Partnerships and collaboration drive innovative graduate training in materials informatics

Angela M. Slates[1,2]†, Shana L. McAlexander[3]†, Jennifer Nolan[4], Juan de Pablo[4,5], Junhong Chen[4,6]*, Harley T. Johnson[2,7]*, L. Catherine Brinson[3]*

Holistic and intentional training prepares next-generation materials informatics leaders and workforce for expedited materials discovery and design.

Materials informatics (MI), which integrates artificial intelligence/machine learning (AI/ML) and computational methodologies with materials science for rapid materials discovery, understanding, and design, has emerged as a pivotal approach to transforming materials innovation. As industries increasingly recognize the potential of MI to revolutionize materials discovery and development cycles, an urgent need has arisen to equip the workforce with essential competencies in this highly interdisciplinary field. The US National Science Foundation (NSF) funds several related graduate training initiatives through the NSF Research Traineeship (NRT) program. This work aligns with NSF's Big Idea: Harnessing the Data Revolution and the Materials Genome Initiative.

CURRENT STATE OF MI AND KEY PROFESSIONAL COMPETENCIES

Several current trends underscore the timeliness of the MI approach: Growing databases of labeled materials data, significant advances in high-performance computing and infrastructure, rapid development of AI/ML algorithms, escalating demand for faster, efficient, and sustainable materials solutions in sectors such as energy, electronics, sustainability, and healthcare, and the transition towards Industry 4.0, characterized by smart, high-throughput manufacturing and data-centric operations.

Recent review articles have highlighted emerging areas in MI including AI for catalysts, photovoltaic materials, batteries, nanomedicine, and environmental and electronic materials along with advances in generative models, high-throughput screening, and autonomous labs to support materials discovery (*1–4*). We have identified critical competencies for successful development of the future materials workforce—including technical skills in data analytics, experience in custom AI/ML applications in materials science, and domain-specific knowledge in materials science. Combined with professional skills in team science/convergence/interdisciplinary collaboration, communication, career development, and ethics, we advocate for a comprehensive curriculum and activity design at the graduate level to train next-generation leaders in MI.

PROGRAM INSIGHTS

Development of convergent graduate programs with interdisciplinary and collaborative cores is crucial to the MI vision. NSF NRT programs, like Data-Enabled Discovery and Design of Energy Materials at Texas A&M University, paved the way for this next generation of collaborative effort (*5*). The following offers insights gained from our experiences in implementing three ongoing NSF NRT programs dedicated to the MI vision. Shared characteristics across these programs include rigorous foundational coursework in both materials science and AI/ML methods, interdisciplinary collaborations at the intersection of AI/ML and materials science, multidisciplinary cross-training, and a focus on professional development. Given the highly convergent nature of MI, a key strength for all programs lies in the participation and commitment from industry and national lab partners, solidifying a symbiotic relationship crucial for both the training and application phases.

Peer teaching and learning

The Data and Informatics Graduate Intern-traineeship: Materials at the Atomic Scale (DIGI-MAT) program was conceptualized in response to the growing demand for graduates with expertise in both data science and materials science. As a result of this increasing demand and recognizing the absence of formal programs preparing students at the intersection of data informatics and materials science, the University of Illinois Urbana-Champaign (UIUC), Grainger College of Engineering (GCOE), and the National Center for Supercomputing Applications (NCSA) initiated a novel approach with DIGI-MAT, leading to the first of the three NRT programs.

