A Novel Multidisciplinary Graduate Education Program in Data Science

Anthony Kuh, Daniel Port, Narayana Prasad Santhanam, London Thompson
University of Hawaii
Honolulu, HI, USA
kuh@hawaii.edu, dport@hawaii.edu, nsanthan@hawaii.edu, london.thompson@hawaii.edu

Mary Lee
Pacific Policy Research Center
Honolulu, HI, USA
marylee@pprchawaii.org

***Abstract*—There has been an explosion of growth in using AI, data science, and machine learning in all aspects of our daily life. There is a global competition among governments, industry, and academic institutions to lead research and development in this area. This paper discusses a novel multidisciplinary graduate education and research program at our institution to help develop a trained workforce to meet the demands required to understand and develop AI, data science and machine learning technologies. The program brings together faculty and students in engineering, computer science, and social science to build a traineeship program where cohort teams study fundamental and applied data science research, using compact modules across courses to personalize instruction and prepare each trainee with skills tailored to their prior experience and future career goals.**

***Index Terms*—multidisciplinary data science, module-based learning**

I. Introduction

Academic institutions around the world have developed multidisciplinary AI, data science, and machine learning research and education programs to meet the increasing demands of creating a trained workforce integrating AI technologies into our society. In the United States the National Science Foundation (NSF) has provided considerable funding for core programs and special solicitations in AI, data science, and machine learning. This paper discusses a recently funded NSF National Research Traineeship (NRT) program at University of Hawai'i at Mānoa (UHM).

A. NSF initiatives

The National Science Foundation (NSF) have invested considerable resources to funding AI and data science research and education. In fiscal year 2018 NSF introduced the "Ten Big Ideas" to build on the foundation of NSF funded research [1]. One of the ideas introduced was "Harnessing the Data Revolution". This consisted of three solicitations: data science fundamentals, 'Transdisciplinary Research in Principles of Data Science (TRIPODS)", data science applications, "Institutes for Data-Intensive Research in Science and Engineering", and data science education, "Data Science Corps (DSC)". Another solicitation that has had several funding rounds is the National Artificial Intelligence Research Institutes [2]. This solicitation led to funding 25 large AI projects in many areas from AI theory to AI for agriculture to AI optimization to AI for dynamic systems. Other solicitations focused on the theory and fundamentals of AI and data science including the "Mathematics Foundations of AI" [3]. The institutions that were awarded grants formed communities of multidisciplinary groups that advanced the theory, applications and education of AI, data science, and machine learning.

The NSF puts an emphasis on graduate education and research through their National Research Training (NRT) solicitation that is an annual competition that "supports interdisciplinary, evidence-based traineeships that advance ways for graduate students in research-based Masters and Doctoral degree programs to pursue a range of STEM careers" [4]. The goal is to build up graduate programs around the country to address workforce development, emphasizing institutional capacity and broadening participation in STEM education. The NRT program aims to shape and support innovative training of STEM graduate students in high priority convergent research areas while increasing the capacity of US graduate programs to produce varied cohorts of trained STEM professionals for careers in industry and academia. With NSF's investment in AI and data science, they had special NRT tracks focused on AI and data science. Awarded grants are for five years with a budget up to $3 million. Motivation for the program comes from several studies examining graduate education in the United States [5].

B. NRT DESCARTES program

At UHM we put together an interdisciplinary group of faculty and students to conduct research and education in AI and data science. The group put together a project entitled NRT Data in Engineering and Society: Converging Applications, Research, and Training Enhancements for Students, (NRT-DESCARTES) that was funded by the NSF starting in August, 2023. The theme of NRT-DESCARTES is AI for dynamical, real-time and online systems that are critical to engineering infrastructure and healthcare. The project trains a new workforce to address cutting edge problems in the above areas. Our novel multidisciplinary graduate program in data science distinguishes itself through its specific focus on AI for dynamic engineering infrastructure systems, a comprehensive integration of a flexible module-based curriculum powered by the Morea framework, a strong emphasis on

The authors acknowledge support from the NSF NRT grant 2244574 and NSF grant 2142987.

convergent research across diverse disciplines and research paradigms, and an innovative micro-credentialing system for personalized multidisciplinary certification in graduate-level data science while also expanding our undergraduate data science programs. The project provides convergent research in two dimensions; across disciplines (engineering, computer science, business, and social science) and across research emphasis (theory to use inspired).

