Integrating Generative and Experimental Platforms for Biomolecular Design

Contents

1 Workshop Summary

Motivation. Biomolecular design seeks to develop new solutions and lower developments costs to medical, industrial, and environmental challenges. Researchers now routinely engineer proteins, ligands, and nucleic acids (DNA, RNA) to perform functions beyond those found in nature. Generative machine learning (ML) has recently shown remarkable promise in creating novel, functional biomolecules, enabled by decades of accumulated experimental data and the rapid growth of discriminative and generative modeling techniques. Yet a critical gap persists: **most ML work still optimizes for state-of-the-art results on static benchmarks, divorced from experimental practice**. This disconnect produces misaligned evaluation metrics and leaves many high-impact biological problems underexplored. With generative modeling now central to ML, the urgent need is not just better models, but models directly integrated with experimental workflows. Collaboration between biologists and ML researchers is therefore *essential* to ensure that generative design translates into validated, real-world biological advances.

Overview. Our proposed workshop will **connect real-world biological problems with generative machine learning**, explicitly fostering collaborations between experimental and computational researchers. We focus broadly on biomolecular and cellular design (spanning small molecule, proteins, nucleic acids, and engineered cells) reflecting the rapid rise of "virtual cell" technologies that now demand generative models capable of reasoning across multiple scales of biology. We will highlight sequence-based and structure-based approaches as well as emerging multimodal methods that unify different representations, with particularly a emphasis on works that have experimental validations. The program centers on three themes:

1. Generative ML for biomolecular and cellular design.
    - *Inverse design.* Generative models can now propose diverse and novel biomolecules, yet designing for *specific functional properties and/or constraints* remains unsolved. We will invite advances across discrete and continuous generative methods for diverse applications such as small molecule modulators, therapeutic proteins, genome editors, and programmable cell states.
    - *Modeling complex biological data.* Biological systems span sparse single-cell datasets to abundant high-throughput assays. We will welcome domain-aware generative models that capture causal mechanisms in these systems. Progress here requires leveraging sequence-based (e.g., protein language models), structure-based (e.g., geometric diffusion), and multi-modal approaches.
2. Integrating generative ML into experimental workflows.
    - *Adaptive experimental design.* A persistent challenge is how to optimally leverage experimental results to inform subsequent generative proposals. We will solicit work on active learning approaches such as Bayesian optimization and reinforcement learning that provide frameworks to iteratively refine biomolecule or cell designs in partnership with the lab bench. This includes interfaces with computational agents, robotic labs and cloud experimentation.
3. Biological problems ripe for generative ML and the development of robust *in-silico* oracles.
    - *Problem settings.* Many high-value areas in protein engineering, small molecule design, RNA therapeutics, genome editing, and cell engineering are underexplored by generative ML despite abundant data or high-throughput experimental pipelines. The workshop will spotlight these opportunities and invite contributions that pair algorithms with new, deployable assays.
    - *In-silico benchmarks.* With the increasing use of *in-silico* oracles, it is critical to assess how well they align with wet-lab outcomes. We will encourage benchmarks that are computationally tractable, biologically grounded, uncertainty-aware, and designed to track real experimental success.

Publication. The workshop is once again partnering with *Nature Biotechnology*, one of the most high-impact journals with a 5-year Journal Impact Factor of 59.5, where select papers will be invited for fast-track submission. This collaboration aims to elevate impact of the presented work and encourage top-tier submissions, especially from the biology sector, which is classically underrepresented at ML conferences. See Section 3 for details.

Previous Editions.
GEM@ICLR 2024 (Vienna). The inaugural GEM Workshop received 108 submissions from 142 institutions across 22 countries; 56 papers were accepted, including 7 orals, and 7 were fast-tracked to *Cell Systems*. We awarded 8 travel grants to students from underrepresented communities. Attendance exceeded 300 participants, with strong engagement from academia and industry. The program featured cutting-edge work on protein, RNA, and biological circuit design, as well as applications in drug discovery. Highlights included two dedicated poster sessions, a dynamic panel of young leaders, keynote talks from established researchers, and trainee presentations.

GEM@ICLR 2025 (Singapore). The second edition cemented GEM as a premier AI $\times$ Bio venue to exchange new ideas. We received nearly 130 submissions, accepting 75; 8 were presented orally and another 8 highlighted as spotlight posters. In a major collaboration, we partnered with *Nature Biotechnology*, and 10 papers were downselected for fast-track submission, many now under review. Overall, we provided 8 travel awards, and **our sponsors contributed more than $200,000 in GPU credits to attendees**. We launched a seed-grant competition, which awarded $\$5,000$ prizes to interdisciplinary collaborations that formed during the workshop, one of which has already led to a publication [2]. We are planning to invite winning teams to present their results, reinforcing GEM's role as a scientific forum and a catalyst for new collaboration.

