train a reliable AI/ML model and therefore make it more generalizable. The biological sciences would benefit greatly from this, as expensive in-person measurements can limit the usefulness of AI/ML models in different scenarios, and finding surrogates for such measurements or eliminating redundant variables would increase the impact of the model. Nevertheless, in any AI/ML application, reducing the number of features is key to a more efficient and generalizable model.

Data sparsity requires us to think particularly deeply about the problem we are trying to address and how it relates to our dataset. How does the data sparsity correlate with the signal that you are trying to find? How does this answer impact the method you chose to address the data sparsity? What uncertainty are you introducing into the system and will you still be able to answer your original science question given those effects? When confronted with data sparsity, the focus should be on the science question you are asking rather than the predictive power of your model.

DATA HETEROGENEITY

Challenges of Data Heterogeneity in NASA Science
Data heterogeneity comes in many different forms. A simple example would be the different formats of data collected from various sensors that need to be put on a common time cadence before usage in an AI/ML model. But it can be much more complicated than this. For example, when tracking astronaut health, one must ask whether astronauts have filled out their food frequency questionnaires with the same level of detail each time and if not, why? The earlier example of spacecraft telemetry was not only a data sparsity problem, but also a data heterogeneity problem: the different reporting triggers and error handling need to be harmonized in order to translate the dataset into a cohesive package where each report means the same thing.

Labeling is another source of data heterogeneity. Different people will label the same dataset differently due to their own understanding, assumptions, and motivations. This concept of heterogeneity is often tackled by having multiple people provide labels for a given example and then taking the average or mode of the labels and assuming this gives a more pure label. However, this is not always the case, as the concept may be multimodal, rendering a simple average or mode an invalid concept itself.

Additionally, specific expertise may be required to identify all the dimensions of heterogeneity. Datasets may appear homogeneous to non-experts (or even to experts), but in reality they require additional processing to truly harmonize them. This problem arises in Earth observing systems where measurements of “temperature” of the same place and time, but collected by different spacecraft and/or instruments, can be significantly different due to different calibrations or data recording formats. This is particularly acute when attempting to marry NASA satellite data with commercial satellite data, especially from commercial satellite constellations, as each satellite in the constellation may not be inter-calibrated. This problem can quickly become overwhelming, especially if you want to do things like assign confidence intervals to harmonization and ensure reproducibility.

3.2 Implications for Data Heterogeneity in NASA Science
Standardizing labels and associated metadata can be particularly helpful for addressing data heterogeneity. There are some best practices being established in different fields for standard labels (e.g. the Cancer Genome Atlas or JECAM for crops), but it can be very challenging in science research since more details need to be included than simple labels in order to do science; in many fields, such standards are not yet established and going back and re-labeling datasets after the fact can be prohibitively time consuming for researchers. One approach to address the complexity problem is to establish encapsulating classes that are separate from the more detailed scientific classification in order to simplify things enough for efficient labeling and input into an AI/ML model.

Another approach is to use AI/ML to harmonize the data by using it to create a transformation that maps data from different spaces together (e.g. harmonizing observations from the Landsat-8 and Sentinel-2 Earth observation satellites). The problem is that this needs to be done for each pair or set of datasets and for cases where satellite constellations are not inter-calibrated; this could require constructing AI/ML models and maps for each pair or group of satellites. Additionally, AI/ML solutions should not always be viewed as the final answer and rather can be seen as guides for decision making by a human expert; the AI/ML model can give us the information we need to make the decision, rather than make the decision for us.

One key thing to keep in mind is recognizing when data harmonization is not possible. Sometimes, the data really should belong to two different projects, and knowing when to stop attempting to harmonize a heterogeneous dataset is important.

Future of Data Heterogeneity in NASA Science Research
As commercial companies and other space faring nations become increasingly important players in space science research of interest to NASA, harmonization of datasets will extend to those of different space agencies as well as commercial partners. If these data harmonization challenges can be overcome, then combining such differing datasets could also help to address data sparsity issues by using complementary datasets to fill in the gaps found in individual datasets. Overcoming these challenges will likely require the standardization of labels that can be used across datasets; while this has worked on smaller scales (e.g. between two specific datasets), scalability across multiple datasets from different space agencies and commercial partners will require careful discussions and clear agreement on the derivation and meaning of each label.

CONCLUSION

When dealing with data sparsity and heterogeneity issues, science researchers need to think deeply about the appropriateness and choice of the AI/ML model and the impacts of the methods they are using to address the sparsity or heterogeneity issues. This will depend on the problem being asked, resulting in a need to focus more on the scientific problem or question being addressed and less on the predictive power of the model. This memo presented common problems that scientists working on research of interest to NASA are facing, their tried and true solutions, and ideas for the future.

FOCUS AREA 3

UNCERTAINTY AND BIAS

Bias in training data is a pernicious impediment in creating trustworthy AI workflows, unique to supervised learning techniques. How might we better learn how to plan for it and other sources of uncertainty, quantify, and exploit it?

KEY CONCEPTS

As the complexity of our requirements grow, the science community is increasingly turning to machine learning (ML) methods to crack difficult scientific problems and manage scientific equipment. Analyzing the biases of our datasets and quantifying the uncertainty of data and models is needed to obtain meaningful scientific results. The data is said to be biased if there is a systematic difference between the actual phenomenon and the model output, while uncertainty quantification is the total variance between predictions and reality, including instrument error, sample bias, computational errors, poor convergence of the training process, etc.

WHY DOES IT MATTER?

Quantifying the uncertainty of our predictions and the biases of our datasets is critical in determining the validity and credibility of scientific results. But uncertainty can also be an important tool in identifying gaps in our understanding (or data sources) informing strategic decisions, such as instrument or mission design and improving models and workflows against benchmarks. Decisions involving large investments or management of resources require a clear understanding of the uncertainties involved. Uncertainty quantification is thus key to informed decision making and risk evaluation.

IMPLICATIONS

Should managing uncertainty and bias be a strategic capability?

Developing and standardizing a set of methods and guidelines for quantifying uncertainty may significantly increase the confidence in the validity of the research being developed across NASA's SMD. Additionally, will confidence in ML enable proper use and for science to advance faster?

WHAT CONTRIBUTORS SAY

"I've seen presentations where they talk about 85% confidence in their forecast or their prediction but that's really compared to the sample data that they have. They don't talk about the inherent biases in the sample population that they're evaluating in order to create training data. If they pick the wrong set of training data, they can get what looks like a very highly accurate reproduction of the model forecasts compared to in the validation process of the machine learning workflow, but it's not really 85% confidence that they're accurately predicting."

"If you can't understand the uncertainty you won't be able to accurately understand the results."

"Uncertainty can be a source of insight."

TECHNICAL MEMORANDUM: FOCUS AREA 3

UNCERTAINTY AND BIAS

Authors: V. Ashley Villar, Columbia University
Michael Little, NASA Goddard Spaceflight Center

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

The analysis of uncertainty and bias is growing in importance to the scientific machine learning community. Many of the ideas and concepts are derived from experimental physics or statistics but require unique characterization to have the right impact on Earth and space sciences in NASA's Science Mission Directorate (SMD). In the current state of these scientific discipline areas, analysis of uncertainty and bias have uneven application. Some papers discuss them carefully and others not at all. Even the vocabulary varies across the domains and the observational and modeling data. Consistent multi-divisional policy is elusive, but the discussion would yield far more insight into the nature of those differences.

Uncertainty and bias considerations are important both from a scientific point of view but also to the credibility of the work and its use in applied science. The value of measurements and the need for improvements in those measurements can be derived from the evolution of the uncertainty and bias in them. How those measurements, models, and analytic results can be used for the next cycle of investigations is directly determined by the stated uncertainty and bias in them. Three things make the use of machine learning in scientific investigations untrustworthy. First, the results of modern scientific investigations challenge conventional wisdom in many areas and make the public and management skeptical about it's validity; explainability is important to credibility. Second, the mathematical underpinnings of machine learning are complicated and difficult to understand without a long and specialized academic background which makes it difficult to access by Earth and space scientists; finding people who can work in both camps and are contributing members of domain science teams is challenging. Third, recent publicity in two areas has created additional credibility issues. The recent publicity for misuse of machine learning in political and sociological activities and the hype of the scientific press has created considerable distrust of the results. A healthy scepticism is important for well-founded science, but an objective resolution of this tension must be based on experimentally derived facts about reality, not speculation and bluster.

The participants in the workshop identified three actions for SMD as multi-divisional advances:

- additional training for NASA SMD technologists, managers, engineers, and scientists;
- specific requirements in research solicitations for a plan to treat uncertainty and bias and specifically, in the selection of labeled training data, followed up during periodic reviews; and
- inclusion of adjunct processes in the emergence of scientific MLOps to illuminate uncertainty and bias.

DIGGING DEEPER

Purpose and Goals
The purpose of the uncertainty and bias session was to identify key issues in characterizing uncertainty and bias in the use of artificial intelligence in scientific investigations. For the purposes of this session, uncertainty is defined to be the quantification of the spread in observed and/or measured properties. This includes aleatoric uncertainty (statistical uncertainty, which can be reduced through more observations) and epistemic uncertainty (systematic uncertainty). Bias is defined to be the systematic difference between the actual value and the value determined by some estimator. Specifically, strategies to deal with bias and uncertainty fall under the large umbrella of uncertainty quantification (UQ).

UQ is increasingly an important aspect of AI/ML methodologies. Given the intrinsic degenerate nature of neural networks, epistemic uncertainty can be challenging to quantify in a Bayesian sense (i.e. one cannot marginalize over the many neuron weights). Bias becomes particularly challenging to quantify outside the bounds of the training set used to train these models. UQ is particularly important in the physical sciences, in which repeated observations and stochastic systems naturally result in probabilistic measurements. Combining a Bayesian framework with ML methods is a rapidly growing field. This session focused on the needs of the NASA community for UQ, the state of the art in UQ for ML, and the future innovations necessary to fully incorporate ML approaches into traditional physics-based pipelines.

This session was led off by a talk by Dr. Brain Nord of Fermilab, whose presentation is available. Dr. Nord suggested the discussions consider the following provocations:

1. AI is untrustworthy: UQ (including bias) is the most important challenge for AI applications;

2. we lack, but need, a unified approach to solving this problem across AI developers and practitioners; and

3. language and jargon are not unified across stakeholder groups, and communication across practicing communities is a key bottleneck for development.

The subsequent discussions were divided up into three phases, each primed with questions for discussion, as described in Table FA3-1.

TABLE FA3-1
QUESTIONS FOR DISCUSSION IN UNCERTAINTY AND BIAS

Understanding the problem	• How does uncertainty in data and models impact your work? • How can we design experiments that capture bias in data collection and analysis?
Understanding the problem	• What would a good uncertainty plan look like? • How could we determine how much the training data impacts our results?
Understanding the problem	• What could be the guidelines to help identify sources of potential error? • What do you find difficult when trying to include uncertainty quantification and bias into your research plan?
Suggestions to improve things	• How could the process of uncertainty quantification be standardized? • Do you have a concrete example of how uncertainty quantification would impact the quality of your research/application?
Suggestions to improve things	• Would uncertainty estimation help the reproducibility of the results and how? • Could the adaptation of the uncertainty quantification methods for specific use cases be reduced and how?
Suggestions to improve things	• How can we communicate uncertainty and biases to end users for decision-making processes? • What do you think is needed for uncertainty quantification to become common practice?
Imagining a future world	• What initiatives could help scientists from different divisions to understand uncertainty and how to work with it? • How could we involve domain experts on the discussion and on different projects where uncertainty quantification is needed?
Imagining a future world	• Should there be a mandated uncertainty plan for proposals and why? • Could we develop a methodology for uncertainty quantification that could be valid for different research domains? Can we think of what it would take?
Imagining a future world	• What concrete points of action can we take to increase the use of uncertainty quantification methods across the different divisions for research projects? • Should there be an effort to create a set of tools and resources for uncertainty quantification?

UNDERSTANDING THE PROBLEM

Origins of Uncertainty and Bias
In designing an experiment to create insight about a natural phenomenon or physical process, the difference between the physical state and the reported value from the experiment represents the sum total of error. There are many contributors to this error; some are stochastic, and some are systematic. It is essential for the scientist to identify and aggressively eliminate or properly compensate for these contributors in order to get the best possible explanation of the phenomenon. In some cases, this is an iterative process alternating modeling and analysis with making new observations. In some cases, the latter step also involves improving the instruments used to make the observations in ways illuminated by the modeling and analysis. Where machine learning is used as an analytic or modeling device, this is much more than simply verifying the model's ability to predict the data in the validation subset. While an important step, the uncertainty and bias must ultimately be against the actual phenomenon.

The origin of uncertainties and biases in scientific observations can be broken into four broad sources:
- experiment design;
- instrument and test technique;
- science data processing; and
- science data analysis and modeling.

An understanding of each is critical to using the scientific data created by the observation. A thorough analysis of the sources of uncertainty is necessary because, to date, estimates are often too low.

Current Practice
The discipline of uncertainty quantification in the Earth and space sciences has uneven application to modeling and observations. In Earth science, in particular, there are widely varying definitions of terms, and typical characterizations are of only one aspect of the total picture. For example, some remote sensing datasets only go so far as to identify missing measurements in the datasets. Others qualify the measurement with qualitative adjectives (e.g. excellent, poor, etc.). These practices help the instrument team characterize and validate instrument performance but compound the problem of characterizing the uncertainty and bias in machine learning workflows. workflows in which quantification of error is essential. Even if uncertainty is quantified, the standard drastically varies across subdisciplines. In some cases, a Bayesian full posterior may have been estimated. In others, a Gaussian is always assumed with the mean and variance always reported. UQ is recognized as an important issue, but the lack of well-known and trusted techniques makes it still an area of research and much further discussion.

Furthermore, there is a language barrier across subfields. This is especially true when directly applying new methodologies from data sciences. Mathematical and AI-specific jargon prevents space scientists from readily adapting new methods for their own purposes (e.g. the common Lagrange multiplier used in physics is replaced with "regularization" in AI contexts). Some discussions of uncertainty convey a negative connotation, implying wrongdoing, whereas uncertainty quantification often is a mechanism for identifying the path to a balance between resources available (time, funding, and staff) and precision needed to adequately test a hypothesis.

As we try to encapsulate our understanding of natural phenomena and physical processes in the physics-based and data-driven models, validation of the output compared to observational data becomes a measure of how good the models are. Uncertainty and bias becomes a key element of diagnosing the differences. In fact, some researchers indicate that the investigation of key drivers in physical model errors is enabled by machine learning techniques in comparing the two.

The consideration of uncertainty and bias plays a key role in experiment design to incorporate appropriate techniques and processes. The choice of instruments are often dictated by the acceptable uncertainty in understanding the phenomenon. For example, some radiometers have high accuracy (~1% error) but are very expensive to construct and to maintain. Others have higher error bars but are much more affordable. Climate modeling requires high accuracy, whereas weather forecasting is less stringent. If an understanding of the phenomenon does not require higher accuracy, the experiment design would select instruments with higher UQ to invest in other aspects of the campaign.

The deliberate identification and elimination of systematic bias in experiment design extends into the data analysis phase. First, the collection of the right observations is critical. Experiments that fail to collect a sufficiently broad set of data create a bias that misses some behavior. Experiments that collect too much data in one regime and not enough in another similarly create a skew in the results that limits the value of the conclusions. Confirmation bias is a common experience in the selection of data that tends to support a poorly formed hypothesis instead of disproving it.

This also pertains to the selection of sample data for use in training models. Because of the shortage of well constructed labeled training data sets, practitioners are sometimes desperate enough to use available sets without adequately reviewing for how well they represent the observations under consideration. Techniques for labeling sample data must consider their impact on UQ. The use of Bayesian distributions from multiple annotators instead of a single label reduces the UQ; often accurate labelling has a more significant impact than dataset size in ML-aided tasks. This is well documented in Gebru et al. in the context of a language model but is equally as deceptive or misleading in scientific observations.

Similarly, the re-use of data in analysis for which its collection process was not designed, requires careful consideration of the uncertainties and biases to ensure the use is appropriate. Without uncertainty and bias characterization, it is easy to use data inappropriately and draw fallacious conclusions.

Why is Consideration of Uncertainty and Bias Important?
There are two primary implications of the evaluation of uncertainty and bias. First, in conducting a campaign to understand the nature of the phenomenon, analysis of the progress in reducing uncertainty and bias gives an indication of how much further it must progress before the topic has been adequately characterized. This depends upon what we define as reliability: is it our ability to classify everything correctly, or is it our ability to accurately predict the rate at which we classify a set correctly? Those are two different measures. Too many predictions without UQ are not scientifically valid. Take for example detection of trends in a time series; if the model doesn't have an uncertainty range, there is no way to interpret the trend result.

Second, the credibility of the scientific results are directly affected by the uncertainty and bias associated with the results. In explaining results to both scientists and laypeople, if we have a clear estimate and explanation for the uncertainty and bias, it strengthens the credibility of the model and the results/ forecasts. However, as previously mentioned, UQ is often either adhoc or overly qualitativer, or qualitative yet difficult to understand without proper training of the scientific community, because the uncertainty and bias are poorly addressed or not at all. The contribution of questionable results makes little progress towards the overall understanding and are often not trusted by other competent scientists. How error, uncertainty, and bias propagate through deep learning models is unclear when you can't interpret the intermediate stages and methods to develop interpretable and statistically rigorous ML models are ongoing. The former is particularly important when results are counter-intuitive.

SUGGESTIONS TO IMPROVE

Communication and creation of a standard practice are consistent themes in suggestions for improvement. We highlight here specific suggestions to (1) create a common language and simple examples; and (2) set similar expectations of UQ reporting.

An Uncertainty and Bias Plan for experiments would improve the quality of the experiment. It should define uncertainty and bias in the context of the experiment or campaign. It should also include a thorough analysis of sources and a way to compare the results to ground truth. Considerations of explainability and physical-based knowledge and conservation laws need to be factored into the models. It should also address compensation for sparse or mixed data. In physics, the conventional wisdom is that more data yields a lower statistical uncertainty, yet systematic error may always dominate in AI/ML studies., Other considerations include:

- if data augmentation can be used to mitigate observational biases, and if there exists prior (physical) knowledge which can be used to generate realistic simulated data;
- the prevalence of epistemic uncertainty in modern supervised learning, and how this uncertainty can be quantified by having a strong human feedback loop over real-time predictions, and periodically re-iterating on the model; and
- if we can say, in the design of an experiment, what the minimum amount of data is to produce a reasonable uncertainty estimate (the size of the training data is not as important as making sure it is representative of the underlying phenomena being classified or detected).

UQ is a developing field in ML. There are currently, loosely, three ways of probing UQ in neural networks in particular.

1. **Probing the model as a blackbox**—the more common method. Neural networks are largely treated as a black box built into a scientific model. Uncertainty is probed via ad hoc methods appropriate for the problem (e.g. Monte Carlo resampling the inputs based on measured errors). Ablation studies (removing components of the neural network) is another common tactic. Cross validation of samples is one of the most commonly used techniques to test typical uncertainty of the model.