The motivation for this project comes from both the nation (Industries of the Future from the Office of Science and Technology Policy [6]) and the state (Hawai'i Science and Technology Plan [7] to promote AI). The proposal also aligns with the UHM Data Science Institute (DSI) (development of an undergraduate data science program) and Hawai'i IT workforce needs in AI [8]. Our vision is to provide an AI-enabled engineering workforce able to tackle the difficult problems in dynamic systems that are critical for engineering infrastructure and healthcare while considering socioeconomic, ethical and legal concerns. The project will have a diverse group of NRT students examining both technical and social science perspectives using a novel fine-grained interdisciplinary learning curriculum based on modules, that lead to systematic understanding of data science methods as practiced across different disciplines. The project has four goals:

1) A productive and immersive student experience: Students will gain a unique experience participating in NRT-DESCARTES based on integrating data science principles from engineering, computer science, business, and social science, using a carefully chosen research agenda.
2) The research focuses on AI for dynamic systems and online solutions that are critical to engineering infrastructure and healthcare by considering both theoretical and use inspired research from technical and social science perspectives.
3) The program is building a student body through targeted recruiting and mentoring. Our students come from varied backgrounds to promote convergent research.
4) The program will assist in building a sustainable data science program at UHM by working with NRT faculty and students with support from our constituents including the UHM community, industry, and external stakeholders.

The paper discusses NRT-DESCARTES and is organized as follows. In Section II we present our collaborative personalized education program. In Section III we discuss our integrated multidisciplinary research program. Section IV summarizes our initial work implementing the program including preliminary evaluation of the program. Section V provides a summary and discusses future directions for the program.

II. Collaborative Personalized Education

Our program is motivated by national and Hawai'i needs and the increasing number and variety of career opportunities for graduate students across various disciplines, including engineering, social sciences and computer science [9]. Students conducting research in data science from engineering and computer science increasingly find opportunities in economics, business, finance, healthcare, public policy, and other social science disciplines. At the same time students conducting research in data science from social science increasingly find that they enter jobs that require knowledge of mathematical and quantitative approaches and understanding of engineering problems.

To accommodate NRT students from different backgrounds and train them for the diverse careers in data science, we require each student to obtain a T-shaped background with breadth and depth in data science. Students coming from an engineering, social science, or computer science background will obtain breadth through taking core courses outside their disciplines and through a finer-grained approach to course structures through modules. Breadth will also be obtained through engagement in research through their cohort group, mentoring, presentations, and workshops. Students will obtain depth by taking follow-up courses (primarily in their department) to prepare them for their research. The NRT-DESCARTES program will have M.S. and Ph.D students obtaining degrees in their respective department and students will also work towards different micro-certificates in different aspects of data science that are being developed at UHM.

-

A. Summary of Enhanced Learning Experience

The program's main objective is to develop a convergent education and research program to train students for industry and academic careers. The program will have students engaged in a novel education program to obtain a broad, engaged, and enhanced series of learning experiences. The following gives a summary list of features of the program.

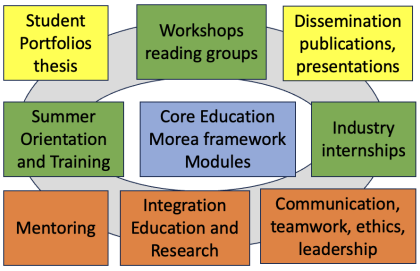

Fig. 1. NRT Education: core education at center in blue; other activities in green, student products in yellow, student skills in orange

- **Morea Framework using Modules** At UHM there are currently many data science and machine learning courses spread over different colleges and departments. The NRT-DESCARTES provides a mechanism for taking these courses and additional new courses to create a multidisciplinary data science graduate curriculum using the **Morea framework**, [10]. The Morea framework and modules helps break down the curriculum into finer grained structures than courses and offers numerous advantages to development of our data science program that can meet the needs of the students from different disciplines.
- **Student Portfolios** The traditional "deliverable" for graduate education is a thesis or a capstone project. NRT-DESCARTES students in the degree program will document

their progress through the modules, courses and research progress (using GitHub) as a "portfolio". Every activity in the program will contribute artifacts to the portfolio, which can be a mixture of reports, badges and micro-credentials, software code, hardware designs, photos, data sets, or video, providing a multi-media perspective on student progress.