2 Logistics

2.1 Format

Our workshop will **be held in-person** with **virtual arrangements** for those unable to attend in-person. It will be a **large-attendance talk** format enriched with **two poster sessions**, a **panel discussion**, **contributed talks** for best papers at the workshop, and multiple **networking sessions**. We believe this format will strike a balance between delivering scientific advancements and building connections between the computationalists and experimentalists attending the workshop. We have planned multiple poster sessions and networking events to discourage attendees from breaking off into small, familiar groups and instead establish relationships with potential collaborators.

Virtual engagement. We will offer virtual engagement options for those unable to attend in person. Attendees will be able to access recorded sessions, posters, slides, and papers through our workshop website. We will require all accepted submissions to send us their posters and camera-ready papers as PDFs. In addition, we will provide links to papers published at our workshop and successfully fast-tracked in *Nature Biotechnology*.

2.2 Accessibility

Website. We will re-use our website[1] for the workshop. We will use OpenReview for the review process. Accepted ML-track papers and biology-track abstracts will be released on our website before the workshop and maintained afterwards. The final schedule will be provided on ICLR's and our websites. Talk title and abstracts will be provided on ICLR's website. The sessions for each poster will also be made available as a public Google Sheets document.

Resources. To promote mutual understanding between experimentalists and computationalists for productive and insightful discussions, this year, we will provide tutorial videos covering key concepts in biomolecular design and generative ML. Furthermore, presenters will receive guidance on making their content more accessible and engaging for all attendees. We also will organize social mingles to promote networking between our diverse audience.

Travel awards. To foster diversity, equity, and inclusion (DEI), our sponsors will provide free workshop registration and fund travel for selected applicants using sponsorship funding, where the priority will be given to *students and minority groups based on DEI*.

2.3 Anticipated Audience & Size

Our audience would be diverse given the interdisciplinary nature of our workshop. Based on our previous year's experience and discussions with related workshop organizers, we anticipate around **300 attendees**. Our primary goal is to attract researchers working at the intersection of machine learning and biology. We also seek to engage pure ML researchers who are looking for applications for their work, as well as biologists who are exploring new machine learning techniques to address their problems. Additionally, we aim to draw in industry researchers actively involved in this field.

2.4 Seed Grants

Last year, the GEM workshop set fundraising records with over $40,000 of sponsorship from a variety of organizations, and one sponsor contributed over $200,000 worth of compute credit to the attendees. This year, we have secured funding from Nvidia, Braidwell, and FutureHouse, and are actively engaging additional sponsors such as Genentech and Microsoft Research, leaders at the intersection of generative ML and biology. This reflects a shared commitment to advancing innovation in biomolecular design.

With these funds, we are launching competitive seed grants ($1,000–$2,000 each), with the final amount awarded depending on the judges' evaluations. The judging panel will consist of our invited speakers and panelists. These grants are intended to catalyze new collaborations between experimental and computational researchers attending the workshop, providing initial resources for high-potential projects aligned with the workshop's themes.

To facilitate team formation, we will host a dedicated "matchmaking" social event during the main conference, where participants with overlapping research interests can connect and brainstorm proposals. Teams will then be invited to submit concise half-page proposals outlining their project idea, collaboration plan, and expected outcomes. Awards will be based on novelty, alignment with the workshop's themes, feasibility, and anticipated impact. Funds will be disbursed after the workshop to the awardees' institutions to support project initiation. To further incentivize follow-through, awardees will be invited back to present short talks at the potential 2027 edition of GEM, showcasing results and inspiring future collaborations.

2.5 Adaptyv-Sponsored Protein Design Competition

In partnership with Adaptyv Bio, we will run a new competition track, *Applied AI Protein Design*, to challenge participants to devise novel strategies for binder design. The format will be interactive and team-based: participants will be presented with a protein design problem at the workshop and asked to form teams, identify compelling biological targets (e.g., diagnostic applications lacking reliable binders), and propose creative generative design strategies.

[1] URL https://gembio.ai/

The top 5 teams, based on their proposals, will be awarded experimental validation of their approaches in Adaptyv's lab. Adaptyv has committed to sponsoring a total of 300 binding assays, which will be divided among winning teams (5 teams with 60 assays each). These assays can accommodate a broad range of protein designs, including peptides, mini-proteins, VHHs, scFvs, and Fabs, provided they are under 300 amino acids in length. The winning teams will have the opportunity to design proteins in the weeks following the workshop, with experimental results delivered within 2–3 weeks afterwards. This initiative ensures that participants' generative ideas translate into tangible wet-lab validation, further strengthening the bridge between ML and biology.

