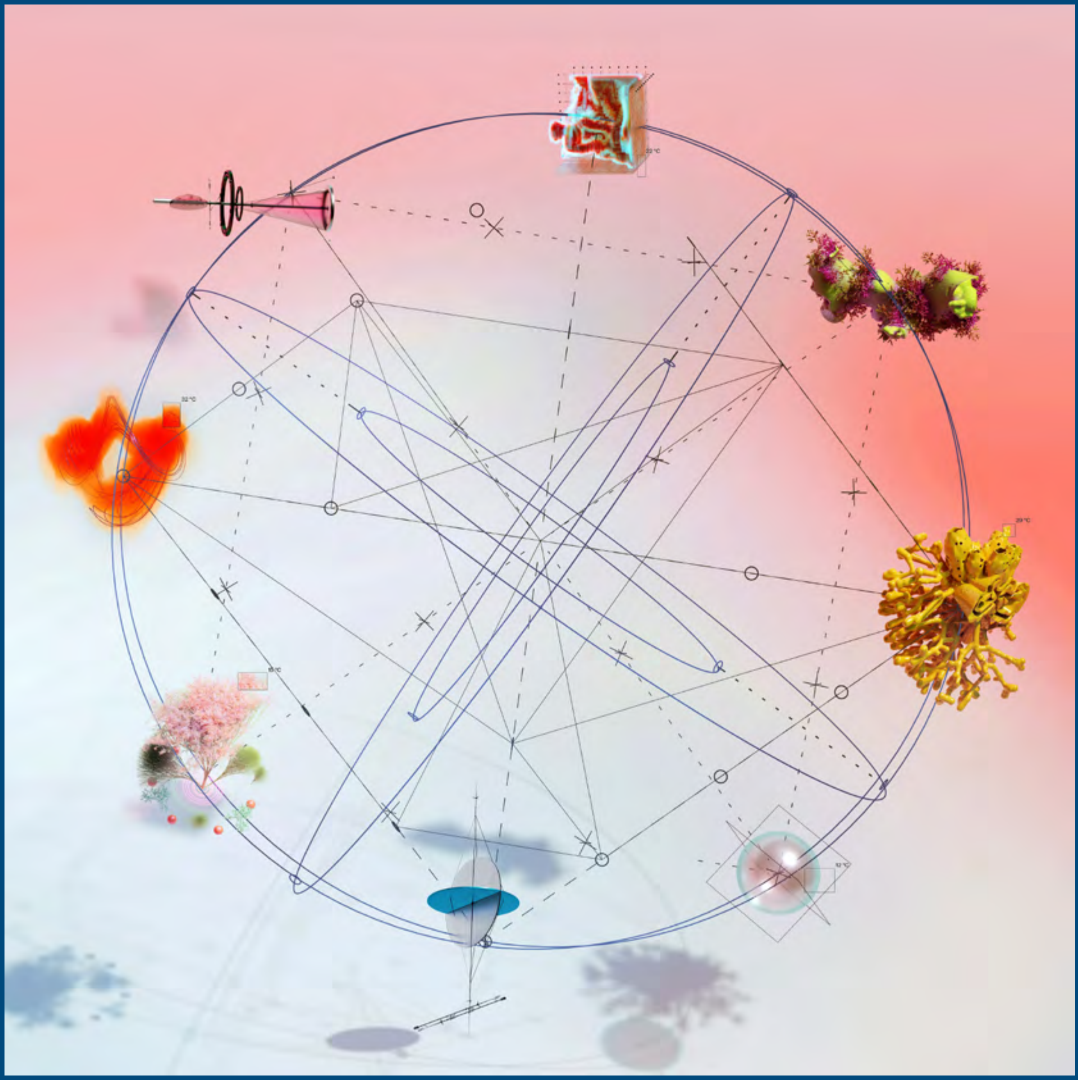
AI for Discovery

New Frontiers, Challenges, and Opportunities for Scientific Research



A report by

SCIENCE & SOCIETY
aspen institute

With gratitude for support from

Google.org

The Aspen Institute is a global nonprofit organization igniting human potential to build understanding and create new possibilities for a better world. Founded in 1949, the Institute drives change through dialogue, leadership, and action to help solve the most important challenges facing the United States and the world. Headquartered in Washington, DC, the Institute has a campus in Aspen, Colorado, and an international network of partners. In 2019, the Aspen Institute launched the Science & Society Program with seed support from the Alfred P. Sloan Foundation and Johnson & Johnson.

The Aspen Institute Science & Society Program strengthens the relationship between scientific progress and public life. Drawing on the Aspen Institute's convening power, cross-sector networks, and public platforms, the program helps people understand science, engage with and shape how science is used, and make decisions grounded in evidence. Through work spanning public understanding, civic engagement, and decision systems, Science & Society connects narrative, institutions, communities, and leaders across sectors to improve how scientific knowledge informs collective choices.

THE SCIENCE & SOCIETY PROGRAM WOULD LIKE TO THANK THE FOLLOWING ENTITIES FOR THEIR SUPPORT:

PROJECT SPONSOR

Google.org

PROGRAM SUPPORT

mbcn Metastatic Breast Cancer Network — Educating • Empowering • Advocating

Laura & Gary Lauder Family Venture Philanthropy Fund — Cultivating Leaders • Launching Innovation

Daniel H. Weiss
Rick Stamberger
Nick DeWolf Foundation
Paul Beirne Foundation
Marc E. Elovitz

Photo credits: Cover: Google DeepMind / Unsplash; Section headings: Anastasiia Ornarin / Unsplash; Page 2 and back cover: A.C. / Unsplash

CONTACT: Please keep in touch with us at aspeninstitute.org/science, and for questions and comments, please write to Aaron F. Mertz, Aspen Institute Science & Society Executive Director, at aaron.mertz@aspeninstitute.org

Table of Contents

Editors' Note

Artificial intelligence is reshaping the practice of scientific discovery.

The development of AlphaFold—an artificial intelligence system capable of predicting the three-dimensional structures of proteins with unprecedented accuracy—marked a watershed moment in this trajectory. Beyond recognizing the tool's contributions to the understanding of life, the awarding of the 2024 Nobel Prize in Chemistry to AlphaFold's creators signaled the arrival of AI as a powerful engine of scientific discovery.

Building on this success, society must now determine: *(1) which scientific challenges now stand to benefit the most from AI and (2) how AI itself will evolve using breakthroughs across scientific domains?*

Researchers understand these questions better than anyone and deserve to actively inform the technologies, research infrastructure, and investments that support their discoveries. With this purpose in mind, we at the Aspen Institute Science & Society Program hosted ten closed-door roundtables with researchers working across biodiversity, climate and Earth science, biological inference, genomics and human variation, and neuroscience and human behavior. Our program has a proud history of convening leaders across disciplines and sectors to address questions that no single community can answer alone. These discussions were no exception; participants represented universities, research institutes, industry, philanthropy, and nonprofit organizations across North America, Europe, Africa, Asia, and Australia.

Comparing perspectives across these five scientific fields revealed shared themes and opportunities, as well as common obstacles.

Each discussion kept scientific inquiry at the forefront. Participants described major unanswered questions, barriers to progress, and opportunities where artificial intelligence could accelerate discovery. Conversation naturally shifted from AI capabilities to the practical work of science: scoping the literature, testing ideas in the lab, and interpreting evidence.

Researchers repeatedly described AI as accelerating the process of discovery, often emphasizing contributions to model development, hypothesis generation, and experimental design. Participants also described exciting opportunities to synthesize diverse evidence streams and to reveal patterns that would otherwise remain hidden to human researchers alone.

The conversations also reinforced enduring scientific standards such as experimental validation and reproducibility. Moving from prediction to mechanistic understanding remained the goal across every discipline. Participants repeatedly returned to the foundations that make AI-enabled science possible, including high-quality data, shared infrastructure, and scientific workforce development.

The conversations presented in this report identify priorities that can help guide future research, investment, and collaboration. Scientists developing AI methods, organizations supporting scientific research, and institutions building shared infrastructure all face choices that will shape future discovery (and perhaps Nobel Prizes to come!). We hope this report provides a roadmap for those decisions—because better decisions begin with deeper understanding offered by those at the center of scientific research.

Sejal Goud – Program Associate, *Aspen Institute Science & Society Program*

Jylana L. Sheats, Ph.D., M.P.H. – Associate Director, *Aspen Institute Science & Society Program*; Clinical Associate Professor, Social, Behavioral, and Population Sciences Department, *Celia Scott Weatherhead School of Public Health & Tropical Medicine*

Amelia Hanbury – Program Assistant, *Aspen Institute Science & Society Program*

Aaron F. Mertz, Ph.D. – Founder and Executive Director, *Aspen Institute Science & Society Program*

The editors used Google Gemini as a drafting tool to assist in converting detailed outlines, manually prepared by the editors from the roundtable transcripts, into prose. All analysis, editorial judgment, and final wording were reviewed and substantially revised by the editors. At the editors' request, subject-matter experts at Google reviewed the report for technical accuracy. The Aspen Institute Science & Society Program retained full editorial independence and control over the final report.

1. Ecosystem & Biodiversity Science

A. Participants

We aim to synthesize and share perspectives from the discussions as a whole rather than to attribute any quotations or viewpoints to specific individuals. Participants with whom we engaged in roundtable discussions are listed below (alphabetically by last name):

- **Tanya Birch** – Geo Sustainability Nature Lead, *Google*
- **Jarrett Byrnes, Ph.D.** – Associate Professor, *University of Massachusetts Boston*
- **Tom Crowther, Ph.D.** – Professor in the Department of Environmental Systems Science, *ETH Zurich*; Chair of the Advisory Council, *United Nations Decade on Ecosystem Restoration*
- **Muki Haklay, Ph.D.** – Professor of Geographical Information Science, *University College London*
- **Avery Hill, Ph.D.** – Systems Ecologist, *California Academy of Sciences*
- **Kate Jones, Ph.D.** – Director, *People & Nature Lab*; Professor of Ecology and Biodiversity, *University College London*
- **Jenna Kline** – Co-Lead, *WILDLABS Edge Computing Group*; Doctoral Candidate, *The Ohio State University*
- **Jingjing Liang, Ph.D.** – Associate Professor of Quantitative Forest Ecology, *Purdue University*
- **Karen Lips, Ph.D.** – Deputy Director General, *International Institute for Applied Systems Analysis*
- **Scott Loarie, Ph.D.** – Executive Director, *iNaturalist*
- **Dan Morris, Ph.D.** – Senior Staff Research Scientist in AI for Nature, *Google*
- **Nicholas Murray, Ph.D.** – Associate Professor in Global Ecology and Conservation, *James Cook University*
- **Harini Nagendra, Ph.D.** – Director of School of Climate Change and Sustainability, *Azim Premji University*
- **Joseph Nguthiru** – Founder, *HyaPak*; Young Champion of the Earth, *United Nations*
- **Michael Simeone, Ph.D.** – Associate Research Professor, *Arizona State University School for Complex Adaptive Systems*
- **Charlotte Stanton, Ph.D.** – Senior Program Manager, *Google Research*
- **Ben Strong, Ph.D.** – Science and Machine Learning Lead, *Earth Genome*
- **Jens-Christian Svenning, Ph.D.** – Director, *Aarhus University Center for Ecological Dynamics in a Novel Biosphere*
- **Dave Thau, Ph.D.** – Data and Technology Global Lead Scientist, *World Wildlife Fund*

B. Introduction

At a time when habitat degradation and species decline often outpace human capabilities for documentation, monitoring, and restoration, the conservation community faces a fundamental dilemma: it is difficult to protect what humanity cannot see, measure, or understand. Speaking to this challenge, this chapter explores how artificial intelligence can be mobilized—in the words of one roundtable participant—to "supercharge our description of life on this planet, so we don't lose it before we even know that it's there."

