NRT-WoU: Team Science and Science Communication for AI and Multimessenger Astrophysics

1 List of Core Participants

Role	Name	Department, Institution	Speciality
PI	Prof. Daniel Whiteson	Physics & Astronomy, UCI	Physics, Commun.
Co-PI	Prof. Pierre Baldi	Computer Science, UCI	AI, Deep Learning
Co-PI	Prof. Simona Murgia	Physics & Astronomy, UCI	Astrophysics
Co-PI	Prof. Steve Barwick	Physics & Astronomy, UCI	Astrophysics
Co-PI	Prof. Maritza Salazar Campo	Organizational Behavior, UCI	Team Science
Key Per.	Prof. Steph Sallum	Physics & Astronomy, UCI	Exoplanets
Key Per.	Prof. Stephan Mandt	Computer Science , UCI	AI, Gen. Models
Eval.	Jessica Martone	The Mark USA	Evaluation
Key Per.	Bri McWhorter, MFA	Activate to Captivate	Communication
Key Per.	Kara Ward	Access, Outreach, Inclusion	Diversity, Inclusion

2 Theme, Vision, Goals

2.1 Research Themes

Over the past few decades, physicists have come to an astonishing realization: most of the matter and energy in the Universe —fully 95%— is unknown to us, and is not comprised of the familiar particles and forces. The picture that has emerged is of a dark Universe, which until recently has been mostly hidden from our eyes and scientific probes. This reflects the enormity of our ignorance, and a tremendous opportunity to uncover the true nature of matter, energy, and space.

The dominant form of matter in the Universe is dark matter (DM), which is not made of any known particle and does not generate or reflect light, allowing it to remain hidden. Physicists have additionally discovered that we lack an understanding of most of the *energy* in the Universe, which is neither matter nor radiation, but some unknown form. This mysterious, yet dominant, component is accelerating the Universe's expansion rate, creating new space between galaxies, and ultimately dominating its fate. Physicists refer to it as dark energy (DE), but know nearly nothing about it, other than that it comprises most of the energy of the Universe, approximately 70%.

These mysteries sit in the context of an even deeper question, about the nature of space itself. Our modern theory of gravity, General Relativity (GR), has been eminently successful at describing space as a smooth fabric, capable of bending and wiggling and making black holes, but is in stark conflict with the quantum mechanical view that space is discrete, not smooth.

Making progress on these fundamental puzzles requires integrating our observational tools (photons, cosmic rays, and gravitational waves) into a holistic program of investigation, a multi-messenger network where each observation channel adds unique information and which together can spot signals from events which would be overlooked by one channel. The NSF has identified the era of *multi-messenger astrophysics* as one of its 10 Big Ideas and an area of high national research priority. In the context of this training grant proposal, we interpret astrophysics broadly, from analysis of telescope data to astroparticle experiments in space and on the ground, and including the full range of photons, from radio to visible to X-rays and gamma rays. All of these exciting topics are fertile ground for the training of students in the interdisciplinary research skills needed to unravel cosmic mysteries.

Analysis of astrophysical data has become increasingly complex, as modern observatories, complemented by particle collider experiments aiming to reproduce the conditions of the Big Bang, have become sources of vast quantities of digital data, which require sophisticated statistical tools to harness their full scientific power. At the same time, there has been a dramatic burst of progress in Artificial Intelligence, specifically deep learning [1]. This progress has led to breakthroughs not only in traditional areas such as natural language processing and computer vision, but also in the natural sciences, including particle physics, often increasing the statistical power of difficult-to-collect data [2]. AI enhances the power of physical sciences by, for instance, preserving information during reduction from high-dimensional raw datasets, learning powerful classification and regression models, learning fast solutions to otherwise slow computational calculations, or identifying anomalies in datasets. However AI cannot be applied blindly to physical data: it must be combined with domain expert knowledge. Hence the key to success is to create highly interdisciplinary teams that are passionate about both disciplines and about creating new knowledge and bettering the human condition. Importantly, when the power of AI has been brought to bear on a new set of scientific investigations, the unique nature of the statistical questions has required and sparked novel AI strategies. Thus, AI is not in service of physical science, but a full synergistic partner.

Clearly, similar opportunities exist in astrophysics, to optimize the analysis of data from telescopes or astroparticle detectors and better extract information from individual messengers: photons, cosmic rays, or gravitational waves. In addition, artificial intelligence (AI) methods may help tie these messengers together. By providing faster identification of the early signatures of a multi-messenger event, AI-powered triggers for a global alert system can direct other facilities to take advantage of rare observing opportunities. Astrophysical signals can benefit, individually and holistically, from the application of techniques from AI, and offer novel challenges which will spark innovation in AI.

However, application of AI for astrophysics and the development of novel AI solutions to the unique challenges of astrophysics is still significantly underdeveloped. A major challenge is that graduate education in astrophysics and AI has not yet provided the educational infrastructure and curricula to support development of the necessary suite of skills: basic knowledge of the foundations of the two fields, and interdisciplinary research skills in communication and team science. Students in astrophysics or AI tend to be siloed into distinct academic units, such that students in astrophysics are aware of the power of AI but lack the statistical or computational knowledge to know where to begin, and students in AI lack the knowledge of the underlying physics to devise appropriate statistical solutions. In addition, little attention is given to training in communication and team science, vital skills for effective collaboration between two fields that often use similar concepts but with separate vocabularies and notations due to their parallel, but separate, historical development. This gap in the educational program leaves students without the explicit training in the scientific, team science and communication skills needed to bridge the fields and capitalize on the opportunities at their interface. **A coherent training program at the interface of artificial intelligence and astrophysics, which provides students with the communication and team science skills to perform effective interdisciplinary work, is needed to foster and disseminate fundamental research advances and promote effective professional training for graduate students in this family of disciplines**. Developing a coherent training program meets a crucial area of need for graduate training, as graduate students are a main driver of astrophysical research.

2.2 Program Vision

The proposed NRT program will equip the trainees with basic knowledge of both fields, as well as the communication and team science skills to bridge the gaps between AI and astrophysics, allowing them to take advantage of opportunities for interdisciplinary and convergent science. We envision a sustainable program that will train primarily PhD students, as well as some Masters students, in the team science and communication skills needed to identify and capitalize on the rich interdisciplinary research opportunities at the intersection of astrophysics and AI, and to share the results and their excitement effectively with the public. The added value that a training program brings to current degree programs is that trainees will be explicitly educated in the scientific foundations of both

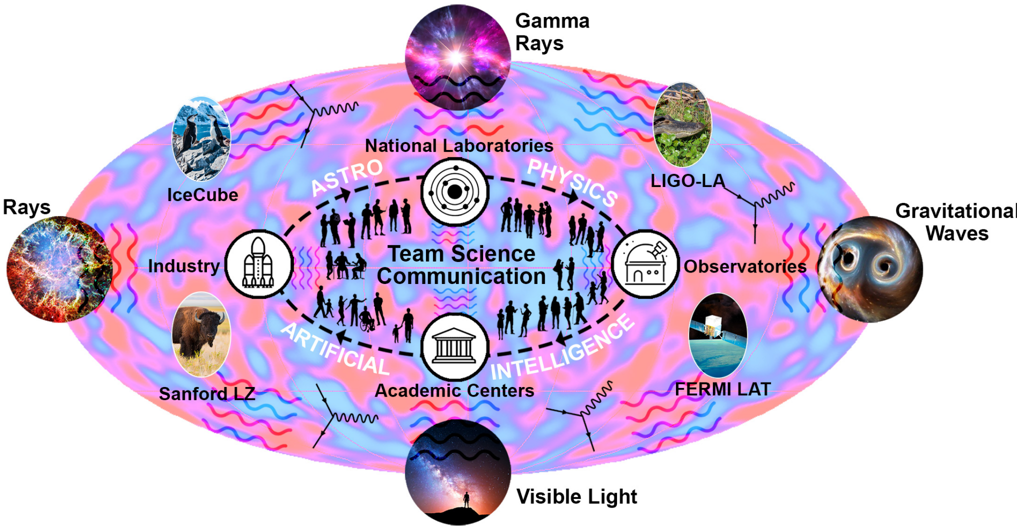

Figure 1: The program will embrace team science and novel science communication to create synergy between emerging artificial intelligence methods and exciting developments in multi-messenger astrophysics. Diverse students, faculty, industry and national lab collaborators will make breakthroughs on fundamental questions using visible light, gravitational waves, gamma rays, and cosmic rays generated by cataclysmic events and captured by observatories across the planet.

astrophysics and AI, giving them the background needed for intellectual investigation. The training program will also develop a rich pedagogy in the strategies of interdisciplinary work, explicitly teaching students how to explain their work to their colleagues in other areas, how to communicate the broad strokes to a lay audience, and how to work effectively in a team where each member plays a unique role. The training program will create a cohort of students who have the technical, interdisciplinary and leadership skills needed to identify and exploit new research directions and share the results broadly, with their peers and the public. An annual public event will feature students explaining their work to a general audience, using traditional (e.g., slides, chalk talks) and innovative media (e.g., cartoons, animation, dance).

Astrophysics students participating in the program will work on fundamental science problems that involve large and complex observational data sets and that demand significant algorithmic data-driven thinking and AI methodologies. Similarly, AI students participating in the program will work on new algorithms and techniques that are directly motivated by, and applicable to, large-scale scientific data generated in astrophysics. In this way, the program will foster interdisciplinary synergisms emerging from ongoing research in AI and astrophysics.

Evidence suggests that, in many cases, interdisciplinary science teams do not achieve the goal of successfully integrating knowledge [3]. Drawing on the insights from diverse scientific disciplines is often elusive due to differences in terminologies and approaches, the makeup of the collaborative teams, and lack of understanding about best practices for participating in and managing these collaborations. The consequence can be a costly investment in scientific endeavors that does not produce the expected benefits. Growing interest and investment in interdisciplinary research, and the need to understand factors that hinder or enhance interdisciplinary collaborative outcomes, have given rise to a rapidly growing field, the science of team science [4, 5, 6], which researches strategies to unify researchers with heterogeneous skills into an effective and harmonious team. For these reasons, team science will be at the foundation of our program, designed by leading researchers and practitioners in the field.

The communications element of the traineeship is built on the recognition that translating the motivation and challenges of research is essential, not just for sharing the excitement of a project with the general public and policymakers, but also to effectively connect with potential interdisciplinary colleagues [7]. Additionally, this traineeship is built on the foundation of storytelling, which is at the heart of all science and is the first step in scientific inquiry. Led by core member Bri McWhorter, an experienced professional communications specialist, and motivated by the PI's strong track record of public-facing [8, 9, 10, 11] and inter-department communication [12, 13, 14, 15, 16, 17, 18], students will learn to explain the motivation for their work in understandable terms, and to find the right words (or art or movements) to express the opportunities and challenges of their research. This will allow them to share the excitement of their work with their home communities as well as convey to their colleagues where there might be opportunities for synergy. Graduates of the program will contribute to research in AI and astrophysics, but will also be on the cutting edge of team science and science communication. Each trainee will be required to include a chapter on team science or science communication in their PhD Thesis.

UC Irvine has the unique set of skills and resources to execute this interdisciplinary program.

Faculty Research: UC Irvine is an international leader in the proposal's research themes, with core faculty who are widely recognized experts in astrophysics as well as in AI, with a strong track record of interdisciplinary work between them (e.g. Baldi with Whiteson [12, 19, 20, 13, 14, 18, 15], Murgia, or Barwick [21], Mandt with Whiteson [16]). UCI has broad strengths in astrophysics, AI, communication and team science beyond the PI and Co-PIs; see Table 2.

