Report on Workshop on Artificial Intelligence in Strategic Planning and Science Prioritization

*Virtual NASA Workshop*

*May 12-13th, 2020*

**Table of Contents**

[Executive Summary 3](#_Toc67644989)

[1. Introduction: National Science and Technology Prioritization; An Opportunity for Artificial Intelligence 4](#_Toc67644990)

[2. Workshop on Use of AI for Strategic Planning 5](#_Toc67644991)

[3. Results 6](#_Toc67644992)

[**a.** **Key Challenges** 6](#_Toc67644993)

[**b.** **Test Cases** 6](#_Toc67644994)

[**c.** **Best Test Case Descriptions** 7](#_Toc67644995)

[**d.** **Test Case Analysis** 10](#_Toc67644996)

[4. Observations 11](#_Toc67644997)

[APPENDICES 13](#_Toc67644998)

[**A.** **Agenda** 13](#_Toc67644999)

[**B.** **Catalog of Challenges** 16](#_Toc67645000)

[**C.** **Catalog of Test Cases** 20](#_Toc67645001)

[**D.** **Participants** 21](#_Toc67645002)

[**E.** **Sponsors, Partners & Contributors** 22](#_Toc67645003)

[**F.** **Organization Committee** 22](#_Toc67645004)

[**G.** **Acronyms** 23](#_Toc67645005)

**List of Tables**

[**Table 1. Key Challenges** 6](#_Toc67644976)

[**Table 2. Summary of Best Test Cases** 7](#_Toc67644977)

[**Table 3. Analysis of Best Test Cases** 11](#_Toc67644978)

[**Table 4. Catalog of Test Cases** 20](#_Toc67644979)

# 

# Executive Summary

This report details the observations from a two-day virtual workshop, held May 12-13, 2020, focused on whether, and how, artificial intelligence (AI) could assist humans in strategic planning, specifically in science and technology prioritization. The participants identified several “key challenges” that AI might tackle in this area. To further understand the value of these key challenges the workshop then developed related test cases that would demonstrate specifically how AI/machine learning (ML) could provide assistance to humans. Approximately 40 subject matter experts (SMEs), with backgrounds in AI, strategic planning for science, and scientific data, were gathered for the conference. This report collates the details of the output of the workshop.

The “best” test cases include (in no particular order):

* Use of AI to assist in selecting Decadal Survey priorities.
* Use of AI to identify new, or previously unidentified, science topics for prioritization.
* Using AI to better label and increase discoverability of scientific literature and proposals.
* Use of AI to enhance current observation capabilities for scientific missions.
* Using AI to mitigate biases in selection of proposal reviewers and membership of advisory committees.

Examination of these test cases indicates that Natural Language Processing (NLP) is a common capability found in most of the ”best” (top-rated) test cases and is a valuable, multi-purpose tool which enables ML in this area.

The organizers thank NASA’s Office of the Chief Information Officer (OCIO), the Office of the Chief Scientist (OCS), the Office of the Chief Technologist (OCT), and the Science Mission Directorate (SMD), for support of this workshop.

# Introduction: National Science and Technology Prioritization; An Opportunity for Artificial Intelligence

Science and technology strategic planning and prioritization are literally “billion-dollar endeavors” as the consequences of these studies guide NASA and other government agencies’ investments for years and often decades.

The process of prioritization is, of course, essential, as there are usually many more candidate areas for investment than there are resources available. Successful investments can produce technologies and missions that yield discoveries that change established fields, establish professional reputations for individuals and teams, and guide funding agencies. This may create a virtuous cycle, as consistent high-quality results from successful investment strategies will reinforce the motivation for the federal government to continue investing in these areas. There is thus a great deal of pressure to “get it right”.

The National Academies of Sciences, Engineering, and Medicine (NASEM) Decadal Surveys (DS) are emblematic of strategic planning in science and are considered a ‘gold standard’ in this area. There has been sustained and broad support for the Decadal process within the funding agencies, Congress, the Administration, and the scientific communities. Indeed, international science agencies have, to some degree, followed the DS model for their own science and technology funding. That said, the very success of DS may have contributed to why there have been few changes to their structure and processes for over half a century, in the case of astronomy and astrophysics.

The main goal of any DS remains the same, which is to review the current breadth of scientific research activity in its domain and gauge its impact, in order to predict desirable areas for future work. Much of the process is still directly executed by humans and relies on a primarily manual process of information synthesis. However, the amount of information to synthesize has grown substantially, making this process more and more challenging. The number of practicing scientists and publications is many times larger today than it was decades ago, and much more international in scope. In essence the amount and complexity of the information to be analyzed by the DS panel, is larger and more complex than ever. It takes little work to see that these issues and trends may also impact other areas of science prioritization, such as panel reviews, mission science prioritization, mission operations, and planning.

Artificial Intelligence (AI), specifically Machine Learning (ML) and Natural Language Processing (NLP), may prove to be useful tools to alleviate these obstacles. AI techniques have been demonstrated to excel at understanding large and complex volumes of information, including text. Large corporations are utilizing and developing AI-based software technologies to improve search and discovery in text (ex. Google reformer[[1]](#footnote-1), Google Cloud Platform[[2]](#footnote-2) Microsoft Azure Machine Learning[[3]](#footnote-3), AWS Machine Learning Platform[[4]](#footnote-4)), as well as the needed hardware and tools to power these approaches (for example, NVidia/GPU[[5]](#footnote-5), Google/TPU[[6]](#footnote-6), IBM/Power9[[7]](#footnote-7)). The Federal Government is taking an interest as well. For example, the Office of Naval Research is using ML-based clustering algorithms to pull promising technologies from published literature and other sources (Ryan Zelnio, Chief Analytics Officer Office of Naval Research, private communication).

The maturity of these techniques is demonstrated, yet strategic planning and the prioritization of science goals represent a relatively unexplored area, where the inputs for making these decisions (e.g., peer-reviewed literature) have been growing much faster than the number of participating SMEs. This situation would seem to be an ideal arena to apply the increasingly capable techniques of AI. This motivated us to consider what the value may be for NASA to use these technologies in this way. To explore the possibilities in this area, our team organized a NASA workshop to bring together SMEs in AI, and practicing scientists in the fields of astronomy, planetary science, and heliophysics, to discuss and assess the value of using these increasingly capable tools to assist scientific planning.