Specifically, the insights that follow address how AI may help illuminate ecological processes that remain beyond human observation; how machine learning can unlock new scientific questions and integrate multiple heterogeneous data streams in meaningful ways; how AI-driven tools can be applied to scenario planning, policy decision-making, and global ecosystem mapping; and what frameworks are needed to ethically support community knowledge and Indigenous data sovereignty.

Underlying this discussion is the recognition that spatiotemporal data has historically developed in isolated silos. Recent breakthroughs in generalizable AI models, bioacoustics, satellite remote sensing, and edge computing offer unprecedented capabilities to bridge these divisions. However, turning computational potential into lasting conservation outcomes requires overcoming persistent data coverage disparities, liberating un-digitized physical archives, moving from correlational to causal modeling, and centering community leadership to ensure non-extractive, sovereign engagement.

<u>Core Takeaways</u>

- **Multi-modal integration unlocks new ecological horizons:** Multi-modal approaches combining remote sensing, bioacoustics, camera traps, and environmental DNA (eDNA) offer unprecedented views of species occupancy and habitat dynamics.
- **Data accessibility remains the primary operational bottleneck:** Unlocking paper archives, standardizing Essential Biodiversity Variables (EBVs), and enabling natural-language retrieval are urgently required to make existing data machine-readable and legible.
- **Process-based causality is essential for decision-making:** Purely predictive AI risks failing under novel climate conditions. Causal and counterfactual architectures must be integrated into ecological models to evaluate interventions with confidence.
- **Ethical conservation demands community sovereignty and federated AI:** Local and Indigenous communities must retain data ownership through local-first architectures, avoiding extractive practices and pro-forma involvement.
- **Philanthropic capital should prioritize data acquisition over model building:** Funders achieve the highest return on investment by supporting sensor deployment, open-access datasets, and ground-truth validation rather than continuing to develop AI algorithms.

C. Areas of Excitement and Progress

Advances in Generalizable Models and Multi-Modal Data Integration

Ecosystem and biodiversity research has historically lacked a unified computational architecture. Unlike climate science—which has benefitted from standardized open-source modeling frameworks like the Coupled Model Intercomparison Project (CMIP)—conservation researchers have tended to use standalone systems, leading to differences across institutional and geographic settings. Recently, however, the emergence of large, generalizable AI models has begun to reshape the field. Roundtable participants specifically noted the utility of computer vision architectures like SpeciesNet and bio-acoustic models such as BirdNet and Perch, which have dramatically scaled the automated processing of visual and audio feeds from field sensors. Simultaneously, spatial models trained on satellite imagery and remote sensing data are revealing macro-level natural patterns, illustrating ecosystem distributions and shifting boundaries across temporal scales.

One prominent advancement highlighted by roundtable participants is Google DeepMind's AlphaEarth Foundations. By integrating diverse Earth-observation data into a unified representation, AlphaEarth provides a general-purpose geospatial layer that can be combined with *in situ* observations for downstream ecological applications. Emerging open models like Tessera and Clay similarly exemplify this trend. Initiatives such as Global Nature Watch, supported by Google.org, demonstrate how increasingly diverse environmental datasets can be brought into accessible digital platforms for non-technical users.

One contributor highlighted how these new modes of analysis reflect an emerging paradigm: "In ecology, we advance our understanding of different systems, primarily ecosystem by ecosystem, and I feel like we're on the verge of now being able to connect that knowledge across many systems together, and an example of that might be global monitoring systems for ecosystems, which typically focus on one or two mangroves or forests, will increasingly start to account for many different systems, together and simultaneously, and we'll start to understand much more the connected nature of the ecological systems on Earth through these types of advances."

Andy Wang / Unsplash

This shift in multi-modal biodiversity analysis represents a fundamental movement away from the historical model of isolated, siloed data streams. While remote sensing satellites, *in situ* camera traps, bioacoustic recorders, and eDNA genomic samples previously yielded fragmented maps, local spreadsheets, and standalone audio archives, next-generation AI models can unify these inputs into integrated representations. These architectures allow general models to be fine-tuned for specific localized microhabitats and enable triggered actions, such as an acoustic sensor dispatching an aerial drone upon detecting an anomalous signal. These activities then feed directly into natural language dashboards, high-resolution species distribution maps, and habitat suitability charts—making integrated data accessible and queryable.

These architectures possess a transformative dual capability: unifying heterogeneous data modalities into a single platform and enabling generalist models to be fine-tuned for specific localized microhabitats. This dynamic accelerates the generation of high-resolution species distribution maps and habitat suitability estimates. Considered a "holy grail" for field ecologists, these unified layers allow researchers to formulate novel questions regarding how species and ecological communities dynamically respond to perturbations over time. On the hardware front, the deployment of Edge AI within smart multi-modal sensor networks—such as camera traps that autonomously trigger aerial drones upon detecting anomalous signals—is revolutionizing real-time field observation.

Data Standardization and Cost Reductions

Parallel to algorithmic breakthroughs, structural progress in data governance and hardware accessibility is resolving long-standing logistical bottlenecks. The conceptual maturity and gradual implementation of the Essential Ocean Variables (EOVs) and the Essential Biodiversity Variables (EBVs) concepts provide long-overdue standardization across messy, heterogeneous ecological data types. By establishing agreed-upon parameters for observation through rigorous discussions among scientists, EBVs and EOVs are poised to catalyze systematic data collection, helping close persistent gaps in fundamental ecological knowledge.

Concurrently, the manufacturing costs of open-source field-deployable computing hardware—including Raspberry Pis, NVIDIA Jetsons, and Nano boards—have fallen dramatically. This cost reduction democratizes localized data management and edge processing directly at the point of collection, mitigating the bandwidth and cloud storage constraints that previously stymied remote field research.

New Analyses and Community Applications

The intersection of civic science communities with new AI workflows has yielded research-quality datasets and real-world conservation applications. Citizen science platforms like iNaturalist now directly feed observations into research-quality databases like the Global Biodiversity Information Facility (GBIF), even leading to the discovery of new species. Researchers are also leveraging models like Gemini for novel biodiversity applications, such as prototyping a framework that provides insights into how global supply chains drive environmental degradation.

Over the last five to ten years, integrated AI systems have radically transformed forest resource management across four core domains:

- **Forest Change Detection:** Tracking micro-disturbances, deforestation, forestation, and broader forested disturbances in near real-time.
- **Taxonomic Enumeration:** Estimating the total number of tree species present and distributed across global forest biomes.
- **Biodiversity Mapping:** Quantifying and modeling the distribution of biodiversity in forested areas.
- **Carbon Accounting:** Estimating and quantifying the distribution of biomass and carbon in forested areas around the world.

Applying remote sensing to biodiversity assessment and ecosystem process modeling has traditionally been technically challenging. AI applications are making these forms of analysis more accessible. The emergence of intuitive, AI-assisted tools allows activists, local communities, and non-technical land managers to teach themselves how to apply these insights to their daily work. As one roundtable participant summarized, there are major opportunities "for large language models to grease the wheels and data information flows between community scientists, decisionmakers, and analytical scientists." Another expert noted the "huge open space in terms of how we collect as much as we possibly can, as quickly as we can. Behind it all, people can only do so much. With the growth and development of AI-based tools, I'm seeing member after member of our [roundtable] starting to find ways to leverage that to cover some of the enormous data gaps that we have out there."

D. Bottlenecks, Blindspots, and Opportunities in Ecosystem Science

Observing Remote and Under-Studied Locations

Despite rapid technological advances, notable spatial and ecological blind spots persist. While certain conservation hotspots have been well-documented, vast ecological frontiers remain understudied due to physical logistical constraints and underfunding. Deep marine zones, mesophotic reefs,

Getty Images / Unsplash

tropical cloud forests obscured by dense canopy cover, subterranean networks (such as fungal biodiversity, soil microdata, and internal tree dynamics), and frontiers where ecological processes meet terrestrial systems (such as coastal environments) are difficult to monitor. Furthermore, a stark geographic coverage disparity persists when comparing data from the Global North and the Global South; biodiversity records across Africa, the tropics of Southeast Asia, and Oceania lag significantly behind.

Regarding data characteristics, platforms like iNaturalist have produced species range and distribution maps that outperform older expert-led taxonomy methods, but there remains a massive gap where humans and ecosystems do not frequently interact. Consequently, the field possesses rich presence-based data, but abundance-based data remain lacking.

To bridge these gaps, AI tools that pair remote sensing with *in situ* data can help identify where undescribed species likely exist, detect small organisms invisible to satellite imagery, and generate potential species lists for ground-truth validation. Similar analytical challenges affect complex coastal frontiers. For instance, EarthGenome has begun working on the seagrass mapping problem to address a lack of understanding regarding the plant's dynamics over time, carbon sequestration capabilities, and underlying ecosystem processes. Bringing together information on shifting tides alongside biotic and abiotic data via AI provides a structured entry point for research questions. Furthermore, untapped opportunities exist in decoding non-human communication—expanding bioacoustic models beyond listening to bird calls to interpret whale vocalizations, plant physiological reactions, and other species interactions.

Geographic, Trophic, and Temporal Scale

The nested spatial, temporal, and trophic scales of ecological phenomena pose significant modeling challenges. Because climate change impacts manifest primarily at the localized population level rather than the species level, monitoring systems must observe changes across multiple sites and populations concurrently. As one roundtable participant noted, "You would actually be able to see how climate change [is] affecting this species, across very different landscapes or ecosystems or habitats, different times, different seasons? I think it helps in questions regarding migration, dispersal, spread of invasive species, or diseases."

Compounding these scale challenges is the structure of academic research—in which researchers often study a single species within the temporal and geographic limits of a doctoral project—and limited data across trophic levels. More broadly, the field currently suffers from a lack of quality time-series data that is "contextually sensible, locally relevant, and globally accessible."