2.6 Align Foundation Tournament

We are also collaborating with The Align Foundation to promote their annual Tournament series, which highlights socially impactful protein engineering challenges. This year's challenge focuses on designing PETases for plastic degradation. While the official competition runs independently, we will advertise the Tournament prior to and during GEM submissions, encouraging participants to engage and contribute. This partnership broadens the scope of opportunities for attendees, providing another venue for applying generative ML methods to pressing global problems. Together with our seed grants and the Adaptyv competition, the Align Tournament ensures that GEM is not only a forum for sharing results but also a catalyst for impactful, real-world discovery.

2.7 Broad Outreach

Our organizing and program committee members will leverage their extensive academic networks to increase workshop awareness in both generative ML and biology. This includes sharing event details within their institutions and with peers in the field. We will also collaborate with our sponsors and our connections in industry research labs, such as Amgen, Genentech, Intel, Microsoft, Nvidia, Recursion, to disseminate information about this workshop. As we successfully did last year to extend our reach, we will employ various social media platforms, including Twitter, Facebook, LinkedIn, WeChat, and blog posts to foster interactions and engage with the general audience of the workshop.

2.8 Timeline

Main workshop deadlines:

- Workshop submission deadline: February 1st, 2026.
- Workshop accept/reject notification date: March 1st, 2026.

We will then follow-up with selected papers for fast-tracking to *Nature Biotechnology*, where decisions are tentatively scheduled before May 2026.

3 Submissions

Submission tracks. The workshop submission is designed to attract high-quality original papers at the intersection between biomolecular design and generative AI. We will provide three submission tracks for topics described in Section 1:

- **Machine learning track**. This track will feature generative machine learning advancements for biomolecular design where results are entirely **in silico**. Topics include inverse design, biomolecular data modelling, adaptive experimental design, and the development of *in silico* benchmarks.
- **Biology track**. This track will consist of papers which have **wet lab** experimental results. We will welcome hybrid works employing ML for experimental biomolecular design problems, as well as biologically grounded problem settings that reveal challenges or opportunities for generative ML (e.g., high-throughput techniques, single-cell analysis) that are relevant to generative ML.
- **Competition track**. With Adaptyv, this track will highlight teams participating in the *Applied AI Protein Design* competition (see Section 2.5). Submissions will consist of short proposals outlining novel generative strategies for protein binder design. Selected teams will pitch their ideas live at the workshop, and winners will be awarded experimental validation of their designs in Adaptyv's lab. This track ensures that the workshop directly catalyzes real-world testing of generative models.

By providing a biology track and partnering with *Nature Biotechnology*, we hope to attract researchers from hybrid labs, as well as biologists who are pursuing state-of-the-art ML techniques for their experimental research. The biology track contained 20 papers last year, featuring diverse wet lab experiments in conjugation with generative ML.

Submission guidelines.

- **Length & format.** Both ML and biology tracks will accept submissions up to 5 pages in length (excluding appendix). The biology track will also consider extended abstracts (up to 2 pages), similar to standard biological conferences. The competition track will only consider extended abstracts (up to 2 pages), similar to typical proposals. The ICLR template will be used.
- **Anonymity.** Submissions are double-blind and cannot reveal author identities in the manuscript, appendix, code, or acknowledgments. Preprints are allowed but must not be cited in a way that reveals identity.
- **Reproducibility.** We strongly encourage releasing code, model checkpoints, and (when permissible) data or synthetic data generators. Include a brief reproducibility checklist covering datasets, splits, metrics, baselines, and compute.

- **Archival.** All tracks are *non-archival*. Dual submission is permitted, but papers previously published at an archival venue will be rejected.
- **Ethics and safety.** Manuscripts must not include procedural details that materially enable misuse or unsafe experiments; any wet-lab work must comply with institutional and national regulations.
- **Use of large language models.** We follow ICLR policies on LLM usage. LLMs and other AI tools *cannot* be listed as authors or reviewers. Any assistance must be disclosed in the paper's statement; AI-written reviews or manuscripts are not permitted.
- **Encouraging short papers.** We align with ICLR's encouragement of short/tiny papers: **all tracks accept short papers**. To lower barriers for junior researchers and experimental labs and to encourage late-breaking ideas, the Biology and Competition tracks explicitly support **extended abstracts/tiny papers** to allow the lower barriers for experimental groups and late-breaking ideas.

Review process. The review process will be double-blind, and they will be conducted through OpenReview. We anticipate around *200* submissions, based on our previous year's experience and discussions with organizers with related workshops. We aim to accept around *75* papers. We will recruit up to *100* reviewers from diverse backgrounds (see Section 10). Each submission will receive *3* reviews and each reviewer will review up to *3* submissions.