- **Summer Training and Orientation** New NRT students come from diverse backgrounds across Engineering, Economics, Business, Mathematics/ Statistics, and Natural Sciences, with different goals and expectations with AI, Data Science, and Machine Learning. An orientation program at the beginning of the school year will train new NRT students about our curriculum approach using modules and the Knowledge Map. This includes a focused one week course on Data Science and Machine Learning basics and engineering applications for students who are completely new to the area or want a refresher course. The class will be coordinated by an NRT investigator, but taught by returning NRT students.
- **Workshops** Each semester will focus on a specific research area where an NRT investigator will coordinate a seminar series in the area. Over each winter break, we plan to hold workshops among NRT students and faculty, the executive advisory committee (EAC), and the NRT evaluator to engage with the external community. The workshop will feature keynote talks from distinguished speakers from industry and academia, panels led by NRT investigators featuring key research and educational developments, and a student research poster session. We will also hold mini-workshops involving NRT investigators and external researchers.
- **Mentoring** In addition to learning through courses and conducting research, students become involved in mentorship of students in (undergraduate and graduate) NRT courses. This will enhance their overall learning experience as they give back to NRT-DESCARTES.
- **NRT Internships and industry partners** The NRT Internship program will support our students interested in AI/ML applications in industry, provide financial support for students, connect students with potential employers for practical experience as well as assist with their technical, teamwork, and communication skills. The program integrates the internship into the curriculum either during summer or part-time during the academic year.

B. Morea Framework and Modules

A **module** is a portion of a course, focused typically on one topic and covering between a week or two of regular coursework. Morea will host the material for the modules—lecture videos, readings, associated programming material, and assessment instruments. Every module will have a host department and instructors that manage their content. Some features of modules are:

- Modules are easily designed and can be tailored to different instructors. The modules are dynamic where content can be easily, added, enhanced, and pruned.
- Modules will be shared across all departments, reducing duplication of content. Students will have access to all modules during their time in the program.
- The modules allow for more **flexibility** over courses and traditional silos that exist in departments. Development of new courses and cross-listing courses takes time and requires administrative approval. Modules are easily introduced and modified by NRT faculty with consultation by the NRT team without the overhead of administrative approval. A module can be introduced into a number of different courses across different departments. A module can also be introduced and be standalone for a possible introduction in a course at a later date. The module could serve as a short course or introduced at a workshop as a tutorial. This is important in a data science program as the field is rapidly evolving. Examples of new module topics include Federated Learning and Machine Learning for Pandemics.
- The modules are **adaptable** and can be changed depending on instructor for both content and methods of delivery (traditional classroom instruction or more virtual online modes). The modules can also be personalized to different student requirements.
- Information about the curriculum will be made available through a **Knowledge Map** that is being developed. The Knowledge Map uses Large Language Models (LLMs) trained on our modules to provide essential information about modules (their content and interrelationship with other modules). As we progress, we will use LLMs to resolve multiple notations, suggest similarities in modules, and give students a better insight into how content they are learning is used across more advanced modules.
- Morea is critical for our **evaluation and assessment** of the progress of the NRT program. Both students and instructors will participate in the evaluation and assessment.

Example module: A fundamental topic in data science and machine learning is linear regression that appears in Engineering, Computer Science and Economics. This important topic is presented from different perspectives for different courses. In ECE 345 (Linear Algebra and Machine Learning) the core module for the linear regression problem is presented by forming and solving the least squares problem. A senior course, ECE 445 (Introduction to Machine Learning) revisits the core module, adds regularization (ridge and LASSO), and expands on the significance of features. Probability courses discuss connections to linear regression using maxumimum likelihood and maximum a priori estimates for Gaussian distributions. A graduate course, ECE 645 (Machine Learning) expands on the core module going deeper into advanced topics such as reproducing kernel Hilbert spaces (RKHS) and kernel ridge regression, connections to Bayesian models, and learning theory. From the Economics Department, there is an econometrics sequence Econ 425 (covering data exploration and regression) and Econ 427 (focusing on prediction and causal inference) where the module would be adapted to economic topics. From the Information and Computer Science

(ICS) department, there is ICS 434 where linear regression is discussed, but with more emphasis on implementation using Python. Modules from the medical school include applications in fMRI that identify graphical models using regression as one of their core building blocks for identifying protein and gene networks.

In our NRT program, an instructor would have access to the different variations of the modules for linear regression and can tailor their course to use or modify the linear regression model to their course and preferences. They could use or adapt material, applications, code, and demos. Instructors and students would then note the wide range of approaches and applications of linear regression in various areas.

C. Recruiting and Mentoring

NRT-DESCARTES will have students that are supported by the NSF NRT grant and include students supported by other funding or teaching assistantships. The goal is to build a community of faculty, students, and staff to develop an innovative and sustainable graduate data science program. The NRT supported students are required to be US graduate students.

Recruiting: To recruit US graduate students, we will reach out to students on the U.S. mainland and in Hawai'i. We will work with the UHM administration which is putting an emphasis on developing programs in data science and developing more multidisciplinary programs and also look to our community to recruit students. This includes women students, Native Hawaiian and Pacific Islander (NHPI) students, and students from the military.