# Workshop on Use of AI for Strategic Planning

NASA had two goals to explore through the workshop, in order to explore the value of AI for strategic planning:

* *Identify several “Key Challenges” in strategic planning where AI might be applied at NASA.*
* *Develop a set of compelling demonstrations or “test cases” associated with these challenges, which the Agency might consider for a follow-up study.*

Because of the COVID-19 pandemic, the workshop, originally scheduled as a face-to-face event in mid-March, was held as a virtual meeting over two days, May 12-13, 2020. Since there has been little AI application in science planning to date, we selected participants who either were experts in AI, or some aspect of science planning and prioritization.

We kept the workshop participation small (~40 attendees) and limited the meeting to partial days (~5 hrs. each). Simultaneous breakout sessions were held on the second day, to address specific challenges and associated questions (in the form of test cases) and to better engage the attendees. All of the discussions and breakout sessions were governed by the modified Chatham House rule, where speakers may be quoted, although not identified, to allow free-flowing discussion.

Our four breakout session chairs were: Dr. Heidi Hammel (Executive Vice President for Science, The Association of Universities for Research in Astronomy (AURA)); Dr. Shirley Ho, Adjunct Professor, Carnegie-Mellon University, and the Flatiron Institute; Dr. David Leisawitz, Chief of the Science Proposal Office, NASA Goddard Space Flight Center; and Dr. David Spergel, Professor of Astronomy, Princeton University, and Director of Computational Astrophysics, the Flatiron Institute. Each breakout session focused on a subset of the key challenges selected by the session chairs and developed a set of test cases based on that key challenge. The participants formulated the test cases and were asked to specifically address the availability of data, the value of the case to NASA, and the likelihood of success.

The Workshop Agenda Appears in Appendix A.

# Results

## **Key Challenges**

We identified candidate key challenges, to be discussed during the breakout discussions, by first asking for ideas via a panel discussion that included the session chairs, followed by an open session in which further challenges were generated. In total, 34 challenges were generated during ideation sessions of the workshop (see Appendix B). Some of the challenges had significant overlap and were merged after discussion. After significant discussion we synthesized individual opinions to identify the most important and interesting, or “key”, challenges within the group. As several of the original generated challenges significantly overlapped in their content with one another, we elected to merge these with the new id becoming the merger of the former two, (for example challenges 2 and 27 became “2/27”).

We thus arrived at the top seven most interesting challenges. Three additional challenges that the organizing committee felt were important, but had not been previously supplied by the participants, were then added, rounding out our list of 10 “key challenges”. ***Table 1*** shows these key challenges compiled.

|  |  |  |  |
| --- | --- | --- | --- |
| **Challenge Id** | **Description** | Notes | |
| **2/27**  [merged] | How to find people with subject matter expertise for review panels and white paper reviews? | |  |
| **21** | Can we use data from over a decade ago to go back and predict today’s hot science topics? | |  |
| **13** | How can AI be used to enhance the capabilities of missions/projects? | |  |
| **9/25**  [merged] | How to best automate the extraction, classification, and prioritization of scientific themes and investigations for Decadal Surveys, important conferences, etc., and compare against NASA science roadmaps. | |  |
| **29** | How can we better tag keywording/classifying for proposals | |  |
| **3** | Synthesize and analyze research evolution. Questions: How to understand and visualize the evolution (e.g., growth, decline) of research areas. How to also observe evolution of techniques and methodologies within these research areas? | |  |
| **5** | What biases exist in scientific planning/prioritization? | |  |
| **10/19/23**  [merged] | How can we pick the optimal science missions for given funding/resources and strategic criteria? | | Added by Org. committee |
| **11** | How to understand what mechanisms of funding deliver the best value for achieving science priorities? | | Added by Org. committee |
| **31** | How can AI be used to spot and mitigate emerging problems in mission planning and guide us to solutions based on lessons learned from past missions, projects, programs, and including lessons from across organizational boundaries? | | Added by Org. committee |

**Table 1. Key Challenges**

## **Test Cases**

Test cases are example studies, or assessments, that might be funded by NASA and/or NSF to evaluate the feasibility of AI in responding to a challenge. We identified over 20 partially to fully described test cases arising from workshop discussions. Where test cases were partially described due to time limitations, we contacted the session chairs and workshop participants to gather additional input. In some cases, different challenges generated similar test cases. When this happened, the test cases were merged to reflect their similarity, and a new id was assigned (for example test case 2/27c was merged with 5b to become “5b & 2/27c”).

To better assess the value of the test cases, we developed a set of categories to classify them into commonly understood groupings. Members of the organizing committee and session chairs discussed which test cases were the “best” based on how well each satisfied the following criteria:

* Technical Feasibility
* Data Availability and Sufficiency
* Explainability[[8]](#footnote-8)
* Value to Customer

A review of the discussion shows that there was little to differentiate the test cases from one another for 3 of the 4 criteria. Only the Technical Feasibility (TF) criteria indicated a possible distinction could be made among test cases. Based on this distinction we identified 5 best test cases (Table 2). A summary description of each of these follows below.