Team Science: UC Irvine has a unique team of faculty participants in team science. Prof. Salazar, a core participant, is a highly regarded thought-leader and researcher in the study of team science and technical collaboration. Together with the Team Science group at UCI, she has developed a variety of tools to measure and evaluate the capacity of interdisciplinary science teams to combine team members' diverse expertise effectively [22, 23, 24].

Communication: Bri McWhorter is a professional communications consultant, with many years of experience collaborating with academics. PI Whiteson has extensive experience in science communication, both inter-academic (with AI colleagues [12, 19, 20, 13, 14, 15, 16, 17, 25, 18], theorists [26, 27, 28], and astrophysicists [29, 30]) and with the general public (widely translated popular science books [8, 9], a highly-ranked physics podcast [10], and a science show on PBS [11]).

Previous experience: In a previous NRT program at UCI (MAPS), team science supported interdisciplinary work between AI and chemistry, earth sciences, and physics. In this proposal, we capitalize on the successful elements, but identify the areas of the MAPS program which were not as effective. While the team science component helped build a trainee peer group that fostered informal discussion and support, the MAPS program did not sufficiently emphasize communications training, which is a centerpiece of this AI/Astrophysics proposal. While MAPS sought to connect AI with a very broad set of physical science domains, it was less effective in earth sciences and chemistry than in physics, where the data are more vast and the questions are more often phrased in statistical terms. To improve on the MAPS program, we emphasize communication skills and align this proposal with the NSF's research theme of astrophysics, where UCI has internationally recognized strength in neutrinos, gamma-rays, dark matter, black holes, neutron stars, galaxy formation, and planetary formation, and strong connections to gravitational wave programs at the nearby campus of California State University Fullerton (CSUF).

2.3 Program Goals

Goal 1: Produce transformational astrophysics research and innovative new AI algorithms. Aim to accelerate research at the interface of astrophysics and AI, generating new scientific knowledge about black holes, dark matter, neutron stars, exoplanets, etc., as well as new AI methodologies to be broadly applied to other scientific domains.

Goal 2: Increase the training of a diverse group of graduate students with transferable science communication and team science skills. The program will produce a new workforce of graduate student researchers who have the skills to identify and capitalize on opportunities to produce research breakthroughs at the interface of astrophysics and AI, in support of

Goal 1. Students will be equipped with the skills to share the excitement of their work with their communities and the public, and to work effectively in heterogeneous teams in their future careers.

Goal 3: Build new institutional support for interdisciplinary research. Develop and expand institutional capacity to support the training of graduate students in this program, both during the program and sustainably beyond its end. We outline below specific mechanisms (a new Professional MS program, fellowships from the UCI Graduate Division) to ensure that the momentum from the NRT program will be maintained beyond the 5-year window of NSF funding.

3 Organization and Management

PI and Project Coordinator: The program will be directed by PI Whiteson, in coordination with an Executive Committee and supported by a Project Coordinator (PC) (50% time) with experience in training programs, including an NSF NRT. The PI and the PC will run daily operations, with the PC reporting directly to the PI. The PC will be responsible for executing all day-to-day program activities, including the logistics of organizing meetings, workshops, and seminars, coordinating the mentorship and internship activities for student fellows, promoting the program, and coordinating with diversity offices. The PI will oversee the NRT program, using it to support and tie together existing related efforts, such as the Communications seminar series (founded by the PI), and the nascent "Physical Sciences Machine Learning Nexus," which grew out of the previous NRT proposal. Meeting monthly with the deans of the Graduate Division, Physical Sciences, and Computer Science, the PI will provide program reviews and work closely with the administration to plan for the program's continuity past the 5-year window. The Deans are strongly committed to our vision, having long supported science communication (Whiteson's Communication seminar series) and interdisciplinary work (via fellowships for the Nexus).

Core Teams: The core participants have extensive experience in AI (Baldi, Mandt with established track records of interdisciplinary collaborations), astrophysics (Whiteson, Murgia, Barwick, Sallum), team science (Salazar) and communications (McWhorter).

Executive Committee and Advisory Faculty: Profs. Baldi, Barwick, and Murgia will be responsible for advising the PI and PC on the program's research and educational aspects. Prof. Salazar and McWhorter will oversee the program's team science and communications aspects, and Ward and Diggs-Yang will oversee the diversity and recruitment aspects (see below). This Executive Committee will meet quarterly with the PI to review the program's general progress and direction, including review of applications for NRT fellowships, reports from committee members about ongoing activities, plans for upcoming events, and general strategic discussions related to the program. Additional meetings may be scheduled on an as-needed basis. The Student Chair of the Student Executive Committee will also be invited to attend in an advisory capacity but will not vote. Each academic year, the final quarterly meeting (in May) will be held as an annual meeting, with the external evaluator attending and with a broad invitation extended to all participating faculty, Department Chairs, and Deans.

Student Executive Committee: To encourage ownership of the program by the participating students, we will have a Student Executive Committee, with a Student Chair, and a committee consisting of the current set of funded fellows plus selected honorary fellows. This committee will meet monthly during the academic year to discuss the program and to plan events: e.g., selecting topics and speakers for monthly lunch meetings and organizing student-led events. These meetings will be student-run, students will have the option of inviting any of the core or participating faculty members to specific meetings, but the idea is to give the students an opportunity to discuss the program on their own without faculty intervention.

Social Networking and Events: To foster community spirit, we will schedule quarterly social-networking events (such as lunch or evening receptions) that all NRT student fellows (both funded and honorary) and all NRT core personnel (including the project evaluator) will be expected to attend, with invitations to all other staff and faculty involved in the program. These quarterly events will be either stand-alone (with short talks from one or more participants) or coordinated with other activities, such as a lecture from a visiting speaker after a team-science workshop. In addition, we will have a monthly meeting (all 12 months, all 5 years) for NRT student fellows, with speakers discussing their research in a relatively informally structured meeting to encourage interaction and discussion, with lunch provided. The project coordinator and at least one member of the Executive Committee will attend each meeting, but the intent is to have the students be the meeting's primary organizers. Each year in March we will organize an NRT Symposium Day, where students in the program (full and honorary) will present their research in a day-long symposium, advertised across campus to encourage broad attendance and will be followed by a networking reception to encourage interaction. Representatives from external research labs will also be invited to participate in this annual symposium. Both the monthly meetings and the annual symposium were components of our previous NRT program and were very effective in building an *esprit de corps* among the participating students.

Relevant PI Experience: PI Whiteson has the scientific expertise to lead the proposed traineeship program, having a well-established track record of conducting related research in physics and AI, and conducting interdisciplinary work. He has significant past experience with the NSF NRT program, as Co-PI of an earlier program at UCI (2016-2021), working closely with the PI of that program, Prof. Padhraic Smyth. Whiteson has experience organizing educational programs, having served as Departmental Vice Chair for Undergraduate Affairs in Physics and Astronomy (2012-2018) and as a founding executive committee member of the Joint Degree Program with San Diego State University, which established a new PhD program in Computational Science and required navigating the bureaucracies of two large public institutions. Whiteson is dedicated to public science communication, co-authoring two books for general audiences [8, 9], co-hosting a popular science podcast [10], and co-creating a science television show for PBS KIDS [11].

4 Education and Training

Our education and training program will center on an intensive program to develop transferable interdisciplinary skills, using team science and communications training to equip students with the necessary skills to translate the context, motivation, and challenges of their work to their academic colleagues and to the public. These transferable skills will allow them to share the excitement of their research with their home communities and with the general public, to work fruitfully at the intersection of AI and astrophysics, and to identify and capitalize on future opportunities for effective interdisciplinary work. Our training program will augment the existing graduate student education, filling in gaps in the current STEM graduate student education program, including formal courses in the complementary discipline, team science training sessions, hands-on communications training and practical applications, summer internships, seminars, and career counseling.

4.1 Components of the Traineeship

Figure 2 shows a sample trajectory of activities for a funded fellow in our program, over the students' five years from admission to PhD completion. Once admitted to the NRT program, full fellows will be required to participate in a variety of courses, workshops, seminars, team science activities, and mentoring, in order to maintain their NRT fellow status and funding. Honorary unfunded fellows will be required to participate in a subset of the activities but are eligible for all activities.

To gain admission to the NRT program, whether as fully funded or honorary unfunded fellows, students will submit an application to be reviewed by the program Executive Committee, with applications reviewed quarterly. The NRT application will require a description of the student's proposed research, the students' transcripts from their undergraduate and graduate education, and recommendations from two faculty mentors (one each from a faculty member in AI and astrophysics) who will mentor the student on the proposed research project.

Co-mentorship: To foster critical interdisciplinary teaming skills, the program will provide opportunities for students to work with cross-disciplinary faculty advisors and fellow doctoral stu-

AI Electives for Astrophysics Students	Astro electives for AI Students
CS 273A: AI	PHY 136 Introduction to Particle Physics
CS 274A: Probabilistic Learning	PHY 137: Introduction to Cosmology
CS 274C: Neural Networks and Deep Learning	PHY 212A: Mathematical Physics
CS 216: Image Understanding	PHY 138: Astrophysics of Galaxies
CS 274E: Deep Generative Models	PHY 139: Observational Astrophysics
CS 172C: AI Frontiers: Technical, Ethical, and Societal	PHY 144: Stellar Astrophysics
STATS 225: Bayesian Statistical Analysis	PHY 145: High-Energy Astrophysics
STATS 230: Statistical Computing Methods	PHY 242: Astrophysics Fundamentals
STATS 245: Time Series Analysis	PHY 61C: Introduction to Astrophysics

Table 1: Elective courses for participants in the proposed NRT program

dents in other fields on joint projects, including exploring cutting-edge AI approaches and working on relevant datasets. In our prior NRT grant, students reported immensely enjoying regular interaction in these smaller cross-disciplinary projects as they gained access to and understanding of terminology, approaches, and disciplinary-based frameworks for understanding. In our evaluation, however, the previous program may have been overly ambitious in trying to span very broad gaps among fields across the physical sciences. In this proposal, we focus on astrophysics, which produces data that are well suited to analysis in statistical and AI terms.

Courses: NRT fellows will be enrolled in their home department's MS or PhD degree program and will be required to complete the regular MS or PhD course requirements for their degrees. Students will also complete at least three course electives outside their home school that are relevant to their proposed research project, from a list approved by the Executive Committee; this list of courses will be updated annually. Table 1 provides an example of an initial course list.

Team science: Typical graduate student programs focus on teaching the core scientific concepts of their area of specialty. However, in an era in which interdisciplinary research is producing rapid and dramatic gains, what is lacking is explicit training in how to build and succeed within a heterogenous scientific team. From the time graduate students are selected as trainees, they will be immersed in a rich research environment where team science fosters creativity and innovation. We will supplement this environment with explicit training in effectively conducting productive cross-disciplinary collaborations as emerging team science leaders (details below).

Communications: While communication is an essential element of pedagogy (via instruction) and of the scientific community (via papers, seminars, conferences), most graduate programs offer little to no training in these vital skills, expecting students to absorb them informally. This traditional approach presents barriers to students who many not have had an opportunity to spend time in research communities and absorb such skills. The traineeship will offer specific training in communication skills, with an emphasis on how young scientists can speak to members of academia outside of their specialty, and how they can speak to the general public (details below).

