# STEM Leadership and Training for Trailblazing Students in an Immersive Research Environment

Marisel Villafañe-Delgado*, Erik C. Johnson*, Marisa Hughes, Martha Cervantes, and William Gray-Roncal
Johns Hopkins University Applied Physics Laboratory, Laurel, MD, United States
marisel.villafane-delgado, william.gray.roncal@jhuapl.edu   *Authors Contributed Equally

*Abstract* - **Educating the workforce of tomorrow is an increasingly critical challenge for areas such as data science, machine learning, and artificial intelligence. These core skills may revolutionize progress in areas such as health care and precision medicine, autonomous systems and robotics, and neuroscience. Skills in data science and artificial intelligence are in high demand in industrial research and development, but we do not believe that traditional recruiting and training models in industry (e.g., internships, continuing education) are serving the needs of the diverse populations of students who will be required to revolutionize these fields. Our program, the Cohort-based Integrated Research Community for Undergraduate Innovation and Trailblazing (CIRCUIT), targets trailblazing, high-achieving students who face barriers in achieving their goals and becoming leaders in data science, machine learning, and artificial intelligence research. Traditional recruitment practices often miss these ambitious and talented students from nontraditional backgrounds, and these students are at a higher risk of not persisting in research careers. In the CIRCUIT program we recruit holistically, selecting students on the basis of their commitment, potential, and need. We designed a training and support model for our internship. This model consists of a compressed data science and machine learning curriculum, a series of professional development training workshops, and a team-based robotics challenge. These activities develop the skills these trailblazing students will need to contribute to the dynamic, team-based engineering teams of the future.**

*Index Terms* – artificial intelligence, cohort-based research, internship, robotics, trailblazing

## INTRODUCTION

As the pace of technology continues to advance, equipping a workforce with the skills necessary to succeed in the 21st century is a major challenge. Emerging technologies, such as machine learning and artificial intelligence (AI), are increasingly critical and rapidly evolving. Addressing critical challenges will require a workforce that is not only adaptive but diverse [1-3]. Several key groups remain underrepresented or underserved [4]. Moreover, traditional training models and recruitment pipelines are still unable to provide the tools needed to ensure access and success for these populations.

The Cohort-based Integrated Research Community for Undergraduate Innovation and Trailblazing (CIRCUIT) program aims to convert potential energy into kinetic energy by providing trailblazing populations with the opportunities and resources needed to succeed [5]. We define trailblazing students as those who are underserved (e.g., first generation, low income) and underrepresented (e.g., African Americans, Hispanics, Latinos, women) in science, technology, engineering, and mathematics (STEM) fields. These include high-achieving students who face barriers in achieving their goals or finding research opportunities. Student challenges are multifactorial and often include feelings of inadequacy [6], lack of access to research experiences [7], the social class achievement gap in first-generation students [8], and assumed family support roles [9]. Low-income students experience opportunity and achievement gaps, which can result in a lack of access to technological resources or even failure to complete a bachelor's degree [8]. Other barriers encountered by trailblazing students include lack of confidence, limited access to mentors, implicit bias, lack of educational opportunities, and limited professional networks. These barriers, which are faced by many students, result in a more challenging recruitment process for companies seeking a diverse workforce.

CIRCUIT provides a combination of resources to help students catalyze their efforts [5]. The research program takes place at the Johns Hopkins University Applied Physics Laboratory (APL), which is a University Affiliated Research Center (UARC). Students are immersed in the critical research problems and challenges typical of this environment, working on cohort-based research with the support of a team of staff members with different levels of expertise (i.e., new hires and subject-matter experts), and their work typically results in a peer-reviewed publication. Sample project topics include applied neuroscience, precision medicine, and AI applications. Students join CIRCUIT from different academic majors, academic levels (from sophomores to seniors), and trailblazing backgrounds. Our training considers student diversity and aims to provide students with tools to overcome challenges and contribute to a peer-reviewed publication. We seek to provide students with the skills required to succeed in cutting-edge research projects (Figure I). Each cohort has a lead mentor who sets the direction for their research topic. CIRCUIT's structure of many-to-many mentoring provides students with broad perspective, expert help in developing key cross-disciplinary skills, and support for common barriers.