We will recruit NRT graduate students through the Vertically Integrated Projects (VIP) Consortium The (VIP) Consortium, led by the Georgia Institute of Technology, consists of more than 30 institutions. VIP programs seek to foster long-term, in-depth, project-based learning to engage students and better prepare them for future careers. VIP teams are vertically integrated: they consist of a faculty mentor, graduate student researchers, and UG students. The projects are long-term and are often based on an externally funded research topic. We will recruit NRT graduate students from the VIP program as they have experience as UG students conducting research, working in team environments, and have leadership skills.

The Native Hawaiian Science and Engineering Mentorship Program (NHSEMP) was established in 2001 and in its first five years doubled NHPI students enrolled in Engineering to 109 students. NHSEMP has been supported by NSF through the Islands of Opportunities Alliance and the Louis Stokes Alliances for Minority Participation (IOA-LSAMP). The program has reversed problems of low attraction and retention for NHPI student enrollment at UHM. Our NRT program will work with NHSEMP to provide information and opportunities to underrepresented students from local K-12 schools through joint outreach programs. Information sessions will also target students around the country who are looking to integrate their academic and career interests at UHM.

The Society of Women Engineers (SWE) is a large student chapter at UHM. SWE has been recognized with several national awards, including 2012 and 2013: Outstanding Collegiate Section - Silver Level. SWE provides peer mentoring, outreach activities, and networking with industry professionals. SWE members are mainly undergraduate US students who are potentially another pool of prospective graduate students. We have also worked with SWE on outreach activities to interest K-12 students in STEM education.

The military has many bases in Hawai'i and we are using our contacts between our institution and the military to recruit both military service members and significant others to enter the NRT program (both supported and unsupported). UHM has a liason (Office of Veteren Student Support) that has been particularly helpful in reaching out to military students. We will also look to get students from industry (especially our former undergraduate students).

Mentoring: Faculty advisors will provide dedicated mentorship and academic support. Current students supported by NHSEMP have indicated that in addition to the financial support, the most important factor in their success is the mentoring they receive from their faculty advisor. Each NRT student will be mentored by multiple faculty members. Underrepresented students will also be supported by NHSEMP and SWE. The advisors will meet with the NRT management to ensure that all students are getting adequate mentoring and are making progress towards their degree.

Students will not only learn, but they will also mentor other students as they become more seasoned students. This strategy has been successfully used for undergraduate VIP programs. All NRT students will be involved in mentoring students and outreach activities. Ph.D. students will be involved in teaching recitations on specific advanced topics, assisting in developing and teaching modules, and leading project cohorts. This will give experience to students interested in academic and teaching careers. It will also help in developing soft skills including leadership, communication skills, teamwork, and ethics.

D. Micro-Credentials for Personalized Multidisciplinary Data Science Certification

Building upon the foundation of our modular curriculum and the Morea framework, we envision the integration of micro-credentials as a pathway to provide personalized multidisciplinary certification in data science. Micro-credentials represent modular, focused learning experiences tailored to specific skills or competencies. These credentials can be accumulated over time, offering learners the flexibility to adapt their education to their evolving career goals and the demands of the job market.

1) Granular Skill Recognition: Similar to the Morea modules within our curriculum, micro-credentials will focus on specific, well-defined data science skills and knowledge areas. Each module within our multidisciplinary data science curriculum could be designed and offered as a stand-alone micro-credential. This direct derivation from the modular curriculum ensures a tight integration between learning units and recognized skills. For instance, the "linear regression" module discussed earlier could be formalized as a micro-credential,

potentially with versions adapted for different disciplinary perspectives such as engineering, computer science, and economics. This aligns with the concept of *incremental credentials* that can be of any size or level. Traditional certificate programs, while providing a broad foundation, may not offer this level of granular recognition for specific competencies acquired within individual learning units.

2) Personalized Learning Pathways: Students entering NRT-DESCARTES with diverse backgrounds (engineering, social science, computer science, business) will be able to construct personalized learning paths by selecting and completing micro-credentials relevant to their prior expertise and desired career trajectories. The *breadth* required for a T-shaped background can be achieved by pursuing core micro-credentials from disciplines outside their primary area, while *depth* can be gained through more specialized micro-credentials within their research focus. The Knowledge Map, designed to illustrate the content and interrelationship of modules, will serve as a valuable tool for students in selecting pertinent micro-credentials.

3) Stackability and Credential Building: Micro-credentials are inherently incremental and can be *stacked* to build towards larger credentials, such as a graduate data science certificate. This "Credential As You Go" highlights how incremental credentials can act as bridges to larger qualifications. Micro-credentials earned through NRT-DESCARTES modules can thus provide formal recognition for specific skill acquisition and contribute to a broader certification in data science.