**Table 2. Summary of Best Test Cases**

|  |  |
| --- | --- |
| **Test Case Id** | **Title** |
| 9/25a & 21b | Can AI Predict Decadal Survey Priorities? |
| 5a | Can AI Identify New or Previously Unidentified Science Topics/Areas for Prioritization/Funding? |
| 5c, 29a & 29c | Can AI Improve Keyword Assignment for Scientific Papers and Proposals? |
| 13 a | How Can AI Enhance Current Observation Capabilities Using New Machine Learning Techniques? |
| 5b & 2/27c | Can AI Identify People Who Have Technical Expertise in a Particular Area/Topic? |

## **Best Test Case Descriptions**

***Test Case 9/25a & 21b: Can AI Predict Decadal Survey Priorities?***

This merged test case utilizes AI to “predict” the 2010 and 2020 Astronomy Decadal Surveys. Natural Language Processing (NLP) may be used to extract key phrases, terms that will supply “document features” that can be utilized to topic-model research literature. Emergent topics can be examined and compared to Decadal Survey reports and submitted whitepapers to determine if the algorithms are consistent with a human opinion on which topic areas are the most important. Input data, in the form of astronomy scientific papers, are readily available and include the prior 10+ years of scientific literature via the Astrophysics Data System (ADS)[[9]](#footnote-9), Decadal Survey panel reports, and the solicited Decadal Survey whitepapers. The AI algorithms can be used to reveal changes and trends in published research activity. The AI-developed models may be utilized to determine if they correctly predict recommendations of the 2010 Decadal Survey by applying them to the panel reports. Subject matter experts (SMEs) can then validate these comparisons. This process potentially offers a new, independent means to augment significant Decadal Survey processes and recommendations.[[10]](#footnote-10)

***Test Case 5a: Can AI Identify New, Previously Unidentified, Science Topics/Areas?***

The purpose of this test case is to find gaps in the funding of science topic areas. Similar to test case 9a & 21b (above), it will utilize available literature from several sources, such as the ADS, Web of Science, Google Scholar, and the NASA Scientific and Technical Information (STI) Repository (NTRS). As before, advanced NLP may be used to pull text features from the gathered science and technology papers and use these features to identify scientific/technical topics.

These derived models could then be applied to funded NASA research grants to determine their topic representation. Those topics which have sparse representation (e.g., those with few research grants associated with them) may represent underfunded opportunities. Human SMEs could examine these underrepresented topics and determine whether they are relevant to NASA priorities, or align with NASA strategic planning, and whether they represent areas of potential future research opportunity. The value to the customer is potentially high, as it can result in identifying important science and technical work that would otherwise be underfunded, or unfunded altogether. Validation is performed by human SME evaluation, as well as analysis tools with graphical interfaces. A knowledge graph, a network diagram that quantifies the relationships between concepts such as science and technical research areas, could be constructed for the prior 10 years of modelled topics. The priority for discovered ‘missed opportunity’ topics could be gauged based on how closely they are associated with other topics of known strategic importance to the agency (as determined by human SMEs with an understanding of strategic priorities).

***Test Case 5c, 29a & 29c: Can AI Improve Keyword Assignment for Scientific Papers and Proposals?***

These merged test cases involve studies of how well people choose keywords for their research papers or submitted proposals, versus how AI might choose them, presumably with less bias. The central question is: *Do the AI-generated journal keywords reflect the research topics better than the human-chosen keywords?* This study has readily available data and can utilize abstracts and associated keywords available from the ADS. As has been mentioned in other test cases, NLP may be used to extract significant terms from the abstracts and used to group them. Using these groups, the most common human keywords may be compared to the most common AI-identified terms, to judge performance of human versus AI term-labeling. In the case of proposals, human SMEs would have to prepare a list of suggested keywords to label/group the proposals. Access to relevant proposals may be challenging, but an external source of proposals, such as those available from the National Institutes of Health (NIH), could be substituted for an early study. In either case, the value of improving labelling includes increased discoverability of relevant information, and the possibility of identifying emerging fields, and/or cross-disciplinary fields, of interest. The suggested technical approach is well-established and feasible, although access to proposal data is more difficult than for scientific papers.

The value to the customer for this is high. Done well, this should reduce work for program officers and panelists, and potentially uncover missing subjects/trends; which promotes a better understanding of submissions and better guards against panelist bias. Metrics could be kept on approvals (by topic) to examine the sentiments of a panel with regards to particular topics, which may/may not be justified. Furthermore, effectively tagging the content of proposals can help program officers assign proposals to the proper panels, thereby reducing workload.

Finally, it would assist in NASA communicating investments across different disciplines in a measurable and systematic way and allow identification of proposal pressure and burgeoning themes that might be missed in the push to organize panels.

***Test Case 13a: How Can AI Enhance Current Observation Capabilities Using New Machine Learning Techniques?***

Astronomical observatories are critically important to the scientific community because the demand for telescope time far exceeds availability, efficient use of time is paramount. As an example of the problem, consider that for the Hubble Space Telescope (HST), some 10,000 to 30,000 observations are scheduled per year, and each is subject to a large number of operational and scientific constraints. A variety of constraints on exposures are required in order to achieve desired scientific goals. HST scheduling thus is a constraint satisfaction problem (CSP) consisting of a set of variables, each with a domain of discrete values, and a set of constraints. The Space Telescope Science Institute (STScI) has already developed SPIKE[[11]](#footnote-11), an AI-driven tool, to handle HST scheduling issues.

When one considers how these observations may be scheduled to provide simultaneous observations, by other missions or ground-based observatories, to enhance our scientific return, the problem becomes even more acute. Are we getting the best science from the current ensemble of observational capabilities? Are we missing opportunities? Can we gauge the scientific value of transient or novel phenomena quickly enough to capture the needed data by multiple telescopes / spacecraft and uncover new exciting scientific data?

This test case proposes the development of new AI techniques to improve the efficiency of mission observations, over what can be currently achieved, when considering multiple outside constraints such as other missions or priorities, for various types of astrophysical phenomena/observables. Data for this test case might be obtained by utilizing the operational observing plans, orbital ephemerides, and related scheduling data, for 2 or more science missions with complementary science goals. These constraints might be paired with a list of priorities for astrophysical phenomena, such as ”maximize observations of transients in the plane of the Milky Way”, along with catalogs of actual activity in the sky, and estimated times of these phenomena. A deep learning solution, building off of experience with SPIKE and related software, can be developed to handle this list of expanded constraints. SME expertise can be used to determine if the new software solution can better identify these opportunities for observations by multiple missions. The value to the customer is clear; expanded capability to maximize the science return for phenomena which are best observed by multiple missions.