Seminars: We will plan a quarterly series of high-visibility external speakers from experts in science communication and interdisciplinary research. These seminars will be distinct from the typical scientific seminar in that they will include a strong focus on the speakers describing the *process* of the work and their path to their current positions. Building on the "Science Communication" seminar series founded by PI Whiteson in the School of Physical Sciences, which hosts well-known science communicators (for example, *New York Times* science journalists, social media science influencers), the seminars will describe the nature of the speakers' work and their career arcs. Fellows will attend the symposia and also have the opportunity to meet the speakers in person.

We will have an annual NRT symposium and poster day in March/April of each year, where students in the program will present their research to a campus-wide audience. This symposium will also be attended by representatives of external organizations interested in summer interns from our program (LBL, LLNL, BNL, LANL, ANL), including presentations from the external labs on activities and opportunities for students to meet with lab representatives face-to-face.

Monthly Meetings: Monthly lunch meetings for all participating fellows (both funded and honorary), will be organized by students and moderated by participating faculty members, where students will be selected on an ongoing basis to give short presentations on their research. This was a very successful element of our previous NRT program and is an excellent way to foster an interdisciplinary spirit. Separately, fellows will meet monthly in the Student Executive Committee.

Internships: Funded student fellows will spend at least 1 summer at a national lab working on a research internship that is related to their thesis work. Honorary fellows who are PhD students will also be eligible for internship placement if additional internships are available. The Project Coordinator will organize the placement of students for summer internships and coordinate a 1-2 day pre-internship workshop at the end of the academic year to provide NRT students with background on the national labs' history and research culture. Students will be required to communicate the results of their internship upon their return to UCI: e.g., at one of the monthly NRT lunch meetings. NRT participating faculty will assist in identifying suitable mentors at the labs and will encourage these mentors to visit UCI, and if appropriate, to engage in research collaborations with the students with whom they are being paired (for example, by becoming a member of students' PhD thesis committees). In this manner, the program plans to foster strong links between the UCI NRT program and various research labs, links that will last beyond the NRT program's five-year duration. We have strong support for our proposal from several national labs: cf. support letters from LBNL, LLNL, BNL, ANL, and LANL, expressing willingness to provide summer internships for, and ultimately hire, students from our program.

4.2 Sustainability of the Program

To sustain the program, we will create a **Professional MS in Data Science**, run jointly by the UCI Schools of Physical Sciences and Information and Computer Sciences, to teach interdisciplinary skills using didactical approaches developed by the NRT fellows. The new degree program will provide teaching positions for AI-Astrophysics students, and tuition revenues to fund graduate student research positions for AI-Astrophysics students. Based on other similar MS programs at UCI, these efforts are expected to support no less than 10 graduate students per year. Moreover, following the formal period of the grant, the UCI Graduate Division will provide funding equivalent to 25% of the NRT positions (see institutional letter). Together with additional research positions funded by grants from the participating faculty, this will ensure the sustainability of the program.

4.3 Trainee Population and Timeline

Our proposed NRT program will have both (a) full NRT fellows (fully funded by the NSF NRT program) as well as (b) honorary NRT fellows (not funded by the program). We plan to have 8 fully funded fellows in the program at any one time, as well as 15 to 20 honorary fellows who are not directly funded by the program but may be candidates for funding in subsequent years. Masters students will only be eligible to be honorary fellows. Students will be able to enter the program from the Schools of Physical Science and Computer Science, with the majority of NRT students expected to come from the Departments of Physics & Astronomy, Computer Science, and Statistics. Students will be able to complete their home programs within the nominal five years. A timeline is shown in Fig. 2.

4.4 Career counseling

The program will provide trainees with counseling for both research and research-related careers, within and outside academia through the structured use of an Individual Development Plan (IDP).

The IDP is an evolving document, which allows the student to communicate their career goals (academic or otherwise) to their advisors and helps ensure that the student's work is aligned with their goals. The process of developing the IDP allows students to perform a self-assessment of skills, identify their academic and career needs and have structured opportunities to communicate them with their advisors, and to set goals and follow through.

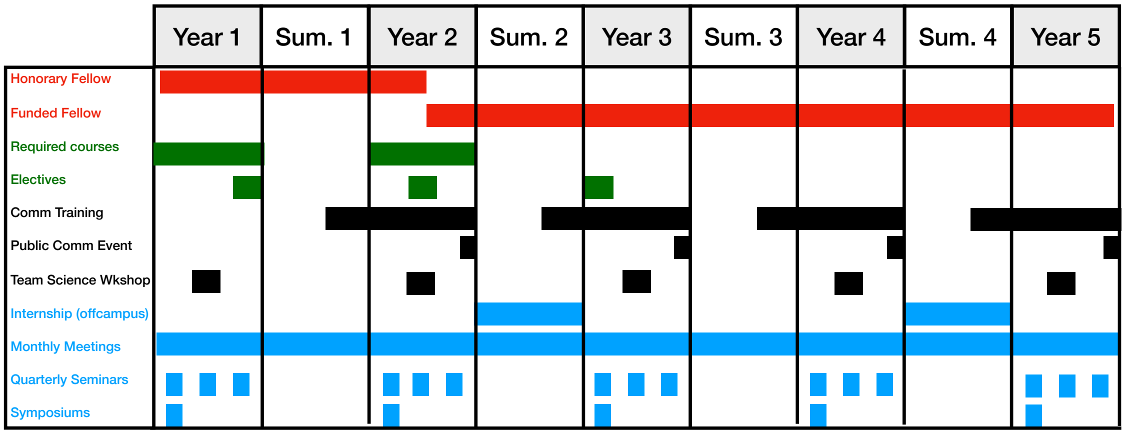

Figure 2: A typical trajectory for an NRT fellow, including coursework, training, seminars, public communication events, and summer internships.

In a recent study, one-third of University of California graduate students were "less than satisfied" with the level of mentorship and career guidance they received from their advisors [31]. But those with IDPs tended to be more satisfied and make more progress toward their goals [32]. Thus, by encouraging students to define and articulate their career goals early in their training program, the IDP helps mentors guide the students' paths through the program, incorporating professional development opportunities when they are available. Students will complete an IDP in the first full year in the program and will review it at least annually with both mentors.

5 Required Skills and Competencies

5.1 Communication

Trainees will receive an intensive course of training in science communication, giving them transferable skills to explain the challenges and opportunities of their research to fellow students, and to communicate the context and broader impacts of their research discoveries to policy makers and the general public. The communications program will run on an annual cycle, beginning at the end of each summer with a three-day bootcamp, led by communications specialist Bri McWhorter. Students will then work through the year on a collaborative science communication project that is to be presented at an event for the public.

Bootcamp: The goal of the summer communication bootcamp is to provide necessary communication tools for students across disciplines to clearly and confidently articulate their ideas and research. Instruction will focus on content creation, organization, and delivery. Since the goal of the NRT is to encourage and create active discussions, these workshops will be in person and interactive. Prior to the bootcamp, the students will record an initial elevator pitch video (three minutes) explaining their research or interests. The bootcamp will be organized as three 1-day workshops beginning with programs to overcome nerves, then focusing on creating a memorable and repeatable story and turning passive presentations into active conversations, and concluding with ways to connect with people outside of your field and to bring ideas to life. During each workshop, students will learn general skills but also apply the concepts to their elevator pitches. The bootcamp will conclude by focusing on interpersonal communication skills and having the students polish their pitches. Bootcamp Director McWhorter will meet with each student after the bootcamp to give them personalized notes on how to improve their communication skills. Two weeks later, students will submit the final version of their elevator speeches for evaluation.

Public Event: During the year, students will work on a final project that will be presented in a public event at the end of the year and shared via YouTube (with permission). The project will

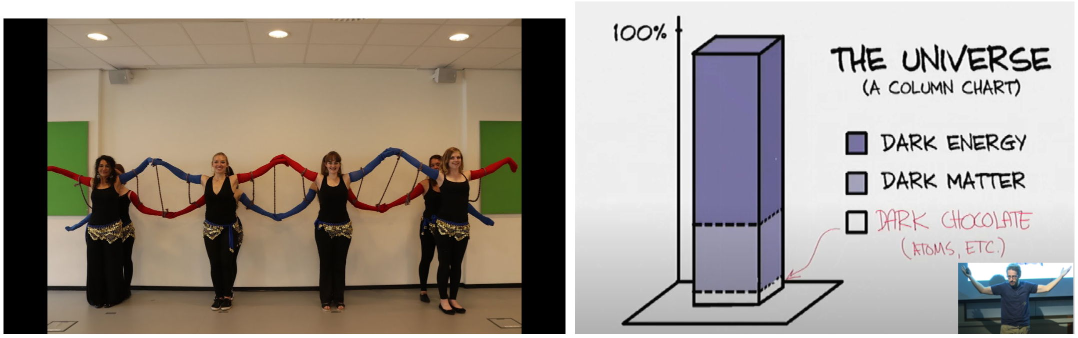

Figure 3: Left, an example of science communication through dance, from *Dance Your PhD*. Right, an example of communication through live cartooning by Jorge Cham of a talk by PI Whiteson.

be a short presentation by a small group of students in which they communicate the excitement and importance of their research area. Presentations may be of any suitable form, including traditional slideshows, but may be in potentially more novel forms such as dance performances, cartoon live-drawing, or radio plays; see Fig. 3. PI Whiteson has experience organizing such events with large attendance beyond UCI, drawing from local high schools and community colleges.

Students will select partners and develop their initial ideas, in consultation with McWhorter and Whiteson. McWhorter will provide private coaching and feedback for each group over several sessions. During the spring, the student groups will meet more regularly with McWhorter as they work toward finalizing their presentations into personal, unique, and compelling projects.

5.2 Teamwork

In years 1-2, NRT fellows will participate in a team science workshop that will prepare them for cross-disciplinary collaboration. The workshop will extend existing team-based instructional strategies to focus on the needs of NRT fellows in the formative stages of their academic development. It will contain elements of how to build leadership capacity, mitigate cross-disciplinary divides, and build social identity such that members can work within and across disciplinary boundaries.

With a focus on the project proposed by the scholar, involving mentors from two distinct fields (AI and Astrophysics), interventions will focus on project planning and execution. Interventions to foster knowledge integration are planned to consist of techniques to launch interdisciplinary collaborations through effective team design and the alignment of goals and objectives. NRT fellows will learn to set project milestones and to develop leadership skills to facilitate project completion. Moreover, NRT scholars will learn nine key core concepts and skills, including the science of team science, team assembly and composition, cross-disciplinary trust, leadership, fostering inclusive collaborative environments, project management, proposal writing, conflict resolution. Also, fellows will learn to develop project implementation plans and authorship agreements, and to track their own professional development on critical team science skills using an individual development plan (IDP). Ongoing team science consultation will also help to provide feedback and guidance when needed to ensure progress is being made toward scientific aims.

Drawing on expertise developed during our past NRT project, our NRT Trainee toolkit consisting of instructor materials, workshop materials, and measurement tools will be refined and used with this cohort. Upon completion of the monthly education seminars, reading assignments, and experiential training on team science, NRT fellows will be able to manage projects, work collaboratively, be innovative, engage diverse audiences and partners, educate and mentor others, and balance multiple commitments.