In addition to their main research task, students participate in technical and professional skills training, which is the focus of this paper. We follow a multifaceted curriculum—consisting of data science and machine learning, professional skills development, and a robotics challenge—to prepare students for the AI workforce.

## RELATED WORK

Prior and recent work documenting the outcomes of research experiences for undergraduates (REUs) have demonstrated how these experiences benefit the career development and goals of students, including students from underrepresented populations [10, 11]. These benefits include increasing retention, helping students to define their career goals, and providing students with experiences and opportunities that they otherwise would not have. REUs also provide the benefit of increasing student interest in pursuing graduate studies, often up to the doctoral level. Moreover, students greatly benefit from being part of cohort-based teams such as those used in CIRCUIT and the Brandeis Science Posse program [12].

Many REUs occur in academic settings, where students are directly supervised by a principal investigator and collaborate with members of a laboratory. Other settings include national laboratories and research centers, where students engage with professionals outside the academic setting. In many programs, students are expected to join the program with the background needed to work on their projects and then receive project-specific training. Various programs have introduced structured teaching tracks to provide students with the skills needed to succeed in their research [13, 14]. In many of these experiences, all projects follow a similar topic, and students learn narrow skills specific to that theme. Course-based Undergraduate Research Experiences (CUREs) are another method aiming to integrate research and teaching [15, 16]. CUREs typically occur during the academic year and introduce ways to apply research methods to problem solving in classes.

## DATA SCIENCE AND MACHINE LEARNING CURRICULUM

CIRCUIT projects were selected across broad themes incorporating an aspect of data science, machine learning, modeling, or Artificial Intelligence. However, trailblazing students come from diverse, nontraditional backgrounds, and most machine learning and AI courses are taught at the upper-level undergraduate or graduate level. Therefore, we implemented a curriculum to accelerate student progress in mastering these topics and necessary prerequisites.

The curriculum development process began by surveying staff project leaders to understand the software used in their projects and anticipated technical scope. Few of the students had extensive data science experience, so most needed additional training; by incorporating training into our model we could recruit for capability rather than previous opportunity. We created a set of short courses covering introductory programming, software development tools, data science, applied mathematics, machine learning,

software engineering, and project planning. Although learners are unlikely to achieve mastery of these diverse fields in one summer, the goal of the curriculum is to introduce the subjects to the students and give them the background, references, and examples they need to begin applying these tools to their ongoing projects (Figure II).
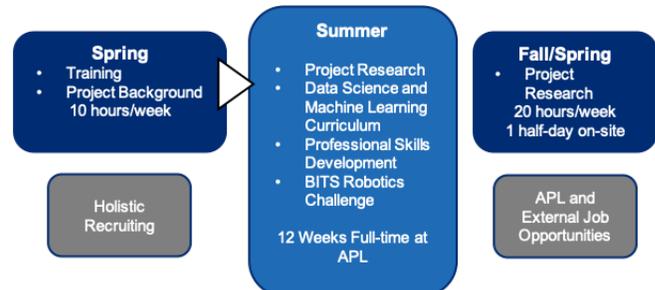


FIGURE I

OVERVIEW OF THE 14-MONTH CIRCUIT RESEARCH INTERNSHIP EXPERIENCE, HIGHLIGHTING SPRING TRAINING, THE SUMMER INTERNSHIP, AND FOLLOW-ON RESEARCH THROUGH THE SCHOOL YEAR

Instruction began with a Massively Open Online Course in Data Science Using Python Programming, provided by Coursera. Students spent eight to ten hours a week on this material during the spring, before arriving for the full-time experience. This approach leveraged existing materials for online education, minimized the teaching load for mentors, and gave the students an option to receive a recognized certificate through Coursera. Beginning in the first week of the internship, students participated in a series of lectures and interactive workshops taught by APL staff. This series included an introduction to all the projects, interactive workshops on software development tools, and lectures in applied mathematics and machine learning topics. For the lectures, examples and exercises were provided to serve as references and starting points for project work and independent, optional problem sets. Given the breadth of student backgrounds, the lectures were designed to complement academic studies by providing broad overviews of topics that typically encompass entire courses, placing a special emphasis on practical applications and connections between topics. As one example, students learned how to combine their software capabilities with core lessons in linear algebra, graph theory, and probability to create probabilistic graph generators. Such fundamental connections are difficult to incorporate into undergraduate curricula while maintaining flexibility in course scheduling and sequence. To connect and contextualize academic training for the workforce, we taught the students software engineering and project management best practices.