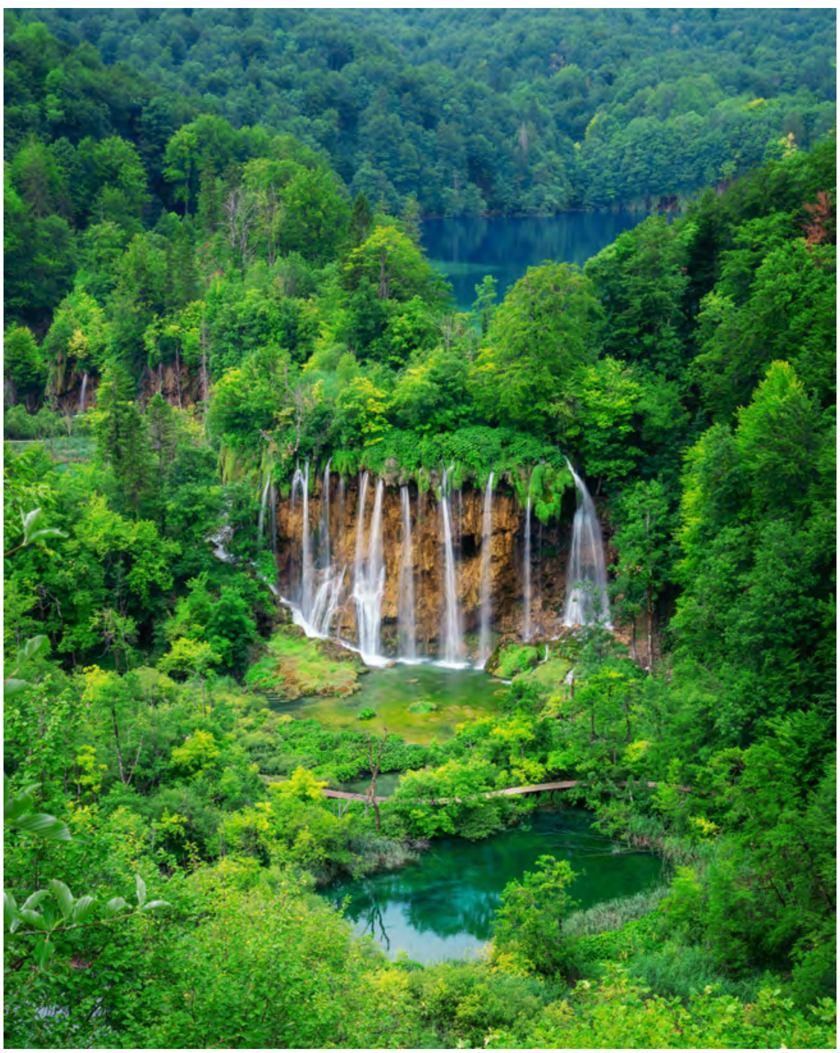
Management and Accessibility of Existing Data

Data gaps arise from disparities in collection and are exacerbated by the inaccessibility of many existing datasets. Decades of valuable ecological records remain locked up because many natural history museum collections and herbarium specimen labels have never been digitized.

One participant highlighted the ability of AI to extract information from unannotated paper and visual archives, describing a collaboration with a student analyzing a quarterly sample of marine rock walls dating back to the 1970s. A large portion of the dataset had remained unannotated in file drawers. Using computer vision, they extracted and segmented decades of dormant biological data. The resulting long-term time series completely overturned their core ecological hypotheses. Through this work, long-standing assumptions regarding system stability were proven inaccurate—a discovery made possible solely by converting legacy physical images into machine-readable datasets.

Retrieval-Augmented Generation (RAG) architectures offer a promising approach to making existing data easier to search and use. By placing natural-language querying capabilities on top of existing databases, data stop being "locked up" and users receive responses based on real observations. Similarly, keeping up with continuous streams of sensor data—particularly visual feeds—is extremely taxing for researchers. To manage this task, organizations like the World Wildlife Fund (WWF) use Wildlife Insights to automatically identify camera trap imagery, significantly enhancing organizational productivity and allowing researchers to ask new questions from previously unprocessed data.

However, systemic barriers remain. Early-stage startups in the climate space face persistent difficulty obtaining satellite datasets from sources such as NASA, Google, and SpaceX. Furthermore, geospatial data remains largely inaccessible to large language models (LLMs). While technical progress continues, the field must still determine how non-experts can intuitively interact with geospatial data layers.



Getty Images / Unsplash

AI Literacy, Transparency, and Domain Expertise

Technical fluency gaps between software engineers and field ecologists present institutional friction. Although experts within ecology and biodiversity sciences are technically savvy in their respective fields, researchers often require additional support to become AI-proficient. Roundtable participants expressed a strong desire to understand how AI systems function on a granular level, articulated to their domain knowledge. Specifically, researchers called for transparency tools that illuminate input sequence evaluations, layer-by-layer processing, and the explicit relationship between training data and outputs.

As one roundtable participant emphasized, "Until [these literacy and transparency gaps are addressed], none of these researchers are really empowered to imagine new frontiers in the work that they do in the way I think that we would ideally like. Until then, they are only consumers of tools that other people design and give to them. And that is not the worst place to be with how powerful some of these capabilities are, and I'm certainly not bemoaning it, but I think if you're thinking about a grand ambition to kind of unlock people's imagination, to ask better questions for decades, then I think that is a fundamental gap that has to be addressed."

Providing this depth of training will also help domain researchers produce data in standardized, AI-ready formats, helping overcome inertia in traditional methodology. It is worth noting that remote-sensing researchers and ground-level field ecologists tend to ask fundamentally different questions on a day-to-day basis. Furthermore, researchers must be equipped to navigate and select from a variety of distinct AI model architectures.

Building Causality into Models

A central limitation in current ecological AI deployment is determining whether a given model is designed for statistical prediction or underlying causation. Moving into novel environmental regimes with a purely predictive model is dangerous; without process understanding baked into the architecture, researchers are left with statistical associations that cannot generate actionable intervention plans.

To understand underlying mechanisms and forecast accurately under novel climate conditions, AI systems will likely need to incorporate counterfactual thinking. As one participant stated, "Unless you have a causal or process-based model underlying the entire structure, you're not going to be able to efficiently or accurately, with confidence, forecast into a novel scenario. I'm really heartened to see things like causal variational autoencoders and other things coming about, where at their core, there is a process-based causal model."

For example, small-scale, highly rigorous experiments—such as long-term ecological research sites in the state of Minnesota—have helped scientists understand how biodiversity directly drives ecosystem services. AI could extend this work by conducting quasi-experiments, identifying treatment and control sites across global biomes to quantify the value of biodiversity for ecosystem services. The scientific community will also need improved methods to add robustness checks to model outputs. These safeguards are increasingly important as low-quality "slop" content enters online repositories such as Wikipedia and risks degrading the shared information resources on which scientific research depends.

E. High-Value Applications for Research and Decision-Making

Integration of Socioeconomic and Political Data

As one roundtable participant expressed, AI is not fundamentally superior to conventional human intelligence; rather, the real question is how to use the technology to accelerate and improve the accuracy of ongoing work. While AI cannot deploy physical sensors on its own, it excels at parsing unstructured human data in the public domain and connecting it with ecological data. There is an urgent need to understand socio-economic dynamics across scales, as connections between nature data and social metrics are important for making restoration work resilient, durable, and beneficial for communities.

Integrating ecological and human data streams involves fusing primary biological inputs with social and policy metrics. On the ecological side, systems ingest satellite imagery, bioacoustic recordings, camera trap video feeds, and eDNA profiles. On the human side, platforms process census data, historical newspapers, upcoming conservation legislation, social media sentiment, white papers, and narrative descriptions of weather changes—sources that traditionally would not be cited in an ecology paper but matter tremendously for understanding historical drivers of change and assessing future threats. Synthesizing these streams through causal AI platforms can better support resilient, community-aligned restoration strategies and policy decisions.

Because culture is dynamic and human beliefs represent a primary barrier to conservation, real-time monitoring of public attitudes toward specific nonhuman species enables conservationists to take targeted action in areas with high support or poor understanding. Momentum should also continue around integrating satellite imagery with community science databases like iNaturalist and non-traditional visual feeds like Google Street View. AI could also analyze the history of scientific literature surrounding specific ecosystems. Realizing this potential, however, requires the academic community to take the lead, as tech companies face legal restrictions against scraping proprietary academic literature.

James Lee / Unsplash

Scenario Planning and Evidence-Based Policymaking

AI-enabled ecosystem models provide vital infrastructure for scenario planning, supporting decisions faced by communities, conservation organizations, governments, and land stewards. However, making the most of these capabilities requires strengthening lines of communication to ascertain the precise informational needs of decision-makers. Specifically, AI empowers leaders by quantifying trade-offs—such as calculating where the highest return on investment for reforesting exists and designing mechanisms like carbon pricing by pairing sequestration processes with economic data.

Predicting environmental outcomes like habitat degradation allows organizations to equip stakeholders with tangible resources before damage occurs. Roundtable experts highlighted Global Fishing Watch—which advocates for Marine Protected Areas (MPAs)—as an excellent example of how improved information leads to better policy outcomes by showing how regional fishing yields can increase if natural nursery areas are protected. Another major effort is the Global Ecosystem Atlas, which maps global ecosystems in a globally consistent and locally relevant manner. This initiative feeds directly into "30x30" policymaking—the global goal for governments to designate 30% of Earth's land and ocean as protected areas by 2030—by providing a spatial typology consistent across all countries that allows local data to tune larger process models. Other real-world applications include helping smallholder farmers in rain-fed economies track and predict crop blights, or modeling landslide risks to inform municipal evacuation timing.

Uncertainty may arise when models trained in one country are deployed globally, or when baseline layers—such as road networks on Google Maps—are updated less frequently in under-resourced regions. Participants emphasized that for AI to be useful in policy, uncertainty should be quantified and communicated transparently. They argued that greater transparency regarding training data origins and inference processes is needed so that AI does not become an "easy scapegoat" for officials seeking to evade political accountability. Participants also stressed the importance of identifying precisely *where* spatial uncertainty exists, including down to the pixel level.

Creating an AI-Native Digital Twin of the Earth

Ecologists currently lack an equivalent to the World Climate Research Programme's CMIP, driving strong demand for a generalizable, open, AI-native "Digital Twin" of Earth's ecosystems. As one participant envisioned, a Digital Twin would "allow us to simulate at a real-time basis, all types of data, presence, and abundance-based data from different types of ecosystems, forests and marine ecosystems as well, so that we can empower people—not only scientists, but decision-makers, local communities, and researchers from the Global South—allowing them to experiment, play with science, play with their ideas, so that not only a handful of elite scientists, but everyone across the world can advance ecosystem sciences together."

For instance, researchers and policymakers would benefit from further knowledge about natural carbon sequestration dynamics, an invisible process in which it remains unclear how much sequestration stems from forest recruitment, tree growth, or tree mortality. AI offers the tools to bridge global reference maps with local reference data, allowing local stewards to fine-tune datasets that feed back into global aggregate models.

Inter- and Intra-Species Interactions

The overarching goal of conservation science is ending the global extinction crisis, which requires an empirical understanding of how species interact across and within trophic levels. Computer vision models can assist by identifying observable species interactions directly from image feeds. Another high-value frontier is analyzing collective animal behavior. While current behavioral recognition models classify macro-states like grazing or walking, researchers seek to understand how collective behavior emerges from individual micro-decisions. Achieving this understanding requires fine-grained kinematics—analyzing head posture, ear movement, and gaze direction. While the field currently lacks the expert-annotated datasets needed to train these models, building this infrastructure is achievable, following precedents set in human behavioral modeling.

Democratizing Conservation Practice

Conservation practice has traditionally been restricted to academic scientists, government agencies, and specialized non-profits, but modern tools are activating wider public participation. AI models can empower private landholders to increase local biodiversity by generating tailored land-management recommendations. For instance, researchers are exploring ways to directly integrate species observation and identification into consumer hardware like doorbell cameras. Initiatives associated with FathomNet and the Monterey Bay Aquarium Research Institute (MBARI), including FathomVerse, are gamifying data collection, inviting citizen scientists to annotate ocean imagery.