Pretests and posttests using surveys, electronic sampling methods and qualitative analysis will measure shifts in the NRT fellow development, and the effectiveness of the interdisciplinary teams associated with the grant by using new and cutting-edge scales recently developed by our research team on transdisciplinary orientation and integrative capacity. In years 3-5, a team science-themed workshop will be made available to other NRT training sites around the United States, drawing from the curriculum established by this and our past NRT, with workshop elements led by UCI NRT fellows, and including training on ethics in research. All participants in this day-long workshop will receive a team science certificate. A Team Science Certificate Program will offer opportunities to make a strong start on developing professional career skills critical for their success. During the workshop, all attendees will complete: 1) three vignettes (mini-courses) on team science (science of collaborative research, team assembly and composition, and cross-disciplinary trust); 2) one training course on diversity and cultivating an inclusive learning environment; and 3) at least one hands-on education activity in three other training areas (leadership, teaching, mentoring, ethics, proposal writing, and outreach). Building on this effort to disseminate knowledge, mid-program NRT trainees will also gain instruction in how to train others in support of creating the Team Science Certificate program, which will be available to other trainee sites.

5.3 Ethics and Additional Skills

The importance of ethics at all levels of research and society cannot be overestimated. We will use a multi-pronged approach for the ethical training of the NRT cohort. All students in the program will be required to take the campus-wide course on the ethical conduct of research. In addition, students will be able to take the course Artificial Intelligence Frontiers: Technical, Ethical, and Societal, recently introduced by Prof. Baldi. The course is primarily a discussion course, based on case studies, articles and videos, and other material fostering open communication and discussion about contemporary issues related to AI and its deployment throughout society. Its main aim is to empower our students to reflect and manage ethical issues in their future professional lives. Finally, ethical issues will also be investigated in the cohort's team science and communication activities through different activities, including role playing.

Trainees will gain important skills in outreach, teaching, and mentorship, and make valuable contacts outside of academia to further their professional development. The strong emphasis on communication, to other academics and to the public, will teach students how to convey the excitement of their work in various outreach activities. By teaching teamwork skills at the Team Science Workshop, students will gain proficiency in teaching and mentorship. Additionally, internships at national laboratories and in industry will offer important new perspectives on research and provide contacts that trainees can use for future professional development.

6 Major Research Efforts

Astrophysics treats the Universe as a laboratory, where messages from extreme events reveal clues about the nature of space, time, matter, and energy. Historically, various astrophysical experiments have been siloed by the type of message they receive: photons, cosmic rays, or gravitational waves. A new effort in *multi-messenger* astrophysics [33] will leverage these disparate channels to observe events overlooked by a single channel and to gain deeper insight. Opportunities to improve the power of individual messengers abound at the interface of AI and astrophysics, as well as at the interface between astrophysical categories, to create AI-powered multi-messenger observations.

6.1 AI

Analysis of complex, high-dimensional, multi-modal astrophysical data is often performed by applying algorithms designed by physicists to extract the relevant knowledge. Due to the limitations of human knowledge, these analysis pipelines are often not optimal, and can be improved by using AI and its statistical AI methods to optimize signal-to-noise ratios. This proposed traineeship program will develop and deploy such methods[1] to analyze astrophysical data with greater statistical power, and to integrate information across different messenger channels, leading to new discoveries. The program will also draw on the field of deep generative modeling, which can be used to accelerate numerical simulations, generate data, or form informative priors for sparse data[34, 35].

At the same time, astrophysics will provide a rich source of data and problems that will challenge and advance AI. Such challenges will include the ability to deal with the entire spectrum of data abundance, from very large to relatively small datasets; the ability to learn from complex, multi-dimensional, multi-channel, variable-size data; the ability to handle group operations such as rotations or permutations; the ability to factor out noise and nuisance parameters; the need for methods that are not only sensitive but also robust and interpretable; and the need to develop methods across multiple learning paradigms, from unsupervised, to self-supervised, to supervised, to transfer, and to reinforcement learning. The program will also challenge the frontier between symbolic and connectionist AI, both in the direction of trying to combine symbolic methods with deep learning methods, and the alternative, biologically-inspired direction of implementing symbolic methods using deep learning methods, for instance, for symbolic regression in physics [36, 37]. After all, the brain is a master at integrating information from multi-messenger channels, such as vision and audition. Finally, at a fundamental level, AI and Astrophysics are at similar crossroads, where the existing methods and theories have been extremely successful, but at the same time are known to be incomplete. The grand goals of general AI and a deeper understanding of the nature of the Universe will resonate with each other throughout the program.

6.2 Cosmic Rays: Neutrinos

The ARIANNA experiment detects high-energy neutrinos from astrophysical sources using the cold, transparent ice of Antarctica. Neutrino interactions can be observed either as radio waves that propagate to the surface through transparent ice, or as light pulses in strings of detectors embedded in the ice itself. Prof. Barwick is an international leader and pioneer of neutrino astronomy since 1991 [38, 39], and PI of the NSF-supported ARIANNA high-energy neutrino detector. The ARIANNA surface technology is currently being envisioned for two applications: an intermediate-scale high-energy neutrino detector at Moore's Bay, Antarctica, called ARIANNA-200[40, 41], and the surface component of a radio detector in a planned IceCube Gen2 program which will be located at the South Pole, Antarctica[42].

In both applications, we expect to include AI techniques to make real-time decisions on the event priority using onboard microcomputers. Most triggers are caused by thermal noise fluctuations, and we have implemented a neural network that rejects 100,000 thermal events for every event that passes, while preserving 95% of neutrino signals. Future work will investigate techniques to speed up the reprocessing of the event data for neural network evaluation by a factor of 10, which will improve the detector's sensitivity by as much as a factor of 1.8[43]. We also propose to explore new microcomputer technologies to reach this goal. To further enhance the capabilities of the ARIANNA-based station, we propose to develop and port AI-based evaluation tools to Field Programmable Gate Arrays. AI promises to boost astrophysical neutrino observations by making real-time decisions to reject noise, enhancing the data's power. The ARIANNA collaboration, in an effort led by Barwick and Baldi at UCI, has promising preliminary results[44].

Concurrently, we have used AI techniques to evaluate the energy resolution and angular direction resolution of the neutrino, starting from the collected radio signals in the time domain from each antenna of the ARIANNA station[45]. The speed of neural networks allows researchers to evaluate and optimize a large range of detector configurations with far less personnel and computational resources. As neutrinos are often the first-arriving messenger from extreme events, a long-term goal is to develop an AI-based alert system for multi-messenger signals, giving other facilities a chance to observe these transient events. Using neural networks, we propose to reduce the latency time to send the alert from fractions of an hour to a few minutes.

6.3 Photons: Gamma-Rays

The *Fermi* Large Area Telescope (*Fermi*–LAT) searches for γ-rays from a wide variety of astrophysical phenomena, including dramatic events such as neutron star mergers, which also produce

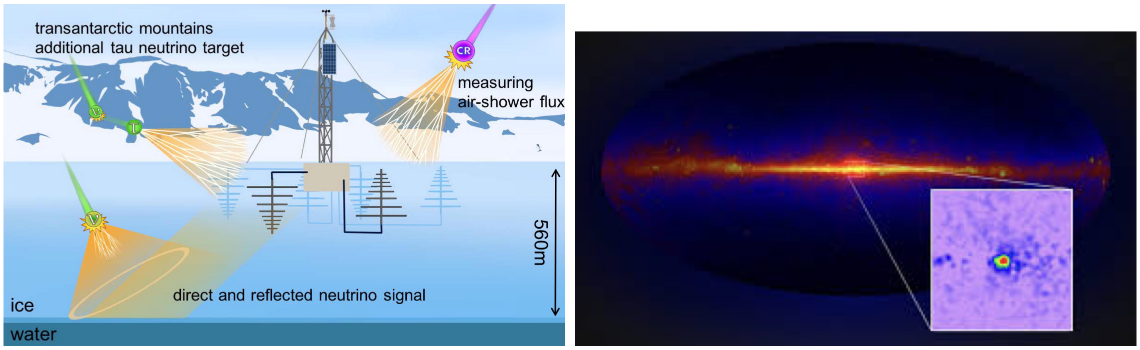

Figure 4: Left, sketch of the ARIANNA high-energy neutrino detector, which will be located at Moore's Bay on the Ross Ice Shelf in Antarctica. Right, view of the galaxy in the gamma ray spectrum, showing the bright spot near the center (Image credit: NASA/T. Linden, U. Chicago).

detectable gravitational waves (Fig 5), generating multi-messenger events which help unravel mysteries such as the source of gamma-ray bursts [46].

In addition, *Fermi*–LAT can observe γ-rays produced by DM annihilations or decays in our Galaxy and beyond [47, 48, 49] and has observed an excess in the center of the Galaxy compatible with a DM signal. To improve on these constraints or, more ambitiously, confirm a signal, improvements are necessary that go beyond the traditional methodologies used to analyze the γ-ray data. Short of an independent detection of a DM signal, such as a sharp peak in the γ-rays energy spectrum, the central challenge is disentangling a DM signal from other, brighter, processes that could mimic it and must be accurately modeled. Currently available models, however, lack the required accuracy because they rely on uncertain extrapolations from observations at other wavelengths of light. Satisfactorily addressing these challenges can be accomplished by judiciously breaking down this problem into more fundamental components, using a combination of AI and additional astrophysical observations to improve each component, and by using AI to properly account for correlations and dependencies. AI is crucial in realizing this because it enables us to exploit a wealth of additional multiwavelength observations that have not been incorporated yet into existing models to ultimately provide robust predictions that could be decisive in searching for a DM signal in the γ-ray sky. As an illustration of this, in recent work, Murgia and Baldi demonstrated that the modeling of the astrophysical background can indeed be significantly refined by employing AI to better resolve fine features in the γ-rays that originate from the interstellar gas via the best available radio observations that trace the gas. These features, if not properly modeled, could conceal a DM signal. Non-parametric approaches such as Gaussian processes or probabilistic modeling can be successfully used to build physically motivated models for these features where radio observations are sparse, to allow for an unbiased extraction of DM signals.

6.4 Photons: X-rays

The collapse of massive stars can leave behind a peculiar object, a neutron star. With a radius of only 10-15km but the mass of our Sun, the density of matter in their cores is far beyond the reach of current laboratory experiments and provides a window into states of matter that balance immense gravity and strong quantum mechanical effects. The spectra of X-rays from hot spots on the surface of the stars can reveal the internal pressure and density of neutron stars but require expensive computational modeling and significant simplifications, sacrificing statistical power from these rare and precious observations. Whiteson and Baldi have partnered with neutron star experts at the University of California, San Diego (Lundblom) and SDSU (Weber) through the Joint Degree Program to use AI to infer the internal state of the neutron star, with promising initial results. The subtle nature of the systematic uncertainties and the expensive computation required to simulate the neutron star internals and the telescope response make this an area where AI could make significant contributions, speeding up core calculations which would allow for use of a richer, less simplified dataset and more rigorous treatment of systematic uncertainties.

6.5 Further Research Opportunities

The faculty at UC Irvine and nearby institutions have an internationally recognized and broad program of research in astrophysics, providing a rich stream of opportunities for new and transformational collaborations with AI experts. For astrophysics faculty from other institutions (CSUF, LBNL, see Table 2), the NRT would support a UCI AI student as part of the cohort, supervised jointly by a UCI AI faculty member and the non-UCI astrophysics faculty member.