Conflict of interest. To prevent conflicts of interest, reviewers will not evaluate submissions from their department or from their collaborators of the past 5 years. Workshop organizers will not give talks at this workshop.

Awards. Accepted papers with exceptional quality will be recognized through Best Paper and Distinguished Paper awards, as agreed upon by review scores and area chairs. Outstanding accepted papers will be selected for a series of contributed talks, offering a platform for further discussion among workshop attendees.

Partnership with *Nature Biotechnology* In collaboration with *Nature Biotechnology*, authors can opt for their paper to be considered for a fast-track review at *Nature Biotechnology*. Based on the workshop reviews, our organizing committee and editors at *Nature Biotechnology* will select high quality papers for an additional round of review at *Nature Biotechnology*. Accepted papers at *Nature Biotechnology* will form a special collection. This will be similar to our collaboration with *Cell Systems* in GEM-2024 and last year's collaboration with *Nature Biotechnology* at GEM-2025.

4 Tentative Schedule

We aim to create interdisciplinary discussions surrounding the formidable challenges in biology and how generative machine learning can tackle them. To create an engaging and inclusive workshop appealing to a diverse audience, our program features a variety of sessions. These sessions encompass keynote speeches (**invited talks**), selected contributed machine learning papers, experimental biology and *in-silico* modelling abstracts (**contributed talks**), engaging **poster sessions**, an insightful **panel discussion**, and **social mingles** in between and after sessions.

Invited talks. We highlight emerging investigators in our talks. Each invited talk is structured with 25 minutes dedicated to the presentation and an additional 5 minutes reserved for questions. Each talk will focus on a theme of the workshop (see Section 1). To ensure a balance of perspectives, we will feature three experimentally focused talks (highlighting advances in data generation, high-throughput screening, and automation) and two ML–oriented talks from young rising leaders at the AIxBio interface. Speakers have been selected for their recognized expertise, diverse scientific achievements, future research potential, and exceptional presentation skills (see Section 5.1).

Contributed talks. For contributed talks, we will employ a rigorous peer-review selection process, guided by the diversity of topics and high reviewer scores, ensuring that we spotlight outstanding and impactful submissions. Our goal is that the talks strike a balance between biology and machine learning.

Poster sessions. Poster sessions follow contributed talks, offering a broader range of topics and a space for more personal and detailed conversations. Discussions around posters will foster connections and idea exchange amongst our participants. We purposely planned multiple poster sessions so presenters can also visit other interesting works.

Panel discussion. The panel discussion will spotlight senior leaders at the intersection of ML and biotechnology to discuss the past and the future of generative AI as applied to biomolecular design. We will identify current misalignment and challenges, unsolved biological problems where generative ML is set to disrupt the stage in the coming few years, as well as AI safety challenges. The panel discussion will be moderated by a workshop organizer, and will include a Q&A session with the audience (see Section 5.2).

Seed grant pitches & announcements Interested researchers can sign up to be grouped into pairs of experimentalists-computationalists at the beginning of the main conference. Each pair will brainstorm a concise research proposal to be shared in the workshop with our panelists/speakers/sponsors during the second poster session. The rapid-fire session will enable researchers to pitch groundbreaking ideas to a panel of established researchers and venture capitalists, garnering immediate feedback and fostering collaboration. The selected proposals will be announced before closing remarks and will receive cash prizes or computational credits courtesy of our sponsors (see Section 2.4).

Social mingles. A key objective of the workshop is to bring experimentalists and computationalists together for collaborations. With support from our generous sponsors, we plan to provide lunch within the venue to foster continued interaction. Additional lunch activities, such as topic-specific, technique-based, or self-organized breakout sessions, can be arranged if the venue layout permits. An after party will also be organized to encourage further networking and idea exchange.

Tentative schedule	
8:50 - 9:00 AM	Open remarks
9:00 - 9:20 AM	Invited talk 1
9:20 - 9:40 AM	Invited talk 2
9:40 - 10:00 AM	Invited talk 3
10:00 - 10:15 AM	Coffee break
10:15 - 11:00 AM	Contributed talks (4 talks)
11:00 - 12:00 PM	Poster session 1
12:00 - 1:00 PM	Lunch break & social mingle
1:00 - 1:20 PM	Invited talk 4
1:30 - 2:30 PM	Panel discussion
2:30 - 2:50 PM	Coffee break
2:50 - 3:10 PM	Invited talk 5
3:10 - 3:50 PM	Contributed talks (4 talks)
3:50 - 4:40 PM	Poster session 2 & seed grant pitch
4:40 - 4:50 PM	Seed grant awards
4:50 - 5:00 PM	Paper awards, closing remarks