At its best, conservation technology has the potential to help bring communities and local ecosystems out of economic competition. As one participant summarized, "Restoration is livelihoods and nature thriving." Another expert noted: "Oftentimes what puts communities in competition with conservation and biodiversity is the need for the resources that are in those ecosystems. If AI is actually on tap, not necessarily to solve these ecology and biodiversity problems directly, but helps alleviate some of the pressure on these ecosystems, that could…model or assist economic strategies and planning."

Matthias Mullie / Unsplash

F. Frameworks for Ethical, Aligned, and Non-Extractive Engagement

Community Engagement, Data Sovereignty, and Privacy

To advance equitable conservation, practitioners must respect historical inequities, including by operating from the foundational principle that ecological and social data do not belong to external developers. One participant emphasized, "At most, you are a custodian of the data on behalf of someone else."

Upholding data sovereignty requires respecting community boundaries and technological limits. Communities must retain the right to say no to digitization. As one expert noted, "I think there just has to be an underlying acceptance that some of this stuff is not good to be digitized or modeled.... Not all knowledge is reducible to AI." Systems must allow communities to benefit from AI while maintaining data privacy, respecting sacred boundaries and ancestral norms rather than forcing data into open repositories. Federated AI and local-first architectures address this by running lightweight open-weights models on local computers under both the FAIR (Findable, Accessible, Interoperable, Reusable) Principles and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) Principles.

Sovereign data governance structures are designed to process sensitive local inputs, such as sacred site locations or eDNA data, using local-first hardware and encrypted open-weights models operating on-site without internet access. This architecture generates two distinct outputs: 1) sovereign local insights, such as custom forest classifications, land tenure maps, and climate adaptation strategies that remain under community control, and 2) anonymized federated updates that transmit general biological trends to global research networks without exposing raw data.

As multi-modal data integration expands, protecting data sovereignty will become increasingly challenging. One participant highlighted the tension between "breaking down silos of data ownership and bringing datasets together, and ensuring that communities, large and small, have sovereignty over their data." Field sensors carry severe risks of privacy violations, requiring simple opt-in and opt-out mechanisms. Sensors must be restricted strictly to ecological monitoring, preventing governments or corporations from using environmental hardware to spy on communities or harass women.

Projects must ask communities directly what privacy looks like to them. Case studies in practice can be found through platforms like Animikii, which focuses on Indigenous data sovereignty, and Abundant Intelligences, an Indigenous-led effort reimagining AI from Indigenous knowledge perspectives. Communities are actively deploying AI to secure land tenure, preserve native languages, adapt to climate change, and label forest types according to traditional classifications ignored by conventional remote sensing. Tech organizations should amplify these self-determined efforts rather than stretching communities thin through pro-forma outreach.

Institutional Trust and Two-Way Governance

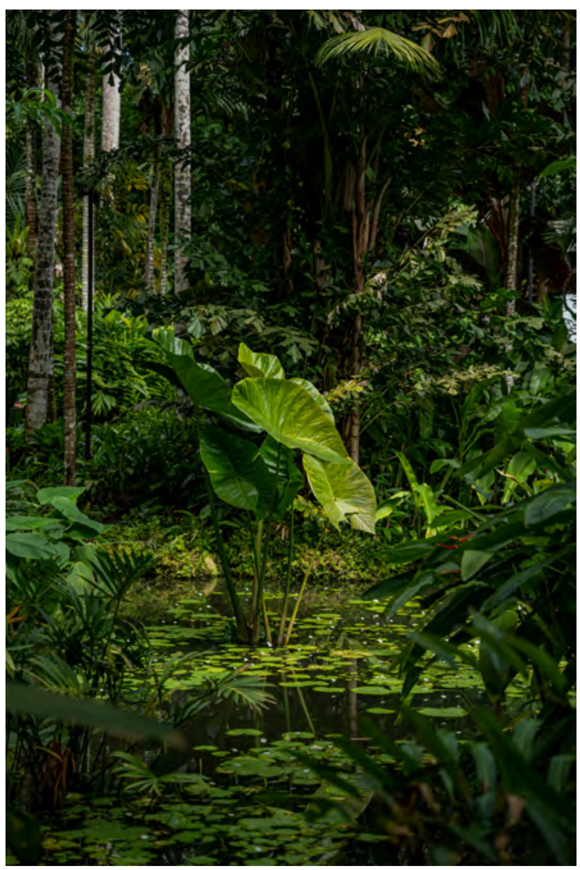
Institutional reputations precede organizations. Many field researchers and local communities willingly partner with organizations like Environmental Systems Research Institute (ESRI) but are reluctant to work with commercial technology companies because of concerns about data governance and privacy. As one participant stated, "If you're auditioning for trust once someone has already come up with the idea for the system, you're already behind." Applications must be designed around local priorities, even if the resulting tool differs substantially from its original design. Furthermore, researchers must avoid assuming Indigenous goals and conservation mandates are always aligned, as local economic goals can occasionally run at odds with strict conservation plans.

Roundtable discussants noted a paradox: many researchers readily trust AI outputs from unverified "black box" AI models while remaining skeptical of data generated through citizen science. Historically, trust-building efforts have focused on asking communities to trust researchers and institutions. Moving forward, trust must operate as a two-way street where researchers also trust community insights. Regardless of whether communities ultimately choose to engage with commercial technology providers, participants stressed the importance of access to clear information about algorithmic design, model limitations, and computational trade-offs—including the water and energy demands associated with AI systems—before deciding whether and how to participate.

Capacity-Building and Human-Centric AI Interaction

As one participant highlighted, global conservation tech faces an equity crisis rather than a capability crisis: "I think it's not an issue of quality but equity when it comes to AI, because AI systems are already wondrously capable. There is already so much about technology to unpack, but the real crisis comes down to the participation gap and the sense that communities and researchers are not the ones benefitting from advancements."

Barriers to equitable participation in AI-enabled conservation include unreliable internet access, limited tool availability, technical barriers to entry, and language constraints, as most AI systems rely



Dario Bronnimann / Unsplash

on written English. Given that many people around the world interact with digital platforms (such as YouTube) primarily through voice, developers should create voice-activated AI spatial and mapping tools. Additionally, because many countries lack comprehensive legal protections such as the European Union's General Data Protection Regulation (GDPR), AI could help summarize data rights, enabling users to better understand what they gain and surrender when agreeing to terms of service.

Experts expressed "qualified excitement," cautioning against overdependence on large-scale AI that lacks ground-truth validation and that is vulnerable to model drift. Technical systems should preserve opportunities for human interaction, local expertise, and field validation. For example, platforms like Merlin Bird ID do not simply produce a species prediction; they ask whether the user saw the bird while iNaturalist offers multiple interpretation pathways. These platforms illustrate how AI can support observation and learning while preserving human judgment and opportunities for verification.

G. Recommendations for Funders

Commercial technology firms are investing billions of dollars in large-scale AI model development, making proprietary software the fastest-obsoleting component of the tech ecosystem. Consequently, philanthropic foundation capital achieves its highest long-term leverage by funding primary data acquisition, open sensor infrastructure, taxonomy, and equitable capacity-building.

1. Think Beyond Direct Work on AI Models

- *Prioritize Data Acquisition and Sensor Hardware:* Shift investments from developing new AI models and toward reducing primary data gaps. As one participant advised: "I work for an AI company, and I think less now about building AI systems. AI is getting better so fast. No non-profit funder can register as more than a drop in the bucket, compared to the amount of money already being spent on AI. Gemini, Claude, and GPT are all getting better so quickly. I am more interested in data acquisition and applications. I would under-index as much as possible on the AI part. The AI part of whatever you do, whatever you invest in, is the part that will get obsolete the fastest. Even if you have a deep belief in AI, which I do, I work for an AI company, it's not the place where I would advise a nonprofit funder to be investing the most heavily."

- *Reduce Cost Disparities in Compute:* Subsidize computational costs and token credits for researchers and non-profits in developing nations, ensuring that high application programming interface (API) costs do not exclude Global South scientists from utilizing advanced foundation models.

- *Leverage Non-Monetary Organizational Assets:* Given the slow pace of government funding, funders should use their influence to secure physical access to unmonitored locations globally, deploying open sensors to generate public data streams.

- *Recognize Macro-Systemic Interconnections:* Target investments at the intersection of energy, water, and agriculture, as transformations across these three sectors will drive shifts in gross domestic product (GDP), environmental degradation rates, and computational energy demands.

2. Help Create a Demand Signal for Open Ecological Data

- *Fund Taxonomy and Data Labeling Infrastructure:* Provide sustained funding for taxonomic classification schemes and human-annotated reference libraries. Automated monitors—including autonomous underwater vehicles, aerial drones, and ocean gliders—require high-quality biodiversity training libraries to function.

- *Map High-Risk, Data-Deficit Environments:* Target habitats under pressure from extractive industries where data is currently monopolized. For example, funding autonomous underwater vehicles (AUVs) to collect bathymetry data and map benthic biodiversity in deep marine zones threatened by deep-sea mining provides open data before destructive extraction occurs.

- *Expand Open Satellite Initiatives:* Replicate open-access data models like the NICFI Satellite Data Initiative, where Planet Labs and the Norwegian Climate and Forest Initiative opened satellite imagery of the tropics to global researchers. Funders should also continue investing in species distribution maps and habitat suitability layers.

3. Broaden and Decentralize Strategic Investments

- *Direct Capital to the Global South and Indigenous Communities:* Address persistent geographic disparities in ecological data by investing directly in capacity-building across Africa, Southeast Asia, Oceania, and Latin America, supporting Indigenous-led sovereign tech infrastructure.

- *Scale Global Innovation Challenges:* Expand global prize mechanisms—such as the XPRIZE Rainforest—that equip multi-disciplinary teams worldwide to rapidly develop autonomous technologies for bio-mapping.

Chirag Saini / Unsplash

- *Fund the Science of Ecological Restoration:* Move beyond funding rewilding projects solely as implementation efforts. Reframe investment in fundamental research as exploring the mechanistic drivers of successful ecosystem restoration and biodiversity recovery.
- *Support Decentralized, Coordinated Monitoring Infrastructures:* Promote global monitoring frameworks that allow individual countries to select monitoring tools tailored to their unique ecosystems while adhering to shared standards, such as EBVs and EOVs. Common infrastructures enable AI models to aggregate time-series data across regions while preserving local flexibility.

H. Conclusion: Toward a Shared Ecological Future

The insights gathered through these convenings reveal a critical juncture for biodiversity and conservation sciences. Artificial intelligence offers unprecedented capabilities to monitor, model, and protect Earth's complex ecosystems. Yet, technology alone cannot resolve the biodiversity crisis. Algorithms are amplifiers of human intent, data infrastructure, and institutional priorities.

Achieving lasting conservation outcomes requires building durable connective tissue between computational scientists, field ecologists, policymakers, and community stewards. By addressing structural data gaps, embedding causal mechanics into predictive architectures, upholding community data sovereignty, and directing philanthropic capital toward primary data acquisition, the global community can construct a transparent, equitable, and highly capable conservation ecosystem. Implemented together, these strategies can ensure that modern technology serves not only to describe life on Earth but to protect it for generations to come.