2. **Directly modelling the probabilistic nature of the system through inclusion of Bayesian inference—**a developing solution. The number of models of this nature is rapidly expanding (e.g. Bayesian neural networks, variational autoencoders). However, there is a significant barrier to non-statisticians to use these models out of the box. Development of deep probabilistic language (e.g. pyro, mc-stan) are providing tools to construct such models, but scientists lack toy examples and tutorials to explain these tools.

3. **Building directly interpretable models**—a state-of-the-art solution. This is the most valuable solution, but the least mature approach. Considerable research is ongoing to develop fully interpretable and Bayesian frameworks but needs expansion to include techniques for assessing and interpreting the uncertainty and bias in the models.

IMPLICATIONS

Further research into the theory and application of techniques for assessing uncertainty and bias is needed. This work would be applicable in all five NASA SMD divisions and would be stronger and better established for it. There was considerable debate as to whether tools or techniques are appropriate. The tool development option makes it easier to introduce the UQ considerations but tends to result in application without understanding. The technique approach is harder to re-use but requires the user to understand what the results mean.

Further staff education in how to apply these techniques to specific experiments and observational campaigns is valuable. Familiarity also can lead to better refereed journal articles and also proposal development and evaluations.

Benchmarks for UQ would be valuable, both in guiding the design and execution of experiments but also in assessing the scientific validity and value of proposals and papers.

Several other techniques to be considered:

- sensitivity analysis which starts with one end of the error/uncertainty of input features and repeats at intervals through the other end, pass each through the model for inference to see how it impacts the output; basically, we sampled from a distribution of inputs to simulate the distribution of output;
- use of validation/quantification techniques such as K-fold cross validation can better quantify the uncertainties in the model and training;
- augmenting datasets to be more robust to noise. In the case of images, for example, one can add white noise and shift/rotate the images;
- visualizing what a model is learning over a dataset or for representative examples (e.g. feature importances, SHAP values, class activation maps, etc.) can help give confidence or identify pitfalls of a model;
- outlining and anticipating worse case scenarios (with either bogus data, or model parameters);
- provide visualizations/heuristics/your testing data together with the model, so users can replicate your validation analysis with their data, and compare;
- defining standard tests (unit, integration, end-to-end) to monitor uncertainty especially in continuous integration settings;
- collaborate across a diverse team from a variety of disciplines and backgrounds to identify sources of bias;
- methods that evaluate the robustness of the model's predictions under the adversarial perturbation can quantify/detect the systematic error of the model or bias;
- make clear the social-curator-post-processing biases and uncertainties, verses the biases and uncertainties within the data itself (how it was collected, limitations, confounds); and
- Feature importance can be used to detect/identify source of bias/error, and automatic approaches such as layer-wise relevance propagation can be used.

SMD AI WORKSHOP
12-14 MAY 2021

Imagining a Future World

In future scientific investigations, uncertainty and bias play an important role in defining the scientific objectives, defining the experiments and campaigns, and in assessing the roadmap for advancing the knowledge of the phenomena. They are included in solicitations to encourage appropriate consideration in award selection and to expand the number of people thinking about this problem. The way to include this without derailing current scientific thinking was the subject of considerable debate.

Analysis of uncertainty and bias are commonly accepted practices. While addition of the requirement to include analysis and error bars in proposals and published papers may be difficult, reviews would accelerate this process by discussing the presence or absence and could increase the visibility of this important aspect of the science.

Review of both proposals and refereed journal articles and conference papers includes assessment of how well they treated uncertainty and bias. They explain the relevant sources and how they compensated or mitigated them and provide an estimate of the residual uncertainty and bias that cannot be affordably remedied.

Development of re-usable training data for supervised learning is enabled through competitive campaigns which encourage demonstration of use in models. The metadata includes an analysis of the uncertainty and bias and the boundary conditions inherent in the data for re-usability. This data is available in open-science repositories and has a pedigree of review for use, including what is outside the boundary conditions.

Tools for analyzing and improving the uncertainty and bias have been validated in multiple domains, vouching for their utility and stability and validity. This may take the form of a library or wiki describing the techniques. ML workflows have been operationalized and validated along with DevOps environments that enable more efficient and effective application.

A noteworthy trend in techniques is the use of MLOps for DevOps for ML, that is standard development practices for how models are constructed, tested, and validated.

KEY THEMES IN UNCERTAINTY AND BIAS

Three key themes recur in discussions of uncertainty and bias:

- sources of uncertainty and bias;
- techniques for identifying, assessing and compensating for uncertainty and bias; and
- means for increasing the scrutiny of uncertainty and bias in the review of proposals and the publication of research.

KEY CHALLENGES IN UNCERTAINTY AND BIAS

NASA SMD needs a consistent model for addressing uncertainty and bias in observations. The lack of a common vocabulary is a major obstacle to multi-domain conversations and re-use. This vocabulary also should avoid allowing the reader to infer negative connotations about uncertainty and bias but rather view it as an important aspect of understanding the data and its limitations.

NASA SMD scientists and engineers need a consistent training experience to perform analysis of uncertainty and bias in observational data and model output. This should include a survey of the various sources of uncertainty and bias and techniques for probing and diagnosing them. Some considerations include:

- the size of data sets and the cost of computation;
- how to deal with non-linear or chaotic situations;
- indentifying the unknown unknowns;
- understanding when bias matters; and
- the difference between addressing bias that is known (gender representation in astronaut health data) and bias that isn't known to the researchers.

The lack of standard and well understood techniques for assessing uncertainty and bias in scientific investigations results in their absence in journals and presentations about the results. Considerable work has already been accomplished, but is not collected in a way that can aid the NASA SMD research and engineering communities.

SMD AI WORKSHOP
12-14 MAY 2021

SUGGESTIONS TO IMPROVE THE FUTURE OF UQ FOR NASA APPLICATIONS

Here we present seven suggestions.

1. **Set a common language for UQ, with simple examples that are directly applicable to space studies.** Currently, ad hoc methods dominate UQ methods, with new techniques developed for each specific application. While this is superior to no UQ, it creates a challenge of the interpretability of UQ and the ability to combine meaningfully studies. A common set of techniques and examples of their applications would be beneficial. Example techniques of interest may include: cross-validation studies, ablation studies, and Bayesian inference methodologies (e.g. with deep probabilistic modelling programs such as mc-stan, pyro, etc). Importantly, incorporating truly Bayesian and/or interpretable methods, which are being actively developed, require a set of simple tutorials for NASA users.

2. **Set similar expectations of how UQ is reported and interpreted in studies.** In this case, a simple solution is to encourage the reporting of UQ and a clear statement of how it is generated and meant to be interpreted. For example, drop out can be used to measure epistemic uncertainty in neural networks but will not account for aleatoric uncertainty. Often there is a mismatch of assumptions between statisticians, NASA scientists, and ML practitioners. Being explicit about limited or ad hoc UQ is especially important for broader interpretation.

3. **Improve NASA researcher awareness and training about uncertainty and bias.** This should include design of experiments, including handling uncertainty and bias. It should provide a framework for evaluating sources of uncertainty and bias, as well as benchmarks. A very effective delivery technique would involve a workshop with sample problems that could solidify the attendees' understanding.

4. **Develop techniques for evaluation of UQ.** This would include a protocol for assessing various sources of uncertainty and bias. Considerable research in this area is being conducted at other federal and ESA research institutions and should be coordinated with DoE, NSF and NIH.

5. **Include specific requirements for consideration of uncertainty and bias in competed research.** Solicitations should require a plan for assessing UQ. Proposal evaluations should specifically address the degree to which UQ is correctly handled.

6. **Journal guidelines for the inclusion of UQ or the creation of UQ benchmarks would aid in the problem of visibility.** Coordinate with publications to include a requirement to address UQ in all papers. This includes requiring referees of papers to assess the robustness of UQ treatment and make deficiencies a required corrective action.

7. **We note that NASA SMD scientists have the opportunity to be leaders in the machine learning UQ field by defining the benchmarks necessary in machine learning algorithms to be useful for NASA-related science.**

CONCLUSION AND NEXT STEPS

Uncertainty and bias in artificial intelligence is an important component to robust AI/ML research and applications. It requires an effective framework and scientifically robust techniques for describing and estimating it. No scientific discovery is fully useful without analysis of the uncertainty and bias and a description of how it propagates through the calculations to blur the results and impact the conclusions. While instrumentation and test techniques have long illuminated the effect on measurements, the emergence of machine learning has fallen behind in characterizing how uncertainty and bias affect such models and analyses.

It is suggested that NASA SMD begin a short research program to provide guides and aids for identifying and characterizing uncertainty and bias in machine learning workflows and use the results to educate the NASA SMD community in using them in their work. Collaboration with NSF-funded work in this area would play a major role in such a research program. Brief educational webinars and training can disseminate summaries of this work to the community at large.

SMD AI WORKSHOP
12-14 MAY 2021

REFERENCES

- Kuleshov, V., Fenner, N. & Ermon, S. (2018). Accurate Uncertainties for Deep Learning Using Calibrated Regression. **Proceedings of the 35th International Conference on Machine Learning**, in **Proceedings of Machine Learning Research** 80:2796-2804 Available from http://proceedings.mlr.press/v80/kuleshov18a.html.

- Srinivasan, Ramya and Chander, Ajay. (2021). **Biases in AI Systems: A Survey for Practitioners**. Association for Computing Machinery Queue. DOI 10.1145/3466132.3466134, Queue, vol. 19, no. 2, Apr. 2021, pp. 45–64., doi:10.1145/3466132.3466134.

- **On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?** EM Bender, T Gebru, A McMillan-Major, S Shmitchell - Proceedings of the 2021 ACM Conference on Fairness, 2021

- McGovern, Amy (University of Oklahoma), Imme Ebert-Uphoff (Colorado State University), Philippe Tissot (Texas A&M at Corpus Christi), Ruoying He (North Carolina State University), Christopher Thorncroft (University at Albany), **Strategic and Implementation Plan for the NSF AI Institute for Research on Trustworthy AI in Weather, Climate and Coastal Oceanography** (AI2ES), 2020. Available from https://drive.google.com/file/d/1AzEidwQXK9-o24pUOTbx9QyK0m16bgye/view

- https://www.nextplatform.com/2018/04/16/gpus-mine-astronomical-datasets-for-golden-insight-nuggets/

- http://edwardlib.org/

- http://pyro.ai/

- https://github.com/pyprob/pyprob

- https://mc-stan.org/

- https://astrobites.org/2021/01/08/meet-the-aas-keynote-speakers-dr-brian-nord/

FOCUS AREA 4

REPRODUCIBILITY

Reproducibility of AI experiments is 'more than just open source code'. What does good reproducibility look like?

KEY CONCEPTS

ML-assisted research is said to be reproducible when the presented results can be reproduced to a similar level of accuracy by a new team with independent research artifacts. Replicability refers to accessing the original data and code and these being made openly available for others to repeat the footsteps of the original researchers. Best practice means releasing well-documented, annotated, and fully open code and data that affords others to independently reproduce the results of a study and extend them (perhaps to a higher TRL).

WHY DOES IT MATTER?

When a ML result is a black box, any results that are not replicable or reproducible can invalidate the results (particularly exciting claims) and casts doubt on ML methods. Clear reproducibility of scientific studies improves the credibility and reliability of the results for integration into ongoing research and is the key for open science, as it allows researchers to set baselines and safely stand on the shoulders of published methods.

IMPLICATIONS

Can we challenge the way we do and share science?

Code semantics are not standardized and usually specific to the author. For this reason, even code that is publicly shared can be hard to understand and reproduce. Standardizing good practices for sharing and making our workflows reproducible will allow other researchers to build on the work, saving time and resources and the temptations of 'not invented here'. The current culture of science is based on publishing and results, but without investment in reproducibility; it doesn't necessarily improve the scientific output and knowledge as a group.

WHAT CONTRIBUTORS SAY

"It is harder to release research code than to publish a paper, this process takes about a year. Code continues to change and by the time it is able to be released it is already out of date."

"One thing that can happen in science is due to low funding levels, you don't always have the size and the team that you want in order to polish things off as much as you want to. I've certainly experienced projects where there's just nobody around to get something polished, and deployable. And so it continues to be a science hack."

"Reproducibility has many facets: sometimes the key issue is data, other times it's code, other times it's all of the above."

TECHNICAL MEMORANDUM: FOCUS AREA 4

REPRODUCIBILITY

Authors: Milad Memarzadeh, NASA Ames Research Center (USRA), Moffett Field, CA, Steven M. Crawford, NASA Headquarters, Washington, DC

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

Reproducibility is a core tenet of the scientific process. Given sufficient information and resources, the results of an experiment should be able to be reproduced. Scientists need to provide sufficient information about how they produced their results in order for them to be reproducible. As new technologies are adopted into the scientific process, the requirements for reproducibility need to be defined for them. Unfortunately, there are a number of reasons that might limit the reproducibility of scientific results derived using machine learning (ML). In this note, we explore the challenges to reproducing scientific knowledge generated through machine learning and solutions to enable further reproducibility of the results.

As Dr. Christine Custis noted during her talk, reproducibility also has different meanings and requirements for different groups. These can include internal accountability, cross divisional transparency, and external reproducibility. The requirements for a model to be reproducible internally to a team can be different for those in a different part of an organization or to groups external to the organization. Further, the needs for a model in a production environment versus an early prototype can also be very different. While these groups have different needs, improved documentation of the entire process can significantly improve the overall reproduction of a model.

In this technical memo, we present the challenges for reproducibility of machine learning along with some potential solutions. The ideas presented in this note are a summary of the discussions held at the 2021 NASA SMD AI Workshop.

For the purposes of this note, we have two definitions to help differentiate between common practices: ML-assisted research is said to be reproducible when the presented results can be reproduced to a similar level of accuracy by a new team with independent research artifacts. Replicability refers to accessing the original data and code and these being made openly available for others to repeat the footsteps of the original researchers.

DIGGING DEEPER

Defining Reproducibility
One of the first challenges is in defining and understanding what is needed for reproducibility.

Studies can often be replicable with all of their software and data available. However, does this mean the study is reproducible? While an independent team may be able to recreate the results using the same software and data, this does not mean that the results may be reproduced given a different data set or software package. For example, the data set used might be unique and extremely large or potentially not well documented. The software might only be available as a black box: an executable without clear information on the underlying algorithm.

There is also a distinction between open-sourcing the code and reproducibility. Just sharing the code with the community does not guarantee reproducibility, and researchers need to go beyond that and provide specific examples and documentations (such as Jupyter/Ipython notebooks) to facilitate the replication and repetition of the experiments.

Another aspect that is a roadblock to reproducibility is difficulty in standardization and translation of information to a greater community, especially in domains such as the medical field where privacy is a big factor.

In many fields, not everything is deterministic. In applications that are dealing with stochastic processes, the reproducibility might mean extra steps to take into account necessary steps for uncertainty quantification.

To help address these concerns, further guidance can be given on what reproducibility means especially when working in ML. Along with the definition of reproducibility, this should include:

1. at a minimum, the software and data should be accessible to assure replicability;
2. to help ensure reproducibility, the entire machine learning lifecycle should be sufficiently documented including obtaining and cleaning data, details of the algorithms used, training and testing of the models, and usage of the models; and
3. best practices for handling stochastic processes in both data sets and models.

SMD AI WORKSHOP
12-14 MAY 2021

Supporting Reproducibility
While reproducibility is a key tenet of scientific research, incentives are rarely aligned with this goal. For example, the hard constraints and requirements for making scientific work and research reproducible and sharing the code among the research community can be a discouragement, especially when the research code is not appreciated as much as publications in some scientific domains. For some, the bar for sharing the lowest-level research code can be very high, to the extent that researchers might find it cumbersome to open-source their code. Moreover, the process of open-sourcing the research code and providing reproducible examples can take a long time (over a year), which by that time, the code has gone under multiple revisions and it is already out-of-date.

Another underappreciated aspect of reproducibility is the extent of documentation that is necessary. Documentation is an often neglected step of the development process with little support for writing or developing it. Outside of papers, researchers are not rewarded or incentivized to write documentation. Services that make documentation easier throughout the entire ML lifecycle such as MLOps or MLflow should be encouraged.

One of the most important aspects of reproducible work/research that is often ignored is the maintenance phase. Sometimes the duration of the maintenance process might be much longer than the project itself and specially the duration might be different for NASA SMD data and code, compared to industry/ commercial products.

As such, the following practices would help increase the support for reproducibility in machine learning:

1. there should be specific allocated funding for maintaining the research code and improving the reproducibility of the research for the time beyond the actual project's phase;
2. Improve the documentation provided which can include training on usage of tools that make it easier, normalizing the expectations for the type of documentation that is provided, and providing support for documentation of processes; and
3. track the usage of the code and data sets developed and published by the researchers, including an official citation of the code usage within a publication, but also how many people have re-used and built on the shared work.

Trust in Machine Learning
There is currently skepticism and lack of trust for AI/ML techniques and their deployment in scientific domains. This is due to the fact that a lot of these techniques are considered a black box model, which lacks reproducibility and explainability. A reproducible AI/ML development can improve the trust of domain scientists and practitioners in these techniques and increase their use and applicability.

Even for replication, sufficient information must be provided including the version of the software and data. Software development and machine learning are evolving rapidly, and without information on the version of the data, the study may be difficult to reproduce. In addition, data sets evolve as well with new processing techniques, updated labels, or additional cleaning. As the underlying training data set changes, the results of the model may change as well.

Using commercial platforms such as MLOps or MLflow can definitely help reproducibility by allowing a proper tracking of the development of data, code, and models. Moreover, having in-house experts in data engineering best practices and reproducibility would facilitate and improve this challenge. Their role would not only be to support the scientists in making their work reproducible but also to design and implement training or educational context for scientists to follow in their day-to-day work that improves the reproducibility of their work.

Supporting the submission of reproducible code along with data as part of the publications of scientific results into the review process may help increase the reproducibility of the manuscript. In this context, reviewers are then not only able to review the papers but also replicate the results presented in the paper. This may, though, result in the review process taking a longer time.

Moreover, journals should reward null (negative) results that preceded success as well. The culture of publishing, which mainly focuses on successful techniques and new discoveries, should shift towards rewarding null results and failures to ensure that knowledge and experiences are better shared.

Furthermore, supporting studies that reproduce the results from machine learning focused projects can help increase the trust in machine learning. With many of the techniques used are quite new, having multiple studies that are not just replicating, but reproducing results using independent analysis helps to increase the trust in the initial results produced.

As such, NASA SMD has an opportunity to help increase the trust in ML produced science by:

1. supporting publication processes including releasing the software, data, and documentation underlying the results including versioning, encouraging the publication of null results, and supporting studies that attempt to reproduce the results;
2. linking the data with the code and the models developed to make inference about the data is an absolutely critical step for reproducibility; and
3. providing environments that make it easy to reproduce the results of ML studies which can include systems for documenting the entire lifecycle and resources for sharing and running the models.