4) Enhancing Multidisciplinary Exposure: By offering micro-credentials that encapsulate data science principles and applications from various disciplinary lenses (e.g., data visualization in social science, in-context learning algorithm optimization in computer science, robustness modeling in engineering), students will gain a richer multidisciplinary understanding of the field. The flexibility of the Morea framework allows instructors to adapt module content, and consequently, micro-credential content, to highlight disciplinary-specific applications and perspectives, further enhancing the multidisciplinary nature of the program.

5) Agility and Responsiveness: The dynamic and rapidly evolving nature of AI and data science necessitates an agile approach to curriculum development. Micro-credentials, being more readily developed and deployed than traditional courses, offer the flexibility to quickly address emerging topics and industry demands.

6) Recognition of Varied Learning Experiences: While NRT-DESCARTES primarily focuses on formal graduate education, the micro-credentialing framework can recognize learning acquired through other avenues, such as participation in workshops, internships, or significant contributions to research projects. This aligns with the broader concept of incremental credentialing capturing learning from various sources. Traditional degree and certificate programs typically focus on recognizing learning achieved through formal coursework, potentially overlooking valuable skills and knowledge gained through practical experiences. The NRT Internship program encourages practical experience into the curriculum and micro-credentials provide a more granular way to acknowledge the specific skills developed during these experiences.

7) Industry Collaboration and Commercial Certification: Our program actively seeks to collaborate with industry partners to ensure the relevance and value of our curriculum. The NRT Internship program connects students with potential employers for practical experience, providing insights into industry needs. Furthermore, we anticipate that the development of micro-credentials can be informed by industry demands, ensuring that the skills recognized are highly valued in the job market. As noted previously, these micro-credentials could also be valuable for industry partners where employees could participate in the program to obtain recognition for specific skills. This aligns with the broader concept of incremental credentials originating from industry and allows for exploring relevant commercial certifications as micro-credentials within our program. This could provide students with a pathway to leverage existing industry-recognized credentials and integrate them into their personalized learning pathways. Further investigation is needed to determine the alignment and equivalency of commercial certifications with our program learning outcomes and the university policy on micro-credentials.

III. Research Integration

The research efforts will consist of fundamental research for dynamic systems including online and distributed learning along with applications of data science research in three areas: decarbonization (focusing on energy and transportation), healthcare, and communications. The research areas will be tied together with each of the areas consisting of a cohort of students and faculty working towards convergent research. The research requires multi-domain knowledge from researchers in different disciplines that provide different perspectives, and it is structured for convergent research across disciplines and research emphasis. The research is integrated with education and across the different research areas. This will be aligned with the overall program of developing students that not only will have broad knowledge of data science (fundamentals and applications), but also have in depth focused work that will be achieved through their thesis research. Each of the research cohorts will conduct collaborative research bringing in students from multiple disciplines. The cohorts have focused research meetings from the different research areas where both students and faculty give presentations including forming reading groups to discuss current fundamental and applied research topics. A goal of these reading groups is to find interesting research topics to pursue bringing in faculty and students coming from different disciplines, offering varied perspectives on the problems. For example, one ongoing reading group connects control theory and robustness approaches in economics (by one of the winners of the 2013 Nobel Prize in Economics).

A. Fundamental Research

The fundamental research is driven by our applications in decarbonization, healthcare, and communications. The applications deal with data that is dynamic (could be described by

a dynamic system and could vary with time). The data also comes from random sources (or is contaminated by random or systemic non-stationary noise), could come from distributed sources (e.g. sensor networks), and could be incomplete.

The research studies the design, performance, and complexity of learning algorithms. Tools from learning theory, signal processing, statistics, control systems, information theory, and economics are used in this analysis. This includes online learning in dynamic environments, classical tuning stability and adaptability of learned models via regularization, and in-context adaptation using transformer neural network architectures. This extends to risk management and adapting hyperparameters in a principled way. Other research includes learning in distributed scenarios using principles of federated learning, online kernel methods, and adapting novel language model architectures for engineering data. Research is also being conducted in anomaly detection using Minimum Description Length (MDL) and new AI based architectures. Other paradigms are being used to study dynamic data including principles of reinforcement learning, semi-supervised learning, transfer learning, large language models (LLM)s, adversarial learning, and causality.

B. Decarbonization Research

Decarbonization focuses on the energy and transportation sectors. Due to Hawaii's geographical isolation, high electricity prices, and reliance on out of state energy resources (petroleum products), Hawai'i faces unique energy challenges. The state has addressed these challenges with the Hawai'i Clean Energy Initiative (HCEI) with the goal of 100% clean energy generation by 2045 [11].