5 Invited Speakers and Panelists

5.1 Invited Speakers

- Emma Chory (emma.chory@duke.edu, confirmed) is an Assistant Professor of Biomedical Engineering at Duke University. An experimentalist, her research career has spanned epigenetics, chromatin remodeling, and synthetic biology, with formative work at Harvard Medical School, the Dana-Farber Cancer Institute, and Stanford. At Duke, her lab combines synthetic biology, chemical biology, automation, and sequence-to-function modeling to build new tools for understanding molecular biology and to accelerate the development of next-generation therapeutics.
- Octávio Luiz Franco (flanco@ucdb.br, confirmed) is a Professor at Universidade Católica de Brasília and Universidade Católica Dom Bosco in Brazil, and a CNPq Researcher 1A. His research focuses on antimicrobial peptides (AMPs), enzyme inhibitors, and molecular strategies to combat pathogenic microorganisms, spanning from fundamental biology to translational therapeutics. More recently, his group has pioneered the use of machine learning for AMP discovery and design, integrating computational models with biochemical and microbiological validation. Franco's work exemplifies how ML-driven generative approaches can accelerate peptide therapeutics, making him a key voice in connecting experimental biology in South America to the global AIxBio community.
- Ben Kompa (ben.kompa@lila.ai, confirmed) is Head of AI Lab Innovation at Lila Sciences, where he leads early efforts in applying AI, robotics, and automation to life, chemical, and materials science. Previously, he has held research roles at Microsoft Research and contributed to AI-driven scientific discovery. His work focuses on scientific "superintelligence" platforms – integrating generative models, lab automation, and closed-loop experimentation to accelerate biomolecular and materials innovation.
- Margaux Pinney (mpinney@berkeley.edu, confirmed) is an Assistant Professor of Chemistry at UC Berkeley and formerly a Sandler Fellow at UCSF. Her lab develops high-throughput microfluidic and enzymology methods to measure catalytic constants and substrate affinities across thousands of protein variants, enabling large-scale maps of sequence–function relationships. By combining these rich biochemical datasets with machine learning, her work probes how proteins evolve new functions and accelerates enzyme design and function prediction.
- Quanquan Gu (qgu@cs.ucla.edu, confirmed) is an Associate Professor of Computer Science at UCLA and a Research Scientist at ByteDance Seed, focusing on bridging fundamental ML with scientific discovery. His domain expertise spans nonconvex optimization, deep generative models, reinforcement learning, and large language models, and he has more recently begun applying these methods within biological and chemical domains.

5.2 Invited Panelists

- Frances Arnold (frances@cheme.caltech.edu, tentative) is the Linus Pauling Professor of Chemical Engineering, Bioengineering, and Biochemistry and the Director of the Donna and Benjamin M. Rosen Bioengineering Center at Caltech. Renowned for pioneering directed evolution, she has been recognized by numerous awards, including the 2018 Nobel Prize in Chemistry. She has co-founded companies such as Gevo and Provivi and served on the boards for companies including Alphabet, Illumina, and Generate Biomedicines. Since January 2021, she has been an external co-chair of President Joe Biden's Council of Advisors on Science and Technology (PCAST).
- Barbara Cheifet (barbara.cheifet@us.nature.com, confirmed) is the Chief Editor of *Nature Biotechnology*. She holds a Ph.D. from Yale University, and spent 7 years at Genome Biology, including 4 years as Chief Editor, before joining *Nature Biotechnology* at the beginning of 2022 and becoming Chief Editor at the end of that year. *Nature Biotechnology*, with a 5-year Journal Impact Factor of 56.9, publishes new concepts in technology/methodology related to biological, biomedical, agricultural and environmental sciences as well as publishing commentary on the societal aspects of biotechnology research.
- Frank Noé (frank.noe@fu-berlin.de, confirmed) is a Professor at Freie Universität Berlin and a Microsoft Partner Research Manager in Microsoft Research (MSR) AI4Science. Frank has co-pioneered the Markov state modeling (MSM) approach for describing the long-time dynamics of proteins and other macromolecules, has developed several deep learning systems for molecular simulation, such as the Boltzmann Generator, and is an advocate of open research and software for the benefit of society.
- Sam Rodriques (sam.rodriques@futurehouse.org, confirmed) is Co-founder and CEO of FutureHouse, where he leads efforts to build AI agents that automate scientific discovery in biology. With a background in physics and bioengineering, Sam invented technologies spanning spatial/temporal transcriptomics, nanofabrication, and brain mapping before transitioning to AI-driven biology. As leader of the Applied Biotechnology Lab at the Francis Crick Institute, he pioneered novel methods for linking molecular and spatial data modalities. At FutureHouse, his team is developing systems like PaperQA2, Aviary, and Bench-Crow agents that read, reason, and propose hypotheses from biological literature and data, aligning with GEM's vision of tightly coupling generative ML and wet-lab experimental cycles.