***Test Case 5b & 2/27c: Can AI Identify People Who Have Technical Expertise in a Particular Area/Topic?***

### 

This test case is a merger of two other test cases which have the common necessity of finding people with a particular technical/scientific expertise. Finding SMEs is critical for any technical field where the latest knowledge is not fully captured in the literature/documentation/etc. Access to the most up to date knowledge prevents wasteful duplicate effort, or fruitless investigation. A knowledge graph of associations between people, capabilities, projects, publications, etc., could be used to locate SMEs for this test case. Technically this is challenging, but potentially can be accomplished (at minimum, initially, by manual construction). One direction which could be pursued, is to use these graphs to search for and locate people with the strongest associations to the types of data representative of a given field of knowledge, which appears in the knowledge graph. Additionally, with the right metadata captured, the graph could be used to identify and side-step selection bias, and thereby obtain greater diversity from a pool of qualified individuals.

In order to be practical, a knowledge graph would have to be generated at least semi-automatically. AI may help do this, and we may also look to advanced NLP (as mentioned in earlier test cases) to extract key phrases from scientific text. They can then be combined with other associated metadata, such as authors, journal ownership, etc. Mined phrases and other metadata may be used to create a graph of knowledge for the document. These sub-graphs, in turn, may then be combined, based on common nodes and associations, to bootstrap a larger knowledge graph. Alternatively, or in conjunction with this, a “seed” knowledge graph might be used to indicate the most important/interesting associations between sub-graphs and facilitate the creation of the final working graph. An alternative approach to knowledge graphs might be to use the aforementioned NLP, along with clustering techniques on the same datasets, creating topic models which coincide with desired knowledge areas, and looking at the paper authorship of documents that fall within one or more relevant topics.

In both cases, Google Scholar, Web of Science, and ADS could all provide papers by researchers. Additionally, NTRS holdings could be leveraged as well. The value to the customer is high, as it would potentially be helpful for people crafting panels, workshops, and other collaborative efforts. Bias may still exist (based on who is publishing more, for example), but this is likely an improvement over other known possible biases, such as gender or race. This approach would be easily explainable and could be tested and validated for a number of subject areas by SME planners to determine if the process did indeed expand the pool of knowledgeable experts.

## **Test Case Analysis**

To better understand these results, the organizing committee reviewed the best test cases, and looked at the type of algorithm/approach the test case would be utilizing, followed with a post-workshop discussion of “Testability” (see ***Table 3*** below).

We examined the expected outcomes of each test case and classified them as follows: “Discovery” means that the test case would reveal new knowledge; “Classification” signifies that the test case would group or classify existing knowledge for greater understanding; “Process Improvement” for when the test case would improve either the speed or efficiency (or both) of an existing process; and “Prediction” to indicate that the test case would attempt to forecast either the past or future.

Testability was rated simply as “Testable,” “Potentially Testable,” or “Not Testable”. “Testable” was assigned if the majority of the reviewing organizing committee members could agree on the test case being testable. “Potentially Testable” was assigned if some of the organizing committees thought the test case could be tested. After review, none of the top test cases were thought to be “Not Testable.”

**Table 3. Analysis of Best Test Cases**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Application Type** | **Technical Type** | **Testability** | **Overall Score** |
| 9/25a & 21b | Prediction | Text Analytics | Testable | 11 |
| 5c, 29a & 29c | Classification | Text Analytics | Testable | 10.7 |
| 5a | Discovery | Text Analytics | Potentially Testable | 10.7 |
| 13a | Process Improvement | Non-Textual | Potentially Testable | 10.3 |
| 5b & 2/27c | Classification | Text Analytics | Testable | 9.5 |

There was significant similarity in the technical approach of the 5 best test cases. The majority of the test cases relied on some use of text analytics (particularly NLP) to extract features (relevant phrases) from text, and to use these to better understand documents of interest (input sources by test case include 9/25a & 21b: Decadal Survey whitepapers and panel reports; 5b, 2/27c: finding experts from technical/scientific papers; 5a: proposals; 5c, 29a & 29c: scientific papers). Only Test Case 13a did not involve using textual data as an input for the AI.

# Observations

The workshop was considered productive, as all of the stated goals of the workshop (section 2) were achieved. Participants generated over 30 challenges (Appendix B), identified ten key challenges (***Table 1***), and created a catalog of 24 test cases (Appendix C) from which several ‘“best” test cases were identified (***Table 2***). An informal poll of the participants at the end of the workshop asking *“How likely is it that AI can be applied to augment activities in science prioritization and strategic planning? Will one or more applications will be feasible?”* resulted in ~85% of the 14 respondents indicating a positive response of “Likely to Highly Likely”.

The virtual meeting format, with truncated hours designed to draw in participants across eight or more time zones, was challenging. Given the amount of generated material over the 10 hours of the workshop, one of the real difficulties was being able to spend adequate time reviewing and refining the material. Examination of the challenges and related test cases showed that for a significant number there was some duplication in thought, and that the material could have been discussed more fully by the workshop.

Nevertheless, we can make several observations about the results of the meeting. It is clear that the workshop attendees had little difficulty finding interesting challenges which AI could tackle in the areas of strategic planning and science prioritization, and we were able to distinguish several key challenges from this population. There is every reason to suspect that, given more time to ideate and discuss, additional challenges would be uncovered. Just this aspect of the workshop indicates that the useful application of AI is clearly an area that it would be timely for NASA to investigate further, to aid the agency.

Analysis of the best test cases (i.e., those highest-rated) indicates that text analytics/NLP often underpins the technical approach. This is unsurprising given that much of the “data” for strategic planning and science prioritization are in the form of documents, rather than scientific data products such as images, spectra, or tabular data. What is interesting is that while there is a common technical basis for these best test cases, there are a broad range of potential applications to solve problems in strategic planning and science prioritization. From this information, we then conclude that text analytics is a valuable, multi-purpose tool in this area.

Examination of session notes, and a debriefing discussion by the organizers, also indicated several generally agreed-upon thoughts by participants, which included:

**1)** General AI usage themes. AI can be used for: a) literature surveys for forecasting promising research areas, and b) summarization and classification of data. A significant minority also felt that there was potential for improved understanding of costs.

**2)** AI cannot be simply a “black box.” The research/planning community must understand it in order to trust it. There must be some validation of the AI process/techniques to foster that trust.