Our program is working with UHM, Hawaiian Electric Company (HECO), industry, and the state to get energy data at different spatial and temporal resolutions. This data will often be distributed in nature and we will apply distributed learning and inference to gather information on the data. This includes forecasting solar energy at UHM using distributed sensors placed on campus. We are also building LLMs that correlate power usage across buildings on campus, adapting to various modes of operation prevalent at UHM, adapting to changing usage, and detecting anomalies. Other research includes a demand side management program using tools from optimization, model predictive control and reinforcement learning. We will also consider the economic impacts of real-time pricing and on how renewable energy penetration affects different communities. In addition, we will quantify quality of service, and provide insights into monetizing the quality of service in power systems. Energy and transportation infrastructure are closely tied with the introduction of electric vehicles (EV)s. In the transportation area we will also work with data sets detecting vehicles for intelligent traffic management and control of highways.

C. Healthcare Research

Research in healthcare integrates data science and machine learning to advance medical diagnosis and public health policies. Areas include heart disease and epidemics. Heart disease is the leading cause of death in the United States, and it affects Hawai'i at higher rates than national averages. The electrocardiogram (ECG) is an essential tool to evaluate patients with cardiac conditions such as irregular heartbeat, bradycardia, tachycardia, congenital and acquired heart defects, as well as cardiac manifestation of certain diseases, such as cancer, lupus, and infections. Here we access ECG data sets combining knowledge from clinicians to apply data science and machine learning principles. This includes developing algorithms for anomaly detection to detect heart disease such as Kawasaki disease from ECG. Epidemics such as COVID19 pose significant challenges to data analysis due to having many heterogeneous data sources, often captured at different time and geographic scales. The spread of epidemics is different in Hawai'i because of its isolated geography than the US mainland. Here we combine domain knowledge and incomplete, possibly biased measurements to clean/pre-process data and develop models for disease propagation.

D. Communication, Network, and Security Research

UHM has a tradition of having a strong communications and networking programs in Electrical Engineering and Computer Science. This goes back to the development of ALOHAnet which was the first successful deployment of a wireless data packet network [12]. The NRT research in communications, networks, and security areas focuses on creating NextG wireless systems with enhanced communication performance, novel sensing functionalities, strong attack resiliency, and high energy efficiency, for uses in Augmented, Virtual or eXtended Reality (AR/VR/XR), autonomous driving, massive interactive real-time applications, advanced industrial, manufacturing, and telehealth. Specifically, we design wireless systems and algorithms to control or explore the wireless propagation characteristics, thus improving signal quality, achieving ubiquitous, low-power sensing, and preventing or reducing attack vectors, such as jamming and eavesdropping.

IV. Implementation and Early Evaluation

NRT-DESCARTES at UHM has been strategically implemented to bridge interdisciplinary research with personalized education in AI, data science, and machine learning. The program's modular structure and convergent research create a dynamic learning environment that adapts to the unique backgrounds and career aspirations of each trainee. Here we present early evaluation findings based on student engagement, program outcomes, and faculty feedback.

A. Implementation Strategy

At the core of NRT-DESCARTES is a modular learning framework, designed to accommodate diverse academic backgrounds while maintaining rigor in AI and data science education. Compact, topic-specific modules are used to personalize instruction and ensure that trainees can build on their existing knowledge while acquiring new, interdisciplinary competencies. These modules are integrated across multiple courses, enabling students from different disciplines to engage in collaborative problem-solving. Interdisciplinary cohort-based research teams are a hallmark of the program. Faculty mentorship is embedded in the process, ensuring that

students receive continuous guidance as they navigate complex research problems.

Beyond technical expertise, the program prioritizes professional development and mentorship. Trainees participate in industry internships, public presentations, and teaching assignments that help refine their communication, leadership, and teamwork skills. Additionally, ethical and policy considerations in AI research are integrated into the curriculum to ensure responsible application of AI technologies.

B. Early Evaluation Metrics and Findings

To assess the effectiveness of the NRT-DESCARTES traineeship, a multi-layered evaluation strategy has been employed, guided by our program logic model and supported by an external evaluator. The Mid-Year Check-In Survey (January 2025) provided key insights into how trainees are engaging with the program, identifying both strengths and areas for improvement. These early findings help inform ongoing program adjustments to optimize the learning experience.

1) Evaluation Methodology: Evaluation questions have been developed to target student and program outcome areas in alignment with the program logic model. The evaluation queries the extent to which NRT trainees: (a) demonstrate fundamental understanding of the methodology and techniques of data science and machine learning; (b) demonstrate ethical and critical thinking about data science applications; (c) synthesize, integrate, and apply different disciplinary methods and techniques in research; (d) communicate across disciplinary boundaries; and (e) demonstrate leadership and collaboration skills when working in teams.