Launched in 2019, DIGI-MAT was introduced as a 5-year multidisciplinary PhD certificate program, recruiting, and cross-training a diverse group of 32 doctoral students across 8 disciplines, including engineering, computer science, physics, chemistry, and statistics, while sustaining the program goal of serving 40% women and underrepresented minorities. As a component of the DIGI-MAT curriculum, "iFridays," launched as a 1-credit hour, weekly seminar, aimed to address various fast-moving topics related to the DIGI-MAT curriculum and the trainees' professional growth. Inviting speakers from industry, national labs, and academia, the seminar covered relevant subjects focused on career

[1]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [2]Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [3]Thomas Lord Department of Mechanical Engineering and Materials Science, Edmund T. Pratt Jr. School of Engineering, Duke University, Durham, NC 27708, USA. [4]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637, USA. [5]Materials Science Division, Physical Sciences and Engineering Directorate, Argonne National Laboratory, Lemont, IL 60439, USA. [6]Chemical Sciences and Engineering Division, Physical Sciences and Engineering Directorate, Argonne National Laboratory, Lemont, IL 60439, USA. [7]Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
*Corresponding author. Email: junhongchen@uchicago.edu (J.C.); htj@illinois.edu (H.T.J.); cate.brinson@duke.edu (L.C.B.)
†These authors contributed equally to this work.

development, thesis and campus support, and future job/internship opportunities.

Building on this success in leveraging graduate student knowledge and rapid skill acquisition, an expanded peer-to-peer (P2P) learning experience was conceived which considers the trainees' desire for teaching experience, improved communication skills and stronger intergroup connections: "iShare" was launched as a professional development seminar specifically devoted to trainees and focused on P2P knowledge sharing.

iShare emerged as a summer replacement for iFridays, creating an opportunity for trainees to teach each key topic of choice, important for their growth as MI experts. The sessions delivered specialized training in areas such as software tools, lab instrumentation and professional development, which evoked increased engagement, autonomy, and connection among trainees. This higher engagement and satisfaction with the iShare model demonstrates what recent research also suggests. While traditional graduate education models position supervisors and faculty members as primary sources of guidance for student development and success, peers and near-peers offer a distinct form of ongoing support. The immediacy and shared experiences among peers create trust and a willingness to learn from each other. The significance of these relationships, while different from those with academic supervisors, can be equally consequential in the overall development and support of graduate students (*6*).

Interdisciplinary teams

The Harnessing AI for Design and Understanding Materials (aiM) program at Duke University focuses on cross-training a diverse group of doctoral students across 10 departments including engineering/physical sciences and math/statistics/computer science. Since 2020, the program has supported 32 trainees with women and underrepresented minorities making up 53% of the trainee population. All students in the program are expected to embark on PhD projects with MI components. This requires training in fundamentals of both materials science and ML and involves collaboration on interdisciplinary teams resulting in joint publications. Trainees begin the program with an orientation and boot camp involving a series of modules that cover disciplinary topics (e.g., materials thermodynamics and statistics) along with soft skills (e.g., science communications and collaboration).

During a team science experience, developed for the Joint-NRT symposium, students rapidly prepare research project proposals across disciplines and institutions. Within 30 minutes, student teams brainstorm project ideas and articulate a research approach. This activity gives students a sense of some of the components required for the culminating year-long team science capstone project course.

For the capstone, students with strong computational backgrounds team with students with strong physical science backgrounds. A faculty team across departments is committed to supporting and advising the projects. Teams create project charters that include roles and responsibilities, objectives/project scope, communication and dissemination plan, proposed timeline, anticipated challenges, and proposed methods. Projects are intended to become a chapter in each team member's dissertation and disseminated through conference presentations and manuscript publications. Among the 14 program projects to date: A team of chemistry and biostatistics doctoral students analyzed databases to improve biomedical polymers discovery using ML (*7*). Another team comprised of civil engineering and materials science doctoral students trained an ML model to predict the mechanical profile of porous materials (*8*).

Partnerships

The AI-enabled Molecular Engineering of Materials and Systems (AIMEMS) for Sustainability NRT program at the University of Chicago (UChicago) trains graduate students to become leaders who have not only outstanding research strengths in AI/ML for molecular engineering toward sustainability but also transferable professional skills to meet their diverse career goals and responsibly solve global challenges in a rapidly evolving environment. Since 2021, the AIMEMS program has recruited three trainee cohorts, including 16 students at various levels of their graduate training, spanning molecular engineering, computer science, physical science, and social science. Women and underrepresented minorities make up 44% of the trainee population. Structured technical and professional skills training are core components of the AIMEMS program. Four newly created courses in advanced materials characterization, AI/ML for molecular engineering, scale-up manufacturing of advanced materials and supercomputing are co-taught by UChicago faculty and Argonne National Laboratory (Argonne) scientists, exposing trainees to Argonne's world class expertise and facilities.