## PROFESSIONAL SKILLS DEVELOPMENT

We emphasized professional development topics, which are typically not covered directly in academic settings; many students learn those skills informally as they participate in extracurricular experiences. For example, teamwork skills are critical for success (e.g., in senior capstone engineering

courses) [17, 18]. In addition to learning specific skills needed to excel in a professional STEM environment, students participated in workshops relevant to the trailblazing CIRCUIT student population.

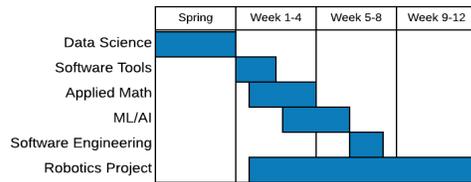| | Spring | Week 1-4 | Week 5-8 | Week 9-12 |
|---|---|---|---|---|
| Data Science | ▮ | | | |
| Software Tools | | ▮ | | |
| Applied Math | | ▮ | ▮ | |
| ML/AI | | | ▮ | |
| Software Engineering | | | ▮ | ▮ |
| Robotics Project | | | | ▮ |

FIGURE II

DATA SCIENCE AND MACHINE LEARNING CURRICULUM—TRAINING BEGINS IN THE SPRING AND CONTINUES IN THE SUMMER WITH SOFTWARE TOOLS, APPLIED MATHEMATICS, AND MACHINE LEARNING.

Our professional skills activities were implemented as interactive lectures, following an informal cooperative learning framework [19]. Students listened to a short lecture and then worked on an activity that gave them a chance to implement what they learned, consolidate knowledge, and ask questions. Students worked within their research cohorts but also had the opportunity to work across cohorts and to apply their knowledge to real engineering situations.

The professional skills courses were grouped into three tracks: Research Process, Science and Professional Communications, and the Professional in the Workplace (see Figure III). Under Research Process, students learned about the basics of scientific discovery, including literature review, data collection and management, and responsible research. The learning outcomes included performing a scientific literature search, using a reference management tool, and responsibly managing data. After completing the responsible research class, students were expected to understand what constitutes fabrication, falsification, and plagiarism and how to respond to these practices. Understanding of all Research Process topics was assessed through student analysis and discussion of case studies.

Next, the Science and Professional Communications curriculum was split into oral and written communications. Oral communications included methods for delivering scientific information, such as the design and delivery of research and poster presentations. Learning outcomes included understanding the basics of effective communications and body language as well as effectively and efficiently telling a research story. Assessments were performed over several practice sessions. Written communications focused on research abstracts and scientific papers. Students learned about different types of abstracts and the basic structure of a scientific paper. By the end of their first semester, cohorts completed a manuscript draft.

The Professional in the Workplace track targeted skills focused on the researcher at work and in the world. These topics included dress codes, diversity and inclusion, teamwork, resume and curriculum vitae development, online identity, networking, and elevator pitches. For most activities, students benefited from teamwork and peer feedback, and they practiced their elevator pitches with speakers who visited and met with students.
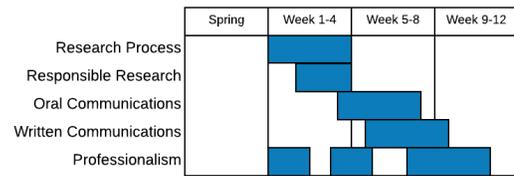
| | Spring | Week 1-4 | Week 5-8 | Week 9-12 |
|---|---|---|---|---|
| Research Process | | ▮ | | |
| Responsible Research | | ▮ | | |
| Oral Communications | | | ▮ | ▮ |
| Written Communications | | | ▮ | ▮ |
| Professionalism | | ▮ | ▮ | ▮ |