Opportunities in Reproducibility
Reproducibility brings accountability to the science result, and it enables continuity of the research outcome and products. When scientific work is reproducible, other researchers can build upon the work more easily and improve the performance of models in the future. Moreover, having the data and code being used by other researchers and further developed by them improves the reliability and quality of the code and its applications. Simplifying and supporting the process to make the data, code, and documentation as accessible as possible will help support reproducibility in machine learning and its application in the scientific domains.

The utilization of the commercial cloud services and opening the data and code to a broader community can also improve collaboration and eventually help achieve a more reproducible work. Reproducible environments can also be supported using technologies such as Docker that allow a snapshot of the software used for a program.

Similar to the data management plans, which is a requirement in the proposal stage of NASA's SMD, making the code and software open source and reproducible should be the requirement as well (i.e. a reproducibility plan). More importantly, this should be a continuous process throughout the duration of the project, the principal investigator (PI) and the researchers should be encouraged to keep this a continuous process. A common mistake is to wait until everything is developed and then spend the last few weeks on reproducibility and code-sharing.

The process of software release at NASA SMD can be difficult which limits the ability to openly share reproducible scientific research. A simple solution could be that the science/research part of the process can be separate and different from the process for code that operates launch vehicles or are mission-critical.

NASA SMD can also make best practices available to the community as guidance to help improve the reproducibility of their work. An example of this from Dr. Custis talk included these questions:
- Have you used a checklist to ensure your code is complete?
- Have you hosted your model files?
- Are you using standardized model interfaces?
- Have you made demos available?
- Are you using leaderboards?

CONCLUSIONS

A key tenet of the scientific process is reproducibility. In this note, some of the challenges to reproducibility were highlighted along with potential solutions. NASA SMD can provide further guidance on reproducibility including definitions and can provide improved policies that would support reproducibility. A key aspect to reproducibility is sharing the data, software, and documentation related to machine learning models, and this means supporting the sharing of the models through simplified processes and supporting technology. Incentivizing reproducibility can help to further the adoption of machine learning in scientific practices.

SMD AI WORKSHOP
12-14 MAY 2021

FOCUS AREA 5

CATALOGING AND SHARING AI READY DATA AND MODELS

What is best practice in effectively managing AI resources?

KEY CONCEPTS

The development of tools for machine learning (ML) (such as developing frameworks and datasets) is rapidly evolving. For this reason, publishing a paper and saving the model and the code is not enough to ensure reproducibility. Cataloging refers to the process of creating metadata representing information about the dataset. Effective cataloging includes subject matter expertise (SME), authors to contact, and other properties of the dataset or the models.

WHY DOES IT MATTER?

The way ML models are shared is not standardized, and they are scattered across the internet. The process of acquiring and preparing a new dataset is time and resource-consuming—with significant impedances for others to adapt to their research. Also, not having effectively curated and catalogued projects results in a lack of baselines for researchers. The consequence of this lack of baselines is a nest of multiple results that are unwieldy to compare and build on.

IMPLICATIONS

Will proper cataloging of models create more reliable, verifiable and reproducible results?

Creating a library of curated, cataloged research outcomes has the potential to help to mature AI/ML models for use by the NASA science community; improving the science we do and helping researchers advance at a faster pace towards a common goal.

WHAT CONTRIBUTORS SAY

“From the commercial side, we had a project where we were dealing with video, specifically and dealing with trying to label things that you saw in video. And we were doing a new software delivery. We didn’t realize that our model was written in pytorch 0.9. And it kept failing in pytorch 1.0, because it was not backwards compatible. It’s not enough to just save off a model, you’ve got to save some information about which framework it was, which version of that framework it was, because things are moving so fast.”

“Training models require good, labeled data. A lot of AI/ML researchers lack good training data, particularly as we get to more science data. Sharing good, curated sets, will help to mature AI/ML models for use by the NASA science community.”

TECHNICAL MEMORANDUM: FOCUS AREA 5

CATALOGING AND SHARING AI-READY DATA AND MODELS

Author: Christopher Lynnes, NASA/GSFC

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

NASA has a long and storied history of openly sharing its science data, the better to recoup the investment required for spaceborne data collection. Data systems in multiple divisions have been providing science data since the early 1990s, from the Planetary Data System to the High Energy Astrophysics Science Archive Research Center to the Earth Observing System Data and Information System. As the Science Mission Directorate evolves from open data to open science, the need arises to share additional aspects and artifacts of the research process. In this session's opening provocation, Dr. Stojnic emphasized the importance of these other aspects, noting as potential shareables:

- tasks (e.g. named entity recognition);
- datasets;
- models;
- code;
- results; and
- papers.

(Stojnic et al., 2020).

The focus area on **cataloging and sharing AI-ready data and models** discussed several technical aspects of the sharing process. However, as the day host Dr. Ramachandran noted, sharing science early in the process is a narrow view of open science, that is necessary but not sufficient. In their recent paper, Ramachandran et al. (2021) defined open science more broadly, as "a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding". The focus area discussions often branched out beyond the technical sharing requirements into the collaborative cultural realms, which come into play in sometimes unexpected ways.

The following section lays out five key areas where concerted action within NASA SMD could achieve significant benefits in the incorporation of artificial intelligence/machine learning (AI/ML) into the NASA SMD's scientific process, both technically-oriented and culturally-oriented. Most of the areas are related to other areas, and in some cases sequencing is important. Although the initiatives need not be addressed serially, there are cases where progress in another area may be key to making progress in the one being considered. Suggested sequencing is laid out at the end.

DIGGING DEEPER

Expand Sharing to Include Other Research-relevant Objects
Expand sharing beyond simply the input data to include other research-relevant objects such as:

1. labeled training and validation data;
2. ML models;
3. code;
4. workflows;
5. results; and
6. papers.

These should be categorized by the task being accomplished by the ML data and model. In addition, it is important to retain the relationships amongst these objects when sharing. The relationships should be both human- and machine-readable, via an API in the latter case.

Benefits. Sharing all aspects of a research project makes the results more reproducible and replicable. In addition, by beginning with working examples and modifying them, many scientists will be able to learn and understand state-of-the-art practice for ML research. Machine-readability aids in reproducibility (and possibly inclusion in ensemble-based research).

Establish Standards, Conventions, and Best Practices for Managing AI/ML-related Objects
NASA SMD or its constituent divisions should adopt (or where necessary, develop) standards, conventions and best practices for sharing the various research objects listed in the previous section. History shows that one of the most critically needed conventions is a system of unique resource identifiers. While this is straightforward for datasets and papers (digital object identifiers (DOI) are widespread for both), similar identifiers are needed for all research objects. For most of the object types, format and metadata standards will be needed. Also, standards or conventions are needed to describe the relationships among resources, especially among resources of different types. While the above standards and conventions are not specific to AI/ML but rather open science as a whole, AI/ML also has a unique area where standards should be adopted, namely in the area of model success metrics which are currently a bewildering (to non-ML experts) array of recall, precision, F1-score, and more esoteric measures. However, rather than down-selecting, it may make sense to include as many as feasible in the standard, as the use case often determines which metric is most useful.

In addition to success metrics, metrics for uncertainty and bias would also be useful. Note that the community may need to acquire some practical experience with the more novel research-object sharing before establishing a standard.

Benefits. Although standards and conventions take effort to adopt or define, their use improves the interoperability and interchangeability of components in a complex system such as the ecosystem of AI/ML components. Over time, this makes the development and adoption of innovations more feasible and cost-effective. Meanwhile, common best practices help participants in the ecosystem to become more efficient and effective.

Develop Reusable AI/ML-relevant Data Management Tools
Once standards and conventions are established, it becomes feasible to develop reusable tools for managing and working with AI/ML-related research objects, such as datasets. This is typically more cost-effective when best practices, conventions, and standards are well established. Potential areas for reusable tools include:

- data repositories;
- data wrangling and conditioning tools;
- workflow tools; and
- portability frameworks (e.g. open neural network exchange-based tools).

Benefits. Reusable tools enhance the efficiency of the entire community. However, even more important for open science, reusable off-the-shelf tools can make AI/ML-based research and applications accessible to a much wider community. They can also enhance collaboration among members of the community.

Begin/Accelerate Workforce Education and Training
Effectively sharing AI/ML-related research objects will require collaboration between the data management and AI/ML data scientist communities. In many cases, participation by domain scientists will also be beneficial, if not required (as is often the case with management of science data). However, in order for these three communities to collaborate, they must have a shared understanding of basic principles, both in data management and AI/ML. Thus, training in data management and domain science will be useful for AI/ML data scientists; training in AI/ML techniques will be useful for data management professionals. It is not essential for a practitioner in one field to attain the same level as their professional counterparts, but to at least understand basic principles well enough to communicate with their counterparts and understand where the main issues are likely to arise. Also, this need not be exorbitantly expensive: capacity building programs (such as ARSET) could be used for training in domain science, while the wealth of massive online open courses in AI/ML might be used for training data management practitioners in basic AI/ML.

Benefits. Cross-training data management professionals and AI/ML data scientists both increases the skill levels in key areas for each group and enables more effective communication between members of the disciplines by providing a common language and basic understanding. The improved communication in turn enables better collaboration and fosters a more cooperative, inclusive culture in that one side no longer feels excluded by the technical jargon used by the other.

Foster an Inclusive, Collaborative Culture in Both Data Management and AI/ML Data Science Practitioners, Within and Across NASA SMD Programs
Effective sharing of AI/ML research objects, from datasets to models and code, to results and papers, will require collaborations between members of the data management and AI/ML data science community. The same inclusiveness that is vital for open science is also fundamental to developing the infrastructure and practices to enable open sharing of AI/ML artifacts. These collaborations require not only shared language (see above suggestion), but also shared open science values. NASA SMD leadership should strive to instill these values in the workforce, both employees and funded investigators. Communications of open science values should be clear, consistent, and continual across the spectrum of channels available to NASA's SMD: scientific and general-audience articles, websites, outreach materials and events, and calls for proposals. In addition to explicit communications about open science and inclusiveness, implicit communications can also help, such as requiring or preferring research teams that include both AI/ML data science and data management professionals.

Benefits. Fostering an inclusive, collaborative culture among data management scientists, AI/ML scientists and domain scientists brings two distinct benefits. Firstly, it allows NASA SMD to leverage the cognitive diversity to be found among these groups. Secondly, methods and practices developed within NASA SMD to cultivate collaboration can often be more themselves utilized to support open science.

CONCLUSION

This is a pivotal time in the physical sciences. Tremendous increases in available science data are helping to drive the utilization of AI/ML methods that can be scaled out; those methods have evolved rapidly in the last decade with the rise of deep learning and available computer power, and open science has come to the fore as scientists seek to understand large, complex system-wide behavior, requiring more multidisciplinary collaboration.

With its long history of openly sharing data and funding large science teams, NASA SMD is in a prime position to take advantage of this confluence. **Cataloging and sharing AI-ready data and models** identified several enabling steps that NASA SMD can take to help effect the adoption of open science and reap its many benefits:

1. **expand sharing to include other research-relevant objects;**
2. **establish standards, conventions, and best practices for managing AI/ML-related objects;**
3. **develop reusable AI/ML-relevant data management tools; and**
4. **begin/accelerate workforce education and training.**
5. **Foster an inclusive, collaborative culture in both data management and AI/ML data science practitioners, within and across NASA SMD programs.**

Several of these steps represent a shift in direction or emphasis from past practice and culture, and thus require both effort and intention in order to achieve. However, the above steps are affordable and can mostly take place in manageable, incremental steps without disrupting budgets or timelines. They require only the decision to move forward.

REFERENCES

- Stojnic, R & Taylor, R. **Papers with Code**, 2020. https://paperswithcode.com/
- Ramachandran, R., Bugbee, K., & Murphy, K. **From Open Data to Open Science**. Earth and Space Science, e2020EA001562, https://doi.org/10.1029/2020EA001562.

FOCUS AREA 6

COMPUTATIONAL PLATFORMS

What are the issues and approaches for training, sharing, and re-using AI data and models in the cloud and HEC?

KEY CONCEPTS

High-end computing (HEC) refers to computing systems that employ large computational power (hundreds or thousands more than a traditional system). Cloud computing is the on-demand service that allows access to computer resources such as data storage or computing power without having to manage those resources. These tools have revolutionized what is possible by enabling the analysis and treatment of massive data from anywhere. However, to design efficient workflows with these tools necessitates co-locating the data and compute to enable the resources to iterate on models and ideas without time or cost concerns—a paradigm shift in the way we think about data archives.

WHY DOES IT MATTER?

These tools can drive the next generation of science at NASA as they provide the capability of accelerating science and discovery, solving problems that cannot be solved in more traditional ways due to the large volume of data and processing power needed. Also, having a shared space for the data and making it fast to access this data effectively facilitates interdisciplinary science.

IMPLICATIONS

Will seamless integration of tools, data, and compute accelerate progress?

Migration to the use of these platforms may both reduce the cost to allow for fast iterations and bolder, bigger scale ideas—potentially accelerating discoveries. Additionally, tactical availability of new platforms (e.g. Spacebourne or quantum) will help researchers open new frontiers to their investigations and build deeper engineering know-how.

WHAT CONTRIBUTORS SAY

"The computing needs of some researchers are outpacing both the standard hardware and software."

"I want to emphasize that science is science and science workflows are different from traditional workflows. Science is about trying things and doing them over and over and over again, and failing. Being able to bring new things into your problem, try things out, throw things away, do things over and over again, is something that we that we have to facilitate with our systems, and on the private systems we can control the cost there, because we've made up capital investment and operational investment, when you move to a cloud, sometimes the costs are a little bit harder to figure out, because if a scientist is trying something out and failing over and over and over in the cloud, they could get quite expensive."

"There is some work being done on edge computing at this point to integrate pre-processing, or some processing on edge computers, and then plug that into the larger scale on premises or cloud computing capability. Edge computing is an interesting concept that should eventually be considered as some type of computational platform."

TECHNICAL MEMORANDUM: FOCUS AREA 6
COMPUTATIONAL PLATFORMS

Authors: Daniel Duffy, NASA Goddard Space Flight Center (GSFC)

SMD AI WORKSHOP
12-14 MAY 2021

EXECUTIVE SUMMARY

For decades, the use of information technology (IT) and high-performance computing (HPC) have been critical to the advancement of science and engineering. HPC has traditionally been leveraged by large-scale applications using massively parallel processing physics-based models to tackle research such as Earth-scale global circulation or the interaction between solar emissions and planetary magnetic fields. Over the past decade in the era of big data, artificial intelligence/machine learning (AI/ML) and commercial cloud computing, high-end IT has become more accessible and is now critical to a much broader range of users than ever before. This, in turn, is spawning a new era of HPC and cloud computing using AI/ML to tackle problems never thought of before.

The purpose of the computational platform focus area was to identify key issues facing not only the more traditional HPC users but also the expanding usage of cloud computing across science and engineering. The goal of HPC and cloud computing is to enable innovation and accelerate science. To do this effectively, the challenges centered around using the current solutions need to be better understood and a discussion about how the future landscape may affect NASA SMD is necessary. Given that every researcher now uses IT in some way to do their work, this was an ideal forum in which to highlight challenges and suggest ways in which to overcome them.

Analysis of the discussions in this focus area identified the following actions for NASA SMD to take in order to maximize the potential for both HPC and cloud computing to meet NASA SMD's mission requirements:

- lower the barrier to entry and expand access to HPC and the cloud by making access to these resources easier and quicker to obtain;
- balance resource utilization based on application requirement, either using on-premises HPC or cloud computing;
- create a collaborative, open science environment where NASA and non-NASA colleagues can come together to accelerate discovery and innovation;
- expand the use of co-design bringing physical scientists, data scientists, and cyberinfrastructure experts together to co-design solutions;
- develop and publish more training data and trained models for AI/ML applications; and
- increase investments in outreach, education, and workforce development.

PURPOSE AND GOALS

For decades, the use of IT and HPC have been critical to the advancement of science and engineering. HPC has traditionally been leveraged by large-scale applications using large parallel processing codes to tackle research such as Earth-scale global circulation or heliophysics models. Over the past decade in the era of big data, AI/ML, and cloud computing, high-end IT has become more accessible and is now critical to a much broader set of problems than ever before. This is spawning a new era of HPC and cloud computing by tackling problems never thought of before.

The purpose of this session was to identify key issues facing not only the more traditional HPC users but also the emerging usage of cloud computing across science and engineering. The goal of HPC and cloud computing is to enable innovation and accelerate science. To do this effectively, the challenges centered around using the current solutions need to be better understand, and a discussion about how the future landscape may affect NASA SMD is necessary. Given that every researcher now uses IT in some way to do their work, this was an ideal forum in which to highlight challenges and suggest ways in which to overcome them.

The subsequent discussions were divided up into multiple groups where each group was given three questions from the following list to guide their conversation.

SMD AI WORKSHOP
12-14 MAY 2021

Understanding the Problem	• Are you aware of use cases in which HEC or cloud computing has amplified impact? How did it make a difference? • What is the role of computational platforms in interdisciplinary science? • What tools do you use for HPC? Is data co-location or choice of silicon an issue (CPU/GPU/TPU/QPU)? • What challenges have you encountered when working with the cloud or 'on-prem'? How did either limit your project? • What challenges have you encountered when working with the cloud or 'on-prem'? How did either limit your project?
Suggestions to Improve Things	• How could we make HPC affordable and accessible to the scientific community? • What are some 'pro-tips' when training models on the cloud (as trying and failing over and over is a large investment)? • If there is spare cloud or HPC resources available, how might that be communicated to the research community? • How is access to cloud/HPC resources managed? Could this be improved? • How can we better understand if HEC solutions improve researchers' ability to get results?
Imagining a Future World	• Is centralized storage and analysis (e.g. the cloud) the future? What are the pros and cons? Where are we headed? • What computational platforms should NASA SMD invest in next? • How is the kind of science we want to do going to inform computational platforms of the future? • How is the kind of science we want to do going to inform the computational platforms of the future? • What are the big themes emerging that require us to rethink our computational platforms? • What should NASA SMD be doing to facilitate and advance more projects with HEC tools? Where will we be in 5 years? 10 years? • What is the future for computational platforms? What science might we see?

KEY THEMES

The following key themes were identified in the conversations and will be addressed in the next sections.

1. Reducing the barrier to entry for HPC and cloud
2. Storage and processor resource availability
3. Usability
4. Cloud adoption
5. Collaboration
6. Co-location of processing with data
7. Education/training/outreach/workforce development
8. Rethinking computational platforms and the future

SMD AI WORKSHOP
12-14 MAY 2021

KEY CHALLENGES

Reducing the Barrier to Entry for HPC and Cloud Computing
Gaining access to NASA HPC and cloud computing resources is difficult and time consuming to the point where users look for alternative solutions. Users requesting access to resources must go through a bureaucracy of rules to request an allocation, obtain an account, get passwords, login to the systems, transfer data, and more. Even across the HEC program, gaining access is not a standardized process. Security requirements add an additional layer of rules and complexity on top of this resulting in a many step process that may take weeks to months to accomplish.