**3)** Most participants agreed that AI is a tool to assist human planners, not replace them. Much of the discussion the first day focused on the key point that, ultimately, humans are the decision-makers, not any “algorithm”. AI is a useful tool, but human scientists (and managers) have the insight and responsibility to consider all aspects and points of view in the strategic planning process.

**4)** Biases exist in the current planning processes but can also exist in AI processes due to the introduction of biases during the training process. This may occur either from the biased selection of input training data, or biases inherent in said data. It is vitally important to recognize this and be aware that biases may be unintentionally introduced into the AI process. There are known techniques for dealing with bias in AI and they should be employed.

**5)** The majority of participants favored the investigation of the use of AI for science priority/planning, which was also evident in the exit poll, as mentioned above. As with anything new, it will take time to prove the value of AI in this context and communicate with the science community to demonstrate that worth and added value. Cultural barriers may be the greatest difficulty, which these techniques need to overcome. Outreach to the community should include cases that verify the utility of AI, without overpromising any particular results.

**6)** The test cases show, and establish, the existence of several opportunities for NASA to conduct short-term studies aimed at demonstrating the usefulness of AI/ML for assisting strategic planning.

# APPENDICES

## 

## **Agenda**

Workshop will be preceded by three pre-meeting web conferences for the participants to discuss background, objectives, and logistics. Discussion and Breakout Sessions will be governed by the modified Chatham House Rule.

**Two Days, 9:30 AM - 3:30 PM US ET (2:30 - 8:30 PM UK)**

**DAY ONE, MAY 12**

|  |  |  |  |
| --- | --- | --- | --- |
| **Login/Socialization** | Informal welcome from Brian Thomas , NASA (Chair) | 9:30 - 10:00 AM | Thomas to greet participants |
| **Introduction**  Welcome/etiquette, instructions. opening Q&A  **Charge to Workshop: What Current Problem Needs to be Solved?**  Requirements/challenges of the customer, goals, deliverables | Alison Lowndes, NVIDIA (10 min)  James Green, NASA (5 min)  Paul Hertz, NASA (5 min)  Brian Thomas, NASA (5 min) | 10:00 - 10:30 AM  (30 min) | Moderator: Thomas  Q&A Wrangler: Thronson  Note-taker: Barbier |
| **Current NASA Strategic Planning and the Academies’ Decadal Surveys**  History, process, deliverables | Rita Sambruna, NASA (20 min)  Heidi Hammel, AURA (20 min) | 10:30 - 11:30 AM  (60 min, w/20 min Q&A) | Moderator: Thomas  Q&A Wrangler: Thronson  Note-taker: Barbier |
| **Examples: Current Application of AI for Prioritization** | Anamaria Berea, GMU (15 min)  Ryan Zelnio, ONR (15 min) | 11:30 - 12:00 PM (30 min, w/10 min Q&A) | Moderator: Crooke  Q&A Wrangler:  Lowndes  Note-taker: Varsi |
| **Current AI Explainability Techniques** | Yarin Gal, Oxford (15 min) | 12:00 - 12:15 PM (15 min, w/ 5 min Q&A) | Moderator: Crooke  Q&A Wrangler:  Lowndes  Note-taker: Varsi |
| **Current AI-Enhanced Technology**  Applications which may be relevant to science and tech prioritization | Scott Penberthy, Google  (15 min)  Animesh Garg, NVIDIA/University of Toronto (15 min) | 12:15-12:45 PM  (30 min, w/10 min Q&A) | Moderator: Crooke  Q&A Wrangler:  Lowndes  Note-taker: Varsi |
| **REFRESHMENT/MEAL BREAK (30 minutes) - 12:45 - 1:15 PM** | | |  |
| **Plenary Discussion of Breakout Session Topics - Challenges**  Identify challenges in which test cases can be developed. | Discussion Leaders (4): Ho (Flatiron Inst.), Leisawitz (NASA GSFC), Spergel (Flatiron Inst.), Hammel (AURA) | 1:15 ~ 3:15 PM | Moderator: Wright  Q&A Wrangler: Diamond  Note-Taker: Memarsadeghi |
| **Organize Sessions for Day 2**  Which “Challenges” for which team? | ALL | 3:15 - 3:45 PM | Moderator: Thomas  Q&A Wrangler: Barbier  Note-Taker: Marquad |
| **After Hours General Discussion** | ALL | 3:45 - 4:45 PM | Moderator: Thronson  Q&A Wrangler: Barbier  Note-Taker: Aziz |

**DAY TWO, MAY 13**

All Times US Eastern

|  |  |  |  |
| --- | --- | --- | --- |
| **Login/Informal Discussion** |  | 9:30 - 10:00 AM | Moderator: Thomas  Q&A Wrangler: Varsi |
| **Review of Day 1**  Charge and Instructions to the Breakout Sessions: Catalog of test cases, Q&A | Alison Lowndes, NVIDIA & Brian Thomas, NASA (30 min) | 10:00 - 10.30 AM | Moderator: Thomas  Q&A Wrangler: Varsi  Note-taker: Thronson |
| **Breakout Sessions (4 Teams)**  Examine test cases: technical feasibility, data available, explainability as critical criteria | Separate breakout “rooms” (120 min) | 10:30 - 12:30 PM | Moderators: A->D : Thomas(A), Crooke(B), Barbier(C), Wright(D)  Q&A Wranglers: Thronson(A), Varsi(B), Diamond(C), Lowndes(D)  Note-takers: Brian and Aziz (A), Marquard (B), Martin (C), Samadi (D) |
| REFRESHMENT/MEAL BREAK (30 minutes) 12:30 - 1pm | | |  |
| **Quick-Look Presentation of Breakout Sessions**: feedback, Q&A | Each team (~10 min) | 1:00 - 1:30 PM | Moderator: Crooke  Q&A Wrangler: Diamond  Note-taker: Varsi |
| **Summary Session**  Key observations and findings: core of written report and presentations, next steps | Each team (~30 min); 30 min general discussion | 1:30 ~ 3:15 PM  (exact time TBD) | Moderator: Thomas  Q&A Wrangler:  Memarsadeghi  Note-taker: Varsi |
| **After Hours General Discussion** | ALL | ~3:15 - 4:15 PM | Moderator: Thomas/Thronson |

## **Catalog of Challenges**

These are the verbatim collected challenges which arose from the first day of the workshop ideation. Our process worked quickly to capture thoughts and then return to them for critical evaluation. Those challenges ranked as most interesting/important by participants are noted as “[High Rank]”. During deliberation, participants identified a number of reasons for dismissing a challenge; reasons that included criteria such as incomplete/confusing description, or modest value of the challenge for the agency. Some challenges were later merged and their id’s combined (ex. Challenges 2 and 27 merged and subsequently identified as “2/27”).