A key feature of the AIMEMS program is the co-mentoring of each trainee by a UChicago faculty member, a scientist from Argonne, and an industry professional. This team of advisers offers guidance on trainee research projects, professional development, internships, and career opportunities. Mentors contribute to trainee development through monthly seminars on professional skills ranging from STEM career development; diversity, equity, and inclusion; teamwork and leadership; entrepreneurship and innovation; and communication. Trainees develop communication, collaboration, teaching, and leadership skills by designing and leading symposia and boot camps in which mentors give research talks, participate in panel discussions, and engage in interactive grand challenge activities.

The co-mentorship model has facilitated transformative experiences for trainees. During a summer internship at Argonne, a molecular engineering trainee successfully ran a flow battery to test membranes and has since begun setting up this testing in house at UChicago. A physics trainee co-authored a journal paper on per- and polyfluoroalkyl substances (PFAS) research in collaboration with program co-mentors (*9*). International partnerships in France, Kenya, and the United Kingdom develop trainees' skills in collaborative global research.

In addition to their research and courses, AIMEMS trainees are dedicated to increasing interest and awareness of MI and STEM fields in-general through K-12 outreach. For example, a student trainee team designed and delivered a full-day outreach event for students from a local predominantly low-income, underrepresented minority high school. A second outreach event with a local middle school took place in Spring 2024, and follow-up events with both schools together are in the planning stages.

Interinstitutional program collaboration

What sets these three NRT programs apart and provides guidance for future MI educational success is meaningful inter-program collaborations. Knowledge sharing and joint activities bolster the interdisciplinary effort and maximize the impact of funding resources. The leadership teams meet regularly to share ideas and troubleshoot mutual challenges, and members of each NRT leadership team officially serve on each other's external advisory boards. Two common ongoing challenges regularly discussed are: effective methods for recruiting diverse populations of students and better engagement of computer scientists into MI. Especially notable is that the programs have brought trainees and leadership teams together for four core activities to date: Chicago Joint-NRT symposium; Hackathon planning and participation; iShare Joint NRT peer teaching seminar series; and professional skills seminars.

At the Harnessing Data for Materials Symposium, a joint event held in Chicago in Summer 2022, program participants and partners convened to share knowledge and develop new research collaborations at the intersection of ML and materials science. Trainees found the rapid proposal development process intense, challenging, and a highlight of the event, with one participating student coining the phrase "collaborate or die!" The benefit of this collaboration is that trainees practiced team science and visualized how their efforts align with the goals of a larger community. Another joint symposium is planned for late Summer 2024.

Following the momentum of the joint symposium, trainees from the three NRT programs met to forge interdisciplinary connections to prepare for the 2023 SSMCDAT Hackathon at Lehigh University. After several iterations of ideas and collaborators, a team of three emerged and submitted a successful proposal on the use of ML methods to bypass density functional theory calculations. Trainees valued the interdisciplinary experience, the collaboration and intense problem-solving using MI during the Hackathon.

The collaboration efforts continued into summer 2023, as the three NRT programs organized an 8-week iShare peer-to-peer teaching series based on the UIUC model. All NRT trainees from all three programs were invited to attend and/or present a topic of their choosing. Trainees taught each other technical skills, demonstrated tools and methods in MI, and offered insights into planning events. The series enabled students to learn from one another, practice their teaching and communication skills, and consider potential collaborations.

Individual institutional partnerships contributed to graduate student development across the NRTs throughout 2023 via online professional skills seminars offered by UChicago and Duke. Students came together virtually to gain knowledge and skills in teaching, technical areas, and career development from a wide array of academic, industry and national lab experts. The connections build upon and strengthen our NRT community at the intersection of ML and materials.