The evaluation also queries the extent to which the NRT program: (f) increases enrollments over the program period; (g) facilitates NRT student entry into academic, industry, or government sector employment upon graduation; (h) promotes trainees' development of values and behaviors associated with lifelong learning; and (i) facilitates safe, supportive, and responsive learning experiences for NRT students.

A mixed methods evaluation strategy has been employed to assess the impacts, effectiveness, and efficacy of the NRT traineeship. The evaluation framework incorporates both quantitative and qualitative data collection methods, including:

- Trainee Pre/Post-Surveys: Distributed to students upon program entry to establish a baseline, and then updated annually for comparison analysis. The survey measures trainees' understanding in AI and data science competencies; interdisciplinary learning, collaboration, and communication; and lifelong learning skills. The post-survey also captures student perceptions of program quality and relevance.
- Portfolio Rubric: A rubric for faculty to score the contents of trainee portfolios, focusing on interdisciplinary learning, applications, and communication.
- Document and Materials Review: Review of program materials/outputs; seminar and faculty meeting notes; faculty feedback on student and program work; and student reported grades earned for completing data science and machine learning related coursework.

Domain Area	Mean (6 pt. max)	% of Responses Strongly Agree/Agree
Lifelong Learning	5.1	81%
Critical Thinking	4.9	71%
Leadership	4.7	70%
Understanding Data Science	4.4	60%
Interdisciplinary Learning	4.3	60%
Communication	4.3	56%

TABLE I
COHORT I PRE-SURVEY RESULTS, FALL 2024

- End-of-Year Trainee and Faculty Focus Groups and Interviews: Administered at the conclusion of each academic year to capture insights into the effectiveness of the modular learning approach, research integration, mentorship structure and quality, and other programmatic aspects. Challenges and barriers to program participation and/or implementation, and corresponding recommendations for program change are also probed.
- Trainee Mid-Year Check-In Survey & Interviews: Conducted at mid-year (December /January) in the "start, stop, continue" tradition to gather formative feedback to inform the program's midcourse adjustments.

2) Initial Findings: To date, the evaluation has collected baseline data from Cohort I NRT trainees through a pre-survey administered in Fall 2024. The survey measured self-reported knowledge and skills aligned with the student outcome areas outlined for the program.

The survey was distributed to 13 trainees, with 9 responding, yielding a 69% response rate. The instrument included 38 items organized across six domain areas listed in Table 1. Descriptive statistics (means and frequency distributions) were calculated for each domain which are summarized in Table 1:

Overall, Cohort I trainees reported stronger baseline capacities in leadership, critical thinking, and lifelong learning compared to data science, interdisciplinary learning, and communication skills.

The evaluation also captured mid-year formative data on trainee experiences with the NRT-DESCARTES program. Complete data for the 2024-2025 annual evaluation report will be collected in May and June of 2025. The following summarizes trainee responses to the Mid-Year Check-In Survey and follow-up interviews conducted in December 2024 and January of 2025 (60% Cohort I response rate). This survey provided insights into how trainees are engaging with the program, identifying both strengths and areas for improvement. These early findings help inform ongoing program adjustments to optimize the learning experience. Trainees highlighted several aspects of the NRT program that most supported their learning and professional growth:

Key Strengths of the Program:

- Interdisciplinary Access and Collaboration (42%):
  - Trainees emphasized the value of working with faculty and peers across disciplines, with meaningful research feedback and opportunities to co-author papers and proposals.
  - Faculty engagement beyond their expertise was particularly noted, creating a dynamic, intellectually open environment.
  - Cohort-based research teams were effective in breaking down academic silos and encouraging peer-to-peer learning.

- Relationships and Camaraderie (29%):
  - Strong peer and faculty relationships enhanced motivation and engagement.
  - The use of Discord for communication was widely appreciated for fostering informal communications.
- Workshops and Research Discussions (29%):
  - AI discussion sessions and data ethics seminars were frequently mentioned as highly valuable.
  - Trainees appreciated the opportunity to engage in interactive learning formats that bridged technical skills with ethical and societal considerations.

Challenges Identified: While trainees generally expressed high satisfaction with the program, early evaluations also identified areas that require further refinement:

- Lack of Clarity in Research Pathways for New Students
  - Some newer trainees struggled to navigate available research opportunities.
  - Greater structure and guidance in onboarding new trainees into interdisciplinary projects is needed.
- Need for Additional Infrastructure Support
  - Trainees requested access to a supercomputer or cloud computing resources to facilitate AI research.
  - Expansion of technical support for computationally intensive projects was a recurring theme in feedback.
- Anticipation of More Structure Over Time
  - As the program is in its early phases, trainees acknowledged that some structural elements are still evolving.
  - There was an expectation that greater program organization would emerge as the traineeship matures.