6 Organizers and Biographies

Website page and email address in colored hyperlink. In bold, we have highlighted previous organizing or related experience.

- Chenghao Liu (chenghao.liu@mail.mcgill.ca) is a postdoctoral fellow at California Institute of Technology and at FutureHouse. He is advised by Frances Arnold. He was a co-founder of Dreamfold, a protein design start-up. He is a chemist by training, and his research is now focused on developing generative machine learning methods for enzyme discovery. He was a co-organizer of the CQMF chemistry conference, and an organizer of the GEM-2024 and GEM-2025 workshops.
- Jarrid Rector-Brooks (jarrid.rector-brooks@mila.quebec) is a PhD candidate at the Université de Montréal and Mila - Québec AI Institute advised by Yoshua Bengio. He is currently a visiting researcher at California Institute of Technology supervised by Frances Arnold. He was a co-founder of Dreamfold, a protein design start-up. His research aims to develop improved generative models specifically for the design of therapeutics with an eye towards high-throughput adaptive experimental design. He was an organizer of the GEM-2024 and GEM-2025 workshops.
- Soojung Yang (soojungy@mit.edu) is a PhD student at the Massachusetts Institute of Technology (MIT), advised by Rafael Gómez-Bombarelli. Her research focuses on modeling protein dynamics by integrating molecular simulations, protein foundation models, and experimental measurements. Her collaborations with Microsoft Research involve the development of BioEmu, a foundation model for equilibrium sampling of proteins. She was an organizer of the GEM-2024 and GEM-2025 workshops.
- Sidney Lisanza (lisanzas@gene.com) is a machine learning scientist at Prescient Design. He develops ML tools to expedite the protein design process both at the lead discovery stage and the subsequent optimization of candidates. He enjoys time with friends/family, being outside, listening to music, and preferably doing all simultaneously. He was an organizer of the GEM-2024 and GEM-2025 workshops.
- Jacob Gershon (jgershon@uw.edu) is a graduate student at the University of Washington within the Institute for Protein Design with David Baker. Jacob is working on developing deep generative models for de novo enzyme design, hoping to someday use these tools to design new materials that support a sustainable future. He was an organizer of the GEM-2025 workshop.
- Lauren Hong (lhong10@seas.upenn.edu) is a PhD student in Bioengineering at the University of Pennsylvania, advised by Pranam Chatterjee. As an experimentalist in a hybrid lab, Lauren focuses on leveraging generative language model-derived peptide binders to post-translationally manipulate proteins for broad-scale therapeutic applications. She has co-organized the Nvidia Duke AI Day as well as the Quantitative Biodesign Seminar at Duke. She was an organizer of the GEM-2025 workshop.
- Pranam Chatterjee (pranam@seas.upenn.edu) is an Assistant Professor of Bioengineering and Computer and Information Science at University of Pennsylvania. Research in his Programmable Biology Group exists at the interface of computational design and experimental engineering, specifically developing novel generative generative algorithms for biologics design. He completed his SB, SM, and PhD from MIT and is the founder of three startups, Gameto, Inc. and UbiquiTx, Inc., and AtomBioworks, Inc., that leverage AI to design the next generation of fertility solutions, cancer therapeutics, and RNA medicines, respectively. He has co-organized numerous AI workshops, including the Nvidia Duke AI Day, ML4LMS at ICML 2024, and SimBioChem at EurIPS 2025. He was an organizer of the GEM-2024 and GEM-2025 workshops.

- Yoshua Bengio (yoshua.bengio@mila.quebec) is a Full Professor in the Department of Computer Science and Operations Research at Université de Montréal, as well as the Founder and Scientific Director of Mila and the Scientific Director of IVADO. Considered one of the world's leaders in artificial intelligence and deep learning, he is the recipient of the 2018 A.M. Turing Award. He is a Fellow of both the Royal Society of London and Canada, an Officer of the Order of Canada, and a Canada CIFAR AI Chair. He was an organizer of the GEM-2024 and GEM-2025 workshops, as well as a founder of ICLR.

7 Diversity, Equity, and Inclusion

We are dedicated to the cause of diversity, equity, and inclusion (DEI) in our proposed workshop. Our efforts span a wide spectrum of academic disciplines, cultural backgrounds, personal experiences, and identities, ensuring an inclusive environment for all. We are committed to creating a space where each individual feels valued and welcomed, regardless of ethnicity, gender, sexual orientation, affiliations, nationality, seniority, abilities, socioeconomic status, religion, backgrounds, experiences, viewpoints, perspectives, and beyond.