1. How to understand the politics around science mission choices?

Politics is an important part of which science is prioritized, and it would be useful to quantify in some manner how likely science will be viewed by politicians, public.

1. **[High Rank]** Identify appropriate subject matter experts for service on proposal reviews and Decadal Surveys (and equivalent). Use AI to select for further consideration likely candidates for selection for panels based on a set of desirable criteria: e.g., breadth of demonstrated research, specialties of research, demographics, etc.
2. **[High Rank]** Synthesize and analyze research evolution. How to understand and visualize the evolution (e.g., growth, decline) of research areas, for example, in number of papers published and other means? How to also observe evolution of techniques and methodologies within these research areas?
3. Perhaps the Hubble double-blind reviews would provide an interesting test data set /challenge. I believe multiple other proposals are also doing this now. I know in Space Physics there are some journals which are also going to double-blind reviews. Perhaps those datasets could be compared with other traditional peer reviewed publications.
4. **[High Rank]** What biases exist in scientific planning/prioritization?
   1. How to overcome / avoid unconscious bias in various parts of the process.
   2. What bias are planners aware of? Not aware of? What are the best ways to quantify bias for scientific planners?
5. How can we educate the stakeholders to the value of AI for strategic planning so they can trust it? This is essential as one of the key properties of Decadal Reviews is the weight they carry with legislators, and that *gravitas* can’t be compromised. What techniques in AI are capable of being validated to the satisfaction of the consumer?
6. How to guard against spurious topical categories that might be generated using AI techniques? Perhaps via human-in-the-loop, semi-supervised clustering targeted to well-defined areas.
   1. Sub-Challenge: At the same time, how does one ensure that important relationships between seemingly disjoint subfields are not ignored, as these may be emerging cross-disciplinary research areas? Perhaps explanation-driven ML that explains why two seemingly orthogonal topics are clustered together.
7. How do we ensure voices feel heard? Obtain consensus in the community, validate results?
   1. Consensus is unlikely to be generated by AI this decade, human consensus requires all in the community to “be heard”, and humans do not derive much satisfaction from being heard by AI. If algorithms are to be used to select participants, or inform analysis, it must preserve this idea if it is to not undermine consensus. The usage and history of entrofy in astronomy provides a case study for an algorithmic approach to representation, including both successes and pitfalls: [[1905.03314] Entrofy Your Cohort: A Data Science Approach to Candidate Selection](https://arxiv.org/abs/1905.03314)
8. **[High Rank]** How to automate the extraction, classification, and prioritization of scientific themes and investigations for Decadal Surveys, key conferences, etc., and compare against NASA science roadmaps. [Combine with 3, 21, and 25?]
   1. This challenge wants to use a wide range of data sources: published materials (papers), websites, Twitter, etc., which will use different terms for the same object/item (e.g., semantics).
9. **[High Rank]** How to understand how much funding NASA is putting into various research areas and whether we should increase/decrease the amounts based on Decadal priorities? Use AI, for example, to evaluate over time the increase in published papers or breakthroughs as a function of funding.
10. **[High Rank]** How to understand what mechanisms of funding are best value for achieving science priorities?
11. How to assess the value of the data needed for AI? What techniques/tools for visualizing, preparing, and sharing. Repository of solutions, techniques, visualizations? Data may come from a variety of sources each of which may have a different ‘data dictionary’ which make it hard to combine and obtain insights. Compare the science priorities, investment strategies, etc. produced by different data sources.
12. **[High Rank]** How can AI be used to enhance the capabilities of missions/projects?
    1. For example: How might we accelerate space missions by pushing AI to the edge on spacecraft or on-instrument: reduction of time to insight, semi-autonomy, channel optimization, etc.?
13. Consider using reinforcement learning if we can characterize the reward properly? Also, is the “loop” in HIL really a loop or more line-like?
14. Engage humans further to improve AI training. Use crowdsourcing and citizen science approaches (e.g., Zooniverse) to develop training models for automated AI for classification.
15. How can we effectively automate some of the more mundane administrative tasks that can free strategic planners, scientists and others whose available time to perform important work is at a premium? AI can be employed to support the development of a recommendation engine/digital assistant capability.
    1. Examples: conducting equipment inventories, ascertaining status of expired IT asset property passes, managing the submission of various Agency forms that can be completed by pulling information automatically from various siloed Agency databases, etc. Scheduling meetings among multiple participants in multiple time zones.
16. What non-textual information can be utilized to aid decision-making for science prioritization? For example: Images, audio and video. Perhaps non-textual information can be used to augment textual data information/modelling or may be used standalone.
    1. Look at Newton’s work on Cumulative Prospect Theory for Reinforcement Learning (RL): <http://proceedings.mlr.press/v48/la16.pdf> for modelling human decisions + ICLR[[12]](#footnote-12) Retweet judgment work by Animesh (reputable people benefit more from social media publication)
17. What are the best proving grounds? The next astro decadal is in 10 years; if we want AI tools to be used then, they must be proved to the community in smaller arenas. I.e., what “missions” will allow us to elevate AI-for-strategy-at-NASA’s “TRL”?
18. **[High Rank]** More bang for the buck? How effective are various approaches of research? What is the right mix to maximize science return? More funding for theory versus experimental/observational research versus technology? Compare breadth of “impact” of different areas of research.
19. Use AI to identify technology priorities that have greatest impact and greatest potential to impact science priorities? Identify the best technologies to fund? Questions: what are the right choices in technology development to optimize science return? Should technology available (or soon available) drive science priorities or should we plan science priorities based on other criteria being more important? What technology areas are accelerating? [Combine with 19?]
20. **[High Rank]** Can we use data from over a decade ago to go back and predict today’s hot science topics? Like gravitational waves or “are we alone”? Example: “Backcasting” AI on the Astro2010 Decadal Survey to assess what it would have recommended for the past decade. [Combine with Challenges 3, 9, and 25?]
21. How might we use AI techniques to make the results of science research more approachable/understandable to the decision makers and public, without sensationalization. For example, what science areas have greatest public impact?
22. **[High Rank]** How can we pick the optimal science missions for given funding/resources? May include criteria as available technology, expertise, risk, etc.
23. How can we optimize surveys? What objects are most interesting to look at? Can we in principle use active learning/reinforcement learning to understand what objects/areas of sky are most interesting to observe with next mission? Can we use AI to define “interesting”? What is “interesting”?
24. **[High Rank]** How can we best assess the past decade of science to determine the best/key science goals? What are best data sources? For example: Can we find gaps in submitted whitepapers to Decadal Survey? Compare an analysis of whitepapers vs prior literature to find gaps (e.g., Hammell pointed out no magnetic whitepapers submitted to previous Decadal Survey).
25. Can we predict what is missing from current scientific research/missions? What are the gaps in missions compared with literature? Brand new wavelength, research area(s) (e.g., gravitational waves)? Where are the best places to investigate?
26. **[High Rank]** Can we use Natural Language Processing methods to match reviewers with right proposal/ white papers? Can we increase the number of reviewers without loss of quality using NLP/other methods? [Combine with 2 and 8?]
27. How can we ensure trust in AI “predictive” or “selections” or “supportive information for decision makers” (i.e., do not want garbage in, garbage out)? How do we ensure that the information going in is high value and accurate?
28. **[High Rank]** How can we better tag/keywording/classify for proposals?

