NEXT-GENERATION BIOCOMPUTING DOCTORAL TRAINING IN THE ERA OF AI

Graciela Gonzalez-Hernandez, PhD
Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA
E-mail: Graciela.GonzalezHernandez@csmc.edu

Steven E. Brenner, PhD
Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

This PSB 2027 workshop examines how doctoral education in health AI, biomedical informatics, computational biomedicine, computational biology and related fields should evolve in the era of artificial intelligence and large language models. Invited perspectives address international AI and digital-health training, computational biology within computer science, biomedical informatics curriculum innovation, health-system-embedded Health AI education, and AI-augmented research practices including vibe coding. The workshop will compare training models, examine tensions involving technical depth, domain expertise, mentoring, assessment, and research ownership, and identify practical experiments that students, faculty, and program leaders can implement in their own settings.

Keywords: doctoral education; health AI; biomedical informatics; computational biology; large language models; AI-assisted research; curriculum design; assessment; vibe coding

1. Introduction

Artificial intelligence is reshaping the scientific and professional environment in which the next generation of biocomputing researchers will be trained. Advances in machine learning, large language models, automated coding, and AI-assisted research are changing how computational methods are developed, how analyses are conducted, and how scientific knowledge is communicated. In health AI, biomedical informatics, computational biomedicine, and computational biology and related fields, these changes intersect with distinctive challenges. Researchers must work with heterogeneous, incomplete, and longitudinal clinical and biological data; connect computational models to disease mechanisms, patient trajectories, and care processes; collaborate across computational, biological, and clinical disciplines; and account for the consequences of error in settings where model outputs may influence scientific conclusions, clinical decisions, or health-system operations.

These conditions create substantial opportunities for doctoral education, while also challenging existing assumptions about what students need to learn and how they should be trained. Biocomputing programs must integrate contemporary AI and machine-learning methods without displacing the methodological, clinical, biological, and socio-technical foundations of their disciplines. They must prepare students to work in environments where AI systems can generate code, summarize literature, propose analyses, and assist with scientific writing, while ensuring that students can evaluate, explain, reproduce, and take responsibility for AI-assisted work. The goal is

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

therefore not simply to add more AI content to already dense curricula, but to develop researchers with the capacity to learn continuously, formulate meaningful problems, collaborate across disciplines, and exercise sound judgment in complex biomedical settings.

The workshop, *Next-Generation Biocomputing Doctoral Training in the Era of AI*, brings together educators, program leaders, researchers, faculty mentors, and trainees to examine how doctoral education should evolve in response to these changes. It focuses on health AI, biomedical informatics, computational biomedicine, and computational biology, with particular attention to how different institutional and disciplinary contexts shape doctoral training. The workshop asks how programs can balance technical depth with domain expertise, formal coursework with self-directed learning, methodological rigor with rapid experimentation, and traditional expectations of independence with new forms of AI-assisted research.

The workshop is intended for PhD students and postdocs, faculty mentors, instructors, program directors, and health-system or industry leaders who contribute to doctoral training. By connecting program-level questions with the everyday practices of mentoring, evaluating AI-assisted work, and conducting research with clinical and biological data, the workshop aims to help the biocomputing community prepare researchers not only to use AI systems, but also to understand, evaluate, govern, and responsibly extend them.

2. Workshop Talks

The program combines short talks on distinct doctoral training models with a guided panel and a practicum-style closing discussion. The talks provide examples of international, computational biology, biomedical informatics, and health-system-embedded approaches to doctoral education. The panel will examine design tensions involving curriculum, mentoring, assessment, research ownership, and the distinctive identity of a Health AI or biomedical informatics PhD. The closing practicum will focus on practical experiments that participants can implement in a course, laboratory, mentoring relationship, or doctoral program, including approaches to AI-use policies, code validation, assessment, and self-curated learning.

The workshop features invited perspectives on distinct approaches to doctoral training in the era of AI. Karin Verspoor from RMIT University will discuss international approaches to education in artificial intelligence, natural language processing, and digital health. Her perspective will highlight how national training initiatives, institutional structures, funding environments, and interdisciplinary programs shape doctoral education and the development of computational health researchers.

Ben Raphael from Princeton University will describe computational biology training within a computer science PhD environment. His perspective will examine a mentoring-centered, method-intensive model in which students develop expertise through research in computational and cancer genomics, including genomic alterations, tumor evolution, single-cell and spatial methods, and AI-driven modeling. This model provides an opportunity to consider how specialization, supervision, and research experience can substitute for or complement formally defined health AI curricula.

John H. Holmes from the University of Pennsylvania will contribute a biomedical informatics perspective grounded in doctoral education and curriculum innovation. His discussion will address how established informatics programs can incorporate modern AI and machine-learning methods while preserving foundational training in clinical informatics, population health, data governance, and responsible research practice. This perspective is particularly important for understanding how mature programs can evolve without losing the disciplinary depth that supports trustworthy biomedical research.