Gravitational Waves: Joshua Smith, Geoffrey Lovelace, and Jocelyn Read (Cal State, Fullerton) explore the nature of gravity and spacetime by studying the ripples in space, gravitational waves caused by massive accelerating bodies. Prof. Smith is a member of the Laser Interferometer Gravitational-Wave Observatory (LIGO) experiment, which observes gravitational waves directly and which can be tied to dramatic astrophysical events through correlations with gamma-rays (see Fig. 5). Prof. Read is a theorist who specializes in modeling the sources of gravitational waves. Prof. Lovelace uses numerical relativity to model sources of gravitational waves, such as merging black holes. In each case, AI tools can optimize the data analysis chain and provide speed boosts for expensive numerical modeling.

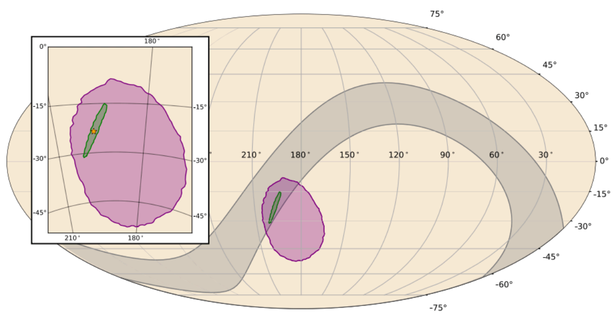

Figure 5: Sky localization of a neutron star merger [46] with multiple messengers: gravitational waves (LIGO, green) and γ-rays (Fermi in grey, INTEGRAL in purple). AI-based algorithms can allow addition of neutrinos observations with few degrees of angular resolution to these kinds of events.

Black Holes: Aaron Barth studies the mysteries at the centers of galaxies, such as supermassive black holes and quasars, using telescopes such as the Keck Observatory and the Hubble Space Telescope. AI tools can improve the analytical power of these observations.

Galaxy Formation: David Buote studies the formation of galaxies, including the nature and distribution of dark matter, using X-ray telescopes such as the Chandra X-ray Observatory. AI can help develop non-parametric models for the backgrounds to such observations.

Dark Energy: David Kirkby seeks to understand the nature of the cosmic acceleration that drives the Universe's expansion, using DESI and LSST to collect optical spectra for millions of galaxies. Kirkby is an expert in the application of statistical AI to astrophysics [50].

Exoplanets: Stephanie Sallum studies the formation of planets using high-resolution imaging and interferometry using the Keck Telescope and other telescopes. Aomawa Shields studies the possible climates and potential habitability of extrasolar planets, comparing climate models to data acquired by ground- and space-based telescopes; she also runs science communication programs for young women. Paul Robertson searches for additional habitable worlds around nearby stars, using the Habitable Zone Planet Finder and NEID instruments. AI can help extract subtle signals and improve the ability to deduce crucial information about distant worlds.

Dark Matter: Peter Sorensen and Scott Kravitz (LBNL) use large underground experiments with chemically pure low-activity liquids (for example, Xenon) to search for the passage of dark

Table 2: List of experts beyond the Core Participants expected to participate in the program.

Name	Department	Institution	Speciality
Prof. Joshua Smith	Physics	CSU Fullerton	Grav. Waves
Prof. Geoffrey Lovelace	Physics	CSU Fullerton	Grav. Waves
Prof. Jocelyn Read	Physics	CSU Fullerton	Grav. Waves
Prof. Aaron Barth	Physics & Astronomy	UCI	Black Holes
Prof. David Buote	Physics & Astronomy	UCI	Galaxy Formation
Prof. David Kirkby	Physics & Astronomy	UCI	Cosmology
Prof. Aomawa Shields	Physics & Astronomy	UCI	Exoplanets, Commun.
Prof. Paul Robertson	Physics & Astronomy	UCI	Exoplanets
Scott Kravitz, PhD	Physics	LBNL	Dark matter detection
Peter Sorensen, PhD	Physics	LBNL	Dark matter detection
Prof. Tim Tait	Physics & Astronomy	UCI	Dark matter theory
Prof. Jonathan Feng	Physics & Astronomy	UCI	Dark matter theory
Prof. Eric Mjolsness	Computer Science & Math	UCI	AI
Prof. Wayne Hayes	Computer Science & Stats	UCI	AI
Prof. Sameer Singh	Computer Science & Stats	UCI	AI
Prof. Gary Olson	Informatics	UCI	Team Science
Prof. Judy Olson	Informatics	UCI	Team Science
Gregory Diggs-Yang, PhD	Access & Inclusion	UCI	Diversity

matter particles, which may rarely scatter off of an atomic nucleus and produce bursts of light. AI can boost the power of these experiments by analyzing the full light waveforms.

Theory: Profs. Jonathan Feng and Tim Tait are internationally recognized leaders in the development of the theories of dark matter production and detection. Their work crucially informs the design of astrophysical experiments and analysis of the data. AI can help explore the large parameter spaces of these theories to identify candidates which are consistent with experiments and observations.

7 Broader Impacts

The NRT program will contribute broadly to the achievement of societally recognized outcomes, including the training of a globally competitive STEM workforce, increasing the participation of underrepresented communities, catalyzing new partnerships among California institutions, making transformational discoveries in accessible areas of great public interest which will be communicated effectively, sparking novel AI strategies which can be applied to other disciplines, inspiring similar training strategies for other STEM fields with interdisciplinary opportunities, and dissemination of the research findings.

Globally Competitive STEM Workforce: The core of the NRT training will equip the students with the transferable communication and team science skills needed to bridge the gaps between fields, allowing them to take advantage of opportunities for interdisciplinary and convergent science beyond the immediate disciplines of astrophysics. Students who are trained in the program will have developed the skills to translate the essential questions of their field, or to explain the power of their novel methods, allowing them to identify areas in their future careers where interdisciplinary approaches can make rapid progress or crack longstanding problems. These students not only will be leaders in their discipline, but as a cohort will become leaders in the Science of Team Science and Science Communication. Students will retain the experience and techniques they used to communicate the nature of their scientific challenges with their colleagues in other academic fields and beyond. Students who go on to become faculty at other institutions can replicate this program's essential structure at their future institutions.

Broadening Participation: The program will directly increase the participation of students from underrepresented communities through our recruitment and retention efforts. More broadly, by teaching communication skills and encouraging outreach in communities with which researchers have existing relationships (that is, their communities of origin), we impact communities beyond those directly participating in the NRT. Many students at UCI, a Hispanic Serving Institution, are first-generation students or from traditionally underrepresented groups. UCI Physics and Astronomy is a leader in efforts to retain such students through the successful NSF S-STEM program, as well as in programs to support women in physics.

Catalyzing New Partnerships: Many of the groups participating in the NRT are aware of the potential opportunities offered by interdisciplinary research but lack the resources or experience to bridge the historical or institutional divides. Explicitly training students in these techniques and partnering two faculty members with two students, one each from AI and astrophysics, will build a connection between the communities which is likely to outlast the NRT-funded work, as successful collaborations naturally lead to follow-up projects and a rapid burst of intellectual productivity.

Transformational Discoveries: The questions probed by the astrophysical research are unusually accessible: *what is most of the Universe made of? What is its long-term fate?* Breakthroughs in these areas are similarly of great public interest. The experience in public communication of the PI and the training provided to the NRT trainees will allow them to share the excitement of these topics with the broader public in an entertaining and accessible manner, producing compelling stories that communicate important scientific information.

Novel AI Techniques: Each field of STEM asks unique statistical questions of its data, driven by the particular scientific motivation or the structure and nature of the data itself. These pose unique challenges for AI, requiring development of new hammers to apply to unique nails, sparking innovation in core AI concepts which are likely to find applications beyond astrophysics, potentially challenging the frontier between symbolic and connectionist AI.

Impact on Graduate STEM Training: Unfunded NRT fellows will benefit from internship opportunities, science communication and team science training, and the vibrant intellectual atmosphere. The potential success of interdisciplinary work in AI and astrophysics will inspire other fields to explore the possibility of similar interdisciplinary research and learn from the successes and struggles of this NRT. Our Team Science Workshops will be open to all NRT cohorts nationwide.

Dissemination: The results of the program will be shared with graduate schools around the country, via the chapters of the trainees' theses, presentations at educational (e.g., ECU or PERC) and science-communication conferences (e.g., ComSciCon or SciComm), and a final report based on the evaluators' summative conclusions. These materials will allow the NRT program to benefit STEM graduate students beyond our NRT trainees and for training innovations to be communicated broadly beyond UC Irvine.

8 Recruitment, Mentoring, and Retention

UCI has a diverse undergraduate population (54% Asian/Asian American, 25% Hispanic, 3% Black, 1% other URM), but the population of domestic graduate students includes only 11% URM. To develop a group of students whose demographics represent the URM population of the U.S. (32%) or California (47%) requires a dedicated program of recruitment and retention of trainees from underrepresented communities, including attention to UCI's institutional culture, inter-institutional partnerships, a focused effort on building critical mass, and maximizing faculty involvement [51]. Our training program will have such efforts built into its recruitment and retention plans.

UCI has numerous programs focused on recruiting, mentoring, and retaining underrepresented community participation in its education, research, and service missions. The NRT program will work with the Office of Access, Outreach and Inclusion (OAOI), in the school of Physical Sciences, and the Office of Access and Inclusion (OAI), in the School of Information and Computer Science, to coordinate the recruiting, mentoring, and retention activities in partnership with other campus-wide initiatives. These offices (OAI and OAOI) offer programs to remove barriers to the recruitment, retention, and graduation of talented students. They provide access to a welcoming and supportive environment that respects and embraces diversity and individual differences to empower everyone to learn. They equip talented students from different backgrounds with critical academic and career skills. Specifically, the offices provide: one-on-one tutoring, academic success workshops, professional development workshops, mentoring programs, scholarships, research and employment opportunities, academic and career coaching, programs designed to support entering graduate students' transition into graduate school, and support for several clubs. PACE (Physics Astronomy Community Excellence) is a graduate student-led mentoring organization by students and for students with the aim of normalizing a holistic approach to success for UCI Physics students while strengthening the community through trained mentors. Key Personnel Prof. Steph Sallum is the faculty advisor for PACE and will be able to connect NRT trainees with PACE mentors.

Core faculty in our NRT program, such as Prof. Baldi, participate in recruitment events such as the California Forum for Diversity in Graduate Education. Other resources for recruitment include The Summer Undergraduate Research Fellowship Program (SURF) and the University of California Leadership Excellence through Advanced Degrees (UC LEADS) program. The SURF program brings URM students with outstanding potential from all over the U.S. to UCI's campus. The UC LEADS program aims to foster high-quality junior and senior-level undergraduate URM students for University of California doctoral programs. Through both programs, prospective students learn to successfully apply to graduate programs while experiencing graduate education at UCI. The Cal-Bridge program, a partnership between the UC and California State campuses, creates opportunities for underrepresented minorities to increase their representation in Physics and Astronomy programs, by receiving intensive joint mentoring by CSU and UC faculty, as well as by having research opportunities and professional development workshops to help scholars apply to graduate school; Prof. Whiteson has been a member of Cal-Bridge since 2018.

Recruitment Plan: Our recruitment practices will include: 1) advertising to broad audiences, with special attention given to working with outlets that reach URM communities; 2) having admissions committee members receive training by UCI equity advisors on approaches to mitigate unconscious bias; 3) monitoring the demographics of the applicant pool relative to the pool of admitted students; and 4) emphasizing campus resources available to support URMs.