2. Climate & Earth System Modeling

A. Participants

We aim to synthesize and share perspectives from the discussions as a whole rather than to attribute any quotations or viewpoints to specific individuals. Participants with whom we engaged in roundtable discussions are listed below (alphabetically by last name):

- **V. Balaji, Ph.D.** – Distinguished Fellow and Climate Institute Lead, *Schmidt Sciences*
- **Annalisa Bracco, Ph.D.** – Senior Scientist, *CMCC (Euro-Mediterranean Center on Climate Change) Foundation*
- **Hannah Christensen, Ph.D.** – Associate Professor of Physical Climate and David Richards Tutorial Fellow at Wadham College, *University of Oxford*
- **William Collins, Ph.D.** – Associate Laboratory Director for the Earth & Environmental Sciences Area, *Lawrence Berkeley National Lab*
- **Simon Driscoll, Ph.D.** – Assistant Research Professor of Applied Mathematics and Theoretical Physics, *University of Cambridge*
- **Pierre Gentine, Ph.D.** – LEAP (Learning the Earth with Artificial Intelligence and Physics) Director, Professor of Geophysics, Professor of Earth and Environmental Sciences, and Professor at the Climate School, *Columbia University*
- **Andrew Gettleman, Ph.D.** – Senior Scientist of Climate and Atmospheric Chemistry, *Pacific Northwest National Laboratory*
- **Karthik Kashinath, Ph.D.** – Distinguished Scientist and Engineer, *NVIDIA*
- **Venkataraman Lakshmi, Ph.D.** – John L Newcomb Professor of Engineering, *University of Virginia*; Hydrology Section President, *American Geophysical Union*
- **Upmanu Lall, Ph.D.** – Global Futures Professor, *Arizona State University School of Complex Adaptive Systems*; Director, *Arizona State University Water Institute*
- **Flavio Lehner, Ph.D.** – Assistant Professor of Atmospheric Sciences, *Cornell University*; Chief Climate Scientist, *Polar Bears International*
- **Christian Lessig, Ph.D.** – Assistant Professor, *Otto-von-Guericke Universität Magdeburg*; Research Scientist, *European Centre for Medium-Range Weather Forecasts*
- **Anna Michalak, Ph.D.** – Director, *Carnegie Institution for Science*; Professor, *Stanford University*; Visiting Faculty Researcher, *Google Research*
- **David Rolnick, Ph.D.** – Assistant Professor of Computer Science, *McGill University*; Founder and Chair, *Climate Change AI*
- **Gavin Schmidt, Ph.D.** – Director and ModelE Earth System Model Principal Investigator, *NASA Goddard Institute for Space Studies*

- **Tapio Schneider, Ph.D.** – Theodore Y. Wu Professor of Environmental Science and Engineering, *California Institute of Technology*; Lead Principal Investigator, *Climate Modeling Alliance (CliMA)*
- **Sabine Scholle** – Machine Learning Student Researcher, *Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research*
- **Aditi Sheshadri, Ph.D.** – Assistant Professor of Earth System Science, *Stanford University*
- **Emily Shuckburgh, Ph.D.** – Professor of Environmental Data Science and Director of Cambridge Zero, *University of Cambridge*
- **Deepti Singh, Ph.D.** – Associate Professor, *Washington State University School of the Environment*
- **Laure Zanna, Ph.D.** – Professor of Applied Mathematics, *New York University Courant Institute of Mathematical Sciences*

B. Introduction

The Earth system operates across vast temporal and spatial scales, regularly straining the limits of human cognition and classical computing. From microphysical cloud interactions occurring in milliseconds to multi-century ocean circulation, modelers have long struggled to bridge the gap between local physical processes and planetary-scale dynamics. In recent years, the rapid integration of artificial intelligence into Earth system science has promised to expand what models can represent, and in turn, generate new paradigms for resilience and decision-making.

However, roundtable experts emphasized that enthusiasm surrounding computational breakthroughs should be tempered with epistemic caution. As one participant reflected at the outset of the conversation, “I would like us to go in with a slightly neutral statement. Yes, we are expecting a lot of breakthroughs to come out of AI, but it is also possibly going to be a setback to the field, because we spend a lot of time trying to figure out how to do these data-driven methods and we [might] forget how to do hypothesis-driven science. So, I would like to leave a little gap for the possibility that AI is not a breakthrough.”

Navigating this tension requires evaluating how machine learning can best be paired with numerical weather prediction, physical conservation laws, sub-seasonal forecasting, and local use cases. While purely data-driven approaches excel at short-term pattern recognition, long-term projections and out-of-distribution climate shifts demand rigorous physical grounding. Participants also drew boundaries between weather and climate predictions, noting that optimal modeling approaches vary between the two objectives.

This chapter explores how artificial intelligence can democratize high-resolution modeling through emulators and data compression; how machine learning can unify existing observations and strengthen knowledge in data-poor environments; how hybrid models can balance physical conservation laws with statistical pattern identification; how AI can bridge the “last mile” to deliver bespoke decision-making applications for communities; and what infrastructure and funding investments are needed to positively shape the field.

Core Takeaways

- **Emulators should accelerate, not replace:** Emulators offer dramatic computational efficiency, enabling massive ensemble generation and rapid model exploration. However, their use should not come at the expense of fundamental numerical physics.
- **Unlocking underexplored data repositories:** Machine learning tools can re-evaluate decades of petabyte-scale satellite and sensor observations, extracting subtle cross-correlations and multi-variate signals that traditional climate models stripped out.
- **Bridging the "last mile" to actionable decision-making:** AI can enable the rapid construction of localized, bespoke applications built on top of broad forecasts, directly addressing specific municipal, agricultural, and industrial decision-making needs.
- **Epistemic risks of structural model errors:** Emulating flawed or uncertain climate models merely reproduces and amplifies underlying structural errors at higher speeds, making rigorous empirical benchmarking essential.
- **Strategic philanthropy must prioritize open infrastructure:** Funders should move beyond funding isolated algorithms and invest in user-friendly data lakes, localized hackathons, and cross-disciplinary training to better the practice of climate science.

C. Areas of Excitement and Progress

Democratization Through Emulators and Data Compression

Roundtable participants expressed optimism about the new generation of machine-learning-based emulators, describing them as a powerful "accelerant, not a replacement" that should be coupled with traditional numerical weather predictions and climate models. Emulation techniques are particularly transformative because they allow scientists to sample the full details of statistical distributions, generate large ensembles, and rapidly explore complex model output—all at a fraction of the cost and time of running a full supercomputer simulation. In particular, attendees noted that the ability to run an emulator on a personal computer "democratize[s] the enterprise," allowing researchers around the world to ask their own questions instead of just looking at published outputs.

Emulators are also opening up the methodological possibility space. "If you have a complex time series dataset or a high-resolution or high-volume dataset, emulation is emerging as a very powerful method to actually be able to represent what that time series does using a differentiable and computationally [efficient] emulator. That has the potential to change the game. Since it's differentiable, we will be able to do inverse methods and 4D variational methods," a participant explained.

Simultaneously, machine learning serves as a high-capacity data compressor. By compressing petabyte-scale outputs from super-high-resolution simulations or other sources, AI can serve refined, actionable data to researchers and institutions that may lack the infrastructure to access this information directly.

Strengthening Knowledge of Data-Poor Environments

Data availability across the globe remains sharply unequal, with participants specifically pointing to critical observational gaps in the Global South. Machine learning offers frameworks to transfer insights between data-rich and data-poor areas that resemble each other on a well-defined set of characteristics. For example, one attendee described inferences between hydrologically similar catchments (determined through variables such as precipitation, soil type, and topography) as an important component of their work. Such efforts were described as moving toward a framework where "we are not using AI blindly as a hammer, but rather deftly as a scalpel."

AI tools also unlock new value from hardware originally built for other applications. For instance, although only a small number of satellites were specifically designed to measure carbon dioxide and methane from space, dozens of other satellites have sensitivity to collateral data. As one scientist explained, "Leveraging AI tools to infer signals about quantities that satellites can measure at least faintly, but that are different from the original intended design of those satellites, is a way to dramatically increase the information that we have to constrain various problems in Earth science."

Unifying Disparate Data and Synthesizing Legacy Observations

Machine learning platforms excel at synthesizing disparate and non-traditional data streams to solve complex optimization problems. While commercial sectors already combine weather forecasts with unconventional variables like flight bookings and hotel occupancy, participants highlighted that equivalent cross-domain frameworks remain underutilized in public climate science. Researchers were specifically optimistic that AI can bridge climate model outputs with frameworks from fields like economics, public health, and water management.

Attendees also noted an exciting opportunity to revisit decades of historical observations with modern AI. Agencies like NASA and the European Space Agency have accumulated hundreds of petabytes of Earth observation data since their founding—a rich, multivariate dataset whose true information content remains wildly underexploited. "If I reflect on how climate scientists have distilled [information], we've sort of removed all the interesting variability and cross-correlations in that data

Getty Images / Unsplash

set in order to look at it one variable at a time in order to diagnose our models. The way that [more traditional] climate models use that data, we have literally boiled out of the data all the cross-correlations, all the high temporal frequency, all the high spatial variability," one participant shared. Revisiting these observations with AI provides the tools to merge across diverse data modalities—combining top-down estimations based on atmospheric data with bottom-up estimations based on greenhouse-gas emissions at the Earth's surface.

By applying AI across these diverse modalities, researchers can uncover physical patterns that traditional approaches missed. For instance, one workshop participant described a machine learning model developed by their lab that is producing skillful hindcasts of phenomena originally missed by European Centre for Medium-Range Weather Forecasts models—even without prior training on those specific occurrences.

Elevating the User Experience and Expanding the User Base

Beyond improving core weather and climate models, machine learning offers major user-facing capabilities, enabling scientists and non-experts alike to interact with large-scale environmental datasets. However, participants cautioned that increased accessibility can carry risks: while AI tools make climate information more available, they can make it harder to unpack nuance in areas where the public has few ways to verify whether AI-generated answers are authentic.

At the daily operational level, participants described machine learning as a "force multiplier" for the scientific enterprise, including by automating code generation, helping streamline software workflows, and assisting researchers in curating relevant scientific literature.

D. High-Value Research Areas and Applications

Accelerating Model Calibration via Emulators

Model calibration and parameter tuning in Earth system modeling has historically been a tedious and expensive process that can span multiple years. Participants explained that machine-learn-

Karsten Wurth / Unsplash

ing-based emulators offer a way to dramatically accelerate this workflow by narrowing parameter uncertainties before running expensive, full-scale forward simulations.

However, joint calibration across multiscale systems remains an open optimization challenge. Attendees commented that when structural or parametric uncertainties within the same model span timescales from minutes to millennia, even newer approaches tend to hit a wall. Calibration techniques are also highly field-specific, requiring tailored innovation and experimentation. “Whereas in the climate models we do not really have least-square calibration of the parameters of the model, in the hydrologic models we do. What I see is that the multi-head calibration versus non-multi-head calibration of even the physics-based models leads to a significant improvement, which the field hasn’t historically done,” one participant shared.