Our organizing committee members come from **7 institutions**[2] with expertise in more than **6 academic disciplines**[3]. Our organizers identify with more than **9 cultural backgrounds**[4] with **2 female organizers**[5] and include first-generation students[6].

We have **2 professors**, **1 postdoc**, **1 industry scientist**, and **4 PhD students** organizers who work on a wide range of problems related to biomolecular design. Our student organizers bring prior experience from organizing conferences or events in various organizations (see Section 6). We provide many industry viewpoints through working or interning at (bio)tech companies such as Genentech and Microsoft Research on scientific applications with ML. Several organizers are founders or scientific advisors to biotech start-ups and established pharmaceutical companies. The experience level of our student organizers span from 3rd to 5th year PhD students while our professors span from assistant to full professors.

We have thoughtfully invited speakers and panelists with DEI in mind. Our panelists and speakers span three continents: Asia, Europe, North America, and South America. We selected speakers and panelists who come from different backgrounds (e.g. computer science, experimental biology, chemical engineering, computational chemistry) and whose topics are orthogonal.

In our commitment to fostering DEI, we are also dedicated to increasing global participation in our workshop. Our sponsors will provide travel grants for students from underrepresented communities. To facilitate international travel to the extent within our control, we will collaborate with the ICLR main organizers and provide assistance in obtaining invitation letters for those who may require them for visa and travel-related purposes. Our aim is to ensure that individuals from across the globe have the opportunity to join us, share their insights, and contribute to our collective learning and growth.

Finally, our workshop will promote a venue for safe, non-judgmental, and respectful discourse. We will set-up anonymous communication channels (phone, email, and form) to report any misbehaviour or DEI related issues.

8 Previous Related ICLR Workshops

An undoubtedly impactful application is combining ML with scientific applications. We list the most related workshops at previous ICLR conferences in order of relevance (first is most relevant).

- Learning Meaningful Representations of Life (LMRL). ICLR 2025. This workshop focused on advancing representation learning for biological sequences and structures, with an emphasis on embedding and latent space methods. Our GEM workshop differs in two key ways: first, we emphasize *generative* models for de novo biomolecular design rather than purely representation learning; and second, we integrate these generative advances with *experimental validation*. GEM is designed to bring experimentalists directly into the ML conversation, something that has been underrepresented in prior ICLR workshops.
- AI for Nucleic Acids (AI4NA). ICLR 2025. AI4NA concentrated on DNA and RNA modeling, focusing on generative and discriminative tasks restricted to nucleic acids. Our GEM workshop is broader in scope: we cover proteins, nucleic acids, and other biomolecules. More importantly, we explicitly emphasize the integration of generative models with *real-world experimental workflows*, ensuring that methods are validated and actionable. Thus, GEM complements AI4NA by extending beyond nucleic acids and by fostering collaborations between computational and experimental researchers.

  Machine Learning for Drug Discovery Workshop. ICLR 2022-2023. This workshop focuses on optimizing and discovering therapeutic candidates. Our workshop focuses on general purpose design of biomolecules – proteins, RNA/DNA – and more so on the generative perspective. The broader scope allows us to explore diverse applications; for example, enzyme engineering can significantly contribute to plastic degradation and gene editing with CRIPSR. Both workshops emphasize bringing ML closer to real-world evaluation and using ML as a core part of biological experimentation. An distinctive feature of our workshop is our aim to encourage the involvement of experimentalists, who have been underrepresented in prior ML conferences and workshops, despite their crucial role in generative biology. To achieve this goal, our workshop offers dedicated tracks and collaboration with a high impact biology journal, and our speakers and panelists come from a wide variety of backgrounds, ranging from experimentation to computational modeling.

[2] Mila, MIT, Caltech, University of Washington, University of Pennsylvania, Prescient Design, and Université de Montréal.

[3] Including computer science, computational biology, biochemistry, physical chemistry, chemical engineering, and bioengineering

[4] Including Canada, China, Finland, France, India, Israel, Kenya, and South Korea

[5] Soojung, Lauren

[6] Jarrid, Soojung

- Deep Generative Models for Highly Structured Data. ICLR 2019 & 2022. This workshop focuses on incorporating structure into generative models for downstream use in real-world data modalities. While proteins and biological data is one application, it is not the focus whereas it is in our workshop. Furthermore, we focus on how to integrate generative models into real-world biological experiments.
- Machine Learning for Materials. ICLR 2023. The goals of this workshop are similar to ours but focused on materials. Here they emphasize identifying the unique challenges with designing useful materials and uncovering the meaningful tasks for novel ML techniques to be developed. We share the common aim of emphasizing the complexity of working with biomolecules and the need for directing ML research towards the most pressing problems in biomolecular design.
- Physics for Machine Learning. ICLR 2023. This workshop focuses on developing ML applications for physics. They also emphasize bringing ML closer to real-world applications in the physical sciences.