Effectively keywording proposals can help program officers reduce the amount of up-front work for program officers organizing panels. In addition, it would assist in NASA communicating our investments across different disciplines in a systematic way, and allow us to identify proposal pressure and burgeoning themes that might be missed in the push to organize panels. All things that should be used to identify and communicate priorities for decadal surveys.

1. The National Academies have about 20 categories of information it maximizes to prioritize choices. What role could AI play to aid human selection/use of these data?

Based on David Spergel’s excellent point in the chat session, “DS: The panel selection process currently involves maximizing the following”:

* 1. expertise in relevant areas (X-ray astronomy, cosmology,)
  2. regional diversity
  3. members from large public universities
  4. government labs
  5. private universities
  6. small universities
  7. gender diversity
  8. range of career stage
  9. ability to listen to others
  10. willing to do the work
  11. Works well with others
      1. The National Academies have about 20 categories and tries to look at all of the elements in the selection.
      2. Given this, is there a role that AI could play in this? Can AI Address Some of These Aspects?

1. **[High Rank]** How can AI be used to spot and mitigate emerging problems and guide us to solutions based on lessons learned from past missions, projects, programs, and including lessons from across organizational boundaries? [See 3, 9, 21, and 25]
2. How can AI be used to conduct better-informed risk trades?
3. How can AI be used to explore multiple possible futures? (What if? scenarios)
4. Extrapolate and apply lessons across organizational boundaries, from program to program, and from project to project

## **Catalog of Test Cases**

**Table 4. Catalog of Test Cases**

|  |  |  |  |
| --- | --- | --- | --- |
| Challenge ID | Test Case | Teams | Descriptions |
| 5 | a | A, D | Can AI identify new (previously unidentified) science topics/areas for prioritization/funding? |
|  | e | C | How would varying proposal peer review panel composition affect selection outcome? |
|  | f | C | Identify undesirable bias in proposal peer review |
| 9/25 | b | A | Compare last planetary decadal survey whitepapers (201?) vs NASA science roadmap (for planetary) |
|  | c | A | Compare last earth science decadal survey whitepapers (201?) vs NASA science roadmap (for earth science) |
| 10/19/23 | a | D | Compare scientific return of Balloon vs Sounding Rocket missions |
| 13 | a | B | Use AI to enhance current observation capabilities using new machine learning techniques |
|  | b | B | Identification of AI to enhance mission capabilities |
|  | c | D | Can we improve efficiency of Hubble scheduling and operations? |
|  | d | D | Can we look at missions/grant proposals data to identify the semantic content of the proposals with a goal of identifying what missions predicted for their most important science outcomes? |
|  | e | D | Onboard AI Autonomy |
| 2/27 | b | B | Can AI provide better proposals for review for a panel? |
| 29 | a | A | Look at the NASA proposals submitted (to a specific ROSES[[13]](#footnote-13) call; submitted TOTAL, not just awarded) and compare AI results of suggested keywords (subjects) for panels to actual keywords (subjects) panels which were created. |
|  | b | A | Combine keywords from 2 different domains (communities) to automate understanding of intersections between domains and uncover proposals which should be considered. |
|  | c | C | Keyword assignment/classification of topics in proposals. |
| 31 | a | C | Could the Columbia accident have been foreseen (by AI)? |
|  | b | C | Could the Apollo 1 fire have been foreseen (by AI)? |
|  | c | C | Hubble mirror anomaly - would AI have caught the problem? |
| *~~~~ Merged Test Cases ~~~~* | | | |
| 5c, 29a & 29c | -- | A, C | *Can AI improve keyword assignment for scientific papers and proposals?* |
| 2/27 a  &  5d | -- | A, C, D | How do we gauge the research performance of an award? / Look at what was proposed vs. what came out in the literature / Compare missions (Balloon vs Sounding Rocket) |
| 9/25a & 21b | -- | A, C | Compare last astrophysics decadal survey whitepapers (Astro2010) vs NASA science roadmap (for astrophysics)/ Backcasting of the Astro2010 DS.  Predict breakthroughs of 2020 using data from 2010. |
| 2/27 c  & 5b | -- | A, B, C | Use AI to optimize (diverse) the selection of panelists for review panel / Compare peer panel structure in historical records (human-generated) with AI-suggested structure based on submitted proposals and the solicitation. / Can AI suggest experts to identify people which have technical expertise in a particular area/topic. |