Retention Plan: DECADE (Diverse Educational Community and Doctoral Experience) is a U.S. Department of Education (ED) grant in partnership with the UCI ADVANCE Program. Using the ADVANCE Equity Advisor model to improve diversity and climate among the faculty and graduate student populations, the NRT program will provide support for the retention of URM groups through programming that focuses on professional development, networking opportunities, skills building, and improving campus climate within the graduate community. Our students will have the opportunity to participate in the DECADE Student Planning Committee, where they may take a leadership role in planning future programming and serve on an advisory committee to DECADE administrators. Moreover, they will also function in an advisory role to UCI administrators about salient diversity and climate issues and can inform future school recruitment and retention initiatives. NRT faculty participation in this program will be tracked as part of the program evaluation. The Black in Physical Sciences group was formed in 2020 and offers an opportunity for Black individuals to connect with one another in an informal setting, meet colleagues and potential mentors in their fields, and exchange resources and ideas. The Womxn in Physics and Astronomy (WiPA) group at UCI strives to create a community for female physicists and decrease the attrition rate of female physics majors. The WiPA initiative includes a mentoring program that pairs undergraduate student mentees with graduate student/postdoc mentors. In 2021, the WiPA launched a Magnifying Voices in Physics speaker series, which highlights women and non-binary individuals of color in Physics and Astronomy.

9 Demographic Tables

The table below provides a comparison between the percentage of UCI PhD students who are female or URM in 2015-2020 to the percentage of PhDs granted nationally to female, Black, and Hispanic students in 2017 [52].

	Physics	Computer Science
Female students	28.5% UCI, 30.9% US	20.6% UCI, 22.6% US
URM students	12.7% URMs at UCI, 5.8% Hispanic in US, 2.9% Black in US	2.0% URMs at UCI, 5.3% Hispanic in US, 7.3% Black in US

The table below provides a detailed breakdown of the demographics of UCI PhD students.

Years 2015-2020	Total Applicants				Female Applicants				URM Applicants				International Applicants			
	Applied	Admitted	Enrolled	Time to degree	Applied	Admitted	Enrolled	Time to degree	Applied	Admitted	Enrolled	Time to degree	Applied	Admitted	Enrolled	Time to degree
Physics & Astronomy Department (0EB, 272, 666)	2248	437	165	5.49	501	126	50	5.38	247	66	25	5.62	954	60	35	5.61
Computer Science Department (201, 20N)	3492	522	248	5.18	801	119	54	5.64	80	13	6	5.33	2915	377	180	5.13
Statistics Department (891)	840	97	56	5.12	310	41	21	4.89	24	7	4	5.00	636	34	22	4.81
Mathematics Department (540)	1164	322	122	5.51	263	78	22	5.58	105	20	13	5.75	559	132	45	5.18
Chemistry Department (153, 325)	2541	980	281	5.10	983	447	130	5.17	290	111	36	5.44	828	136	37	4.91
Informatics Department (06G,19H)	752	202	95	5.26	378	104	47	5.25	49	18	12	4.33	516	107	47	5.45

Years 2015-2020	Total Applicants	Female Applicants	URM Applicants	International Applicants
	Enrolled/ Applied %	Enrolled/ Applied %	Enrolled/ Applied %	Enrolled/ Applied %
Physics & Astronomy Department	7.3%	10.0%	10.1%	3.7%
Computer Science Department	7.1%	6.7%	7.5%	6.2%
Statistics Department	6.7%	6.8%	16.7%	3.5%
Mathematics Department	10.5%	8.4%	12.4%	8.1%
Chemistry Department	11.1%	13.2%	12.4%	4.5%
Informatics Department	12.6%	12.4%	24.5%	9.1%

Years 2015-2020	Attrition of Female PhD Students	Attrition of URM PhD Students
Physics & Astronomy Department (0EB, 272, 666)	3	3
Computer Science Department (201, 20N)	3	1
Statistics Department (891)	2	
Mathematics Department (540)	6	2
Chemistry Department (153, 325)	14	8
Informatics Department (06G,19H)	4	

10 Diversification Strategy

UCI is committed to maintaining a high standard of excellence for its graduate programs, while increasing overall diversity and the number of incoming students each year. To achieve these goals, only predoctoral applicants with exceptional records of academic excellence and research, and outstanding letters of recommendation, are "pre-admitted." All other applicants are interviewed during recruitment weekends before a formal letter of acceptance is issued. During a potential student's visit to UCI, they have extensive contact with current graduate students, which has been an effective recruiting tool, particularly for applicants from underrepresented (UR) groups (underrepresented minorities [URMs], students with disabilities, and students from disadvantaged backgrounds), as they are able to interact with our diverse student population. In this context, the UCI Graduate Division performs a range of recruitment activities designed to identify and attract traditionally UR students to increase the diversity in UCI graduate programs. UCI Graduate Division offers five diversity fellowships: Eugene Cota-Robles Fellowship (ECR), Graduate Opportunity Fellowship (GOF), UCI Diversity Recruitment Fellowship (DRF), Faculty Mentor Plan (FMP), and President's Dissertation Year Fellowship (PDY). UCI recruits students from a range of institutions and graduate preparation programs such as the McNair Scholars Program, the Minority Access to Research Careers (MARC) program, and the Minority Biomedical Research Support (MBRS) program. Specifically, the Graduate Division regularly participates in campus-based graduate-recruitment fairs at campuses such as UCLA, UC Davis, and Cal State Los Angeles, which have large populations of students. Through these programs the NRT program will participate in national recruitment fairs/research conferences that target UR students in the STEM disciplines, such as SACNAS (Society for the Advancement of Chicanos and Native Americans). The effects of the efforts outlined earlier have resulted in significant increases for URM applicants and admissions at UCI. Over the last decade, URM enrollments in the Physical Sciences have increased over 30%, and in ICS they have increased by a factor of nearly 3.5. At the graduate level URM enrollments have increased by nearly 24% university-wide, with an excellent retention rate (83% over 5 years). Overall, UCI has raised graduate URM enrollment to over 20% of student enrollment.

11 Performance Assessment/Project Evaluation

Jessica Martone, Director of Research and Evaluation at the evaluation firm The Mark USA, will serve as the senior evaluator to conduct front-end, formative, and summative evaluations. A mixed-methods approach of quantitative and qualitative data sources and analysis will comprise evaluation procedures [53]. Evaluators will work closely with the NRT leadership team to establish baseline data and measurable targets, and to collect evidence to determine annual progress made toward program goals. Evaluations will track the early cohorts of participants beyond their involvement in the project, and project evaluation will be consistent with strategies identified in NSF guidelines [54]. The evaluator will complete one formative deliverable and one summative report per year, to be reviewed and used to improve project implementation. Evaluators will meet annually with the leadership team and present at Independent Advisory and leadership meetings.

Front-end Evaluation organizes the project. The evaluator will work with project leaders to develop benchmarks and timelines, instruments, and data collection procedures.

Formative Evaluation monitors the quality of project activities and provides feedback to the leadership team to strengthen implementation. The following questions will guide the formative evaluation: Are leaders implementing effective strategies and activities to successfully achieve project goals? Is the project on schedule and expected to meet output targets? Are obstacles limiting successful implementation of the project? If so, how can these obstacles be overcome? The formative evaluation will be conducted using two assessment methods: *1. Evaluation of project activities*: Participants complete Likert scale [55] and free-response evaluation forms on an annual post-survey to assess satisfaction, usefulness, and achievement of activity objectives. The summer, core curriculum, and lab internships components will be evaluated. Annual interviews will be conducted to assess project effectiveness and to identify areas where improvement is needed. *2. Tracking outputs compared to targets*: Evaluators will collect tracking data to monitor outputs compared to identified targets regarding the number of trainees recruited, the number of honorary fellows, the attendees of the Team Science Workshops, the number and participation level of monthly lunch meetings, communication seminars, annual symposia, and public science events. Evaluators will record the participation of URM students and women in all categories. Additionally, observational rubrics and surveys will be used to assess progress toward development of competencies in team science, communications, and ethics; see details below.

Concepts measured	Assessment method	Data timeline	Comparison (analysis)
1) Advance scientific knowledge and skills: Production of new knowledge at the interface of astrophysics and AI?			
• Development of research products (e.g., publications, proposals, conference abstract submissions/presentations)	Tracking data	End of acad. year	Y1 to Y5 (descriptive stats)
	Progress survey	End of acad. year	Baseline-post statistics
• Improved technical skills (e.g., data science, machine learning tools) • Improved interdisciplinary collaboration • Improved confidence in research skills	Trainee annual progress survey	End of acad. year	Baseline-post statistics
2) Workforce Development: Preparation trainees for the workforce in the fields of astrophysics and AI?			
• Improved communication, team work and ethics skills.	Annual survey	End of acad. year	Baseline-post statistics
	Comm rubrics.	Bootcamp, Public Ev.	Baseline-post statistics
• Increased trainees' career confidence • Increased interest in research career	Trainee annual progress survey	End of acad. year	Baseline-post statistics
• Number of URM and women trainees	Program tracking data	End of acad. year	Y1 to Y5 (descriptive stats)
• Employment, 1 year post-graduation	Follow-up survey	End of Year 5	Descriptive statistics
3) Institutional capacity building: Expansion of institutional capacity for interdisciplinary collaboration?			
• Diversification of faculty knowledge, teaching, and research skills • Increased partnerships, communication • Institutionalization and sustainability	Faculty survey	Years 3, 5	Compare year 3 and end of program (Descriptive statistics)

Summative Evaluation assesses achievement of project goals and the broader impacts on graduate students, the university, and the STEM community. Guiding evaluation questions are aligned with the goals of the NSF NRT program and are based on the goals of this project. Has this NRT project: (1) produced transformational astrophysics research and innovative new AI algorithms, (2) increased the training of a diverse group of graduate students with interdisciplinary skills, and (3) built new institutional support for interdisciplinary research?

Specifically, the target core competencies will be evaluated for each student: skills in team science, communications, and ethics. For team science and communication evaluations, rubrics developed specifically for this task will be adapted from those created during the previous NRT. For the communication evaluations, an example of a measure may be having attendees at a student presentation score it on its content, the student's delivery, and the visuals developed for the presentation. Each scoring category will be described in detail to allow for a reliable and systematic quantification of an audience member's experience about the communication quality, and provides for specific feedback which can be given to the student to focus their future efforts. The ethics evaluation will be done using surveys, where students will be asked to evaluate choices to be made in various realistic scenarios. The Program Sustainability Assessment Tool [56] will identify paths to sustainability.

12 Independent Advisory Committee

An independent advisory committee (IAC) will join the annual evaluation meetings, as well as in several meetings during the year to discuss progress of the NRT and gain advice on open issues. Members of the IAC will have experience working in communications, and at the interface of physics and AI. Examples include: Rich Caruana (Microsoft), Larry Smarr (UCSD), Lise Getoor (UCSC), Jesse Thaler (MIT), Sarah Demers (Yale), Peter Elmer (Princeton), and Gordon Watts (UW).