FIGURE III

PROFESSIONALS SKILLS TRAINING—TRAINING PROGRESSES THROUGHOUT THE EXPERIENCE FROM INTRODUCTORY TO MORE ADVANCED TOPICS

## BITS ROBOTICS CHALLENGE

We designed Breakthrough Investigation of Terrestrial Swarms (BITS), a series of robot swarm navigation challenges (see Figure IV). This was a chance to apply training to the hard robotics problem of effective swarm control while demonstrating mastery of key concepts in a project-based setting. The trainees were divided into teams that were different from their research cohorts, allowing them to interact with a wider range of students.
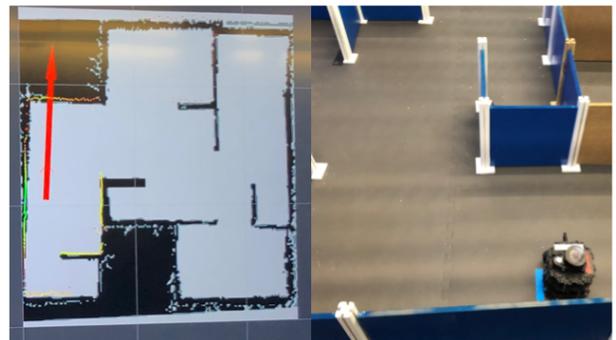


FIGURE IV

STUDENT TEAMS CONTROLLED ROBOTS IN A NAVIGATION CHALLENGE. (LEFT) A VISUALIZATION OF THE ROBOT LOCALIZATION AND THE ROS (ROBOT OPERATING SYSTEM) INFRASTRUCTURE. (RIGHT) A TURTLEBOT IN OUR MODULAR MAZE ENVIRONMENT.

Students worked with TurtleBots (which they built themselves from kits) to (1) navigate a robot from a start to a target point in a known map; (2) navigate a team of up to three robots from a start to a target point in a known map; and (3) use a team of up to three robots to explore a known map to locate targets at unknown points. The third challenge simulates the real-world problem of using a robotic swarm to find victims during a disaster response. Teams developed Python solutions to the challenge, using a custom ROS (Robot Operating System) environment and a modular maze. Specific techniques taught in the technical curriculum (such as linear algebra, graph traversal searches, and reinforcement learning) were directly applicable to these challenges, allowing students to demonstrate the transfer of knowledge from the lectures to a real-world environment. Student teams had approximately four hours a week to work on this challenge with mentors, giving the students a chance

to interact with a new project team, understand the engineering design process, practice teamwork and communication skills, apply their knowledge in a real-world domain, and achieve a second technical accomplishment for the summer.

On the final competition day, solutions were scored on the basis of time and navigation accuracy. Teams also received points for their project plans and software requirements. This activity emulates, in a controlled way, the nature of the modern research engineering environment, where staff often work in multiple, multidisciplinary teams to solve a diverse range of critical challenges. This effort can play a catalyzing role in developing these workforce-critical skills, which are difficult to practice in a traditional internship environment.

## DISCUSSION

We have developed and implemented CIRCUIT, which provides a unique research internship for trailblazing students; as the program has matured in its mission and focus, we have been working with stakeholders to generalize and formalize the curriculum. Our model maps this teaching paradigm onto an AI workforce development model that reaches a wider range of learner backgrounds. We offer an innovative approach to support high-potential students with limited research experience and immerse them in important, emerging research areas. We expect incoming students to be strong in learning speed and motivation. Our approach has the potential to extend broadly to internship experiences in AI and other domains. Anecdotally, we have seen great success in helping to inspire students and launch their careers as engineers and scientists; our future work will formally evaluate these approaches and link our methods to existing literature and models.

## ACKNOWLEDGMENT

## REFERENCES

[1] B. Ozkazanc-Pan, "Diversity and future of work: inequality abound or opportunities for all?" Management Decision, 2019.

[2] T.C. Dildine and L.Y. Atlas, "The need for diversity in research on facial expressions of pain," Pain, vol. 160, pp. 1901-1902, 2019.

[3] S.M. James and S.R. Singer, "From the NSF: The National Science Foundation's investments in broadening participation in science, technology, engineering, and mathematics education through research and capacity building," CBE—Life Sciences Education, 2016.