C. Next Steps for Program Enhancement

- Enhanced Research Pathways & Onboarding Support
  - Developing a clearer system for matching new trainees with research projects and mentors.
  - Establishing an interactive research project database that trainees can use to explore ongoing interdisciplinary work.
- Expansion of Technical Infrastructure
  - Exploring partnerships with AI labs for resource-sharing.
  - Securing access to high-performance computing resources to better support AI model training.
- Refinement of Structured Learning Pathways
  - Building additional modules based on trainee feedback.
- Strengthening industry and external collaboration opportunities to enhance workforce preparedness.

Based on these early evaluations, the above steps are being prioritized to strengthen the NRT-DESCARTES program. As the program continues to evolve, these refinements will ensure that it remains a cutting-edge, interdisciplinary, and student-centered AI traineeship. The ongoing use of surveys, faculty feedback, and structured program evaluations will further refine the approach, ensuring the long-term success of trainees.

V. Summary and Further Directions

This paper summarizes our NSF NRT-DESCARTES program that is a current NSF funded program to advance graduate education and research in data science. It develops an innovative graduate program that brings together engineering, computer science, social science, business and medicine to harness the transformative power of data science. This cross-disciplinary program will train a new generation of graduate students incorporating tailored course modules and cohorts working on convergent research.

The NRT program is evolving as the Morea framework and modules are incorporated throughout the curriculum. Information about the curriculum is through the Knowledge Map that is being developed through AI technologies. A key to the program is continual improvement of the educational experience of students through implementation and thorough evaluation of all aspects of the program. The program is dynamically changing as we make changes based on our evaluations and incorporate new faculty and students into the program. We also look at developing micro-credentials for students as they complete certain requirements, courses, and modules. We also look to work and collaborate with other NRT programs in AI and data science along with other NSF AI and data science programs (e.g. AI Institutes and TRIPODS).

References

[1] NSF's Ten Big Ideas. Webpage: https://www.nsf.gov/about/history/big-ideas, NSF, 2016.

[2] NSF: National Artificial Intelligence Research Institutes. Webpage: https://www.nsf.gov/funding/opportunities/national-artificial-intelligence-research-institutes, NSF, 2020.

[3] NSF: Mathematical Foundations of AI. Webpage: https://www.nsf.gov/funding/opportunities/mfai-mathematical-foundations-artificial-intelligence, NSF, 2024.

[4] U.S. National Science Foundation Research Traineeship Program. Webpage: https://www.nsf.gov/funding/opportunities/us-national-science-foundation-research-traineeship-program, NSF, 2024.

[5] C. Wendler, B. Bridgeman, F. Cline, C. Millett, J. Rock, N. Bell, and P. Mcallister. The path forward: The future of graduate education in the United States. Technical report, Educational Testing Service & Council of Graduate Schools, April 2010.

[6] OSTP report on the industries of the future act. Technical report, White House, https://www.whitehouse.gov/wp-content/uploads/2022/04/04-2022-OSTP_IOTF_Report.pdf, April 13, 2022.

[7] HSTC. Harnessing science and technology in support of Hawai'i's future. Technical report, University of Hawai'i, http://hawaii.edu/epscor/wp-content/uploads/2021/11/Hawaii-ST-Plan-08.05.2021-1.pdf, 2020.

[8] SMS Research. Hawaii IT workforce needs report. Technical report, Chamber of commerce Hawaii and UH Community Colleges, https://www.hec.org/itworkforce, Nov. 2021.

[9] Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, V. Galanos, P. V. Ilavarasan, M. Janssen, P. Jones, A. K. Kar, H. Kizgin, B. Kronemann, B. Lal, B. Lucini, R. Medaglia, K. Le Meunier-FitzHugh, L. C. Le Meunier-FitzHugh, S. Misra, E. Mogaji, S. K. Sharma, J. B. Singh, V. Raghavan, R. Raman, N. P. Rana, S. Samothrakis, J. Spencer, K. Tamilmani, A. Tubadji, P. Walton, , and M. D. Williams. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57:101994, Apr. 2021.

[10] Philip M. Johnson. https://morea-framework.github.io, 2021.

[11] Hawaii Clean Energy Initiative. Website report, Department of Business Economic Development & Tourism, http://www.hawaiicleanenergyinitiative.org/.

[12] N. Abramson. The ALOHA system - another alternative for computer communications. Proc. 1970 Fall Joint Computer Conference, AFIPS Press, 1970.