13 Recent Student Training Experiences

Training of students in the last five years by PI and co-PIs is summarized below:

Faculty	PhD Students	URMs	Women	Notes
Whiteson	12	2	5	16 undergrads, 10 URMs
Baldi	11	4	4	
Murgia	4	2	1	3 undergrads
Barwick	8	1	1	
Salazar Campo	5	0	4	Team Science Cert for 30 students

14 Results from Prior NSF Support

Daniel Whiteson and Pierre Baldi:NRT-DESE 1633631: Team Science for Integrative Graduate Training in Data Science and Physical Science. 09/2016 to 08/2021. $2,965,000. *Intellectual Merit* – Whiteson and Baldi were co-PIs on an earlier NRT which trained students to apply AI to the physical sciences. *Broader Impacts* – Promulgation of team science training techniques and graduation of a cohort of interdisciplinary students. **Simona Murgia**: Murgia has no prior NSF support. **Steven Barwick**: NSF grant 1607719: Precision Operation of Hexagonal Radio Array, 2016-2020, $0.5M. NSF MRI: Development of the ARIANNA High Energy Neutrino Telescope Instrument, 2011-2015,$1.9M. *Intellectual Merit* – Development of a novel data acquisition system based on the SST chip *Broader Impacts* – 10 papers in peer-reviewed journals [57, 58, 59, 60, 61, 62, 63, 64, 65, 66]. Inclusion of women and Hispanic graduate students. **Maritza Salazar Campo**: Awards 1262754, 1262745 ($189,000, 2013) Building Resources through Integrating Disciplines for Group Effectiveness in Science *Intellectual Merit* – new team training curriculum for faculty in academia to foster and assess integrative capacity in teams' potential to combine diverse knowledge and insights. *Broader Impacts* – This training grant resulted in two validated training programs focused on team strategic planning and improved cognitive integration in scientific teams.

Facilities, Equipment, and Other Resources

Campus Computing Facilities; Students and faculty of the NRT program will have access to two high-performance computing facilities, associated high-performance data storage, and dedicated research high-bandwidth networking for big data applications. The Green Planet Cluster is sited in the School of Physical Sciences and is comprised of 250 nodes (6000 CPU), data storage of 200 TB, and affiliated support staff (4 FTEs). The HPC3 Cluster is a campus-wide resource comprised of approximately 9000 CPUs, 52 GPUs, nearly 1 PB of storage and affiliated support staff (3 FTEs in the Office of Information Technology).

Participating Campus Research Labs (core faculty labs): *Prof. Daniel Whiteson*: Dr. Whiteson's lab is at the forefront of high-energy particle physics with collaborations that span the globe. *Prof. Pierre Baldi*: Dr. Baldi's lab is renowned for developing machine learning applications across the natural sciences, from physics to chemistry expert systems, and modeling metabolic, signaling, and regulatory networks in systems biology. *Prof. Simona Murgia*: Dr. Murgia is a member of the Fermi collaboration and does research at the forefront of *gamma*-ray astrophysics. *Prof. Steve Barwick*: Dr. Barwick is an international leader in neutrino astrophysics, and a member of the ARIANNA experiment in Antarctica. *Prof. Maritza Salazar*: The Team Science Accelerator Lab is comprised of graduate students who study and research organizational phenomena-team performance, multicultural collaboration and decision making. *Prof. Steph Sallum* : Dr. Sallum studies planet formation directly with high-resolution imaging and interferometry and is involved in Keck and the Thirty-Meter Telescope projects. *Prof. Stephan Mandt*: Dr. Mandt is an expert in deep generative models.

Affiliated Campus Labs Prof. Aaron Barth studies black holes and quasars using Keck and Hubble. Prof. David Buote studies galaxy formation using Chandra. Prof. David Kirkby studies cosmic acceleration using DESI and LSST. Prof. Aomawa Shields studies exoplanet climate and atmosphere using climate models and telescope data. Prof. Paul Robertson searches for habitable exoplanets using the Habitable Zone Planet Finder and NEID instruments. Profs. Jonathan Feng and Tim Tait study theory of dark matter production and observation.

Team Science Accelerator Lab Prof. Salazar Campo leads the Team Science Accelerator Lab, together with Profs. Judy Olson and Greg Olson. They are leaders in research and application of the principles of team science, and offer extensive experience in team science training, including a previous NSF NRT as well as several other major projects.

Off-campus Labs Profs. Josh Smith, Geoffrey Lovelace, and Jocelyn Read, California State University Fullerton, experts in gravitational wave observation and members of LIGO.; see letter of support. Drs. Peter Sorensen and Scott Kravitz, detection of dark matter particles using large underground detectors, members of the LZ collaboration; see letter of support.

Other Campus Resources: The Office of Access and Include (OAI) in the School of Information and Computer Science and The Office of Access, Outreach and Inclusion (OAOI) in the School of Physical Sciences were established to support recruitment, retention, and graduation of undergraduate and graduate students from historically excluded communities who are currently underrepresented in STEM. These offices offer programs to remove barriers to the recruitment, retention, and graduation of talented students. They provide access to a welcoming and supportive environment that respects and embraces diversity and individual differences to empower everyone to learn and achieve to their highest potential. They equip talented students from different backgrounds with critical academic and career skills to be leading scientists, engineers, and information and computer scientists. The OAI and the OAOI are committed to working with the NRT program to provide the necessary support for recruitment, mentoring, and retention of trainees with a particular emphasis on broadening participation of underrepresented groups.

The Joint Degree Program in Computational Science, a joint program between UC Irvine and San Diego State University, provides students at SDSU the opportunity to collaborate with UC

Irvine faculty and obtain a PhD.

Data Management Plan

Overview

The NRT program will produce knowledge, software, and data throughout its lifetime as outputs from its research, education, and workforce development, broadening participation, and collaboration and knowledge transfer efforts. In addition to modeling and software development, the team will conduct program evaluation, focus group, and other interview-type activities that will produce data to be coded for statistical analysis. The type and formats of these outputs will vary widely, and will include the source code for software, training materials, journal and conference papers and, finally, open datasets. In this document, we outline a data management plan for the overall program.

The PI and co-PIs will be jointly responsible for overseeing and verifying the program data management plan. As the program will be a distributed organization across UCI and involving other campuses, the day-to-day activities required by this plan will be executed by the program Executive Board and the participating institutional faculty and their teams. To help verify that the institutional teams are complying with this plan (and to give an opportunity to evolve the plan along with the activities of the program), we will include the management of data products as a topic in the annual planning cycle.

Research Data Artifacts

Websites, Public Outreach, and Presentations: The NRT program will maintain a web presence for the program itself (linking with external partners such as the CERN Open Data Portal or the Open Science Grid). The website and its contents will be kept in a git repository hosted on GitHub (or a similar openly accessible repository). As public outreach material is created, it will be added to this repository and webpage.

A key part of engaging the wider community will be creating posters and presentations and attending meetings. The material presented will be archived in PDF format and included publicly on the website. When appropriate videoconferencing technology is available, meeting recordings will be uploaded to YouTube. Training material associated with any classes taught by the program will be kept in a similar manner; in addition to the traditional materials (slides, webpages, exercises), the environments and software used for training will be handled as software artifacts as described below.

All material will be made available under an open-source license. When possible, all data will be made available under the CC-BY 4.0 [1] open- source license unless limited by the conference or workshop where the material was shown.

Software and Algorithms: Some of the NRT program's activities will result in software: for example, novel AI models or reference algorithm implementations. This work will be kept on GitHub (or a similar openly accessible repository) where possible, including both the software and related development artifacts (discussions, issues, documentation pages). Code will be made available under an open-source license; The NRT program will use the Apache Software License 2.0 [2] by default but may utilize other licenses as necessary to contribute to external or preexisting projects. If The NRT program staff contribute to external software projects, we will encourage our collaborators to work in a similar manner. The NRT program will comply with best practices for software citations. [3] Software released by the project will be assigned a DOI. We will cite software in our papers and articles when possible, and our source code will contain instructions for citations.

[1] https://creativecommons.org/licenses/by/4.0/
[2] https://www.apache.org/licenses/LICENSE-2.0
[3] https://www.force11.org/software-citation-principles

Articles and Reports: The NRT program will disseminate its research through conference papers, journal articles, and other reports. Preference will be given to publishing in open-access journals; pre-prints or final copies will be uploaded to arXiv.org.

Datasets: Reference datasets for training AI algorithms are an essential component of any AI research endeavor. To ensure proper management and to ensure the broadest access to this data, The NRT program will employ the services of a dedicated data curation expert. A central responsibility of the data curator will be to ensure that all the NRT program datasets are made available both within the team and to the broader community in a free, open, portable, and citable fashion. The datasets will be made available under an appropriate open-source license, assigned a DOI, and made available from the NRT program's science gateway. Datasets associated with a journal paper will be published alongside the paper itself when possible.

The NRT program does not plan to keep personally identifiable information, user authentication, or other private data as research data. The data generated by this project will not pose a disclosure risk. All data will be de-identified before posting to the Web site established by the principal investigators. For this project, informed consent statements will use language that will not prohibit the data from being shared with the research community.

The NRT program will preserve its data in non-proprietary open formats. For example, presentations and articles will be made available in PDF format; source files will be made available as simple text files where possible. Reference datasets for training AI algorithms will be kept in a variety of formats, but will always include one format that is easily readable by standard Python-based tools from the AI and data science communities, such as HDF5. The data may also be made available via ROOT file formats, when appropriate (e.g., data related to particle physics experiments); however, software tools will be available to read the formats into Python-based tools. Source code will include integral comments and relevant build code necessary for reproducing the binary artifacts.

Datasets produced by the NRT program will be made publicly available as soon as possible, and the team will be encouraged to work in the open. If collaborations with one of the experimental projects require the use of data or algorithms that are non-public, the NRT program will attempt to keep its efforts separable from the proprietary work (allowing public sharing) or, failing that, create representative sample data that can be shared with others. Raw data belonging to the outside experiments will remain bound by the respective experiments' data sharing policies.

Archiving

EZID, a service from the University of California Curation Center, facilitates data sharing, data reuse, data citation, and data attribution, and creates Digital Object Identifiers (DOI) for research data will be used for the technical reports. The Merritt Repository will make data publicly available. Research data and records will be maintained for as long as they are of continuing value to the researchers and project collaborators. Data will be retained subject to the University of California retention policy.

All public data will be deposited in the Merritt Repository Service from the University of California Curation Center (UC3), which has capabilities to manage, archive, and share digital content. Merritt allows access to the public via persistent URLs, provides tools for long-term data management, and permits permanent storage options. Merritt has built-in contingencies for disaster recovery, including redundancy and recovery plans.

The program will make a best-effort attempt to keep the relevant documentation and software source code together to best enable reproducibility of the results by future projects.

References

[1] P. Baldi. *Deep Learning in Science.* Cambridge University Press, Cambridge, UK, 2021.

[2] P Baldi, P Sadowski, and D Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, 5, 2014.

[3] National Academy of Engineering. *Advancing Diversity in the US Industrial Science and Engineering Workforce: Summary of a Workshop.* The National Academies Press, Washington, DC, 2014.

[4] Robert Croyle. The national cancer institute's transdisciplinary centers initiatives and the need for building a science of team science. *American Journal of Preventive Medicine*, 25, 2008.

[5] Daniel Stokols, Shalini Misra, Richard P. Moser, Kara L. Hall, and Brandie K. Taylor. The ecology of team science: Understanding contextual influences on transdisciplinary collaboration. *American Journal of Preventive Medicine*, 35(2, Supplement):S96–S115, 2008. The Science of Team Science.

[6] Katy Börner, Noshir Contractor, Holly J Falk-Krzesinski, Stephen M Fiore, Kara L Hall, Joann Keyton, Bonnie Spring, Daniel Stokols, William Trochim, and Brian Uzzi. A multi-level systems perspective for the science of team science. *Science Translational Medicine*, 2(49):49cm24–49cm24, 2010.

