Graduate Training in Translational Convergent Research - Harnessing AI and Data Science for Sustainable Agriculture (GRAD-AID for Ag) NSF-National Research Traineeship

Dr. Catherine E. Brawner, Research Triangle Educational Consultants

Catherine E. Brawner is President of Research Triangle Educational Consultants. She received her Ph.D.in Educational Research and Policy Analysis from NC State University in 1996. She also has an MBA from Indiana University (Bloomington) and a bachelors degree from Duke University. She specializes in evaluation and research in higher ed STEM disciplines.

Dr. Kimberly Bourne, North Carolina State University at Raleigh

Kimberly Bourne PhD is a project coordinator with the GRAD-AID for Ag (Graduate Training in Translational Convergent Research – Harnessing AI and Data Science for Sustainable Agriculture) program at North Carolina State University, Raleigh, North Carolina, USA. She focuses on developing tools, methods, and workshops to increase convergence capacity in cross-disciplinary teams. She is also interested in using team science, novel modelling, and stakeholder engagement methods to improve environmental decision-making.

Dr. Terri Long, NC State University

Dr. Terri Long is a Professor and University Faculty Scholar in the Department of Plant and Microbial Biology at North Carolina State University. She earned her Bachelor of Science in Biology from the University of North Carolina at Chapel Hill and obtained her Ph.D. in Molecular Genetics from the University of Georgia in 2005. Following her doctoral studies, Dr. Long conducted postdoctoral research at Duke University in Dr. Philip Benfey's lab. She then served as an Assistant Professor in the Department of Biology at the University of Illinois at Chicago before joining the faculty at NC State in 2011. Dr. Long's research group employs molecular and systems biology approaches to investigate the molecular mechanisms underlying plant nutrient stress responses. Her teaching primarily focuses on the fundamentals of plant physiology. She served as the Platform Director for Education and Workforce Development for the North Carolina Plant Sciences Initiative, where she spearheaded the development and implementation of strategic programs to cultivate the next generation of leaders in plant science, 2022-2025. Dr. Long received the Presidential Award for Excellence in Science, Mathematics and Engineering Mentoring (PAESMEM) from US White House in January, 2025. Additionally, Dr. Long serves as Lead Principal Investigator for the NSF Research Traineeship program Graduate Training in Translational Convergent Research - Harnessing AI and Data Science for Sustainable Agriculture (GRAD AID for Ag), an initiative to equip doctoral students with interdisciplinary expertise at the nexus of plant science and artificial intelligence, fostering innovation in sustainable agriculture. She currently, serves as Head of the Department of Plant and Microbial Biology at N. C. State University.

Graduate Training in Translational Convergent Research - Harnessing AI and Data Science for Sustainable Agriculture (GRAD-AID for Ag) NSF-National Research Traineeship

Background

Social and economic pressures on food and agricultural systems impact the ability to sustainably grow crops to feed our population. While advances in both fundamental and applied research have driven innovation in agriculture, there is a disconnect between fundamental and applied plant biology research - fundamental plant science discoveries made using plants grown in incubators and greenhouses rarely hold up in the field [1], [2], [3]. Solutions have emerged such as stress-responsive molecular mechanisms, predicting crop yields, guiding plant breeding strategies, automating disease detection, and evaluating crop quality by using artificial intelligence (AI), machine learning, and other data driven approaches [4], [5], [6], [7], [8]. Despite the wealth of data generated in the respective domains, cross-domain data analytics remains underutilized due to the separation of knowledge and historical disciplinary silos, particularly in student training. The common thread linking these domains is data. By utilizing AI to facilitate the creation of models that merge data from fundamental and applied plant biology, a pathway emerges to enhance the relevance of lab experiments and translate outcomes into field-applicable results more efficiently.

A new generation of scientists with the skills, knowledge, and experience to bridge gaps in experimentation, analysis, and field applications is needed to bring AI advances into agricultural fields by spanning boundaries between disciplines and sectors. Recognizing the increased need for transdisciplinary innovation in plant science and Ag, graduate training curricula have evolved in recent years to better integrate data science, engineering, and modeling approaches with plant science [9]. In this paper, we introduce GRAD-AID for Ag, a National Science Foundation National Research Traineeship where graduate students engage in convergent research using AI and plant biology research to address the needs of the agricultural industry, and provide a progress report on the first year of student activities.

Enhancing Graduate Training Through Convergent Research

Figure 1 summarizes how GRAD-AID for Ag addresses the talent gap of scientists trained in both plant and data science. Program goals are to: 1) recruit and build two collaborative cohorts consisting of diverse PhD students from basic and applied plant biology, and data science and computer engineering programs; 2) increase student proficiency and core competency in AI and basic and applied plant biology by leveraging cross-training curriculum that also trains students on the societal implications of AI in Ag; and 3) create an interdisciplinary, diverse, and sustainable Ag workforce with skills across plant science, AI, and engineering using new research and professional engagement tools developed with industry, grower, and academic partners that enhance student professionalization and preparedness.