Once access is provided, users must understand how to efficiently utilize the system for their applications. HPC systems are rather fixed and slowly changing, resulting in users having to modify their workflows and applications to the HPC environment. This results in significant time spent on user requests for assistance in porting applications to the system, installing libraries, and testing their applications. On the complete opposite end, cloud computing has so many different possible configurations and options, it is difficult to know where to start. Many users in the cloud mimic what they do on NASA SMD resources resulting in a lift-and-shift approach that is not cost effective.

Resource Availability
Even with the emergence of commercial cloud computing as a resource for NASA SMD science, there remains a need for on-premises HPC and storage resources built to enable global-scale or cosmological-scale simulations. Physics-based, tightly coupled, large-scale applications remain as a necessity to meet NASA SMD's strategic science goals. These applications can require significant amount of dedicated compute, storage, and networking capabilities that may be too costly to obtain in the cloud and difficult to obtain in a timely manner[1]. However, the refresh cycle for on-premises resources can be very slow, while the commercial cloud capabilities are evolving much more quickly. The performance gap between private HPC resources and commercial cloud computing is closing quickly.

NASA SMD's HPC resources are engineered and operated primarily in support of these large-scale applications. Priorities are given to the more computationally intensive over the data intensive applications. This has resulted in a growing imbalance of resources and not necessarily having adequate available resources for all applications. As a specific example, the HEC program is installing GPU systems that are primarily being used for AI/M but not to accelerate physics-based applications. A better balance of resources to applications is needed.

Usability
One of the key challenges identified is the move from local desktop computing resources into the HEC or cloud environments. In many cases, users develop on their local computing resources before moving to larger environments. If users are not familiar with the high-end capabilities, there are several issues that can

[1] HPC in the cloud is becoming more accessible. Commercial cloud providers are providing high-speed networks and file system capabilities comparable to on-site resources.

cause inefficiencies and slow down science. As mentioned above, the inflexibility of a standard, shared HPC environment requires that users port and migrate their application and workflow to the system; the user interface is different and tools and libraries have incompatibilities. Different tools are used to understand file system performance, and the use of batch processing changes how to submit and manage jobs. Security is also a major limiting factor of the usability of the system and in enabling more opportunities for NASA collaborations with non-NASA co-researchers.

As data sets have grown in duration and resolution and as data analysis grows to require more data, the inefficiencies of moving data into and out of computational resources (HPC or cloud) has become a limiting factor. Not only is it difficult to stage data into the HPC or cloud, but it is also important to understand how the data is stored near the compute resources. In other words, file system or object storage performance is important to understand to optimize the applications. In most cases, this is non-trivial and requires an understanding of the underlying hardware, which is often outside the scope of the scientist's expertise. People should not have to be experts in HPC or cloud in order to effectively utilize those systems.

Cloud Adoption
Commercial cloud computing has the potential to provide access to high-end computing resources without requiring access and credentials for on premises HPC resources. This solution can reduce the barriers to entry by providing a more collaborative environment to facilitate open science for both NASA and non-NASA researchers.

However, there are many challenges when trying to efficiently utilize commercial cloud computing that need to be addressed. These include the following:

- gaining access to cloud computing needs to be streamlined;
- fully understanding the cost is a key factor in using cloud computing; on premises computing does not require funding from the researcher, whereas cloud computing does, and the concept of a project paying for the resources could limit science and development;
- migration, start up, and efficient utilization of commercial cloud computing can be challenging for projects; clouds have such a wide variety of services, and it can be daunting for scientists to understand where to start and how to cost effectively use the cloud;
- without taking a cloud-native approach to applications, a lift-and-shift approach to using cloud computing can end up being significantly more costly than on premises resources; and
- there is a perception that there is **not** an advantage to using cloud computing for HPC or AI/ML, either in terms of cost or performance.

SMD AI WORKSHOP
12-14 MAY 2021

Collaboration
The capability of working closely with other experts across and outside of the agency is an essential part of open science and critical to the success of NASA SMD's mission.

Primarily due to IT security requirements, NASA does not facilitate effective collaboration across virtual environments. Access to on-premises IT requires a NASA identity which can take weeks or months to establish, and when speed of research is essential, this can cause major disruptions and delays in funded research. The sharing of data can also be very difficult, especially when the data sets have become very large. Traditional ways of sharing data, such as FTP, have been deemed insecure and shut down causing more disruption to collaboration.

Co-location of Data
The co-location of data with other data and compute resources was widely recognized as a key to future success of NASA. There are activities that are moving in this direction and will take some time to get there. Currently, though, one of the challenges include co-locating multiple and disparate data sets. With the onset of AI/ML, larger and more diverse data sets are being brought together to train models and make predictions. The data is not federated across the HPC centers or clouds, and data exchange is difficult across these platforms.

The issue of data formats and ontological alignment was also mentioned as a challenge. For example, temperature in one data set may be measured in different units than in another data set. Consumers of the data must be alert to these differences and careful to take these issues into account. The lack of analysis ready data or AI/ML training sets is a significant shortcoming. Overall, researchers continue to spend significant amounts of their time co-locating data and converting data to more optimal formats.

Education/Training/Outreach/Workforce Development
This is seemingly a large category to consider, and there were many challenges identified in the workshop that cut across these categories. In many cases, it was clear that researchers do not even know what capabilities exist across NASA for HPC and cloud computing, which puts them at a disadvantage. The existing workforce is falling behind in certain areas, such as AI/ML, while the emerging workforce needs more NASA specific training. The convergence of science with IT in both HPC and cloud computing is not a common set of tools that comes from fresh graduates. These skills have to be grown and learned by practice over time.

The actual usage of HPC and cloud computing resources can also be very inefficient. Researchers are often scientists and engineers, but not necessarily computer scientists or experts in HPC or cloud computing. This results in selection of the wrong size resources, inefficient workflows, and inefficient applications all of which adds to the cost of execution.

Rethinking Computational Platforms and the Future
What does the future of computing hold and how can NASA SMD leverage that to its advantage? This was recognized as a very difficult topic to discuss, as the future remains extremely fluid. There are not enough investments being made in the following: (1) update and modernize applications and codes, and (2) test out emerging platforms for potential use. As an example, most weather and climate models are written in Fortran, which very few if any recent graduates know. The lack of skills for this model of programming combined with the decrease in vendor support for this language puts applications at risk. That, combined with the emerging exascale platforms, also implies that significant code modernization of applications is needed. Funding is needed to upgrade or rehost these codes.

The IT hardware and software landscape is rapidly evolving, seemingly faster than ever before. The onset of AI/ML over the past few years has resulted from major investments by computing vendors and cloud providers. Prior to the rapid emergence of AI/ML, the big data explosion also resulted in major capability increases. These fast changes are extremely difficult to keep up with.

SMD AI WORKSHOP
12-14 MAY 2021

SUGGESTIONS TO IMPROVE COMPUTATION PLATFORMS

Reducing the Barrier to Entry for HPC and Cloud
NASA cannot afford the time wasted in the current bureaucratic process; therefore, gaining access to HPC resources and cloud computing must be simplified and accelerated. While there is a recognition that a governance model is needed to ensure resources are allocated and used appropriately, the overall process should be simple, easy, and standardized.

Once access is provided, it should be much easier to get applications up and running. This can be obtained through a variety of mechanisms, including the following:

- expanding use and coordination of container technologies across the HPC environments, placing an emphasis on interoperability between systems[2];
- providing easier methods to install libraries and code dependencies[3];
- providing access to highly relevant data sets, including machine learning training sets;
- promoting the use of HPC and cloud through computational grants for smaller groups that may not be able to afford these resources;
- providing training and use case examples for how to efficiently utilize resources, either on premises or in the cloud, that are readily available and accessible to the users; and
- standardizing the interfaces across HPC and cloud computing to make moving between the systems easier for users[4].

Resource Availability
Many researchers continue to leverage the local desktop or laptop resources and need an easier path to migrate from local resources to the HPC environments or cloud. This could be addressed by providing a better balance of utilization between large-scale compute and data intensive applications. As data has grown exponentially and with the onset of more analytics-based AI/ML applications, data-intensive models are emerging which require massive amounts of compute and storage. The NASA HEC program must provide adequate and balanced resources for these requirements.

Computing architectures are changing rapidly, and the NASA HEC program must keep pace with these emerging system architectures. Recognizing the challenges of procurement, facilities, and integrations, a hybrid-approach is needed where

[2] Container technologies include Singularity (https://sylabs.io/singularity/) and Docker (https://www.docker.com/). In HPC environments, Singularity is preferred due to security considerations; however, in commercial clouds, Docker is more common. This makes the interoperability between on-premises HPC resources and commercial cloud computing more difficult.

[3] One approach is to utilize capabilities like Spack (https://spack.io/) which allow the users to define their own modules that they can maintain on HPC and cloud systems. Rather than only have the HPC operations team install libraries, users would have the ability to do this themselves and share those libraries with others.

[4] The use of a web-based approach like Jupyterlab (https://jupyter.org/) has the potential of providing users with a common interface between platforms. It further provides uses with an environment in which applications can be run and output data can be quickly analyzed and visualized.

there is easy access to both on-premises HPC and commercial cloud computing. Striking a balance between easy platform switches with well-tuned efficient use of the resources is a real challenge. Tools, libraries, interfaces, and data sets need to be similar enough to enable rapid migration between on-site HPC and cloud computing depending on the application requirements and project timelines. Validation of results on each of the different platforms is essential to establishing confidence in the results.

Usability

In many ways the HPC environments need to be more flexible and adapt more readily to the user requirements than they do currently. While the cloud is an extremely adaptable system (perhaps with too many choices), the current HPC systems are hard to change for the application. This includes the use of common interfaces from desktop to HPC to cloud that can adapt and modify to meet users' needs[5]. Users need easier ways to modify the environment, install software, and get support from HPC and cloud experts. In many ways, a co-design approach would benefit both the service providers and the science teams. Co-design, in this case, means having HPC and/or cloud experts side-by-side with the science teams to create HEC and storage solutions. This is not the typical approach where a set of requirements is listed and solutions are provided; rather this is an agile approach to creating usable and efficient solutions for science teams. This has been demonstrated to rapidly infuse an understanding of the new platform with some projects.

HPC and cloud must provide a balanced approach for large data volumes, both read and write. It is not sufficient to just co-locate data and compute if the data is stored inefficiently or is not in an analysis ready state (either hardware or software). There is also a need to recognize and support the different requirements for data storage and data management. Currently, the HPC and cloud solutions typically only provide storage solutions with very little data management capabilities; data management is then left up to the end user. More support is needed in this area, as is done at the observational data repositories, to not only control the growth of data but to fully harness the potential of the data. As with observational data, a review should be conducted before release of any files to ensure they are worth preserving and have sufficient metadata to be reusable.

A solution is also needed to create collaborative, open-science environments that enable both NASA and non-NASA researchers a place to innovate. The information technology security requirements inhibit access to NASA resources and data; a more balanced approach is needed which accounts for both security and data sharing.

Cloud Adoption

As mentioned before, easier access to get into the cloud will encourage more adoption. To overcome the potential cost barrier, NASA should offer credits for commercial cloud computing to projects. This could be accomplished in several different ways: (1) partnerships with commercial cloud providers, (2) funding through various working groups across NASA[6], or (3) support through the HEC program[7].

[5] See the footnote about Jupyterhub, web-based interfaces.
[6] The SDMWG AI/ML working group funded AI/ML projects to use cloud computing in FY21.
[7] Over the past year, the HEC program has been providing access to AWS commercial cloud for applications.

SMD AI WORKSHOP
12-14 MAY 2021

In addition to providing funding, additional support is needed to create cloud-native solutions specifically designed for science applications. For most scientists, the transition from internal HPC or desktop computing to commercial cloud services is not a straightforward endeavor. Cloud-native implies that applications are utilizing native commercial cloud services, rather than re-engineering these solutions by hand in the cloud. Examples of these services include relational databases, parallel clusters, load balancers, and much more. This implies that these services must be available for use and not limited by security concerns.

Cloud-native solutions should be built using the concept of infrastructure-as-code which is a process for creating and maintaining solutions through configuration and definition files; examples of this include CloudFormation Templates[8] in AWS or a more general approach that has some portability across clouds using Teraform[9]. These templates should be stored within code repositories and shared across NASA.

In addition to sharing solutions, lessons learned should also be shared across different use cases. As an example, the use preemptable machines on commercial clouds (e.g. spot instances on AWS) and provide significant cost savings. The use of object storage in the cloud can provide additional cost savings but has limitations with respect to file system type access. Furthermore, the use of containers can greatly simplify the transition from on premises resources to the cloud, provided that the containerization services are compatible. In general, example use cases that have lessons learned applied should be made available.

Collaboration

As mentioned before, the speed of access to NASA IT is critical to collaborative and competitive research. NASA SMD should invest in platforms, such as in the commercial cloud, that can facilitate open science with access to large data sets where the barriers to entry are extremely low. Can partnerships be set up with commercial cloud providers to better facilitate this capability?

Once these environments are established, data and models should be easy to access. While recognizing the need to maintain some level of privacy due to the competitive process of grants and writing papers, NASA should publish their data sets as soon as possible, including training data, trained models, and environments in which these models can be easily reproduced.

NASA should also partner more with other agencies and universities to establish more cross-utilization of HPC and cloud resources. There are many HPC centers throughout the US with different capabilities that might be more appropriate to specific applications. These partnerships could accelerate science by maximizing the utilization of these resources.

[8] https://aws.amazon.com/cloudformation/resources/templates/
[9] https://www.terraform.io/

Co-location of Data
Better tools and faster access are generally needed even when data are co-located; this includes better search functions, ontological alignment tools, meta-analysis of what data is of most interest or common across a community, and easier data migration tools. Overall data management is a major challenge, and even though this is being addressed for the Earth observation data, scientists are left managing their own data with rudimentary tools. Tools to assist researchers in moving data between platforms and migrating data sets into analysis ready formats, such as Zarr[10], would greatly benefit the end user.

NASA is currently co-locating Earth observation data into AWS, and there is a danger of being locked into a single commercial cloud. It was recognized that NASA is looking at this and the suggestion is to have data stored and federated across multiple clouds vendors.

Education/Training/Outreach/Workforce Development
Outreach and training were highlighted as key investments that could greatly benefit the research community. Starting with outreach across NASA to educate the existing workforce on the capabilities available to them would go a long way. A comprehensive training program that would continue to invest in the development of the workforce is also needed[11]. It should also be recognized that these training programs need to evolve as new capabilities, languages, and environments become available. They should be delivered in a professional and stimulating way.

In addition to more training, documentation, and representative use cases, NASA SMD could make a call to seed applications in HPC and cloud computing. This could be an investment in short term projects where domain science researchers are coupled with experts in cyberinfrastructure and data science with funding to cover labor and for the compute and storage resources. The goal of these projects would be to provide a foundation for these applications to efficiently scale to tackle the next generation problems for NASA SMD.

More partnerships with universities, colleges, and even K-12 school systems could be made to introduce students to NASA relevant data and analysis as soon as possible. Investments in bootcamps, hackathons, and fellowship programs to develop the next generation workforce were also suggested. This also aids in recruiting talented and motivated individuals.

Rethinking Computational Platforms and the Future
Overall support for NASA researchers to keep more up to date with emerging technologies and providing them support to evolve their applications and models is critical. Moreover, NASA needs to invest in a better understanding of the following:

- the climate and environmental impact of the compute and storage requirements being used across NASA; we do not want to affect the climate by studying the climate;
- domain specific languages, like Kokkos[12] or GridTools[13], to provide higher level programming frameworks that will make porting across different platforms much easier and make applications more maintainable;
- investments in augmenting physics-based models with trained model components that can run faster on smaller resources for specific applications; there will still be a need to run the large-scale physics-based models, but in many cases, trained model components may be more than adequate;
- how to effectively leverage the Internet of Things (IoT); smart cities are emerging as major sources of data, and NASA applications should leverage this data;
- edge computing is a capability that is common, but more work needs to be done as to how to effectively deploy and integrate this for NASA missions, including autonomous missions;
- neuromorphic platforms, which are systems designed to accelerate AI/ML training and inference and are modeled after the human brain; and
- quantum computing, which still may be a decade away from real world applications, but investments need to be made now to understand how this could be leveraged.

CONCLUSIONS AND NEXT STEPS

The future success of NASA's mission needs HPC and cloud computing to address the next generation questions and to design, build, and launch remote sensing platforms for the future. Science and engineering are more tightly coupled to IT than ever before, and therefore, NASA must make these capabilities easier to access and use. To simplify the diverse comments from above, the following are a set of next steps for NASA to consider:

- lower the barrier to entry and expand access to HPC and the cloud by making access to resources easier and quicker to obtain;
- balance resource utilization based on application requirement, either using on-premises HPC or cloud computing;
- create a collaborative, open-science environment where NASA and non-NASA colleagues can come together to accelerate innovation;
- bring domain scientists, data scientists, and cyberinfrastructure experts together to co-design solutions;
- develop and publish more labeled raining data and trained models for AI/ML applications; and
- increase investments in outreach, education, and workforce development.

By taking the above steps and acknowledging issues with HPC and cloud computing, science and engineering can be greatly accelerated. At the convergence of high-end information technology and science, NASA is driving the future understanding of our world and universe.

[10] https://zarr.readthedocs.io/en/stable/
[11] As an example, the NASA Center for Climate Simulation (NCCS-Code 606.2) at Goddard provides a series of python training sessions which also includes AI/ML capabilities.
[12] https://kokkos.org/
[13] https://gridtools.github.io/gridtools/latest/index.html

FOCUS AREA 7

CROSS DIVISIONAL PROJECTS

Cross-divisional science's time has come. Why does it matter? What are the barriers? What are the universal challenges?

KEY CONCEPTS

Cross-divisional science can be sharing methodologies, techniques, datasets, and tools across different divisions. It also refers to the collaboration of researchers from different backgrounds to investigate a common challenge that sits between the demarcation lines between divisions. Machine learning is particularly suited to interdisciplinary science due to its ability to extract insights from multiple datasets.

WHY DOES IT MATTER?

It is often observed that the most impactful scientific and research progress occur at the intersection of domains. Examining both the benefits of and the barriers to cross-divisional science is vitally important to NASA and other agencies with multiple research domains, potentially reducing redundancies across the different divisions and encouraging exploration of phenomenologies that are outside traditional domains—or problems that require skills from multiple areas of expertise.

IMPLICATIONS

Can ML-enabled cross-divisional projects be a way to unlock new science questions and outcomes?

The way we do research is sometimes conditioned to our heritage and certain inertia from our area. Researchers from different backgrounds coming together can shine new light on problems by breaking away from traditional approaches. Specifically for AI, having ML experts work together with domain experts can help push the frontiers of what we thought was possible.

WHAT CONTRIBUTORS SAY

"JPL is putting in some internal resources to take data from the planetary data system, imaging data of Jupiter and using it as a backdrop for trying to look for exoplanets and so forth. We can use our own solar system as a baseline for being able to then do exploration of industry astronomy data sets and you can begin to define some similarities. This just begins to connect to archives, but then, you know, you've got all the work to figure out how to put that data together in a way that you can start to create a more integrated data set for exploration."