[7] Baruch Fischhoff. The sciences of science communication. *Proceedings of the National Academy of Sciences*, 110(Supplement 3):14033–14039, 2013.

[8] J Cham and D Whiteson. *We Have No Idea.* Riverhead/PRH, 2017.

[9] J Cham and D Whiteson. *Frequently Asked Questions About the Universe.* Riverhead/PRH, 2021.

[10] J Cham and D Whiteson. Daniel and jorge explain the universe, 2018.

[11] J Cham and D Whiteson. Elinor wonders why, 2020.

[12] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for Exotic Particles in High-Energy Physics with Deep Learning. *Nature Commun.*, 5:4308, 2014.

[13] Alexis Romero, Daniel Whiteson, Michael Fenton, Julian Collado, and Pierre Baldi. Safety of Quark/Gluon Jet Classification. 3 2021.

[14] Michael James Fenton, Alexander Shmakov, Ta-Wei Ho, Shih-Chieh Hsu, Daniel Whiteson, and Pierre Baldi. Permutationless Many-Jet Event Reconstruction with Symmetry Preserving Attention Networks. 10 2020.

[15] Daniel Guest, Julian Collado, Pierre Baldi, Shih-Chieh Hsu, Gregor Urban, and Daniel Whiteson. Jet Flavor Classification in High-Energy Physics with Deep Neural Networks. *Phys. Rev. D*, 94(11):112002, 2016.

[16] Jessica N. Howard, Stephan Mandt, Daniel Whiteson, and Yibo Yang. Foundations of a Fast, Data-Driven, Machine-Learned Simulator. 1 2021.

[17] Jacob Hollingsworth and Daniel Whiteson. Resonance Searches with Machine Learned Likelihood Ratios. 2 2020.

[18] Chase Shimmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Søgaard. Decorrelated Jet Substructure Tagging using Adversarial Neural Networks. *Phys. Rev. D*, 96(7):074034, 2017.

[19] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Enhanced Higgs Boson to $\tau^+\tau^-$ Search with Deep Learning. *Phys. Rev. Lett.*, 114(11):111801, 2015.

[20] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. Parameterized neural networks for high-energy physics. *Eur. Phys. J. C*, 76(5):235, 2016.

[21] A. Anker et al. White Paper: ARIANNA-200 high energy neutrino telescope. 4 2020.

[22] Gary M Olson and Judith S Olson. Distance matters. *Human–computer interaction*, 15(2-3):139–178, 2000.

[23] Amanda L Vogel, Kara L Hall, Stephen M Fiore, Julie T Klein, L Michelle Bennett, Howard Gadlin, Daniel Stokols, Linda C Nebeling, Stefan Wuchty, Kevin Patrick, et al. The team science toolkit: enhancing research collaboration through online knowledge sharing. *American journal of preventive medicine*, 45(6):787–789, 2013.

[24] Maritza R Salazar, Theresa K Lant, Stephen M Fiore, and Eduardo Salas. Facilitating innovation in diverse science teams through integrative capacity. *Small Group Research*, 43(5):527–558, 2012.

[25] Jacob Hollingsworth, Michael Ratz, Philip Tanedo, and Daniel Whiteson. Efficient sampling of constrained high-dimensional theoretical spaces with machine learning. 3 2021.

[26] Jeong Han Kim, Kyoungchul Kong, Benjamin Nachman, and Daniel Whiteson. The motivation and status of two-body resonance decays after the lhc run 2 and beyond. *Journal of High Energy Physics*, 2020(4), Apr 2020.

[27] Nathaniel Craig, Patrick Draper, Kyoungchul Kong, Yvonne Ng, and Daniel Whiteson. The unexplored landscape of two-body resonances. 2016.

[28] Mohammad Abdullah, Kevin Bauer, Luis Gutierrez, John Sandy, and Daniel Whiteson. Searching for spin-3/2 leptons. *Physical Review D*, 95(3), Feb 2017.

[29] Arvind Rajaraman, Tim M. P. Tait, and Daniel Whiteson. Two Lines or Not Two Lines? That is the Question of Gamma Ray Spectra. *JCAP*, 09:003, 2012.

[30] Daniel Whiteson, Michael Mulhearn, Chase Shimmin, Kyle Cranmer, Kyle Brodie, and Dustin Burns. Searching for ultra-high energy cosmic rays with smartphones. *Astropart. Phys.*, 79:1–9, 2016.

[31] UC Office of the President. Graduate student well-being survey report. 2017.

[32] Geoff Davis. 3. improving the postdoctoral experience: An empirical approach. In *Science and engineering careers in the United States*, pages 99–128. University of Chicago Press, 2009.

[33] B. P. Abbott et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophys. J. Lett.*, 848(2):L12, 2017.

[34] Fabian Jirasek, Robert Bamler, and Stephan Mandt. Hybridizing physical and data-driven prediction methods for physicochemical properties. *Chemical Communications*, 56(82):12407–12410, 2020.

[35] Yadong Lu, Julian Collado, Daniel Whiteson, and Pierre Baldi. Sparse autoregressive models for scalable generation of sparse images in particle physics. *Physical Review D*, 103(3):036012, 2021.

[36] Brenden K Petersen, Mikel Landajuela Larma, T Nathan Mundhenk, Claudio P Santiago, Soo K Kim, and Joanne T Kim. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. *arXiv preprint arXiv:1912.04871*, 2019.

[37] Samuel Kim, Peter Y Lu, Srijon Mukherjee, Michael Gilbert, Li Jing, Vladimir Čeperić, and Marin Soljačić. Integration of neural network-based symbolic regression in deep learning for scientific discovery. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[38] M. G. Aartsen et al. Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector. *Science*, 342:1242856, 2013.

[39] M. G. Aartsen et al. Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data. *Phys. Rev. Lett.*, 113:101101, 2014.

[40] A. Anker et al. White Paper: ARIANNA-200 high energy neutrino telescope. *arXiv:2004.09841*, 2020.

[41] S. Barwick for the ARIANNA collaboration. Science Case and Detector Concept for the ARIANNA High Energy Neutrino Telescope at Moore's Bay, Antarctica. *PoS(ICRC2021)1067*.

[42] S. Hallmann, B. Clark, C. Glaser and D. Smith for the IceCube-Gen2 collaboration. Sensitivity studies for the icecube-gen2 radio array. *PoS(ICRC2021)1183*.

[43] A. Anker for the ARIANNA collaboration. A novel trigger based on neural networks for radio neutrino detectors. *PoS(ICRC2021)1069*.

[44] Astrid Anker et al. Improving sensitivity of the arianna detector by rejecting thermal noise with deep learning. in preparation, 2021.

[45] C. Glaser, S. McAleer, P. Baldi and S. Barwick for the ARIANNA collaboration. Deep learning reconstruction of the neutrino energy with a shallow askaryan detector. *PoS(ICRC2021)1051*.

[46] B. P. Abbott et al. Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *Astrophys. J. Lett.*, 848(2):L13, 2017.

[47] M. Ackermann et al. Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data. *Phys. Rev. Lett.*, 115(23):231301, 2015.

[48] Simona Murgia. The fermiÄìlat galactic center excess: Evidence of annihilating dark matter? *Annual Review of Nuclear and Particle Science*, 70(1):455–483, 2020.

[49] M. Ajello et al. Fermi-LAT Observations of High-Energy γ-Ray Emission Toward the Galactic Center. *Astrophys. J.*, 819(1):44, 2016.

[50] Shahab Joudaki, Manoj Kaplinghat, Ryan Keeley, and David Kirkby. Model independent inference of the expansion history and implications for the growth of structure. *Phys. Rev. D*, 97(12):123501, 2018.

[51] Stacy-Ann A Allen-Ramdial and Andrew G Campbell. Reimagining the pipeline: Advancing stem diversity, persistence, and success. *BioScience*, 64(7):612–618, 2014.

[52] NSF. 2020.

[53] Michael Quinn Patton. Enhancing the quality and credibility of qualitative analysis. *Health services research*, 34(5 Pt 2):1189, 1999.

[54] J Frechtling, MM Mark, DJ Rog, V Thomas, H Frierson, S Hood, and G Hughes. The 2010 user-friendly handbook for project evaluation. national science foundation, 2010, 2010.

[55] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4):396, 2015.

[56] Annaliese Calhoun, Avia Mainor, Sarah Moreland-Russell, Ryan C Maier, Laura Brossart, and Douglas A Luke. Peer reviewed: Using the program sustainability assessment tool to assess and plan for sustainability. *Preventing chronic disease*, 11, 2014.

[57] S. W. Barwick et al. Radio detection of air showers with the ARIANNA experiment on the Ross Ice Shelf. *Astropart. Phys.*, 90:50–68, 2017.

[58] S. W. Barwick, E. C. Berg, D. Z. Besson, G. Gaswint, C. Glaser, A. Hallgren, J. C. Hanson, S. R. Klein, S. Kleinfelder, L. Köpke, I. Kravchenko, R. Lahmann, U. Latif, J. Nam, A. Nelles, C. Persichilli, P. Sandstrom, J. Tatar, and E. Unger. Observation of classically 'forbidden' electromagnetic wave propagation and implications for neutrino detection. *JCAP07(2018)055*.

[59] S. W. Barwick, E. C. Berg, D. Z. Besson, T. Duffin, J. C. Hanson, S. R. Klein, S. A. Kleinfelder, K. Ratzlaff, C. Reed, M. Roumi, T. Stezelberger, J. Tatar, J. Walker, R. Young, and L. Zou. Design and performance of the ARIANNA HRA-3 neutrino detector systems. *IEEE Transactions on Nuclear Science*, 62(5):2202–2215, oct 2015.

[60] S. W. Barwick et al. A first search for cosmogenic neutrinos with the ARIANNA Hexagonal Radio Array. *Astroparticle Physics*, 70:12–26, 2015.

[61] S W Barwick et al. Radar Absorption, Basal Reflection, Thickness, and Polarization Measurements from the Ross Ice Shelf. *J. Glaciol.*, 61(227):438–446, 2015.

[62] S. W. Barwick et al. Time-domain response of the ARIANNA detector. *Astroparticle Physics*, 62:139–151, 2015.

[63] A. Anker et al. Targeting ultra-high energy neutrinos with the ARIANNA experiment. *Advances in Space Research*, 64:2595 – 2609, jun 2019.

[64] A. Anker, S. W. Barwick, H. Bernhoff, D. Z. Besson, N. Bingefors, D. García-Fernández, G. Gaswint, C. Glaser, A. Hallgren, J. C. Hanson, S. R. Klein, S. A. Kleinfelder, R. Lahmann, U. Latif, J. Nam, A. Novikov, A. Nelles, M. P. Paul, C. Persichilli, I. Plaisier, T. Prakash, S. R. Shively, J. Tatar, E. Unger, S. H. Wang, C. Welling, and S. Zierke. Neutrino vertex reconstruction with in-ice radio detectors using surface reflections and implications for the neutrino energy resolution. *Journal of Cosmology and Astroparticle Physics*, 11:030–030, nov 2019.

[65] A. Anker et al. A search for cosmogenic neutrinos with the ARIANNA test bed using 4.5 years of data. *JCAP*, 03:053, 2020.

[66] A. Anker et al. Probing the angular and polarization reconstruction of the ARIANNA detector at the South Pole. *JINST*, 15(09):P09039, 2020.