RECOMMENDATIONS

As we move forward with development of extensible methods to educate the future MI workforce, the priority remains to sustain and amplify collaborative energy, ensuring that academia, industry, and national laboratories remain interwoven in this journey. By sharing best practices from our respective programs, we hope to provide a blueprint for other institutions to create similar programs to cross train and collaborate to create the MI workforce.

While these NRT programs necessarily focus on graduate level training, we encourage use of our work as inspiration to prepare students at all levels for interdisciplinary collaboration at the intersection of AI/ML and materials. Key steps include 1) providing core training in foundational areas of both AI/ML and materials, 2) breaking down department-based silos to encourage productive teaming, and 3) prioritizing partnerships with academia, industry, and national laboratories. We also found that significant team-based student projects leverage the core foundational skills and promote multidisciplinary learning and research that drives both discovery and student success (*10*). Partnerships promote building an integrated, interdisciplinary workforce that can effectively tackle global materials challenges and drive innovation. We urge academic institutions to evaluate and address departmental silos while providing students with the interdisciplinary mentality, skills, and knowledge necessary for success in the 21st-century MI workforce.

REFERENCES AND NOTES

1. R. J. Hickman, P. Bannigan, Z. Q. Bao, A. Aspuru-Guzik, C. Allen, Self-driving laboratories: A paradigm shift in nanomedicine development. *Matter* **6**, 1071–1081 (2023).
2. C. Duan, A. Nandy, H. J. Kulik, Machine learning for the discovery, design, and engineering of materials. *Annu. Rev. Chem. Biomol. Eng.* **13**, 405–429 (2022).
3. Z. Wang, A. Chen, K. Tao, Y. Han, J. Li, MatGPT: A vane of materials informatics from past, present, to future. *Adv. Mater.* **36**, e2306733 (2024).
4. B. H. Ryu, L. Q. Wang, H. H. Pu, M. K. Y. Chan, J. H. Chen, Understanding, discovery, and synthesis of 2D materials enabled by machine learning. *Chem. Soc. Rev.* **51**, 1899–1925 (2022).
5. D. A. Fowler, R. Arroyave, J. Ross, R. Malak, S. Banerjee, "Looking outwards from the "central science": An interdisciplinary perspective on graduate education in materials chemistry" in *Educational and Outreach Projects from the Cottrell Scholars Collaborative Undergraduate and Graduate Education Volume 1*, R. Waterman, A. Feig, Eds. (American Chemical Society, 2017), pp. 65–89.
6. D. Lorenzetti, L. Nowell, M. Jacobsen, L. Lorenzetti, T. Clancy, G. Freeman, E. O. Paolucci, The role of peer mentors in promoting knowledge and skills development in graduate education. *Educ. Res. Int.* **2020**, 8822289 (2020).
7. S. M. McDonald, E. K. Augustine, Q. Lanners, C. Rudin, L. C. Brinson, M. L. Becker, Applied machine learning as a driver for polymeric biomaterials design. *Nat. Commun.* **14**, 4838 (2023).
8. W. Lindqwister, J. Peloquin, L. Dalton, K. Gall, M. Veveakis, Predicting the mechanical response profile of porous materials via microstructure-informed neural networks (2023); http://dx.doi.org/10.2139/ssrn.4605136.
9. S. Dasetty, M. Topel, Y. O. Tang, Y. Q. Wang, E. Jonas, S. B. Darling, J. H. Chen, A. L. Ferguson, Data-driven discovery of linear molecular probes with optimal selective affinity for PFAS in water. *J. Chem. Eng. Data* **68**, 3148–3161 (2023).
10. National Academies of Sciences, Engineering, and Medicine, *Graduate STEM Education for the 21st Century* (National Academies Press, 2018).

Acknowledgment: This work is supported by the National Science Foundation under grant numbers: DGE-2022040, DGE-1922758, and DGE-2022023.

10.1126/sciadv.adp7446