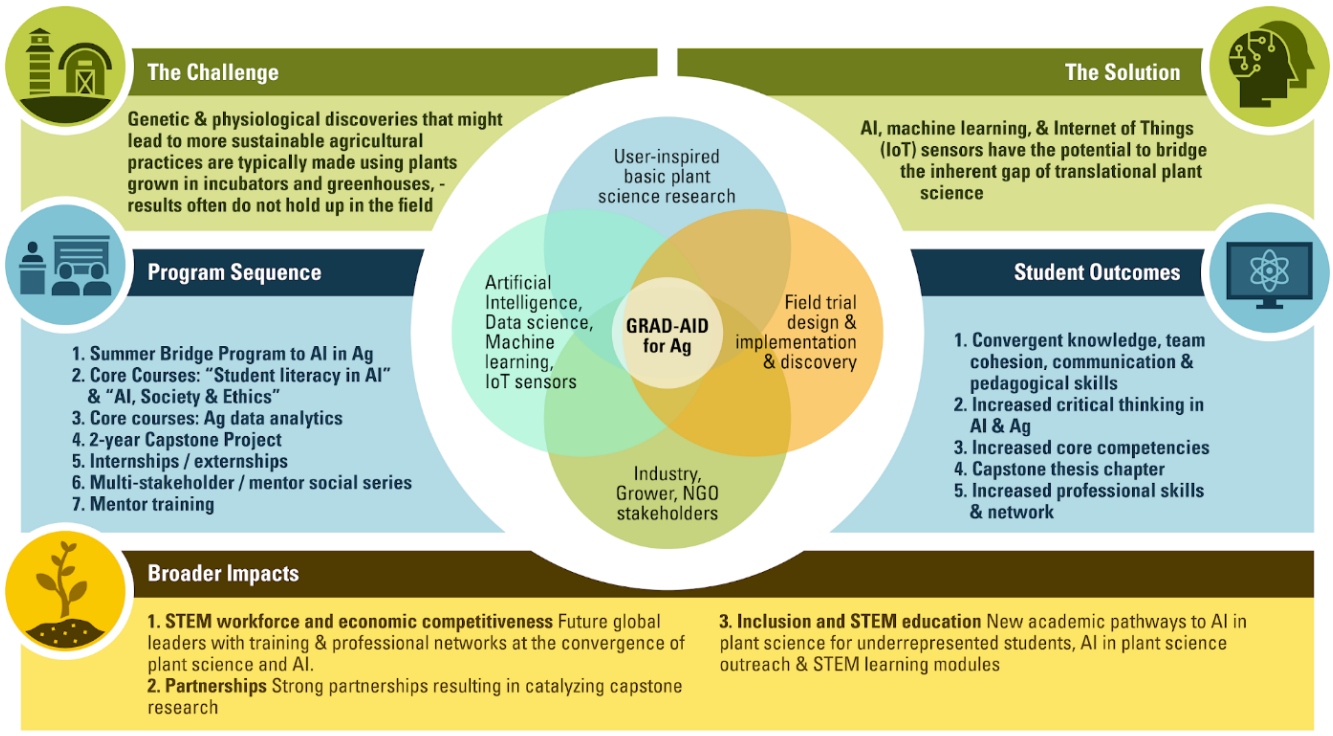

Figure 1. Overview of GRAD-AID for Ag

Project Components

Each of the two planned cohorts of students will begin their two-year fellowship with a 3-week summer bridge experience. Over four semesters, they will take a one-semester exploratory course in convergence and research development and four additional core courses – Student literacy in AI; AI, Society, & Ethics; statistical methods, and Ag data analytics – which culminate in an Ag Data Science Certificate. Other project components include faculty mentor training, professional development and networking opportunities, and partnerships for teaching and learning with Fayetteville State University. During the project's first year and a half, the first cohort of students was recruited, they completed the summer bridge experience, and took both Student Literacy in AI and the exploratory course where they developed the project proposals for the research to be conducted for their capstone project over the following three semesters.

Student Recruitment and Selection

Students were recruited through professional networks of the project team, posts on social media, the NC State Graduate School, and through the graduate program directors of the relevant programs. Entering students and those in their first through third years of graduate study were eligible. Each application was scored using a rubric that assessed 1) motivation and alignment with program goals; 2) perseverance, passion, determination, and resilience; 3) career goals and impact of program on them; 4) academic competitiveness; and 5) qualitative measure of enthusiasm for the student by the rater. The final cohort included eight women and one man in the disciplines of Electrical and Computer engineering (2 students), Biological and Agricultural engineering, Bioinformatics, Biomath, Soil Science, Crop Science, and Genetics (2 students).