"A lot of times people begin to think you can just use the exact same software tool or something like that and some of the words we use often as "methodology transfer", in particular, to say that sometimes what you're doing is looking at how to share or purchase things as a common methodology and how that translates into, to physical software, systems, algorithms, things like that depends on the problem space. But if we begin to recognize the fact that what you're trying to do is look at how to come together and collaborate and take lessons learned. That's an opportunity to also look at reducing redundancy."

TECHNICAL MEMORANDUM: FOCUS AREA 7

CROSS-DIVISIONAL PROJECTS

Authors: Srija Chakraborty, NASA Goddard Space Flight Center, USRA

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

The spectrum of data collected by NASA science missions and NASA Science Mission Directorate (SMD) areas (Earth Science, Planetary Science, Heliophysics, Astrophysics, Biological and Physical Sciences) span diverse areas capturing unique phenomena across each discipline. While reducing, analyzing, and extracting meaningful information across these areas distills out novel information specific to a domain, the process of information extraction and data analysis methods to aid and accelerate science analysis often share similarities across disciplines. Leveraging machine learning (ML) approaches are often ideally suited to facilitate this information extraction process due to the large dataset volumes, complexities of patterns, spatio-temporal heterogeneity, and data sparsity observed across these science areas. This focus area reflects on the challenges and opportunities in enabling transferability of ML methods including generic approaches, code, workflow, and software stack across divisions and disciplines while adhering to domain-specific constraints and are referred to as cross-divisional approaches.

DIGGING DEEPER

Cross-divisional approaches are applicable at varying stages of the data processing pipeline ranging from onboard analysis for improving downlinking to ground-based approaches focused on retrieval algorithms and noise reduction, followed by analyses of higher level products including classification, segmentation, modeling, superresolution, anomaly detection, time-series analysis, forecasting, creating knowledge bases for information retrieval, and question answering, to name a few. Ultimately, this transfer of ML capabilities is crucial in the process of accelerating data analysis and discovery, tackling tasks that are otherwise infeasible even by domain experts and building off of success stories from diverse science disciplines, while avoiding already explored strategies that result in known failures. Moreover, such approaches can also help bring in novel perspectives that are not usually explored by a given discipline leading to further contributions.

Facilitating and advocating for interdisciplinary research lies at the core of utilizing cross-divisional ML approaches across domains. There are six key challenges in designing and adopting this principle.

1. **Determining interdisciplinary overlap:** A barrier to adopting cross-divisional approaches can arise from the task of determining areas that share commonalities. Determining this interdisciplinary overlap can often be challenging, particularly for areas that have not exchanged ideas traditionally and could result in missed opportunities of collaboration with appropriate disciplines.

2. **Disconnected terminologies:** This refers to the challenge of exchanging ideas and ML approaches that are related across divisions, but seem disconnected due to the use of domain specific terminologies, which may also create a further barrier to interdisciplinary collaboration. For example, analyzing light curves (astronomy) can benefit from the broad class of time-series approaches, which is a generic terminology applicable across a large array of disciplines.

3. **Search, retrieval and efficacy:** A further challenge that may arise after determining overlapping areas and potential ML approaches is searching through the ever-increasing volume of scientific literature to be informed about the state-of-art practices to best address a given research question. This can be further compounded by the fact that ML contributions are very often reported on real-world datasets that may not share the same challenges as Earth and space science observations, namely, spatio-temporal heterogeneity, data sparsity, high dimensionality, need for fusing observations from multiple modalities for comprehensive insight into a process, barriers to easy crowd-sourcing for labeling large data volumes due to the need for domain specific knowledge, and scenarios relating to unknown unknowns in areas such as planetary exploration, any of which may render the direct application of ML methods less effective. In such cases, additional steps are necessary to tailor an ML approach for a specific Earth and space science dataset or research question.

SMD AI WORKSHOP
12-14 MAY 2021

4. **Data, sharing, implementation and easy translation:** Although a solution to address the previous problem would be to encourage the development of generic ML methodologies that can be fine-tuned to the problem at hand, adoption of ML approaches across NASA SMD is still limited by data, metadata, workflow (including preprocessing steps), labels, and code sharing practices due to lack of awareness, guidelines or standards, and infrastructure access due to policies. Reproducibility of the environment and software stack can give rise to conflicts that may also discourage the reuse of even openly available ML implementations. Additionally, a vital requirement for prototyping ML algorithms is the availability of cleaned and correctly preprocessed data.

5. **Cultural barriers:** Incorporating ML methodologies and interdisciplinary paradigms into any Earth and space science framework often requires cultural shifts from the traditional approach of seeking a more within-domain exploration philosophy. Furthermore, skepticism over ML approaches is also rooted in a black box perspective of ML and deep learning methods that is viewed as a hindrance by scientists to gaining insight into a process of interest. On the other hand, domain scientists interested in incorporating ML approaches may find it challenging to find skilled and interested ML practitioners or experts to collaborate with, which may also be a barrier in supporting cross-divisional projects.

6. **Practicalities:** Finally, lack of funding, solicitations, and encouragement to explore ML methodologies is a very significant gap on the pathway to facilitating interdisciplinary projects that can be bridged by ML.

IMPLICATIONS

The workshop identified several solutions that could potentially address these challenges and support ML-based interdisciplinary collaborations. These are listed below:

1. **Identify common themes in science questions**: Determining the commonalities in research questions across domains is expected to highlight the natural synergies that stand to benefit from ML-based approaches. This requires an aggregation of experts from both science and ML areas to identify the overarching themes and then explore potential AI/ML solutions to tackle these questions.

2. **Working groups and AI/ML workshops for NASA science:** Working groups and workshops aiming to highlight AI/ML research for NASA science are expected to underline common approaches that can be exchanged across diverse areas while fostering collaborations. Additionally, setting up affinity groups can be useful to further dive into domain-specific challenges while adopting an ML based workflow. Conducting Earth and space workshops at broader ML conferences can be a venue to forge external collaborations and to be informed of generic ML advances.

3. **Breaking down language barriers:** A significant effort needs to be put into revisiting the terminologies in these domains to ensure that ML approaches for overlapping data format or science themes across divisions are identified to support the exchange and flow of ideas.

4. **Improved searching capabilities for visibility of methods, collaborators:** A common repository that indexes AI/ML approaches, advances, projects, and papers across NASA science divisions can be a useful tool to explore potential methodologies and to identify collaborators. Such repositories should be supported with search interfaces that are flexible to retrieve ML-based advances in a given domain (e.g. AI/ML use in heliophysics) as well as solutions to a given class of ML problems (e.g. current projects exploring low shot learning, forecasting) that may be applicable across several science areas.

5. **Open sharing practices and platforms:** Hosting NASA repositories that share data code, metadata, workflow, and strategies to allow easy replication of the environment is crucial to encourage ML use in NASA science. While domain scientists should play a crucial role in guiding science question formulation (suggesting constraints, labeling, and validating) ML practitioners should adopt reproducible workflows, open data, and code sharing policies, as well as diversify the ML pipeline to handle different data formats and interactively initialize the pipeline for a new dataset. In addition, domain-guided models (wherever applicable) and interpretability methods should be given emphasis to explain ML outcomes and increase scientists' trust in AI/ML.

SMD AI WORKSHOP
12-14 MAY 2021

6. **Funding and cultural shifts:** Designing solicitations that explore AI/ML approaches are essential for encouraging scientists to collaborate with ML experts and create interdisciplinary research teams. This process should include both domain scientists as well as ML experts and jointly create solicitations (whenever possible) across related disciplines. Solicitations that encourage the use of ML in a given science area, across a mission, as well as to explore transferability across domains for benchmarking, labeling should be more mainstream to increase AI/ML use. Providing high performance computing resources to researchers, irrespective of funding, is also a vital step to conduct pilot studies, design proposals, as well as complete an ML-based research task successfully. Finally, special issue publications for AI/ML, venues soliciting important negative results should also be emphasized to track both success and failure stories and share the lessons learned.

CONCLUSION

AI/ML approaches have a vital role to play in accelerating analysis and discovery across NASA science areas with a cross-disciplinary approach supporting the exchange of ideas and novel perspectives, reuse of code, workflow, and sharing computational resources. This session identified technical, cultural, and practical gaps in realizing such a paradigm, as well as several strategies to adopt to enable successful cross-divisional projects across NASA science areas in the future.

FOCUS AREA 8

ADAPTING TOOLS AND METHODS ACROSS DOMAINS

What are the ideals and realities of cross-domain adaptation?

KEY CONCEPTS
Regardless of the application, machine learning workflows can include generic machine learning tools (such as data loaders, MLOps or implementations of certain algorithms). These generic tools can be modified and repurposed for different applications. Similarly, curated datasets used in other areas could be the starting point for a different investigation.

WHY DOES IT MATTER?
Understanding commonalities across data and use cases can be key to reducing duplication and leveraging existing workflows, tools, methods, and datasets. Sharing methodologies will also build an understanding of the tools and catalyze trust within the community.

IMPLICATIONS
Could existing workflows from other areas help crack our open problems?

Encouraging these synergies may increase re-use, leverage capacity across the community, and increase the viability and feasibility of approaches. In this way, NASA data will be better leveraged to gain new insights.

WHAT CONTRIBUTORS SAY

"I think more important is the difference in the users in the way in astronomy, astronomers are used to working with, with data, versus the way atmospheric scientists or other members of the science community might be used to working with data is actually quite different. Meeting users where they are is really important and the way we build up the systems has to start with the users."

"(We need to) bring all domains to the same level of ML science and use"

"I think part of this theme of adapting tools and methods across domains is oftentimes getting some scientists in some domains to understand how to trust and understand how to use modern ML methods."

TECHNICAL MEMORANDUM: FOCUS AREA 8

ADAPTING TOOLS AND METHODS ACROSS DOMAINS

Authors: Jeffrey C. Smith, SETI Institute, Mountain View, CA
Amitava Bhattacharjee, Princeton University

SMD AI WORKSHOP
12-14 MAY 2021

EXECUTIVE SUMMARY

NASA SMD recognizes the benefits of adapting tools and methods in machine learning (ML) across multiple domains. Points raised at the 2021 NASA SMD AI Workshop include:

- investments made in code development build on each other to enhance not duplicate;
- tools that are evaluated and used by multiple parties have better quality, elicit stronger confidence in their validity, and have more accurately described capabilities;
- re-use or modification of existing code accelerates the time to science; and
- publication of robust, flexible, and validated re-usable code makes NASA intellectual property available for broader use.

The study groups at the NASA SMD AI Workshop identified and discussed actions needed to successfully share tools and methods across the entire range of programs and research. Some discussion overlapped with focus area 5, cataloging and charing AI-ready data. NASA SMD needs to create an environment that encourages and supports the re-use of software, trained models, and labeled training data within and across the entire Mission Directorate. As the use of ML expands, it is expected this will extend into trained models. Three lines of advance are needed: (1) help developers create re-usable code, (2) help end users identify, evaluate, and adopt other's code, and (3) create a cultural environment that encourages points 1 and 2.

Improvements to the development and publication process involve:

- replacing NASA's software release approach with a more flexible and faster approval;
- a consolidated publication service with a trained staff providing easily maintained metadata to make it easy to select among re-usable software;
- funding the extension of selected codes to new domains;
- preparing clear documentation at an appropriate depth; and
- sharing lessons learned from all software development, including what did not work.

Improvements to the user/consumer include:

- a catalog to help find and evaluate re-usable tools;
- encouraging cross-domain participation in projects for fresh perspective; and
- providing incentives to re-use rather than re-invent tools.

Improvements to the cultural environment include:

- funding education, training, hiring and promotion of developers and scientists;
- encouraging outsider participation;
- assigning a data concierge to each repository who understands the data; and
- raising community awareness of available tools and capabilities.

PURPOSE AND GOALS

NASA provides significant intellectual investment in the development of software. This is growing in software involved with ML. These may be completely new algorithms manifested in research papers, documentation, and code. They may be trained forecast models using common or unique codes, or they may be the labeled training data used to train models. Experience to date has shown the value of re-using software in the form of open-source and commercial tools and libraries such as PyTorch, scikit-learn, Tensorflow, Keros and Sagemaker. This accelerates the time to discovery as well as allows scientists with less ML experience to take advantage of these tools, similar to the way they use statistics without having to derive their own techniques. However, building on someone else's work to advance your own requires that work to be trustworthy and transparent to scrutiny. Thus, generic implementations of convolutional neural nets (CNN) or long short-term memory networks (LSTM) in a validated tool such as Sagemaker can reduce the threshold to entry and allow attention to be applied to the new problem.

Regardless of the application, ML workflows can also include generic ML support tools (such as data loaders, MLOPS, or implementations of certain algorithms). These generic tools can be modified and repurposed for different applications. Similarly, curated labeled sample datasets used in one investigation could be the starting point for a different investigation.

Understanding commonalities across data and use cases can be key to reducing duplication and leveraging existing workflows, tools, methods, and datasets. Sharing methodologies will also build an understanding of the tools and catalyze trust within the community.

NASA SMD's purpose of looking for opportunities and technologies to adapt tools and methods across domains is to move more quickly up the learning curve. By expanding the use of related tools across domains we expect to reduce the costs and accelerate the process of maturing them both in adding functionality and in validating them. Tools are essential to the discovery and understanding of natural phenomena and physical processes. Having multiple viewpoints examining a problem is likely to yield results faster and with higher quality. Validation of the tool's reliability and the quality of the results improves the confidence in the tool. This collaboration also helps to identify the boundary conditions beyond which it is invalid.

NASA SMD has four goals in advancing the work against this problem:

1. making NASA intellectual property available for wider use;
2. creating a community work environment which encourages and aids the development of tools and methods that can be used across the full range of NASA SMD;
3. identifing lessons learned from past successful re-use of workflows, tools, methods, and datasets and to apply them to future work so as to improve re-usability; and
4. avoiding investment in redundant development of new code when the effort could be applied towards improving and validating an existing code performing the desired function.

SMD AI WORKSHOP
12-14 MAY 2021

SUMMARY OF FOCUS AREA 8 DISCUSSIONS

Understanding the Problem
There are a wide range of sources of reusable ML tools and very few sources of re-usable training data. The fact that commercial tools exist and can be purchased is demonstration of the demand for this. There are a wide range of open-source tools in wide use. Lessons can be learned from the way this is done.

NASA researchers may develop a new algorithm and translate it into code for their specific project, but it often continues to be usable only within the context of the specific problem for which it was developed. Frequently, papers are published to describe the science result and only briefly touch on the algorithm or tool developed. Currently, a number of obstacles exist to sharing this code outside the project itself:

- consider whether the emphasis should be on protocol or a software framework; and
- the NASA software release process adds to the labor burden of sharing as well as the delay and cost.

Researchers should always be cognizant that others could use their tools, provided they are designed in a way to enable re-usability. Doing so might just provide the most impactful legacy of their research. There are a number of benefits to developing ML algorithms and codes that can be re-used that are not being realized:

1. reducing the amount of funding for duplicated work and instead invest it in enhancements and improvements built on existing software and datasets accelerating the subsequent times;
2. repeating the application of tools and training datasets allows multiple points of view to examine software and training data for defects and biases; and
3. software that has been validated and qualified by multiple parties improves the confidence that it can be re-used:
    a. produce what it is designed to do; and
    b. calibrates the accuracy of the claims of the developer/users.

Some key questions that help dictate how to take advantage of opportunities:

1. when to use commercial software instead of developing your own; and
2. how to justify building extensions to leverage commercial software instead of developing a competing, open-source standalone tool with a much higher investment and demand on validation.

Jargon conflict can isolate domains. How can this be mitigated? When should we strive to use common terminology and nomenclature to help facilitate the exchange of information? There already are interdisciplinary talks and workshops, but do we fully utilize the opportunities to tear down communication barriers?

SUGGESTIONS FOR IMPROVEMENTS

Three general categories of improvements can be made:

1. improvements to the creation of re-usable software and datasets;
2. improvements to access and availability of re-usable software and datasets; and
3. improvements to the culture and community environment.

Improvements to the creation of re-usable software and datasets are the things on the publication side of the equation:

- collecting a catalog annually of software developed and assessment of ability to re-use it;
- funding dedicated support to codes that are competitively selected as most valuable to share; and
- performing continuous integration of the tools and labeled training datasets as the instrument evolves.

Improvements to access and availability of re-usable software and datasets involves investments infrastructure to help projects identify, adopt and make contributions to the software and datasets.

Improvements to the culture involve investments of funding or attention made in the community environment to give re-use more attention and make some aspects of it more acceptable. Some of these changes reflect discussion in the breakout groups identifying trustworthiness of other researchers work. They are also essential for the support of open science:

- it's okay to fail, as long as you can describe lessons learned;
- it's okay to re-use someone else's work, with credit;
- publish reports on experiences and lessons learned;
- credit researchers for non-publication contributions, both at universities and at NASA, particularly in tenure consideration and performance plans;
- create a mechanism for reducing the language barrier and translating among science and technology domains;
- include lessons learned in mission annual reports and continuation reviews;
- increase the opportunities for formal training in domain science by the data scientists and vice versa;
- recognize that agility is often required in the face of adherence to a plan; and
- encourage attendance at conferences where this is discussed such as pyData, pyAstro.

Another way to convert the current NASA SMD culture to one encouraging developing tools for use across domains is to find commonalities among the five Divisions:

- reduce the language barrier by creating a dictionary of synonyms or glossary;
- provide formal training in domain and data science;
- look for practical applications and experiments as well as the theory/ abstractions;
- create opportunities for participation by outsiders to broaden their perspective;

SMD AI WORKSHOP
12-14 MAY 2021

- pitch a difficult problem in one domain to experts in other fields and ask how they would solve it; and
- create funding opportunities to stimulate collaboration across domains and among scientists and engineers and among universities and federal laboratories.

Similarly, significant impact on the culture can be had by making investments in infrastructure:

- platforms to make re-use possible:
  - a repository or at least catalog of labeled training data, standards, and mechanisms for automation; and
  - access to support resources, including documentation, FAQ, and additional support;
- for each data source, invest in a data concierge (similar to the DAAC scientist role) who understands the data, not just the file format;
- raise community awareness of the infrastructure for sharing and collaboration as well as the visibility of codes available for re-use; and
- a unified, official github with full support to encourage migration from the various githubs and limit them to short term use with migration onto the unified site for re-usable code.

Considerable discussion revolved around different approaches to creating data science frameworks instead of relying more loosely on a protocol approach. Protocols may be too weak and frameworks maybe too rigid. Since something more in the middle ground may require a framework be constructed as the basis for re-use but making it flexible and generalized enough to be re-used for different domains. As there was no winning notion selected, this would be an important topic for further discussion.