Graciela Gonzalez-Hernandez from Cedars-Sinai Medical Center will discuss a health-system-embedded model of doctoral training in Health AI. The PhD pedagogy adopted at Cedars-Sinai emphasizes students' capacity to learn, rather than dependence on a fixed body of technical content that may quickly become outdated. Students develop as researchers through an apprenticeship model in which they work closely with faculty and clinicians, learn to formulate meaningful problems, and iteratively connect computational methods to clinical needs. Active learning, clinical immersion, project-based work with real hospital data, and integration with AI governance and evaluation are central to this model. The health-system setting allows students to learn not only how to build models, but also how clinical workflows, institutional priorities, data provenance, safety, and implementation constraints shape responsible Health AI research.

Steven E. Brenner from the University of California, Berkeley, will focus on AI-augmented research and the emergence of vibe coding in computational biology and the life sciences. Natural-language-to-code systems may allow researchers to prototype analyses, explore datasets, and construct computational pipelines more rapidly, including for students without extensive formal programming backgrounds. However, vibe coding also introduces new risks and responsibilities. Doctoral students must be able to validate AI-generated code, recognize hidden assumptions and errors, reproduce analyses, explain computational decisions, and modify generated pipelines when research questions change. These developments challenge traditional assumptions about programming competence, supervision, authorship, and independent work. The talk will consider how doctoral programs should teach students to validate, explain, debug, modify, and take responsibility for AI-generated code.

Together, these talks present complementary perspectives on doctoral training across institutional, disciplinary, and national contexts. They address established and emerging program models, the role of mentoring and clinical immersion, the integration of AI and machine learning into curricula, and the changing expectations surrounding computational competence and independent research.

3. Discussion

The workshop’s central discussion will examine the design tensions that arise when doctoral programs adapt to pervasive AI. One tension concerns depth versus breadth. Students increasingly need familiarity with machine learning, large language models, software engineering, clinical data, responsible AI, and domain-specific science. However, doctoral training cannot simply expand indefinitely. Programs must determine which competencies are essential, which can be developed through self-directed learning, and which should remain the responsibility of formal coursework and supervision.

A second tension concerns methods versus domain knowledge. Health AI researchers must understand both advanced computational methods and the clinical, biological, and organizational contexts in which those methods are applied. Programs that emphasize generic AI skills may produce technically capable graduates who lack sufficient understanding of healthcare and biomedical research. Conversely, programs that focus primarily on domain knowledge may leave students unprepared for rapidly changing computational practices.

A third tension concerns assessment and research ownership. When AI systems can generate code, analyses, literature summaries, and written text, conventional assignments and milestones may no longer reliably measure individual understanding. Doctoral programs may need to place greater emphasis on oral examinations, live problem-solving, code walkthroughs, reproducibility exercises, and the ability to explain and defend an analytical pipeline. These approaches can distinguish productive use of AI as a research tool from reliance on AI outputs that a student cannot evaluate or justify.

These questions will be explored through an open, moderated panel rather than a sequence of independent program descriptions. The panel will ask whether programs have deliberately removed or de-emphasized topics to make room for AI, what makes a Health AI or biomedical informatics PhD distinct from a generic AI or data-science degree, how programs can demonstrate that students own and understand AI-assisted work, and how doctoral education can remain attractive and meaningful when students can enter industry without pursuing a PhD.

The discussion will also consider how programs should establish expectations for responsible AI use. Important questions include how students should document AI assistance, how AI-generated code should be validated, how errors should be attributed, and how mentors can supervise work in which substantial portions of the implementation were generated through natural-language interaction. These issues affect not only coursework and examinations but also dissertation research, authorship, research integrity, and professional preparation.

Finally, the workshop will address opportunities for collaboration and self-curated learning. Participants may benefit from shared curricula, open syllabi, peer-led reading or coding groups, cross-institutional project opportunities, common assessment frameworks, repositories of validated teaching materials, and communities of practice for mentoring in AI-intensive research environments. Because many faculty members and students do not control an entire doctoral program, the practicum-style closing discussion will ask each participant to identify one concrete experiment for the coming year, such as changing an assessment, piloting an AI-use policy, creating a code-validation exercise, or developing a structured learning plan.

4. Conclusion

The workshop highlights that next-generation biocomputing doctoral education cannot be defined solely by adding AI courses to existing curricula. It requires reconsidering how programs combine

methodological depth, domain expertise, clinical and biological context, mentorship, assessment, and research practice.

The invited perspectives provide a basis for comparing different models: international and interdisciplinary training initiatives, computer-science-based computational biology programs, established biomedical informatics curricula, health-system-embedded Health AI programs, and AI-augmented approaches to computational research. Their convergence suggests that future doctoral training will need to support both technical adaptability and deep scientific understanding.

The workshop's intended outcome is a set of practical experiments that participants can pursue over the following year. These may include revised assessment practices, explicit guidance for AI-assisted research, new approaches to teaching and validating AI-generated code, self-curated learning opportunities, and stronger cross-institutional collaboration. The discussion will also support longer-term community efforts, such as a shared resource repository or a perspective on doctoral training in Health AI and biomedical informatics in the era of pervasive AI and vibe coding. More broadly, the workshop aims to help the biocomputing community develop doctoral programs that prepare researchers not only to use AI systems, but also to understand, evaluate, govern, and responsibly extend them.