Three-Week Summer Bridge Experience

The three week bridge program consisted of an "industry week," an "on-farm week," and "Camp Converge." Industry week began with a kickoff where the nine fellows were introduced to each other and the program. They also heard from a panel of start-up entrepreneurs that are supported through the Plant Science Initiative at NC State and a presentation on how to get venture capital funding. Over the next 3.5 days, they visited local businesses that engage in agricultural research or data science and AI. These businesses ranged from small companies focused on researching a single genetic trait for plants to divisions of multinational corporations. The last day of industry week focused on professional development, with sessions on networking and time management.

On-farm week included visits to five farms in the region that grow row crops (e.g., corn, soybeans), visits to four agricultural research stations, and participation in the Blacklands Farm Managers Tour. The students spoke to the farm owners and managers from farms ranging from 600-8000 acres. One farmer leased his land while the other farms were family-owned and operated. They learned that for all of the farmers, the goal is to get the highest yield of crops at the least cost. The farmers differed in their use of data to accomplish this objective. For example, one uses available data to make decisions about which crops and seed varieties work the best in which fields, how much fertilizer to apply, etc. By contrast, another farmer uses trial and error, paper records, walking the farm, and his personal intuition to make the same decisions. At the research stations, students were introduced to researchers who do cutting edge agricultural research and extension agents who support both the research endeavor and the farmers. The Blacklands Farm Managers tour brought all of these groups together with demonstrations by researchers and extension agents as well as seed, pesticide, and fungicide companies.

During on-farm week, the group stayed together in a dormitory-style arts center with a communal kitchen and other common spaces for relaxing and brainstorming. There was also a kayaking trip for recreation. Several members of the leadership team and students' advisors participated with the students in a day or two of activities, shared meals, and spent the night.

Camp Converge was a four-day workshop where students learned basic phenotyping and were introduced to several software applications for data analysis. Camp Converge introduced the engineers and data scientists to basic biology bench work while providing all students an introduction to software tools available through the university.

Explorations of Plant Science Data Course

Following the Summer Graduate Bridge Program, students took the two credit hour exploratory course which had the following student learning outcomes:

1. Design and implement a convergence research project at the intersection of plant science and AI.
2. Use collaborative approaches to solve complex problems.
3. Write an effective research proposal.
4. Present in a collaborative team atmosphere.

During the course, students heard a series of guest lectures managed by the project's coordinator. Each student's advisor was invited to give a brief presentation on the types of methods used by their lab and the kinds of research questions they address. Additional lectures were designed to introduce students to the fields of Implementation and Integration Sciences, Science of Team Science, and Systems Thinking.

The students participated in facilitated brainstorming sessions to develop project ideas in cross-disciplinary teams. They were placed in groups of (1) applied plant scientist, (1) basic plant scientist, and (1) data scientist and asked to design a potential project. After developing a brief proposal, they elicited feedback from mentors of the GRAD-AID for Ag program and their advisors. The projects students proposed in this course will be implemented over the remaining 1.5 years of their fellowship and will constitute a chapter in each student's dissertation.

Students developed three distinct projects. The first explores the genes that regulate cold tolerance in maize using machine learning techniques. Another seeks to predict sweetpotato breeding compatibility by modeling the well-documented success rates of the North Carolina Sweetpotato Breeding Program. The last group aims to design and implement a smart irrigation system using AI to predict optimal irrigation using soil, weather, and plant data.

Formative Assessment

The evaluator fully participated in industry week and on-farm activities with the trainees as a participant-observer. Patton [10, p. 223], describes participant observation as permitting the evaluator to understand a program to a degree not possible from second-hand accounts, such as interviews. Trainees were also asked to respond to short daily surveys about each day's activities and a longer survey at the end of each week in which they reflected on the full week's activities. The trainees also participated in a focus group at the end of the exploratory course to provide feedback about the course itself as well as a reflection on the summer bridge program.

Summer Bridge

Evaluation results show that the summer bridge program was quite successful at introducing students to multiple types of businesses engaged in agricultural research and to growers to help understand their needs. These experiences allowed the students to begin to formulate ideas for how AI can be combined with agriculture research to help solve grower problems. They also found that the time together, particularly in the dormitory and the van rides as well as the kayaking trip, was invaluable for getting to know each other and where each of their interests lay, which proved useful for later team formation. Industry week was quite fast-paced and future cohorts would benefit from introducing networking skills at the beginning of the week so that they could be employed during the company visits and eliminating overlap in the types of companies and their research with one or two fewer visits. Similarly, there were diminishing returns on additional visits to farms and research stations and eliminating one of each would allow the students more time to process what they learned. Camp Converge, was valuable for introducing the engineering and data science students to bench work in plant science and for introducing AI tools to all. However, on reflection, the students suggested that it be replaced with

visits to each of the labs of their advisors where they could learn more about their research and possibilities for convergence.