Lastly, improvements to the culture require improvements to the status of the workforce. Demonstrate the value of sharing tools across domains by making this a feature of hiring and performance plans and tenure offers. Expand existing fellowships such as the Lasker (STSci), Simons (CCA) and other Frontier Development Lab (FDL)-centric post-docs. Create pilot projects, internships, and support early career accelerators like FDL and DEVELOP with multi-domain collaborations. Create more training and summer school programs like JPL's and the Heliophysics Summer School.

Imagining a Future World

The end state desired by the NASA SMD community is one that enables the release and sharing of high quality, validated, and credible algorithms, codes, tools, models, and labeled training data.

From the developer's point of view, there needs to clear criteria for fast-tracking software release. There is a process for selecting and funding software validation and qualification that is comparable to that used in the commercial sector with dedication to completing it. One criterion should be an independent assessment as to its re-usability. Documentation of sufficient quality should be available to help understand its limitations and how to use it. Projects should require all participants to agree to some software release license at the beginning.

From a user's point of view, there should be a point of entry to search for, identify, and select among software packages and tools that can be re-used. This would include those which only work with commercial environments, clearly identifying their limitations and carry with it the past validation as well as notes about enhancements or extensions that are needed.

The community perpetuates itself in support of open science by sharing more broadly than today. It would be easy to find the tools you need, if they exist, and modify them to fit your investigation.

KEY THEMES IN ADAPTING TOOLS AND METHODS ACROSS DOMAINS

The study groups generally settled on several key themes to creating the culture change require to enable tool re-use. First, NASA is not on the cutting edge of tool development and re-use and to be so will require fundamental changes to the most important features of NASA SMD business practices in order to provide the motivation and to demonstrate management seriousness, specifically:

- criteria for selecting projects for funding;
- dedication of resources to supporting maintenance and modification of codes for re-usability; and
- transform software development in NASA SMD from an artisan to industrial process.

KEY CHALLENGES IN ADAPTING TOOLS AND METHODS ACROSS DOMAINS

Without focus, attention, and funding, there is no means to adapt tools and methods to work across domains except good will. Encouragement and motivation require a commitment to hiring people with the skills, including it in the performance or tenure-track plans and rewarding their achievements.
The black-box problem is of particular concern with re-using software, whether open source or commercial; while many sources provide adequate documentation to run the code, the details of the algorithm being implemented may be proprietary or obscured. Also, it should be expected that solutions do not work well out of the box for very specific tasks and require extensive validation and verification. Moreover, years later an investment in a particular solution may result in thinking about the problem in a certain way, which resists more modern or emerging techniques. Some open source software, such as various Apache tools, have a robust maintenance model but rely on contributed effort from the community. Similarly some commercial software companies plow profits back into the improvement of the code and can advance much faster with professional and continuous attention; other companies do less well.

The NASA management systems do not encourage cross-domain work and code re-use. Funding is available for missions and for science research grants, but it is difficult to augment or supplement those funds to adapt tools and methods for use outside the project. Little funding is made available to maintain key codes after the completion of the project.

The availability of trained software engineers and data scientists who can help the domain scientists create tools and implemented methods across domains is extremely limited. NASA SMD will need to grow their own.

The NASA cultural environment has created a highly competitive environment for funding, which tends to reduce the opportunities to find collaboratively-minded individuals and teams. Additionally, the way NASA funds projects within each SMD division makes multi-domain funding opportunities difficult.

Suggestions to Improve Adapting Tools and Methods Across Domains

1. Develop a roadmap for successfully creating an open-source shared software environment. Start with a review of commercial practices to identify success factors and manifest them in the development of a guidebook for successfully sharing software including a roadmap of the considerations at various stages. This roadmap for portability would encourage the use of existing or modification of existing software, preferably open-source software, but not exclude the re-use of commercial software when it makes sense. Only if existing software is unworkable would the choice be to develop new software.

2. Develop a process and allocate funding to examine software projects for funding of extension into re-use and code maintenance and support.

3. Work with standards bodies, like OGC, to develop standards for acceptable codes and labeled training datasets for re-use. The same concern relates to the data and models, both of which also need to be freely accessible when possible and standardization of data formats would be desirable. There are obstacles to this standardization, not the least being academics tend to be quite independent and prefer using their own standards. When do we accept data format differences and when do we impose a universal format? Could some aspects of the data be standardized but still allow freedom for the researcher to prepare the data most conducively to their project?

CONCLUSION AND NEXT STEPS

With significant culture change, it is possible for NASA SMD to achieve its goals in adapting tools and methods for re-use across domains. A comprehensive plan, integrated with other suggestions from the NASA SMD AI Workshop and the NASA SMD survey, is needed to guide investments and changes to infrastructure and business practices. Without appropriate resources and management encouragement, it will be impossible to make the changes required.

FOCUS AREA 9

PRACTITIONERS CHECKLIST AND AI ETHICS

What do good AI ethics look like in data science, from a practitioner's perspective and what do we mean by AI ethics?

KEY CONCEPTS

In previous sessions, we have discussed AI readiness, uncertainty, reproducibility, and many important aspects of AI. As this workshop comes to an end, can we develop a checklist for AI for science? In this final session, we also want to reflect on what AI ethics means and the role it should play in our research. Ethics is the set of values and standards that guide our conduct and the development of technologies but it has not been specifically defined for AI.

WHY DOES IT MATTER?

The application of AI research can be used for purposes that have a large impact such as informing science policy decisions, monitoring conservation, and human actions. Therefore, as more of our decisions and scientific results are based on AI it is important that we ensure the validity and veracity of our results and that we come together as a community to define what ethics means for AI and how to implement it.

IMPLICATIONS

Can we develop a checklist for ethical practice for AI and science?

Building understanding and awareness about good ML practices will avoid the publication of unverified or biased results, which slow the adoption of AI. Further, evaluating the ethics of our research will ensure its quality and integrity. The adoption of guidelines to improve the neutrality, transparency, and explainability of AI will be key to achieve these goals. As a result, we will also boost our understanding of our models and make AI trustworthy and reliable.

WHAT CONTRIBUTORS SAY

"There is no such thing as a 'black box.' Methods to explain and interpret results should be an essential part of any AI project."

"The hope would be that the different disciplines could mostly get something that's systematic, but then where there does need to be fine tuning so it'd be more practicable."

"Calling out subtleties."

"Ethical AI means that you have really brought in a broader context, from what you're doing, you're committed to understanding not just what the data says, but you're also committed to understanding it's imprint and its representation in the greater picture."

TECHNICAL MEMORANDUM: FOCUS AREA 9

PRACTITIONERS CHECKLIST AND AI ETHICS

Authors: Edward L. McLarney, NASA, Hampton, Virginia
B Cavello, TechCongress

SMD AI WORKSHOP
12-14 MAY 2021

INTRODUCTION

With rapid adoption of artificial intelligence (AI) across the NASA science ecosystem, scientists and researchers need practical guidance for ethically applying AI to their work. High level ethical AI topics are being debated at global and national levels even while practitioners are experimenting with and deploying initial AI systems. One way of creating practical guidance is in the form of a checklist of items for ethical AI practitioners to ask themselves to help guide their work. The NASA SMD AI Workshop leveraged v1.0 of the NASA Framework for Ethical AI as a conversation starter for participants to consider how high level principles might apply to science's work. These ethical AI conversations will provide a starting point for a community-generated set of ethical AI checklist items for NASA's science community. This memo serves to document key points from the ethical AI checklist workshop session, providing source material and guidance for creating science-specific ethical AI practitioner checklists as follow-on work.

DIGGING DEEPER

Rosie Campbell, head of safety-critical AI at the Partnership on AI, kicked off the ethical AI checklist session with a thought-provoking topic keynote, highlighting the proliferation of AI, the need to implement AI ethically and responsibly, and suggestions to move ahead with multi-faceted mitigation approaches despite the turbulent solution space. More details from Campbell's keynote are found below. Inspired by the keynote, participants broke into small virtual groups to discuss three main areas of issues regarding ethical AI: understanding the ethical AI problem, suggestions for improvements within the broader NASA community context, and imagining the future. We briefly explore the keynote and the three discussion topics below, followed by a summary.

Campbell's keynote, "Responsible Research and Publication Practices," focused on four key areas. She noted AI as a high-stakes endeavor, with AI becoming ever more powerful and the risk of any powerful tool for good also potentially being wielded as a weapon. She noted four key elements of responsible AI: research integrity, research culture, research ethics, and down-stream considerations. Campbell noted there must be shared responsibility among leaders and practitioners for responsibly and ethically employing AI. She concluded with three areas of emphasis: being transparent about research motivations and levels of AI employed in work, considering downstream consequences, and reporting on the level and quality of training for a given AI model. Her final encouragement to the group was to set a positive trajectory for responsible AI now to set the stage for the best possible future.

Understanding the Problem
From the breakout groups' discussions of ethical AI, several key themes emerged.

- **Trust.** Participants noted many AI solutions are seen as black-box solutions where the algorithm's training or inner workings are not understood. Potential mitigation approaches included creating grey-box AI solutions designed to expose as many inner algorithm details as appropriate, and also trusting systems by conducting rigorous verification and validation and reporting it in open literature and the repository for that model. Another trust issue was how to enable additional personnel to use trusted AI systems built by others.
- **Bias.** Bias issues included the concept of bias entering AI systems via their human creators, via biased data, errors in data labelling, using data not representative of the overall solution space (e.g. models not trained to all skin colors), and more. Bias considerations also came up in human resources use cases, noting humans must double check to ensure personnel regulations are fairly adhered-to. Bias can also enter the system during checking of data sets. It was also noted that bias can occur in a variety of systems that are not about human societies including Malmquist bias in which brighter celestial objects are more readily identified in observational astronomy. Some participants hoped that recognizing other types of bias might help reduce some of the tension researchers may feel in discussing societal bias issues.
- **Unethical uses.** Groups noted privacy as a risk area for unethical use of AI. These risks include facial recognition potentially infringing on privacy rights or high-resolution satellite images enabling people to track the location and movements of individuals or vulnerable communities. Sloppy data validation was also noted as unethical, and even worse was the idea of falsifying data to fit one's purposes. One group noted that any ethical tool can be repurposed by others for unethical use, so it is important to consider the likelihood and dangers associated with that possibility when developing technology.
- **Process harms**. A tangential, yet important, unethical risk is the exploitation of low-wage earners (e.g. students or international workers) to conduct data labelling via crowdsourcing applications. Participants regarded this as a less-often discussed possible harm. Other risks noted regarding the process of AI-enabled science were climate impacts (e.g. carbon footprint from computer power requirements) and considerations about the transparency of decision-making about how compute resources are allocated.
- **Sources of AI failure.** The group noted multiple contributors to AI failures, to include conducting science without including those affected by the results, lack of understanding of the problem, lack of understanding of the data, assuming correlation equals causation, improperly applying AI outside of domain expertise, lack of bias training, using data inappropriate to the problem, imbalance in presentation of the problem/solution, and lack of standards to judge solutions by. The AI failure source discussion was rich and the most-populated, indicating a high need to address risks and prevent failure, to include ethical AI failure. Several participants pointed out science fiction can be mined for ideas of different ways AI, or other technologies, can go wrong. While the community does not need to get lost in science fiction, it can be an interesting source of ideas and inspiration for inquiry.
- **Ethical authority.** Discussion included the idea many existing organizations are not well-equipped to define right and wrong for ethical AI and may consider it outside of their present scope. This implies science must not only strive to define ethical practices, but also lean on the global ethical AI debate and additional participants beyond traditional IT circles, such as lawyers, philosophers, human rights advocates, etc. An organization such as the National Academy of Sciences would be the appropriate level of independence and breadth and the NASA Advisory Council might review internal efforts.

Overall, the workshop's breakout groups had robust, energetic discussions and approached the ethical AI problem from a wide variety of viewpoints. Ethical AI is a challenging, multifaceted topic in the global scientific community, and many participants noted a sense of "having more questions than answers." The topic areas that emerged from this workshop—trust, bias, unethical use, and sources of failure—all resonated with themes in the broader the science community and are reflected in the larger global debate.

IMPLICATIONS

Suggestions for Improvements
Breakout groups did far more than just admire the problem. Thoughtful, creative experts generated ideas around several themes.

- **Data handling.** Some of the areas discussed included pursuing reproducible data, focusing on data quality, citing data sources and rewarding good documentation, appropriately deciding when and when not to report data location, data abstraction when needed, and more. There was also discussion about developing guidelines for sourcing data processing and annotation work with respect to worker well-being and fair wages and discussions of the added sensitivity of human data as in the case of astronaut health.
- **Environmental considerations.** Quantifying emissions from AI workloads, better strategy for using compute resources on the right problems, making code more efficient, reducing travel while developing AI (as demonstrated during quarantine), and other suggestions were provided as possible ways to mitigate or prevent the potential negative impacts of AI research on the environment.
- **Scientific method.** Participants broadly agreed that generating knowledge itself is an ethical pursuit, and many advocated that science (including data science) should not be silenced. A key point of agreement was to apply existing responsible research/science practices with additional emerging considerations and methods for ethical AI. One discussion group summed it up saying “ethical science is good science.”
- **Checklist contents.** Key ideas included: keeping and exposing the data one’s analysis uses; checking work via physical interpretation and extensive real-world validation; and establishing an oath among AI practitioners to “do no harm,” or similar, following traditional design-of-experiments techniques thus seeking to bake in ethical-by-design when creating capabilities.
- **Sharing data and/or workflows.** Ideas included labelling data sets as AI-ready if they adhere to domain-specific standards, using high-level “intent” questions to guide development, and also considering both the data itself and potential uses of the data. Additional ideas included an ethics review prior to data / workflow release.
- **Declaring/disclosing ethics practices.** Discussion included making reproducibility a standard for all proposals, forming a kind of ethical copyright, making disclosures standard practice, having ethics experts review proposals, making time to disclose ethical aspects despite deadlines, and incentivizing all organizations (government, academia, industry) to declare and disclose ethical AI practices.

SMD AI WORKSHOP
12-14 MAY 2021

Imagining the Future Including Ethical AI
Small group discussions envisioned a future where AI is robustly adopted in an ethical manner.

- **Continual aspiration for enhanced ethics.** With AI as an augmentation or amplifier of human capability, participants pointed out AI systems should seek to mitigate undesired human bias rather than amplifying or obfuscating bias. Participants recognized that many ethics issues are not unique to AI and that there are opportunities to learn from other fields. Variety and diversity of ethical opinions was seen as key to robustness of discussions and to continually improving AI’s moral compass including involving stakeholders from outside the traditional science community. Engaging a broader research community is important to creating a culture in which ethics is considered as part of the research.
- **Prioritization of ethical AI efforts.** Participants emphasized a need to ensure focus on the most critical ethical AI questions, while avoiding expending resources on low priority issues or overburdening early AI experimentation with undue overhead. Forward thinkers also mentioned NASA needs to guide and accelerate ethical use of AI, always being on the lookout for too many regulations that over-constrain the problem. At the same time, some participants expressed concerns about their ability to raise concerns under the current NASA structure.
- **Formal methods for assessing ethical AI.** Participants envisioned creating ethical AI benchmarks, quantifying uncertainty for AI systems to make them more trusted and accepted. Experts suggested data, algorithms, analysis, conclusions, and probabilities should all be bundled in assessing AI systems and their impact on the larger environment/system of systems. Other discussions included standardized, accepted mechanisms for establishing baselines for algorithms, such as grading standards for AI systems, along with touchstone test data sets, allowing fair and unbiased AI performance assessment, and encouraging algorithm creators to innovate solutions which continually score better on these metrics. Others suggested a technology readiness level system for measuring ethical AI readiness. At least one group recognized that as important as benchmarking and assessment are, it’s important to communicate about ML models and AI systems with end users and key stakeholders in language that is clear and understandable to them.
- **Systems for mitigating unintended consequences.** Discussion included examples such as flood prediction models having the unintended effect of reducing home values. An abstraction of this example is an envisioned system that continually tests for unintended consequences and actively mitigates/self-corrects for them. Another suggested mitigation was to assess emerging technologies in limited release in order to refine training and troubleshoot them prior to widespread scaling/deployment. It is worth considering a requirement for awards for competed research to include an examination of the unintended consequences of their research as an adjunct to a proposal, either as part of the solicitation or at the kickoff meeting. Many participants noted that it may be impossible to avoid all negative unintended consequences, but that there is opportunity to be strategic.
- **Lifelong ethical AI infusion and experts.** Leading thinkers emphasized that scientific ethics should be part of all elements of education, from high school,

college, graduate studies, and continuing through work life. Another idea was creating a required ethical AI/scientific ethics class to complement existing financial ethics training. Experts also indicated that adding a technology ethics attorney/counsel to NASA's legal team. Groups also discussed the idea of encouraging ethical AI training in the broader NASA community, not just within work.

- **Globally-informed science AI ethics.** One group focused on bringing additional expertise into NASA's thought processes, with ideas including finding ethics experts, inviting them to join NASA discussions, hiring selected ethics experts, respecting their ideas and expertise, engaging with the larger scientific and ethical AI community, having ethical AI cross-training assignments, and generally demonstrating that NASA/science values ethical AI and ethical science expertise.

CONCLUSION

Despite being the last session in NASA SMD's whirlwind virtual AI workshop, the ethical AI checklist session was well-attended, with participants generating a wide range of thoughtful ideas to provide practical guidance to AI adopters/practitioners, even with ethical AI being a challenging open question for the global scientific community. The group discussed multiple facets of the ethical AI problem space, provided a wide range of ideas for creating concrete guidance, and imagined an even-brighter AI-enabled future state for the science community. Topic areas for understanding the problem included: trust, bias, unethical uses of AI, sources of failure, and ethical authority. Idea groupings for improvements were: data handling, environmental considerations, scientific method, checklists, sharing data and workflows, and declaring or disclosing ethical AI practices. Future visions included: continually aspiring for enhanced ethical AI, prioritization of ethical AI efforts, formal methods to assess ethical AI, mitigating unintended consequences, lifelong ethical AI learning, and leveraging the global community to inform science's approach to ethical AI. Based on these learnings, follow-on groups are advised to work iteratively to create initial science-facing checklists for ethical AI practice as-inspired by SMD's AI workshop.

Suggestions for next steps

1. Conduct a series of workshops on ethical AI with the objective to raise the consciousness of a large number of NASA researchers and to collect anecdotal experience in trying to apply ethics;

2. Conduct a National Academy of Sciences study on ethics in science and AI-enabled science in particular;

3. Ascertain what actions other federal R&D organizations have taken, including NSF, NIH, DARPA, ONR, AFOSR, DoE Office of Science; and

4. Using this background, identify specific actions related to competed research that should be included in solicitations, evaluations, awards, and reviews.

SMD AI WORKSHOP
12-14 MAY 2021

REFERENCES

- **The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems Key Information, Milestones, and FAQs about The Initiative,** November, 2020.

- IEEE, **Ethically Aligned Design, First Edition, A vision for prioritizing human well-being with autonomous and intelligent systems,** 2019.

- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. **On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?** In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. DOI:https://doi.org/10.1145/3442188.3445922

- **AAAI Code of Professional Ethics and Conduct.**

- ACM, **ACM Code of Ethics and Professional Conduct,** 2021.