## **Participants**

|  |  |
| --- | --- |
| **Attendee Name** | **Organization / Title** |
| Acromazzi, Alberto | Harvard-Smithsonian CFA / Program Manager Astrophysical Data System |
| Altaf, Naeem | IBM / CTO for Space Tech Industry |
| Ansdell, Megan | NASA HQ/ Program Scientist |
| Berea, Anamaria | George Mason University / Associate Term Professor |
| Brandt, Terri | NASA GSFC / Senior Scientist |
| Campbell, Newton | SAIC / Senior AI team lead |
| Carrol, Mark | NASA GSFC/ Lead Data Scientist Code 606 |
| Cheung, Mark | LMSAL Heliophysicist/ FDL Mentor |
| Crichton, Dan | NASA JPL / Data Scientist |
| Dorr, Bonnie | Florida Institute of Human and Machine Learning Cognition |
| Fong, Terry | NASA ARC / Director of Intelligent Robotics |
| Gal, Yarin | University of Oxford / Associate Professor of Machine Learning |
| Garg, Animesh | University of Toronto / Assistant Professor |
| Guharthakurta, Madhulika | NASA HQ / Heliophysics |
| Halford, Alexa | NASA GSFC/ Heliophysics |
| Hammel, Heidi \* | AURA, Vice President for Science |
| Hertz, Paul | Director of Astrophysics/NASA |
| Ho, Shirley \* | Group Leader, Cosmology X Data Science, Flatiron Institute |
| Hughes, Peter | NASA GSFC / Chief Technology Officer |
| Jain, Shashi | Intel / Strategic Innovation Manager |
| Leisawitz, David \* | NASA GSFC /Astrophysicist and Chief of Science Proposal Support Office |
| Martin, Rodney | NASA ARC / Data Scientist |
| McCauley Rench, Rebecca | SMD/Planetary Strategic Planner |
| Menard, Brice | Johns Hopkins University/ Associate Professor |
| Oza, Nikunj | NASA ARC Data Scientist, Lead of Group |
| Peek, Joshua | STScI / Associate Astronomer |
| Penberthy, Scott | Google / Director of Applied AI |
| Powell, Brian | GSFC / Data Scientist |
| Prockter, Louise | Lunar Planetary Institute / Director |
| Sambruna, Rita | SMD/Astrophysics Strategic Planner |
| Scannapieco, Evan | SMD/Astrophysics Strategic Planner |
| Seablom, Mike | SMD / Chief Technologist |
| Sidehammer, Ted | NASA LaRC / Data Scientist |
| Soderstrom, Tomas | NASA JPL / Data Technologist |
| Spergel, David \* | Flatiron Institute / Director, Center for Computational Astrophysics |
| Thompson, Meagan | SMD/Planetary Strategic Planner |
| Zelnio, Ryan | Office Naval Research/Chief Data Scientist |

\* Session chair

## **Sponsors, Partners & Contributors**

* 1. NASA: SMD, OCIO, OCT, OCS
  2. SETI Institute
  3. NVidia
  4. Google
  5. Flatiron Institute
  6. Universities
  7. AURA/NSF

## **Organization Committee**

Huzna Aziz (NASA HQ)

Louis Barbier (NASA HQ)

Julie Crooke (NASA GSFC)

Bill Diamond (SETI Institute)

Alison Lowndes (NVIDIA)

Erica Marquard (HQ-CM000)

Nargess Memarsadeghi (NASA GSFC)

James Parr (Frontier Development Lab)

Shahin Samadi (INNOVIM)

Harley Thronson (NASA retired)

Brian Thomas (Chair, NASA HQ/GSFC)

Giulio Varsi (NASA retired)

Kenneth Wright (NASA HQ)

## **Acronyms**

|  |  |
| --- | --- |
| ADS | Astrophysics Data System |
| AI | Artificial Intelligence |
| BERT | Bidirectional Encoder Representations from Transformer |
| CSP | Constraint Satisfaction Problem |
| ICLR | International Conference on Learning Representations |
| LDA | Latent Dirichlet Analysis |
| ML | Machine Learning |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| RL | Reinforcement Learning |
| ROSES | Research Opportunities in Space and Earth Science |
| SME | Subject Matter Expert |
| SPIKE | Not an acronym |

## 

1. https://arxiv.org/abs/2001.04451 [↑](#footnote-ref-1)
2. https://cloud.google.com/ai-platform [↑](#footnote-ref-2)
3. https://azure.microsoft.com/en-us/services/machine-learning/ [↑](#footnote-ref-3)
4. https://aws.amazon.com/machine-learning/ [↑](#footnote-ref-4)
5. See publications at https://research.nvidia.com/publications [↑](#footnote-ref-5)
6. http://arxiv.org/pdf/1704.04760 [↑](#footnote-ref-6)
7. https://ieeexplore.ieee.org/iel7/9183768/9188050/09188164.pdf [↑](#footnote-ref-7)
8. Availability of logical or scientific explanations of the results produced by the AI. [↑](#footnote-ref-8)
9. Astrophysics Data System, abstract service https://adsabs.harvard.edu [↑](#footnote-ref-9)
10. Preliminary work on this test case reported at [**https://aas237-aas.ipostersessions.com/?s=A9-A3-DA-5A-0C-CA-90-16-9C-2A-FC-A5-BD-64-A0-55**](https://aas237-aas.ipostersessions.com/?s=A9-A3-DA-5A-0C-CA-90-16-9C-2A-FC-A5-BD-64-A0-55) [↑](#footnote-ref-10)
11. https://www.stsci.edu/~miller/papers-and-meetings/93-Intelligent-Scheduling/spike/spike-chapter3.html [↑](#footnote-ref-11)
12. International Conference on Learning Representations, see <https://iclr.cc/> [↑](#footnote-ref-12)
13. Research Opportunities in Space and Earth Science [↑](#footnote-ref-13)