[4] A. Freire, L. Porcaro and E. Gómez, "Measuring diversity of artificial intelligence conferences," arXiv Preprint, arXiv:2001.07038, 2020.

[5] M. Encarnacion, C. Bishop, J. Downs, et al., "CIRCUIT summer program: A computational neuroscience outreach experience for high-achieving undergraduates via sponsored research," in 2018 IEEE Integrated STEM Education Conference (ISEC), pp. 45-52, 2018.

[6] L.V. Truong, "First in the family: how high-achieving, first-generation Vietnamese-American students navigate the elite college environment, a study using portraiture methodology," Ph.D. dissertation, Northeastern University, 2016.

[7] T.P. Hébert, "An examination of high-achieving first-generation college students from low-income backgrounds," Gifted Child Quarterly, vol. 62, pp. 96-110, 2017.

[8] R. Chetty, J.N. Friedman, E. Saez, et al., "Mobility report cards: The role of colleges in intergenerational mobility," National Bureau of Economic Research, 2017.

[9] R. Covarrubias, I. Valle, G. Laiduc, et al., "'You never become fully independent': Family roles and independence in first-generation college students," Journal of Adolescent Research, 2018.

[10] B.C. Tamer and J.G. Stout, "Understanding how research experiences for undergraduate students may foster diversity in the professorate," in Proceedings of the 47th ACM Technical Symposium on Computing Science Education, pp. 114-119, 2016.

[11] D.J. Follmer, S. Zappe, E. Gomez, et al., "Student outcomes from undergraduate research programs: comparing models of research experiences for undergraduates (REUs)," Scholarship and Practice of Undergraduate Research, vol. 1, pp. 20-27, 2017.

[12] M.S. Kosinski-Collins, K. Godsoe and I.R. Epstein, "The Brandeis Science Posse: Building a cohort model program to retain underserved students in the sciences," in Diversity in the Scientific Community Volume 2: Perspectives and Exemplary Programs, ACS Publications, 2017.

[13] K.M. Booth, B. Dosono, E.M. Redmiles, et al., "Diversifying the next generation of information scientists: Six years of implementation and outcomes for a year-long REU program," in International Conference on Information, pp. 655-664, 2018.

[14] M. Barrow, S. Thomas and C. Alvarado, "ERSP: A structured CS research program for early-college students," in Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education, pp. 148-153, 2016.

[15] K.M. Cooper, P.A. Soneral and S.E. Brownell, "Define your goals before you design a CURE: A call to use backward design in planning course-based undergraduate research experiences," Journal of Microbiology and Biology Education, vol. 18, 2017.

[16] E.E. Shortlidge, G. Bangera and S.E. Brownell, "Faculty perspectives on developing and teaching course-based undergraduate research experiences," BioScience, vol. 66, pp. 54-62, 2015.

[17] N. Ismail, N.A.A. Aziz, C.K. Hong, et al., "Assessing Teamwork Value in Project-based Learning of Capstone Project Course," EasyChair, 2020.

[18] B. Beigpourian, D.M. Ferguson, F.C. Berry, et al., "Using CATME to document and improve the effectiveness of teamwork in capstone courses," ASEE Annual Conference and Exposition Proceedings, 2019.

[19] R.T. Johnson and D.W. Johnson, "Active learning: Cooperation in the classroom," The Annual Report of Educational Psychology in Japan, vol. 47, 29-30, 2008.

## AUTHOR INFORMATION

**Marisel Villafañe-Delgado,** Research Scientist, APL. Ph.D. in Electrical Engineering, Michigan State University. Student Development Lead.

**Erik C. Johnson,** Research Scientist, APL. Ph.D. in Electrical Engineering, University of Illinois at Urbana-Champaign. Technical Curriculum Lead.

**Marisa Hughes,** Assistant Program Manager, APL. Ph.D. in Mathematics, Cornell University. Leadership member.

**Martha Cervantes,** Research Engineer, APL. Bachelor's in Mechanical Engineering, Johns Hopkins University. Project Lead.

**William Gray-Roncal,** Research Scientist, Assistant Research Professor, APL. Ph.D. in Computer Science, Johns Hopkins University. Program Lead.