- **AGU Scientific Integrity and Professional Ethics,** 2017.

- IEEE, 7.8 **IEEE Code of Ethics, IEEE Policies,** Section 7.

- Burton, Emanuelle, J. Goldsmith and Nicholas Mattei. **"Teaching AI Ethics Using Science Fiction."** AAAI Workshop: AI and Ethics (2015).

- V. Vakkuri and P. Abrahamsson, **"The Key Concepts of Ethics of Artificial Intelligence,"** 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), 2018, pp. 1-6, doi: 10.1109/ICE.2018.8436265.

- Judy Goldsmith and Emanuelle Burton. 2017. **Why teaching ethics to AI practitioners is important.** In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17). AAAI Press, 4836–4840.

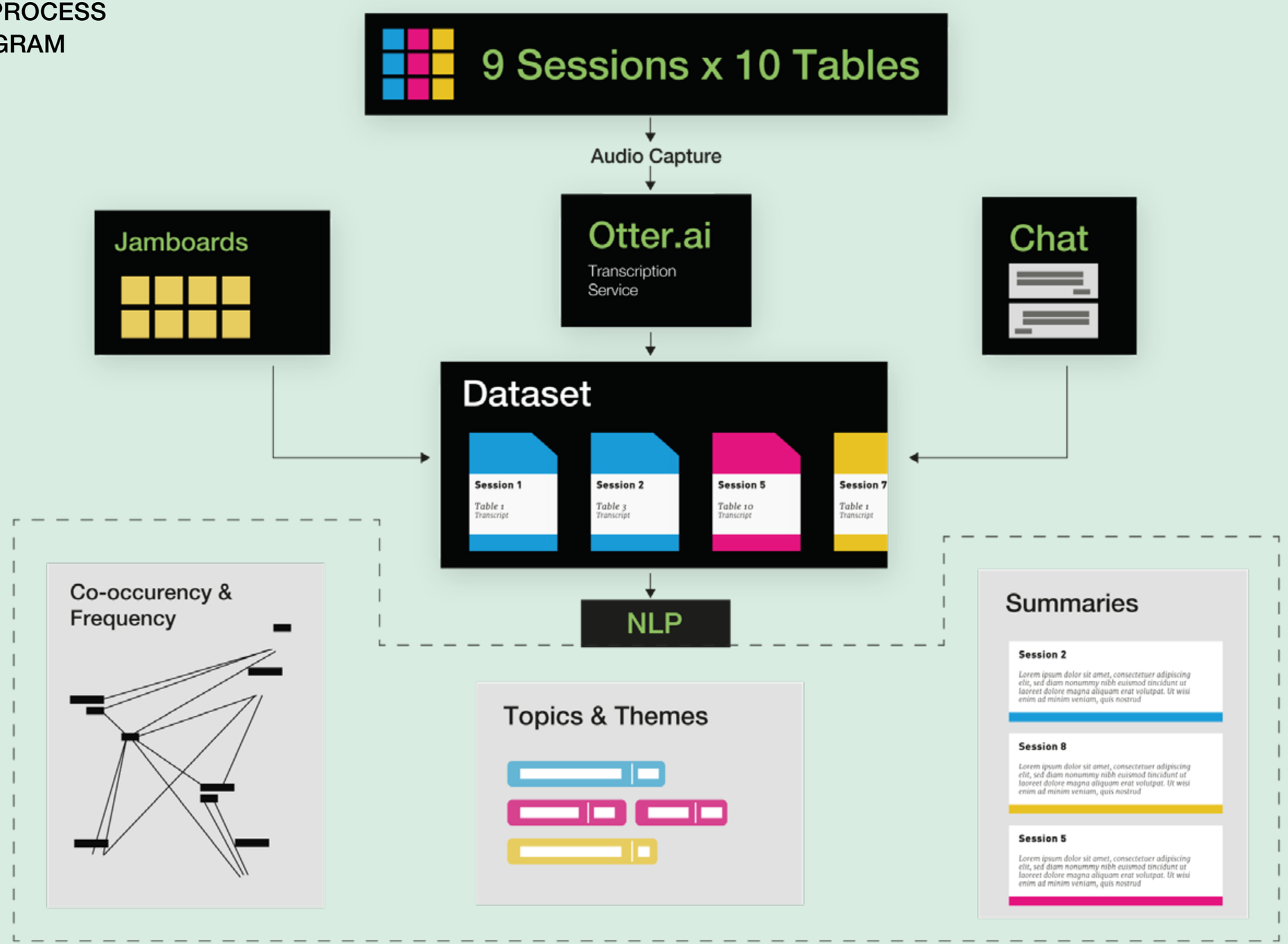
- Michael Anderson, Susan Leigh Anderson. 2015. **The Status of Machine Ethics: A Report from the AAAI Symposium,** Technical Report FS-05-06. Published by The AAAI Press, Menlo Park, California

PART 3

AI AUTO-SUMMARIES OF THE 9 WORKSHOP FOCUS AREAS

SMD AI WORKSHOP
12-14 MAY 2021

AI ASSISTED PROCESS OVEVIEW DIAGRAM



ABOUT THE NLP AUTO SUMMARIZATION PROJECT

NASA SMD AI WORKSHOP NLP AUTOSUMMARY PROJECT

Frontier Development Lab (FDL) / Trillium Technologies Inc
Leonard Silverberg & Frank Soboczenski

The goal of the Natural Language Processing (NLP) auto summarization project was to test the ability of emerging NLP techniques to automatically capture and provide a coherent overview of the major insights and implications from vibrant group discussions between over one-hundred experts running in parallel during the NASA workshop, which generated over 90 hours of insight. A machine learning based speech-to-text service was used (www.otter.ai) to aid the authoring of the technical memos by assigned experts.

It should be noted that this project was run as an experiment to test the potential of these tools and none of the formal conclusions in the NASA SMD summary document were informed by the NLP model.

While the aim of this experiment was to highlight the performance of current state-of-the-art NLP summarization methods the presented results in this report have been edited for minor corrections such as repetition, incomplete sentences and transcription errors by the Frontier Development (FDL) team. A critical element to highlight is that the majority of the above mentioned errors resulted from the automated transcription service not from the developed NLP models themselves as the transcripts after a limited auto correction for filter requests were fed directly into the Machine Learning (ML) pipeline. A comparison of Focus Area 9 has been provided to highlight the differences between the final edited summaries and the NLP generated result (Figure 1). The FDL team as included a metric to highlight the level of overall corrections of the presented automated summaries. The FDL team is aware that those are not standardized metrics for the evaluation of summarization models such as Perplexity, ROGUE / BERT scores and others but the purpose

SMD AI WORKSHOP
12-14 MAY 2021

Figure 1: Example of difference between NLP generated result (left) and final edited summaries (right)

here is to give a brief high level overview of the utility for such models in future live workshop deployments. However, we have seen that the ability to provide highly accurate transcripts is of enormous value in ensuring high fidelity value extraction and provides a foundation stone for next generation knowledge management - and the first ever NLP transformer models trained on space science subjects, to our knowledge. This project therefore is also a learning opportunity for principled use of expert data, future use of developed models and providing such data and models for the community.

The ethical implications of these kinds of techniques informing actual decision making for NASA SMD are yet to be determined. However, there are well established procedures in academia and industry that provide human-in-the-loop ethics evaluation and the team has applied as many applicable elements of such procedures for this experiment. For example, informed consent was sought after from all workshop participants and the option was given to individual participants to opt out of the experiment to name a few. For participants who opted out, a separate workshop session was created where none of their conversations were recorded or used for any NLP model development. NASA SMD may consider applying internal ethical procedures or adopting existing well established frameworks for future workshop experiments of this kind.

PURPOSE

While the presented outcomes in this report only provide a high level overview on the performance and applicability of the chosen NLP methods on dialog data, in this case the NASA SMD workshop, a scientific article published in appropriate venues will provide a more detailed technical insight into the used methods, model details and evaluations.

The purpose of this project was to apply state-of-the-art ML and NLP methods on data collected from the NASA SMD AI Workshop in order to automatically collate insight into the most prominent items discussed during each focus area. Automatic and parallel content capture from group discussions is potentially a useful method for delivering improved workshop outcomes for NASA. Several ML/NLP techniques have been applied to assemble an overview of the most significant topics, word frequency distributions, and graphical relationships. In addition, the entire data was used to fine-tune pre-trained transformer models in order to provide participants and NASA SMD with concise automated summaries of the entire workshop as well as its individual thematic breakout sessions. The developed models and toolkit can aid past and future workshop participants in content and knowledge exploration.

DATA AND LEARNINGS ON BEST PRACTICE

The data gathered for this project was spoken dialog data recorded from workshop participants who were made aware of this undertaking and agreed to their anonymized data being used for this purpose as well as future developments by signing a consent form. Audio (main workshop discussions and breakout rooms) was recorded and automatically transcribed into textual form via Otter.ai (https://otter.ai). Provision was provided for workshop participants who after explanation of the project goals indicated that they did not want their voice being recorded or content integrated into the NLP model, a step known as 'informed consent' a standard requirement for obtaining ethical clearance in ethics review procedures. Chat logs and digital content such as the output of the Jamboard group work were also recorded and digitally transcribed via Optical Character Recognition (OCR) or alternative software. The transcribed recordings were downloaded from Otter.ai by python scripts, that we developed, that can access the Otter.ai Application Programming Interface (API) and retrieve the data automatically in the desired Comma-Separated-Values (CSV) reordered form. After retrieval of the CSV transcripts, limited (automatic) data cleaning scripts removed any coding artifacts, or items selected for deletion (for example names of workshop participants) from the data to create an anonymized, AI ready dataset. The data in total consists of ~70 session transcripts and an overall dataset size of 2.5MB.

ANALYSIS AND NLP METHODS

In order to provide a platform for comprehensive analysis of the workshop data, several methods were applied for different purposes. Those methods range from sentiment analysis, frequency distribution modeling, topic modeling, named entity recognition (NER), visualization and extraction as well as OCR software. Specifically, SpaCy (spacy.io) was used for NER, Google's Tesseract was used for OCR tasks (https://opensource.google/projects/tesseract), and two strategies were used for topic modeling namely Guided Latent Dirichlet Allocation (LDA) with collapsed Gibbs sampling and a probabilistic LDA model based on Pyro (https://pyro.ai/).

SUMMARIZATION MODELS

Transformer models have become the state-of-the-art for tasks such as language classification, language generation, question answering systems, and many others. In addition, other domains such as computer vision recognized the potential for such models as recent advances with vision transformers confirm. The space of transformer models is vast and there are numerous base architectures that can be used however, for the purpose of summarization the T5 and BART architectures were chosen with BART being the more advanced in terms of performance as an internal evaluation has shown. BART is known for its effectiveness in text generation and comprehension. Hence, BART was the final architecture choice for this project. The pre-trained BART model (Wikipedia, Books, and scientific articles) was fine-tuned on the extreme summarization (XSum) dataset. XSum is a dataset to enable one-sentence summarizations from large documents. A critical factor here with the NASA SMD AI Workshop data is that the data does not consist of scientific articles, newspaper reports, or other narrative text but specifically dialog data. Hence, the model was further fine-tuned on the SAMSum corpus which consists of messenger-like conversations and summaries.

The model pipeline was implemented in Python with PyTorch (https://pytorch.org/), PyTorch-Lightning (https://www.pytorchlightning.ai/), and the Huggingface (https://huggingface.co/) transformer library and repository. For performance tracking and evaluation, hyperparameter, search/sweeps Tensorboard, and Weights and Biases (https://wandb.ai) were used. A scientific publication resulting from this work will provide a more detailed technical insight into methods, architectures, parameters

and evaluations than this high level overview.

ADDITIONAL USE AND ETHICAL CONSIDERATIONS

The benefit of fine-tuned large-scale transformer models aside from the workshop summarization use case presented in this report is the ability to further use such models to build additional applications, subject to ethical review. For example, the models based on NASA SMD AI Workshop topics could be leveraged for question answering systems, classification tasks, sentiment analysis, benchmarking for the community to name a few. Additionally, expanding outcomes from future workshops can be ensembled to allow for cross-document summarization techniques or other emerging NLP strategies to be applied; in other words, the corpus becomes more capable every time it is added to. Furthermore, the dataset itself is a unique resource that is potentially very valuable for the NLP community.

NEXT STEPS

Frontier Development Lab is in the process of assembling the working building blocks of individual tasks addressed by the NLP methods into a comprehensive interactive platform accessible via the browser for NASA SMD as well as past and future workshop participants. Additionally, with further workshops where the NLP pipeline is applied the collated datasets have the ability to grow and lead to a refinement of the current models. FDL / Trillium Technologies also plans to transform this solution into an active learning platform where the underlying NLP models train and evaluate newly added data on the fly.

SMD AI WORKSHOP
12-14 MAY 2021

NLP AUTO-SUMMARY

STANDARDS FOR AI READINESS

Reviewed by Frontier Development Lab (FDL)

Scientists are increasingly turning to AI-readiness as a way to improve the quality and reproducibility of their research. However, not having a common set of guidelines that can be shared with other researchers is not enough to ensure uniformity across the scientific community.

Standardizing the definition of an ML-ready dataset will allow scientists to work together without having to repeat large amounts of data or code. Also, it will reduce the time and cost of developing new tools and methods. For this reason, we need to standardize the way we prepare and validate our results. The development of standards for these tools is key to improving the reliability and validity of our science. In this context, creating a global library of pre-validated models could significantly increase the value of what we do and ease the process of making decisions based on multiple datasets.

As a result, there is a significant amount of variation in the types of studies being carried out by different teams. This has the potential to lead to problems that would otherwise be hard to solve due to the lack of homogeneity of the datasets and unrepresentative properties of any study.

Additionally, many of us don't know how to annotate and harmonize our experiments so we often end up publishing thousands of pieces of code that are not replicable or reproducible.

Solving this problem will help unlock the power of Machine Learning (ML) and open new frontiers for others to advance at a faster pace. It will also allow for more collaboration between scientists from different backgrounds, co-operating with each other and accelerate the adoption of new technologies such as machine learning but also enable them to develop new ways of solving challenges that were once considered "out of reach".

Therefore, standardization of AI readiness refers to ensuring the integrity of its predictions and artifacts. Moreover, building a pipeline for AI ready for use is needed to avoid biases and issues that may limit the applicability of your research and have a positive impact on the ability to drive forward the next generation of discoveries.

A clear understanding of best practices should be achieved if we want to move away from traditional approaches and fast access to real-world applications.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 15%

NLP AUTO-SUMMARY

DATA SPARSITY AND HETEROGENEITY

Reviewed by Frontier Development Lab (FDL)

Scientists are increasingly turning to AI as a way to improve the quality and reproducibility of their research. However, having a standard set of guidelines for ML-readiness is not enough to achieve this goal due to the large amount of data that can be stored and stored in different formats.

For this reason, we will also need to develop tools that allow scientists from different backgrounds to co-operate with each other without having to rely on an unvalidated or biased dataset. Additionally, there is no standardization of what types of scientific experiments should be conducted and how to reproduce them. As a result, it is often not possible to ensure the accuracy of our results.

The next generation of researchers will have access to a larger library of pre-processing and validated datasets so that they can advance faster and more reliably. Also, being able to integrate multiple datasets into a study could significantly reduce the time and cost of doing science. In this context, increasing the availability of repeatable, reproducible and interpretation-ready datasets is key to achieving the goals of any new project.

Machine Learning is particularly important because it allows us to build on existing workflows and solve problems that were once out of reach. Solving these issues has the potential to increase the value of your research and ease the development of new models and methods.

Sparsity refers to significant gaps in the current state of the scientific community when dealing with large amounts of raw data. This lack of uniformity means that even the most well-documented studies are prone to bias and uncertainty. A common approach to improving the reliability of its results is needed to unlock the power of artificial intelligence and extend the applicability of computing power.

Sparsity comes at a significant cost if we do not have a shared understanding of machine learning across all divisions. Heavily unannotated and unrepresentative historical data is also limiting the ability to make informed decisions based on real-world predictions.

Creating a framework for sharing and standards for best practices may help to harmonize the way we think about AI and inform decision-based science by reducing the risk of errors and biases within the lab and accelerating the pace of discovery.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 13%

NLP AUTO-SUMMARY

UNCERTAINTY AND BIAS

Reviewed by Frontier Development Lab (FDL)

Scientists are increasingly turning to AI and machine learning (ML) as the frontiers of science. However, when it comes to reproducibility, having a certain level of certainty about the validity of a study is not enough to ensure the accuracy of our results.

For this reason, we need to develop tools that can be shared with other researchers so that we do not have to rely on an unvalidated or biased dataset in order to advance at a faster pace.

In this space, uncertainty refers to the uncertainty of the data being used for research. A large amount of time and resources are spent in making predictions based on a pre-defined set of datasets. The lack of confidence in these methods means that they are prone to bias and biases.

Solving this problem will allow scientists to conduct more accurate experiments without having to repeat the original work. urrently there is no way to measure the value of any scientific results without quantifying their uncertainty. As a result, many of us don't know if our findings are reliable and reproducible. This has the potential to lead to significant gaps in our knowledge and decision-making ability.

Machine learning is particularly vulnerable to uncertainty due to its non-representative nature. It is also difficult to accurately reproduce the results from different datasets and requires great amounts of computing power. Therefore, publishing a paper with high levels of uncertainty is key to improving the quality of decisions made by scientists.

Uncertainty is often the most important factor in determining the suitability of your research and informing policy decisions.

"Unreliable uncertainty" is the absence of trust in the reliability of what you're doing and could significantly increase the risk of error rates within the field. Additionally, unfavourable properties of AI may limit the impact of using AI as a tool - such as limiting the number of ways to make informed decisions and increasing the cost of replicability and co-hundreder than traditional approaches.

SMD AI WORKSHOP
12-14 MAY 2021

These issues are potentially life-enhancers that slow the process of discovery and help unlock the next generation of discoveries. But uncertainty and uncertainty are also barriers to uncovering the benefits of artificial intelligence (AI) because they pose a long-term challenge compared to standardization and predictability. New ways of measuring the relevance of human performance and application are needed to achieve these goals.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 10%

SMD AI WORKSHOP
12-14 MAY 2021

NLP AUTO-SUMMARY

REPRODUCIBILITY

Reviewed by Frontier Development Lab (FDL)

The current state of AI and ML is not reproducible. Reproducibility refers to the ability to reproduce the original work with a set of tools and methods that can be shared by other researchers. For this reason, reproducibility of scientific results is often not enough to ensure reproducibility of existing research.

Having an open code base for other scientists to copy and reproduce will allow them to build on their previous work without duplicating or re-improving it. The way we conduct science is limited due to large amounts of time and resources (such as data from different labs). Additionally, there is no standardization of what is being done at any stage of the process.

Machine learning is particularly prone to artifacts that are not repeatable when multiple studies are carried out. In this context, creating a library of pre-researches has the potential to unclog the gaps in our knowledge and enable us to advance faster and potentially solve problems that have been previously solved.