Exploratory Course

The exploratory course was challenging, particularly for a 2 credit hour offering which students took in combination with a 1 credit hour AI Literacy in Ag course. While the intent was for the courses to run in parallel with project work being advanced through both, in reality, the courses and their assignments were not well aligned resulting in unanticipated extra work. The recommendation is that the two courses be combined into one 3-hour team taught course in future offerings. From a professional development standpoint, the students appreciated the sessions on resumé and cover letter writing and informational interviews. They learned how to write a research proposal, which most of them had never done before. The students felt exceptionally well supported by the course instructor/project manager, citing her flexibility and can-do attitude in response to feedback. The process of developing their project proposals presented a challenge. Although there was one brainstorming session, the students would have preferred more and more structured brainstorming opportunities that incorporated the ideas they initiated during the summer bridge. They also requested more frequent feedback from the project's Research Training Advisory Board to help them refine their proposals to be doable over the two year timeframe. This culminated in a suggestion that project mentors be assigned who would meet regularly with the groups.

Future work

In response to student feedback on the project selection process, each team will have a dedicated mentor, who may or may not be any of their advisors, to guide them as they continue to develop their research and writing. In addition, students have monthly group meetings and will meet individually with the project manager and project PI over the course of the remaining year and a half of their projects. Each student is expected to participate in a summer internship experience between the first two years of their fellowship.

As part of their training, fellows will take 5 courses toward an Ag Data Science certificate. They have already taken an overview of AI in Ag and the exploratory course and will be introduced to ethical AI practices, statistical methods, and advanced analytics in Ag in future semesters. Other aspects of the project to be developed over the coming years include partnering with Fayetteville State University. Trainees will provide guest lectures to FSU undergraduate biology classes. The project will also work to develop a formal pathway between FSU and NC State, which may include undergraduate transfer programs (e.g., 2+2 or 3+2), dual degree programs, certificate programs, or bachelor-master programs.

Together, these multifaceted training approaches will drive innovation in AI technologies, while helping to ensure a skilled, diverse workforce capable of solving societal grand challenges in global Ag sustainability and food security.

This project is supported by NSF Grant No. 2345863

References

[1] M. Reynolds, A. Borrell, H. Braun, G. Edmeades, R. Flavell, J. Gwyn, D. Jordan, K. Pixley and G. Rebetzke, "Translational research for climate resilient, higher yielding crops," *Crop Breeding, Genetics and Genomics,* vol. 1, no. 2, 2019.

[2] S. Mourtzinis, P. D. Esker, J. E. Specht and S. P. Conley, "Advancing agricultural research using machine learning algorithms," *Scientific Reports,* vol. 11, no. 1, 2021.

[3] D. B. Dresbøll, B. Christensen and K. Thorup-Kristensen, "Chapter Five -- Approaches to translational plant science," in *Advances in Agronomy*, vol. 131, D. L. Sparks, Ed., Academic Press, 2015, pp. 305-335.

[4] Y. Xu, X. Zhang, H. Li, H. Zheng, J. Zhang, M. S. Olsen, R. K. Varshney, B. M. Prasanna and Q. Qian, "Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction," *Molecular Plant,* vol. 15, no. 11, pp. 1664-1695, 2022.

[5] A. L. Harfouche, D. A. Jacobson, D. Kainer, J. C. Romero, A. H. Harfouche, G. Scarascia Mugnozza, M. Moshelion, G. A. Tuskan, J. J. B. Keurentjes and A. Altman, "Accelerating climate resilient plant breeding by applying next-generation artificial intelligence," *Trends in Biotechnology,* vol. 37, no. 11, pp. 1217-1235, 2019.

[6] A. L. Harfouche, F. Nakhle, A. H. Harfouche, O. G. Sardella, E. Dart and D. Jacobson, "A primer on artificial intelligence in plant digital phenomics: Embarking on the data to insights journey," *Trends in Plant Science,* vol. 28, no. 2, pp. 154-184, 2023.

[7] T. Van Klompenburg, A. Kassahun and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture,* vol. 177, p. 105709, 2020.

[8] A. Morales and F. J. Villalobos, "Using machine learning for crop yield prediction in the past or the future," *Frontiers in Plant Science,* vol. 14, p. 1128388, 2023.

[9] T. J. Heindel, C. J. Lawrence-Dill, J. A. Dickerson and P. S. Schnable, "An interdisciplinary graduate course for engineers, plant scientists, and data scientists in the area of predictive plant phenomics," in *Proceedings ASEE Annual Conference & Exposition*, Columbus, OH, 2017.

[10] M. Q. Patton, Qualitative Research and Evaluation Methods, 3rd ed., Thousand Oaks, CA: Sage, 2002.