Balancing Observed and Learned Physical Constraints

A prominent debate during the roundtables centered on the extent to which physical laws should be explicitly hard-coded into model architectures versus learned directly from observational data. On the one hand, imposing physics-informed constraints can promote model interpretability and realism. On the other hand, doing so may eliminate possible solutions that AI algorithms would have discovered through pattern-matching. Participants added that the field has a tendency to be overly critical of machine learning models, emphasizing that the primary objective is not achieving zero bias but improving upon the known biases of existing numerical weather prediction (NWP) climate models. Indeed, experts pointed out that even gold-standard Earth-system models rely on artificial fixes—such as the explicit “energy kludge” coded into the Community Earth System Model (CESM) to force conservation—while widely used tools like the Weather Research and Forecasting (WRF) model do not require strict energy conservation to function effectively. “So yes, we would like our machine learning algorithms to conserve energy and moisture, but at the same time, our physical models couldn’t do that, and we fix it in,” one participant commented.

The conversations revealed an important distinction between the work of weather forecasting and climate prediction, with several participants suggesting that these sub-fields must operate under fundamentally different validation regimes. Whereas weather forecasts are continuously verifiable on daily human timescales and were regarded by most participants as “in-sample,” climate projections operate in out-of-distribution regimes that are not verifiable on a humanly useful timescale. As one participant argued, “Machine learning works well when you have full, observational datasets and you can emulate things within that box quite well. When you try and extrapolate outside of the range of the training data, it often fails, quite appallingly. So things like actual climate predictions independently directly from data, as opposed to emulating models that have generality, [are] I think a non-starter.”

These extrapolation paradigms led participants to consider how physical constraints should be applied over multi-decadal timescales. Rather than favoring one approach over another, participants underscored the importance of working backwards from the model’s objective and the available datasets. While short-term weather forecasts do not require strict conservation, systems that deal with longer-term predictions are sensitive to small perturbations, which can manifest as leaks. Establishing community standards to judge physical fidelity requires atmospheric physicists who understand conservation laws to work closely with machine learning developers. One participant’s lab recently examined a highly respected emulator that yielded stable predictions out to a century, yet violated the fundamental laws of energy and water conservation on the timescale of a single week.

As the discussions progressed, participants discussed combining pattern matching with physics-informed constraints, provided that physical models remain independent of the observational data used to train the AI models, thereby reducing the risk of correlated errors. As one contributor explained, "when we literally mean physical laws, I see no harm in encoding those physical laws. I also see no harm in not encoding them, and seeing whether the AI-based models are, in effect, obeying those laws because they have learned about them. I think both of those approaches are perfectly fine and complementary to one another. I think where this approach starts to fail is when we say physical laws, but in reality, what we mean are not truly fundamental physical laws, but things that we think we understand about how one process responds to another process."

Finally, one participant challenged the assumption that training models on other models is inherently flawed. Operating under a structured model hierarchy, they characterized the training of computationally efficient convection parametrizations on convection-resolving models as "perfectly appropriate." In this view, models are representations of physical theory, meaning that emulators can be deployed to test theories by identifying structural model errors when parameters fail to fit observational data.

Shorter-Term Forecasting

While predicting climate shifts a century out remains difficult, short-term and sub-seasonal forecasting represents a domain where AI can deliver immediate, high-value societal benefits. These practical questions operate on timescales compatible with political elections and day-to-day human management. As one participant reflected, it is helpful to think about "the interface between the questions in earth science and climate that AI can address, and those questions that if we answer them would have the most immediate positive impact on societal outcomes and on improving our [management abilities on] timescales compatible with political elections.... That's where the benefits of AI are, so I wouldn't spend too much time today sad about the fact that we don't think AI will

Noaa Kcvlb / Unsplash

tell us what the climate will be a century from now. I think that those more practical questions are actually ones that are incredibly exciting, and that are much more poised for an AI revolution."

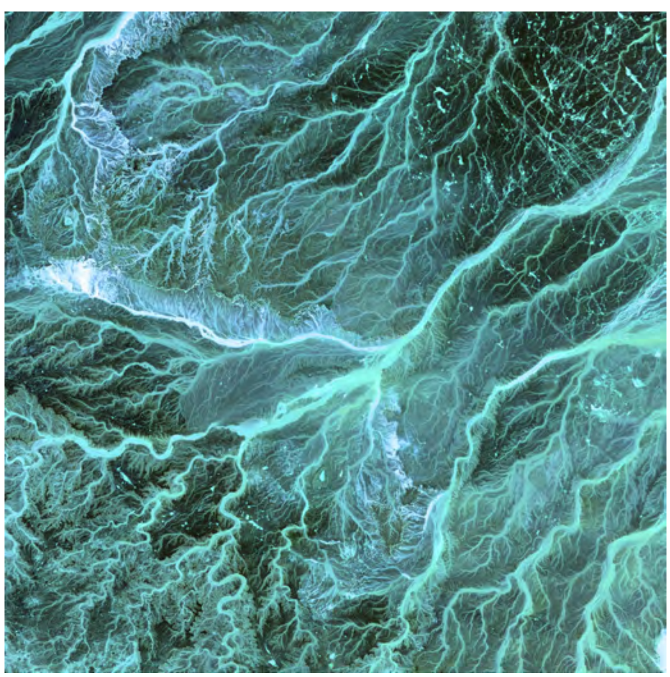
Bespoke Applications for Decision-Making

A primary strength of AI lies in solving concrete, "last mile" problems: converting general predictions into actionable information tailored to a user's needs. Participants described possible applications ranging from agriculture (i.e., predicting the intervals between dry spells and rainfall to better inform irrigation practices) to fleet management (i.e., forecasting localized hail damage that may affect autonomous vehicle operation).

As one expert detailed, "Weather forecasts are useful to a very large number of communities, governments, industries, and so on. In practice, however, the precise piece of information that is useful or actionable is extremely community-dependent or application-dependent.... One of the areas where AI can help is to build a very large number of bespoke applications, on top of existing models, that directly address very specialized and localized user needs, because there will never be enough [separate] investment [in] each of those applications. Each of those applications is complicated but, in a way, discoverable."

Building on this point, attendees cited accounts from resilience policy expert Alice Hill, noting that AI can assist municipalities that lack dedicated climate planning staff to create data-driven plans around needs like stormwater management. Early efforts to build out such infrastructure include California's Cal-Adapt website. When creating bespoke models, scientists should engage in stakeholder dialogue to better understand problems and deliver useful information to decision-makers.

Capturing all of the variables specific to a scenario has traditionally been computationally expensive. Emulators in particular allow researchers to achieve dynamical downscaling at a fraction of traditional costs. As one scientist emphasized, "The beauty of the emulators is that you can now have your cake and eat it, too. So you can get what is effectively dynamically downscaled information



KmebjM / Unsplash

with all the variables that you need." They provided the example of evaluating a building's energy performance, explaining that emulators "allow you to get bespoke without the cost" and to include diverse information about solar illumination, thermal absorption, and wind currents.

Cryospheric Modeling

Ice sheet modeling was described as an "almost a perfect use case" for machine-learning emulation. Ice sheets are governed by relatively well-understood physical processes, including mass input from the top and static and kinetic friction from basal drag. However, modelling remains challenging because velocities can range from centimeters per year in the center of a sheet to kilometers per year near the edges. Emulators allow scientists to "perform systematic attacks on the uncertainty associated with our projections of sea level rise due to ice sheets," addressing what participants noted were large uncertainty ranges in the Intergovernmental Panel on Climate Change (IPCC)'s Sixth Assessment Report.

In addition to narrowing uncertainty, machine learning can aid sea ice observation by bridging disparate satellite sensors. Although synthetic aperture radar (SAR) instruments aboard satellites like Sentinel-1 penetrate cloud cover, this technology is not universal. Attendees found value in transferring learnings from SAR-equipped satellites to those unable to see through clouds, in turn strengthening the field's understanding of variables like melt pond fractions.

Extreme Weather Detection and Attribution

Roundtable experts noted that while prediction is a crucial aspect of climate science, global detection and attribution work also have important policy implications. To disentangle anthropogenic climate signals from natural variability, scientists compare observed events with modeled counterfactual conditions representing a climate without specified human influences. Climate emulators extend these capabilities by generating massive ensembles that would be computationally infeasible with traditional numerical models, allowing researchers to sample the far tails of climate distributions and investigate plausible worst-case scenarios for measures like rainfall and temperature. In one study cited during the discussions, researchers used the Spherical Fourier Neural Operator (SFNO) machine learning model to generate a massive ensemble containing over 7,000 recreations of the summer of 2023, enabling novel analysis of heat extremes and probabilities. Participants highlighted the importance of pairing rigorous validation with complementary theory development, such as Probable Maximum Precipitation (PMP) methods.

E. Persistent Challenges, Limitations, and Risks

Benchmarking and Uncertainty Quantification

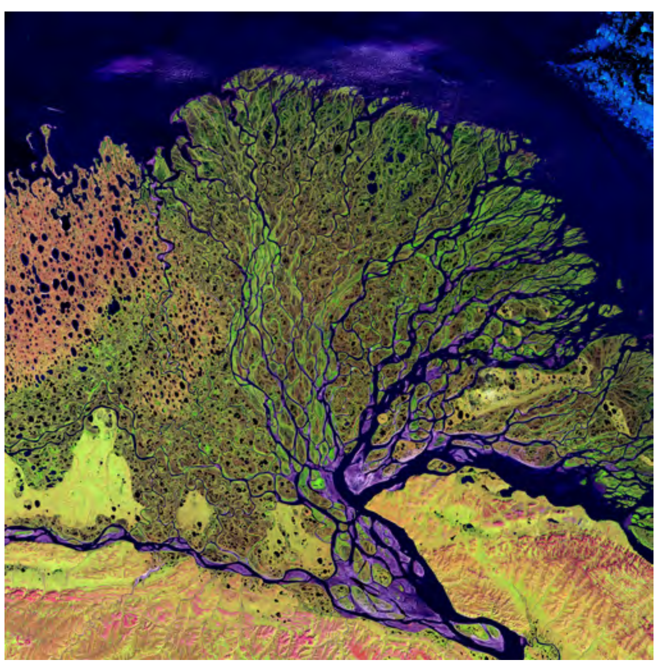
Despite the excitement surrounding ensemble generation and parameter testing, experts underscored that emulating flawed climate models carries significant risks; if an underlying physical model contains large errors or draws on incorrect assumptions, an emulator will merely reproduce those imperfections at higher speeds. As one participant cautioned, "While I completely agree that one of the areas that AI can help with is in building emulators of models in a variety of areas…this is actually one of the areas that I'm most nervous and skeptical about, because we know that these models that we currently have themselves have huge amounts of uncertainty—structural uncertainty, parametric uncertainty. A lot of assumptions are hard-coded into those models. So while I see the value of running those models faster, I'm a little bit skeptical about the degree to which that will really help us answer fundamental questions in Earth science. To the extent that those models have imperfections associated with them, we'll just be doing those more."

Participants also explained that because emulators sample the uncertainty as represented by the model—not the uncertainty inherent in the physical problem itself—scientists should maintain two tracks of dialogue around the concept of uncertainty. One approach to uncertainty involves evaluating machine learning models against empirical benchmarks. However, the process of developing such benchmarks remains an open question. In terms of content, development tends to be more feasible on shorter timescales where observational data is abundant, unlike decadal and centennial timescales that may lack sufficient historical data to appropriately constrain models securely. In terms of source, attendees noted that benchmark development should involve affected communities, open knowledge transfer between researchers, and movement toward public-private partnerships over reliance on big tech companies.