However, if we want to improve the quality of our research, we need to move away from paper-based research and create a shared space for others to cooperate with each other. As a result, publishing a fully annotated version of a study would significantly reduce the time needed to synthesize and extend the validity of its results.

This approach could also help to speed up the development of next generation models and accelerate the adoption of new ways of doing science.

To do this, the first thing to do is to develop a tool that allows other people to replicate and harmonize the results of your research so that they can make informed decisions based on the same values and standards. Solving this problem is key to effectively leveraging the power of these technologies and accelerating the pace of discovery.

A significant amount of effort is required to share the benefits of modern science across the global community. It is also important to avoid repeating the mistakes made by previous researchers and increase the accuracy of one's work. Otherwise, future research will be biased and flawed.

Improving the reliability and veracity of human and machine learning readiness of academic research is critical to unlock the value of science as a whole.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 13%

NLP AUTO-SUMMARY

CATALOGING AND SHARING AI-READY DATA AND MODELS

Reviewed by Frontier Development Lab (FDL)

Scientists are looking for ways to improve the quality and reproducibility of their research. For example, having a catalog of ML-readiness (ML) tools that can be shared with other scientists will allow them to build on previous work without duplicating it as they do not have to deal with large amounts of data or unvalidated code.

Being able to compare and reproduce the results of different datasets is key to improving the accuracy of AI models and methods. The way we conduct science is limited due to the large amount of time and resources involved in training and developing new algorithms.

As a result, there is a significant lack of space for replicability of the current dataset. In this context, publishing a library of pre-researches has the potential to enable researchers from different divisions to co-operate more effectively with each other.

Additionally, open access to an openly shared set of guidelines could help to standardize the way our research is carried out and advance at a faster pace. It is also important to ensure that the scientific results are reproducible and repeatable across all divisions.

Solving this problem would reduce the cost of computing problems and increase the value of discoveries. However, creating a repository of standards for others to reproduce and extend the validity of your work is not enough to achieve this goal. To do so, you need to create a platform for other researchers to share and verify the correctness of what you're doing.

This process is particularly important when dealing with high-scale projects within AI and machine learning (such as those involving large numbers of samples and large datasets). A common approach to sharing and harmonizing the properties of these tools allows to speed up the development of next-generation applications and accelerate the adoption of new technologies.

SMD AI WORKSHOP
12-14 MAY 2021

Heavily leveraging the power of big data may also allow us to rapidly advance the frontiers of our knowledge and solve problems that previously had to be solved by just a handful of researchers.

Currently, many of us don't even know how to annotate and validate our studies. Therefore, if we want to move away from paper-based research, we cannot rely on traditional methods and develop new approaches based on well-pres knocked off the habit of making decisions that lead to better predictions and inform the next generation of decision-makers. Moreover, building a reference base for further understanding is needed to avoid biases and artifacts that limit the impact of multiple experiments.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 4%

NLP AUTO-SUMMARY

COMPUTATIONAL PLATFORMS

Reviewed by Frontier Development Lab (FDL)

The current availability of high-performance computing (HEC) tools is limited due to the large amount of time, resources and cost of these tools. However, having a common set of tools that can be shared across multiple platforms will allow researchers from different backgrounds to work together more effectively.

There is no standardization of what kind of data should be stored on a computer and how to reproduce it. Also, not enough resources are being made available for machine learning research. As a result, many scientists do not have access to an end-to-end platform that allows them to conduct their research in a reproducible way.

The next generation of ML has the potential to solve problems that were once solved by traditional methods. In this context, we also need to develop tools with built-in capabilities so that we can build on the results of previous projects without having to spend time and money on developing new ones.

Solving this problem could significantly increase the efficiency of our research and speed up the pace of discovery. Additionally, leveraging the power of cloud computing would allow us to advance at a faster pace as well as improve the quality of science.

A significant number of applications require large amounts of computing power and computational power. This lack of resources means that other researchers don't have the necessary resources to carry out scientific experiments. Therefore, creating a shared space for interdisciplinary research is key to unlocking the benefits of cutting-edge technologies and reducing the costs of code development.

Moreover, if we want to move away from paper-based research, collaboration with other scientists will help to reduce the adoption of expensive and time-consuming models and accelerate the process of solving difficult problems. Machine learning is particularly important for this mission because it allows to extract insights from unrepresentative data from any dataset and extend the applicability of your research - such as those based on big data and large datasets.

Currently, "cloud computing" refers to pre-processing and analytics that rely on large quantities of memory and processing power but they are not standardized or repeatable when dealing with large workloads. New ways to ensure reproducibility and ease of use are needed to drive forward the frontiers of modern science through the cloud.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 11%

NLP AUTO-SUMMARY

CROSS DIVISIONAL PROJECTS

Reviewed by Frontier Development Lab (FDL)

Scientists from different divisions are increasingly turning to cross-divisional science as a way to advance the frontiers of their research. However, not enough time and resources are being spent on developing tools that can be shared across divisions or between divisions.

It is not often possible to cooperate with other scientists without having to transfer skills from one area to another. The potential for interdisciplinary collaboration is key to improving the quality and reproducibility of our scientific results.

In this context, we will also need to develop new ways to apply AI methods to solve problems that exist in different domains. Additionally, creating a common set of guidelines for researchers comes at a significant cost to both traditional and non-specialist scientists. As a result, many of us do not have time or resources to work together even when we know how to do so.

Solving these issues could significantly reduce the barriers to making informed decisions about our research faster and more reproducible. Also, having a shared understanding of each division's challenges means that we can move away from pre-processing. Combination of the two divisions has the potential to drive forward the next generation of science.

Currently, only a small fraction of US scientists come from outside divisions - such as engineering or astrophysicists - due to lack of time, resources and expertise. Therefore, if we want to effectively extend the reach of what we do and improve our ability to study the Earth by leveraging the power of artificial intelligence we should seek to unclog the gap between our divisions and unlock the benefits of collaborating with experts from diverse backgrounds.

There is no end to gaps in the current development of AI and ML models. Moreover, they pose a challenge to standardizing and harmonizing the way we design and code for these tools. It is also important to ensure the integrity of your research and increase the value of its validity and applicability.

This is particularly important for those with different backgrounds and experience. Otherwise, you risk breaking down the boundaries of any field and potentially jeopardising the credibility of human knowledge. New ways of doing science are not uniform and time-consuming to achieve goals that allow for further exploration and discovery.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 12%

NLP AUTO-SUMMARY

ADAPTING TOOLS AND METHODS ACROSS DOMAINS

Reviewed by Frontier Development Lab (FDL)

Scientists from different backgrounds are increasingly turning to AI and ML tools as they seek to advance the frontiers of their research. However, having a common set of tools that can be shared across disciplines is not enough to ensure reproducible reproducibility of the results.

The development of an open-source tool that allows other scientists to build on the original work has the potential to reduce the time and cost of doing science. Also, being able to share data and code with other researchers will allow for new discoveries to be made without duplicated or unvalidated.

It is critical that we develop tools capable of replicability and co-operating with any other discipline. In this context, creating a shared library of ML-ready tools is key to improving the quality of our research and accelerating the pace of discovery. There is a significant gap in the current availability of these tools due to the large amount of time, resources and costs involved in developing them.

This lack of standardization means that even the most well-established scientific results are out of date and difficult to reproduce. Solving this problem could significantly increase the value of what we do and speed up the process of discovering new ways to solve problems.

Machine learning is often referred to as “hundred-times more accurate” when compared to traditional approaches. It is also extremely time-consuming and expensive to adapt to new technologies such as computing power and processing power.

Therefore, the next generation of machine learning tools will be used to improve the reliability and validity of existing models and extend the applicability of your research by allowing for collaboration between different divisions. A new approach to sharing and harmonizing the way we conduct research is needed to achieve this goal.

SMD AI WORKSHOP
12-14 MAY 2021

Heavily leveraging the power of artificial intelligence may also help unlock the benefits of cutting-edge research at a faster pace and cut the barriers to reach new goals. Accordingly, we need to move away from traditional paradigms and address issues that are not repeatable if we want to progress faster and better understand the challenges of human health and prevent gaps in our knowledge.

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 13%

NLP AUTO-SUMMARY

PRACTITIONERS CHECKLIST AND AI ETHICS

Reviewed by Frontier Development Lab (FDL)

In light of the growing adoption of AI and machine learning methods, it is important to ensure that the results of a study are reproducible with other researchers. In this context, replicability refers to the process of duplicating or reproducibility of an original work as well as creating a new dataset based on previous research.

Having a shared set of guidelines for ML-readiness will significantly reduce the time and cost of code development. However, not being able to reproduce a paper from the original has the potential to lead to unvalidated results that could be used in different ways by other scientists.

As a result, more than half of all AI-related research can be lost due to non-referential issues such as artifacts or biases. Also, there is no way to verify the validity of what is being developed and how it should be carried out.

The lack of transparency means that we do not know whether our results are reliable or accurate. This is particularly true if we want to extend the reach of our research and advance at the same time.

A significant amount of time, effort and resources have been spent in standardizing and harmonizing the properties of these tools so that they stand on the shoulders of experts from different backgrounds. Additionally, ensuring the integrity of scientific results is key to effectively leveraging the power of artificial intelligence and improving the quality of their results.

Solving problems that arise when dealing with AI may also allow others to advance faster and better understand them. Therefore, publishing a fully annotated and verbatim copy of your work is critical to avoid conflicts of interest and potentially invalidating the conclusions of previous studies.

It is also necessary to build trust and confidence in the next generation of ML models. Heavily reliant on repeatability of previously published results and open-interpretation of existing technologies - including those designed for use outside the traditional limits of space and large scale research (such as computer science) but without any guarantees of accuracy or uniformity. Moreover, many of today's AI experiments are biased and prone to bias because they don't reflect the current nature of human knowledge.

SMD AI WORKSHOP
12-14 MAY 2021

ABOUT THIS AUTO SUMMARY

It should be noted that this project was run as an experiment to test the potential of NLP methods in comparison to human generated summaries. None of the conclusions in the auto generated text above should be taken as formal recommendations for NASA SMD or any other of its partners or affiliates.

The summaries presented here are not the raw model output but the copy edited version for presentation purposes. Minor edits have been made such as removing repetition, transcription errors based on semantic understanding of the Otter.ai output and incomplete sentences.

However, the FDL team wanted to provide a metric to highlight the differences and the amount of corrections needed as seen by the percentage below. The percentage is based on word differences between the raw output and human edits. Additionally, as mentioned above a resulting scientific publication of this experiment will provide proper details on evaluation.

Furthermore, the FDL team wanted to highlight that most of the errors that needed correcting resulted from the Otter.ai transcription service that presented a middle layer between the spoken dialog and the NLP models.

AMOUNT OF CORRECTION: 17%

PART 4
WORKSHOP DETAILS

ORGANIZING TEAM

The Strategic Data Management Working Group (SDMWG) AI/ML Working Group was established in 2020 under the charter of NASA Science's SMDWG. The SDMWG lead the development of a new SMD-wide data management strategy plan, will align the advances in information technology with the unique needs of science data systems and computing. This union lets the strategic plan both inform technology investments and provide a roadmap for how SMD can partner with other organizations, within NASA and externally, to enable greater scientific discovery. The SDMWG representatives include:

Dr. Manil Maskey

Dr. Daniel Duffy

Dr. Megan Ansdell

Dr. Madhulika Guhathakurta

Dr. Evan Scannapieco

Dr. Srjia Chakraborty

Yvonne Ivey

Michael Little

Dr. Steven Crawford

Dr. Roopesh Ojha

Dr. Sylvain Costes

SMD AI WORKSHOP
12-14 MAY 2021

PARTICIPANT LIST

Many thanks to everyone who lent their insights and wisdom to the 2021 NASA SMD AI Workshop:

Aaron Piña, NASA HQ (Earth Science Division)
Abby Azari, University of Berkeley, California, Space Sciences Laboratory*
Adam Oberman, Mila and McGill University (Dept of Math and Stats)
Aenor Sawyer, UCSF (Dept Orthopaedic Surgery, UC SpaceHealth)
Akshit Arora, NVIDIA (Solutions Architecture)
Alan Li, NASA Ames Laboratory for Advanced Sensing
Alexander Barrie, NASA
Alexander Lavin, Institute for Simulation Intelligence
Amitava Bhattacharjee, Princeton University (Department of Astrophysical Sciences
Anamaria Berea, George Mason University and Frontier Development Lab
Andrés Muñoz-Jaramillo, Southwest Research Institute
Andrew Westphal, UC Berkeley
Andrew Michaelis, NASA Ames Research Center (TNC)
Anirudh Koul, Pinterest / SpaceML
Anuj Karpatne, Virginia Tech, Department of Computer Science
Ashish Mahabal, Caltech (Astronomy) and JPL (39)
Ashley Villar, Columbia University (Astronomy)
Ashley Pilipiszyn, OpenAI
Asti Bhatt, SRI International (Center for Geospace Studies)
Atilim Gunes Baydin, University of Oxford
Ayris Narock, NASA
B Cavello, TechCongress
Barbara Thompson, NASA GSFC Heliophysics Science Division
Bertrand Le Saux, ESA ESRIN Ø-Lab
Bill Diamond, SETI Institute
Brant Robertson, UC Santa Cruz, Department of Astronomy and Astrophysics
Brian Thomas, NASA/Heliophysics (Code 672)
Brian Powell, NASA GSFC
Brian Nord, Fermilab *
Bruce Bassett, AIMS, SAAO and UCT
Bruno Sánchez-Andrade Nuño, Microsoft
Chedy Raissi, Ubisoft
Chris Bard, NASA
Chris Gottbrath, Facebook AI (PyTorch)
Chris Mattmann, NASA JPL
Christian Reyes, NASA

SMD AI WORKSHOP
12-14 MAY 2021

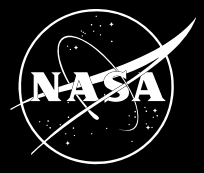
Christine Edwards, Lockheed Martin Space
Christine Custis, Partnership in AI
Christoph Keller, NASA Goddard, GMAO / USRA
Christopher Lynnes, NASA IMPACT project (SMD)
Compton Tucker, NASA GSFC
Daisuke Nagai, Yale University
Dan Crichton, NASA/JPL
Dan Berrios, ARC/BPS (USRA)
Daniel da Silva, NASA/USRA, Heliophysics
Daniel Duffy, NASA Goddard Code 606
David Donoho, Stanford University
David Hall, NVIDIA
Delaney Cosgrove, NASA
Ed Mclarney, NASA HQ OCIO TDD BIO / DT
Ellianna Abrahams, UC Berkeley, Astrophysics and Statistics *
Erik Antonsen, Baylor College of Medicine
Erin Ryan, Booz Allen Hamilton (Astrophysics)
Evan Scannapieco, NASA SMD - Astrophysics Division
Frances Adiele, BAH
Frank Soboczenski, King's College London (Medical Division)
Gautham Narayan, University of Illinois at Urbana-Champaign
Geeta Chauhan, Facebook AI
Graham Mackintosh, BAERI / NASA (ARC-TN)
Hamed Alemohammad, Radiant Earth Foundation
Hannah Kerner, University of Maryland/Earth Sciences
Ignacio Lopez-Francos, NASA ARC
Ingo Waldmann, University College London (UCL)
Irina Kitiashvili, NASA Ames Research Center (NASA Advanced Supercomputing Division) *
Ivan Zvonkov, University of Maryland/NASA Harvest
Jack Hidary, Google X@alphabet
Jacqueline Le Moigne, NASA Earth Science Technology Office (ESTO)/NASA ESD
James Parr, Trillium Technologies & Frontier Development Lab (FDL)
Jeffrey Smith, SETI Institute
Jenna Lang, AWS - Healthcare & Life Sciences
John Dorelli, NASA-GSFC Heliophysics Division
John Moisan, NASA/GSFC Code 610.W
John Kalantari, Mayo Clinic
John Karcz, NASA Ames Research Center, Space Science and Astrobiology Division
Josh Peek, STscI (astro)

Katie Baynes, NASA ESDS
Kevin Murphy, NASA SMD
Laura Carriere, NASA Center for Climate Simulation (SED)
Lauren Sanders, NASA Ames GeneLab
Leonard Silverberg, Trillium Technologies & Frontier Development Lab (FDL)
Lorraine Fesq, JPL/Caltech
Louis Barbier, NASA / Office of the Chief Scientist
Lukas Mandrake, NASA JPL Div 39
Madhulika Guhathakurta, NASA SMD/Heliophysics
Manil Maskey, NASA, SMD, Earth Science
Mark Cheung, Lockheed Martin Advanced Technology Center
Megan Ansdell, NASA HQ, Planetary Science Division
Michael Krihak, NASA ARC
Mike Little, NASA GSFC
Mike Seablom, NASA SMD - Earth Science Division
Milad Memarzadeh, NASA Ames Research Center - Intelligent Systems Division
Muthukumaran Ramasubramanian, NASA IMPACT
Nadia Ahmed, UCI
Nathan Kutz, University of Washington
Nicolas Longepe, ESA ESRIN Ø-Lab
Nikunj Oza, NASA ARC
Pierre-Philippe Matthieu, ESA ESRIN Ø-Lab
Rahul Ramachandran, NASA MSFC Earth Science
Regiuel Days, Google
Rob Reynolds, NASA (Human Health and Performance Directorate)
Robert Stojnic, Facebook AI
Roopesh Ojha, NASA HQ
Rosie Campbell, Partnership on AI
Ryan Scott, NASA Ames; BPS; NASA GeneLab and Ames Life Sciences Data Archive
Sapna Rao, Lockheed Martin Space - CCS
Sara Jennings, Trillium Technologies & Frontier Development Lab (FDL)
Savannah Thais, Princeton University (Physics and Research Computing)*
Scott Penberthy, Google Cloud
Shashi Jain, Intel Corp
Siddha Ganju, NVIDIA, AI Applications
Srija Chakraborty, NASA GSFC, USRA (Earth Science)
Steve Crawford, NASA HQ
Stewart Doe, NASA
Supriyo Chakraborty, IBM Research

SMD AI WORKSHOP
12-14 MAY 2021

Sveinung Loekken, ESA ESRIN Ø-Lab
Sylvain Costes, NASA (Biological and Physical Sciences)
Tianna Shaw, NASA Ames Research Center Space Biosciences Division
Tsengdar Lee, NASA/SMD/ESD
Ved Chirayath, NASA ARC
Victoria Da Poian, NASA GSFC (699)
William (Bill) Miller, National Science Foundation, Office of Advanced Cyberinfrastructure
Yarin Gal, Oxford University
Yvonne Ivey, NASA SMD
Zack Gainsforth, University of California at Berkeley / Space Sciences Laboratory
Zain Masood, Boulder AI (CTO, Engineering)

*These individuals politely requested to opt-out of the data gathering aspect of this workshop after being informed of the intention to create discussion transcripts and auto-summarization. The organization team respects their principled decision and help to establish best practice in AI ethics and knowledge management.



SMD AI WORKSHOP

12-14 MAY 2021

Many thanks to everyone who lent their insights and wisdom to this workshop.