Why this workshop (and why now). It is an exciting time to work at the intersection of biology and ML. AlphaFold [1] undoubtedly shifted the direction of many ML researchers to work on structural biology and bioengineering applications. Experimental biology is reaching a inflection point of data generation becoming cheaper and faster than ever. Works such as RFdiffusion [4] published in *Nature* is a example of novel ML research combined with experimental validation to achieve unprecedented biomolecular design success [4]. The advent of ChatGPT and foundation models have also provided a possible road map to developing interactive systems between scientists and AI for scientific discovery [3]. However, the ChatGPT paradigm cannot be transferred to biology where data cannot be readily annotated or generated at scale. There will need to be important breakthroughs in generative AI tailored to biology and adaptive experimental design to reduce cost and time. We believe the time is ripe to bring together the different disciplines to chart a path towards rapid scientific discovery in the age of AI.

9 Sponsors

Nvidia, Mark III, Braidwell, and FutureHouse have confirmed their sponsorship of the workshop, with additional commitments under discussion from several other industrial research labs. While final contribution amounts are pending, we anticipate raising at least $50,000 in sponsorship, consistent with last year's support. These funds will be allocated with the following priorities:

1. Registration and travel grants to support students and participants from underrepresented groups.
2. Lunch catering to encourage networking and collaboration during the workshop.
3. Best paper awards to recognize outstanding contributions.
4. Seed funding for collaborative proposals initiated at the workshop.
5. A post-conference gathering to foster informal discussions and community building.

10 Program Committee

The role of program committee members will be to help review the workshop submissions. Each organizer will recruit enough reviewers to provide 3 reviews for each submission. Since each organizer comes from a different lab with different personal connections, we expect to have a rich pool or reviewers. We have prepared a list of reviewers who have tentatively confirmed to be reviewers. We have additional ways throughout our diverse connections to seek more reviewers if needed.

Patrick J. Almhjell, Ph.D., Simon Axelrod, Ph.D., Ava P. Amini, Ph.D., Emmanuel Bengio, Ph.D., Joey Bose, Ph.D., Tim-Henrik Buelles, Ph.D., Bianca Dumitrascu, Ph.D., Michael Galkin, Ph.D., Alex Hernandez-Garcia, Ph.D., Riashat Islam, Ph.D., Kadina Johnston, Ph.D., Michal Koziarski, Ph.D., Alex X. Lu, Ph.D., Pablos Lemos, Ph.D., Ge Liu, Ph.D., Sulin Liu, Ph.D., Santiago Miret, Ph.D., Ariane Mora, Ph.D., Nathan Frey, Ph.D., Andrei Nica, Ph.D., Ladislav Rampasek, Ph.D., Seongok Ryu, Ph.D., Lena Simine, Ph.D., Almer van der Sloot, Ph.D., Alexander Tong, Ph.D., Kevin K. Yang, Ph.D., Zichao Yan, Ph.D., Bruce Wittmann, Ph.D., Zach Wu, Ph.D., Tara Akhound-Sadegh, Lucas Arnoldt, Ron Boger, Shahar Bracha, Paul Bertin, Maria Carreira, Tianlai Chen, Itamar Chinn, MinGyu Choi, Felix Faltings, Jacob Gershon, Prashant Govindrajan, Sarah Gurev, Guillaume Huguet, Ian Humphreys, Moksh Jain, Andrew Kirjner, Maksym Korablyov, Daniel Levy, Seokhyun Moon, Sean Murphy, Peter Mikhael, Juno Nam, Lena Nehale Ezzine, Umesh Padia, Andrei Rekesh, Raman Samusevich, Wenxian Shi, Luca Thiede, Brian Trippe, Veronica Tarka, Tony Tu, Pascal Sturmfels, Akshay Subramanian, Allen Tao, Hannes Stark, Sophia Vincoff, Sasha Volokhova, Rachel Wu, Jason Yang, Dinghuai Zhang, Yinuo Zhang, Wonho Zhung.

References

[1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873): 583–589, 2021.

[2] Sawan Patel, Sophia Tang, Yinuo Zhang, Pranam Chatterjee, and Sherwood Yao. Multi-objective-guided generative design of mRNA with therapeutic properties. In *ICML 2025 Workshop on Scaling Up Intervention Models*, 2025. URL https://openreview.net/forum?id=0kg9KhUYSK.

[3] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[4] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.