Even with the existence of benchmarks like WeatherBench 2, uncertainty quantification remains a major challenge in machine learning. Standard outputs often fail to produce well-calibrated confidence intervals, if anything providing a single deterministic value. Participants explained that this represents a mismatch with the full probability density functions that decision-makers hope to see, cautioning that predictive models should not inform policy unless researchers can interpret how confidence operates in that context.

Scientific Publishing, Reward Structures, and Reproducibility

Participants warned that the proliferation of papers produced using generative AI and unverified machine-learning methods is introducing noise into the scientific literature, creating "AI slop" and



USGS / Unsplash

an explosion of "dodgy data." They expressed concern that high-profile journals increasingly publish attention-grabbing AI applications that lack physical substance, weakening long-standing review standards. Experts emphasized that extraordinary results require extraordinarily high levels of proof, adding that the field requires a paradigm shift in reproducibility, moving the focus from raw data to the specific methods used to generate that data. Against this backdrop, they noted that in-person interactions and academic conferences will become increasingly vital for evaluating research quality.

Simultaneously, as literature volumes expand beyond a quantity that labs can realistically digest, research groups are increasingly filtering papers by journal prestige, overlooking other relevant and credible sources from less prominent or open-science websites. Attendees were hopeful that natural language processing tools could help distill findings from across vast published records, including under-utilized repositories.

Finally, roundtable experts called for changes in academic promotion structures to better reward peer review and reproducibility checks.

Uneven Data Coverage and Institutional Barriers to Data and Model Access

While AI can strengthen inference in data-poor environments, algorithms cannot fully circumvent physical observation gaps. Environmental monitoring maps reveal severe geographical data gaps across the Global South. As one expert detailed, "As soon as you look at maps of this kind of [temperature and precipitation] data, it's extremely obvious that in many parts of the world, there are very few surface-based observations.... This means that any AI-enabled climate or weather model trained on this dataset is reproducing model data in many regions which include communities that are more vulnerable, underserved, or overlooked.... You can't get away from the fact that you would need to have better data for those regions." Others agreed, adding that, from a temporal perspective, every unmeasured year poses a loss for longitudinal analysis of processes like land-carbon cycles, permafrost, and tipping points.

Institutional policy barriers further restrict data and model access. Attendees noted, for instance, that river discharge data in India is frequently embargoed by government authorities, while historical hydrological records across many parts of Africa remain undigitized. Participants also described bottlenecks arising from the slow cadence of the Coupled Model Intercomparison Project (CMIP). They noted that the time between successive CMIP model releases can leave coordinated climate projections reliant on older configurations, even as observational capabilities and models developed by individual research groups continue to advance.

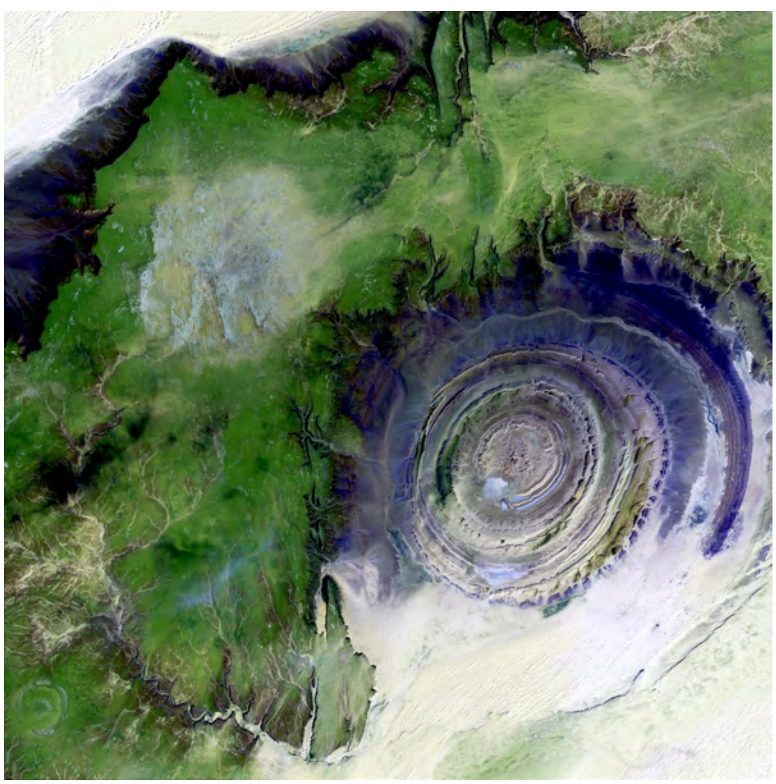
Conflation Between Climate Models and General Purpose AI

Discussions also touched on narrative drift surrounding AI tools. "There is increasingly a conflation of AI with large, agentic language-based systems, which are the products of a select few large tech companies. If you look at everything that [attendees] said, there were only a couple of situations where those were actually the tools that were useful in climate science," one participant explained. Although many labs regularly draw on commercial platforms within their workflows, the breakthroughs shared during the roundtable sessions largely stemmed from problem-specific models built by interdisciplinary teams. “I’m worried that AI for Science [will be] used to sell AI that isn’t actually helpful in science,” the participant continued.

F. Recommendations for Funders

Participants noted that philanthropic organizations and funding agencies frequently view computational models as more “sexy” than physical data gathering. However, they emphasized that primary observational data serves as the essential “fuel for all the machine learning” and encouraged funders to rebalance investments in ways that prioritize data infrastructure, accessible computing, and interdisciplinary training.

1. **Invest in User-Friendly Open Data Lakes:** Although more than 100 terabytes of remote sensing data are collected each day—in large part by government agencies like NASA and the ESA—public availability does not equate to usability. Funders should support centralized, open-access data lakes that allow researchers to explore homogenized datasets and quickly pull down the data relevant to their project. Participants suggested prioritizing a few areas like top-of-atmosphere energy fluxes, microwave radar, and cloud LiDAR (Light Detection and Ranging) to begin. Drawing on burgeoning data infrastructure efforts such as Destination Earth, these repositories should adopt open formats (such as Zarr) and leverage AI as a catalyst to standardize global climate data. Before and during the creation of such data lakes, participants encouraged conversations about why such efforts have been unsuccessful in the past.
2. **Organize Problem-Specific Hackathons:** Funders should support targeted problem-solving competitions such as the Global KM-scale Hackathon. At the local level, hackathons help



USGS / Unsplash

equip creative research teams with right-sized models and data—unlocking access to tools that may otherwise be prohibitively expensive. One participant reflected on their experience organizing a 700-participant tournament, explaining the importance of inviting teams from across disciplinary backgrounds.

3. **Drive Down the Cost of Climate Models:** As high-resolution numerical models grow increasingly expensive, smaller research groups are often excluded from running custom simulations or sensitivity tests, forcing them to rely passively on decade-old CMIP outputs. Funders should support open-source emulators capable of simulating time-series trajectories. As one expert summarized, "One big opportunity here is this idea that you can emulate anything that produces a time series trajectory. If that actually proves to be something that has legs, then I think we can use that to democratize the enterprise so that a lot more people can actually do actual science. That means [you can] ask questions, not just look at forecasts."

4. **Build a Common Interdisciplinary Language:** Achieving transformative breakthroughs requires interdisciplinary fluency rather than surface-level collaboration, yet a persistent cultural gap separates Earth scientists from machine learning experts. Participants emphasized that early-career training programs must educate students to be fluent in both machine learning computation and environmental field work. As one participant summarized, "I think the operative word here should be integration.... That starts with how students are being trained [through cross-linked] graduate programs between computer science, applied math, [and] climate science." They continued, "In the end, we will not win big here with just domain science expertise or just AI expertise."

G. Conclusion: Toward Actionable and Integrated Earth System Science

The integration of artificial intelligence into the climate and Earth sciences represents a pivotal evolution in how researchers model, understand, and respond to planetary dynamics. By accelerating numerical simulations through emulators, extracting hidden signals from disparate sensing tools, and creating bespoke decision-making tools for local communities, machine learning offers unprecedented capabilities to bridge the gap between complex climate data and societal action.

However, realizing the full promise of AI for the Earth and its inhabitants requires maintaining a grounded balance between statistical innovation and physical reality. Data-driven speed cannot be a substitute for physical fidelity, nor can computational models circumvent the urgent need for better primary observations in overlooked regions. By building open, user-friendly data infrastructure, enforcing fundamental conservation laws across timescales, establishing trustworthy empirical benchmarks, and fostering genuine interdisciplinary fluency, the scientific community can ensure that AI serves as a powerful instrument for Earth system understanding and global climate resilience.

3. Biological Inference

A. Participants

We aim to synthesize and share perspectives from the discussions as a whole rather than to attribute any quotations or viewpoints to specific individuals. Participants with whom we engaged in roundtable discussions are listed below (alphabetically by last name):

- **Otto Cordero, Ph.D.** – Associate Professor of Ocean Utilization, *Massachusetts Institute of Technology*
- **Iain Couzin, Ph.D.** – Director, *Max Planck Institute of Animal Behavior*
- **Robert Darnell, M.D., Ph.D.** – Professor and Head of Laboratory of Molecular Neuro-oncology, *The Rockefeller University*
- **Michael Desai, Ph.D.** – Fisher Professor of Natural History and Co-Director of the Systems Synthetic and Quantitative Biology Ph.D. Program, *Harvard University*
- **Polly Fordyce, Ph.D.** – Associate Professor of Bioengineering and Genetics, *Stanford University*
- **Paul François, Ph.D.** – Professor of Biochemistry and Molecular Medicine, *Université de Montréal*; Associate Academic Member, *Mila Quebec Artificial Intelligence Institute*
- **Benjamin Good, Ph.D.** – Assistant Professor of Applied Physics and Biology, *Stanford University*
- **Simon Levin, Ph.D.** – Distinguished University Professor and Director of Center for Biodiversity, *Princeton University*
- **Andrea Liu, Ph.D.** – Founder and Director, *University of Pennsylvania Center for Soft and Living Matter*
- **Madhav Mani, Ph.D.** – Associate Professor of Engineering Sciences and Applied Mathematics, *Northwestern University*
- **Ron Milo, Ph.D.** – Professor of Systems Biology, *Weizmann Institute of Science*
- **Thierry Mora, Ph.D.** – Deputy Director of Laboratoire de Physique, *École Normale Supérieure*; Fellow, *American Physical Society*
- **Daniel Segrè, Ph.D.** – Professor of Biology, Bioinformatics, and Biomedical Engineering, *Boston University*
- **Alex Shalek, Ph.D.** – Professor and Director of Institute for Medical Engineering & Science and Director of Health Innovation Hub, *Massachusetts Institute of Technology*; Institute Member, *Broad Institute of MIT and Harvard*; Member, *Ragon Institute of Mass General Brigham, MIT, and Harvard*
- **Pamela Silver, Ph.D.** – Professor of Systems Biology, *Harvard Medical School*
- **Justin Sonnenberg, Ph.D.** – Professor of Microbiology and Immunology, *Stanford University School of Medicine*

- **Joseph Thornton, Ph.D.** – Professor of Ecology and Evolution and Professor of Human Genetics, *The University of Chicago*
- **LingChong You, Ph.D.** – Director of Graduate Studies and Professor of Biomedical Engineering, *Duke University*
- **James Zou, Ph.D.** – Associate Professor of Biomedical Data Science, Computer Science, and Electrical Engineering, *Stanford University*

B. Introduction

Rapid advancements in machine learning are transforming biology from an observational research area into an increasingly predictive discipline. Despite notable progress across disciplinary niches—from genomes and microbiomes to animal groups and global ecology—roundtable participants underscored that biological inference is inherently plagued by a "curse of dimensionality" that complicates mechanistic understanding. As computational tools become more deeply integrated in the life sciences, the next frontier of discovery will involve moving from pattern recognition to systems understanding: deciphering the underlying principles that govern dynamic, living collectives across scales and using these insights to deliver societally beneficial outcomes.

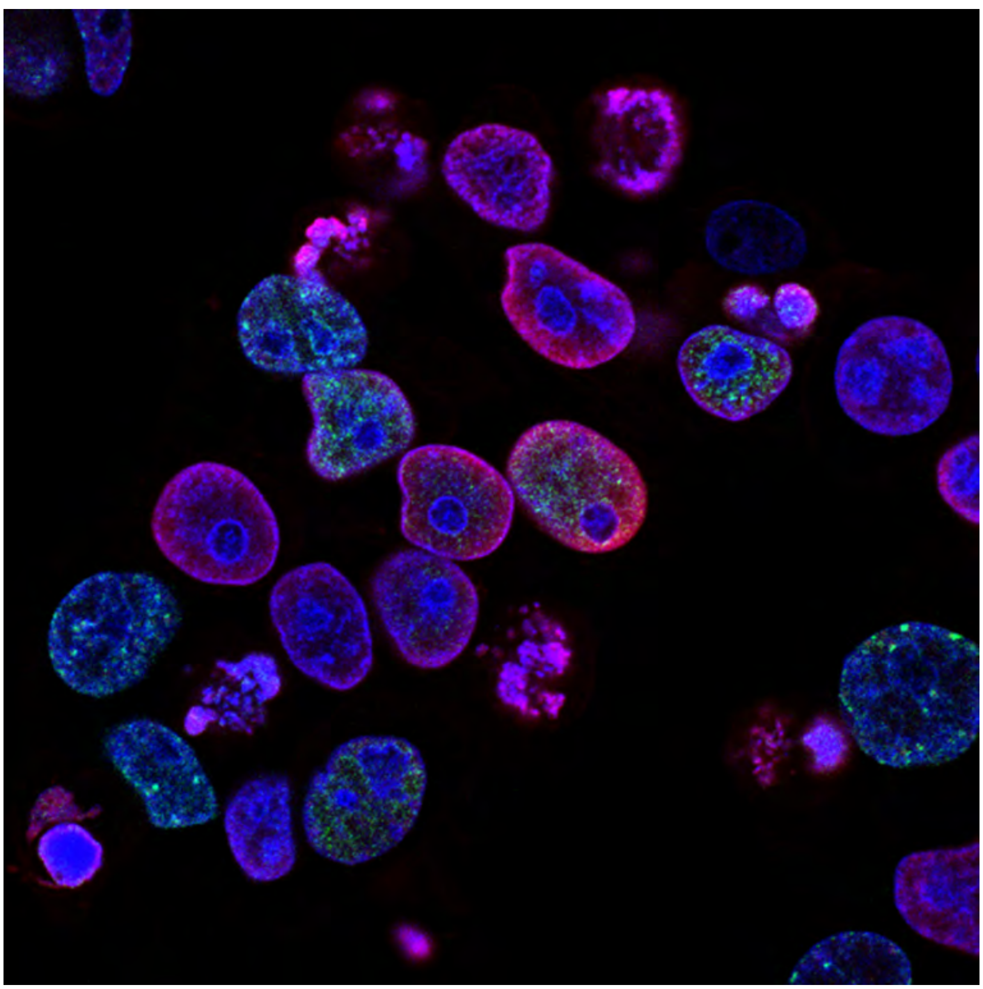
This chapter considers how this transition could unfold with the strongest returns, specifically exploring how AI can enable mechanistic understanding in complex living systems; how machine learning might expand exploration of biological hypothesis spaces; how theory, modeling, and experimentation may be reshaped by AI-assisted discovery; and what constitutes genuine, explanatory insight in an era of increasingly powerful predictive models.

Across biological subfields, roundtable experts identified a central tension: *how necessary is mechanistic understanding, and for what use cases can pure prediction suffice?* Participants noted that AI-driven, black-box outcomes are more "honest" to the observed reality because they report what is seen directly without forcing human assumptions onto a complex system. These models can deliver immediate value in applied domains such as diagnostic screening, encouraging patients to seek further specialized care. However, a majority of participants insisted that setting the criteria for predictive success and intervening safely in dynamic systems require causal understanding. Without deeper knowledge, drug manufacturers may target the wrong mechanisms—leading to harmful side effects—and clinicians may not understand how to best administer ramp-up doses or manage therapeutic progression.

Throughout the convenings, participants also grappled with the very meaning of the term "mechanism." Proposed definitions ranged from operational concepts—such as perturbing a variable to reliably drive a system to another state—to biochemical signaling pathways that can be experimentally validated by removing a gene, observing system collapse, and restoring function via gene re-introduction. Complicating this debate are the standards set by outside institutions. Whereas one participant explained that approval for their work on a diagnostic algorithm for cardiac disease depended on accuracy over mechanism, others shared difficulties in attracting investors and clinical trial participants for a cancer study in which mechanisms remained opaque.

Core Takeaways

- **Prediction does not equal mechanism:** While predictive accuracy delivers immediate value in applied domains like diagnostic screening, safe interventions in dynamic biological systems require causal understanding.
- **"Combinatorial creativity" expands scientific reasoning:** AI lacks human creativity and social understanding, but can still serve as an intellectual conduit by connecting terminology across disparate fields, upending entrenched theoretical biases, and re-parameterizing complex problems.
- **AI-oriented experimental design drives new possibilities:** Coupling AI reasoning engines directly with automated robotic platforms enables long-horizon, autonomous experimentation that overcomes human experimental constraints.
- **Hybrid and interpretable models anchor biological reality:** Embedding physical laws, mass conservation, and kinetic constraints into deep learning models helps bridge the gap between raw data fitting and biophysical truth.
- **Philanthropic capital must target data ensembles and outcome-driven discovery:** Funders should under-index on broad AI development and invest heavily in AI-ready databases, domain-specific foundation models, and solution-focused biological applications.



National Cancer Institute / Unsplash

C. Areas of Excitement and Progress

New Predictive Models

Participants opened the discussions by sharing the various ways AI has enabled new predictive insights across their respective subfields. For instance, researchers of the microbiome have found success in predicting whether a species will be present in the human gut, given knowledge of the other species present. Experts with a background in genomics spoke to the increasing quality of genome regulation models, which are now highly capable of identifying regulatory sites and predicting cell-type-specific patterns of gene expression. Participants also reiterated the landmark importance of models such as AlphaFold, which draws on amino acid sequences to quickly predict 3D protein structure.

Nevertheless, participants consistently expressed a desire for deeper mechanistic understanding. Returning to the microbiome space, researchers are able to predict the effect of certain perturbations—for example, microbiome diversity tends to fall when fiber is removed from the diet but increases sharply when fermented foods are introduced—but still do not understand why or how these changes occur on a mechanistic level. As another participant noted, this lack of general principles "reflects the entire field of quantitative genetics," where it is often possible to predict phenotype from genotype with varying degrees of accuracy across different systems. Similarly, in global ecology, researchers frequently observe patterns where the power laws governing populations, ubiquity, and diversity relationships remain unclear. Finally, participants expressed concern regarding the prediction-understanding divide in medicine. Although clinical trial data help researchers move beyond simple correlations, doctors and scientists still face difficulty explaining the precise mechanisms behind most observed therapeutic outcomes.

Alternative Perspectives and Combinatorial Creativity

Across the board, participants appreciated the ability of AI to integrate multiple data flows, suggest alternative ways of parameterizing problems, query situationally relevant literature, and join together insights from disparate disciplines. Although participants expressed that certain AI-generated outputs are obvious to experimentally focused research groups, they noted that these same results can be a game-changer for labs unaccustomed to computational workflows. Across subfields, experts highlighted that AI tools are helpful for finding the vocabulary needed to connect across different fields, unlocking entirely new possibilities for interdisciplinary teamwork and collaboration.

While experts debated whether AI outputs can represent true human-like creativity—referencing the burst of genius involved in Alexander Fleming's discovery of penicillin—there was relative consensus that the "combinatorial creativity" of AI is immensely valuable for identifying problems that can be solved by mixing and matching principles from two unlikely fields like economics and cell systems. Particularly in fields like quantum physics, where concepts can be difficult for human scientists to visualize and build into models, AI can serve as a conceptual conduit. Participants noted that AI, by providing such combinatorial solutions, is also helpful for parsing out which questions will require truly novel approaches.

The different approaches and disciplinary pairings encouraged by AI can help dismantle deeply entrenched theoretical biases. As one participant commented, the study of collective animal behavior has historically relied on concepts from nonlinear statistical mechanics, such as phase transitions. "Only recently, we've found that many of these fundamental features that we predict…[don't] happen in biological systems. And the papers where we thought it did happen, we've now realized we were wrong. A part of that is that we were so enamored by statistical mechanics, that we kind of saw what we were expecting to see. So I think that having AI can potentially help us become less canalized or less biased in our perspective, and also keep large amounts of data in sight," they shared.

Generating Hypotheses and Prioritizing Experimental Approaches

With the number of possible biological sequences far exceeding the capacity of wet labs to probe exhaustively, participants stated excitement that AI can be deployed up front to design sequence libraries that yield the maximum possible information per experiment. As one expert urged, "I think that we need to flip a lot of the paradigms that we've had—where we acquire a lot of data and then we evaluate AI algorithms based on how well they can guess held-out results—to partnering with some of our computational colleagues at the beginning to design the sequences that are going to be tested instead."

Other avenues for AI agents to help prioritize experimental approaches include providing protocols and cost-benefit analyses alongside the hypotheses they generate, essentially weighing in on how labor-intensive a set of experiments could be versus what a researcher could gain if the tests are conducted successfully and conclusively. Participants also noted that AI is promising for contingency planning; when one step of the experiment yields result X, the system can intervene to write the code for the next part of the model.

Despite the utility of these technologies for narrowing down vast search spaces, participants raised concerns about the development of fully automated researchers, instead stressing the importance of "human-in-the-loop approaches." In these methodologies, AI outputs are coupled with human judgment to evaluate the safety and tractability of experimental procedures. As another participant summarized, scientists should be "thinking about how to shape [AI] systems in a way that isn't just based around increasing knowledge, but is based around the beauty of the scientific endeavor that we all enjoy so much."

In Silico Experimentation

Regardless of how many physical measurements scientists collect, participants noted that the scale of their samples will always remain small compared to the possible sequence space they would like to investigate. For this reason, experts emphasized the role of AI models as "*in silico* oracles," allowing researchers to ask a range of questions computationally and then follow up *in vitro* where the results are surprising. "I think [this] is a nice partnership, where we're able to use the models to maximize our ability to learn, given our capabilities as experimentalists," one participant reflected.

AI-Oriented Experimental Design

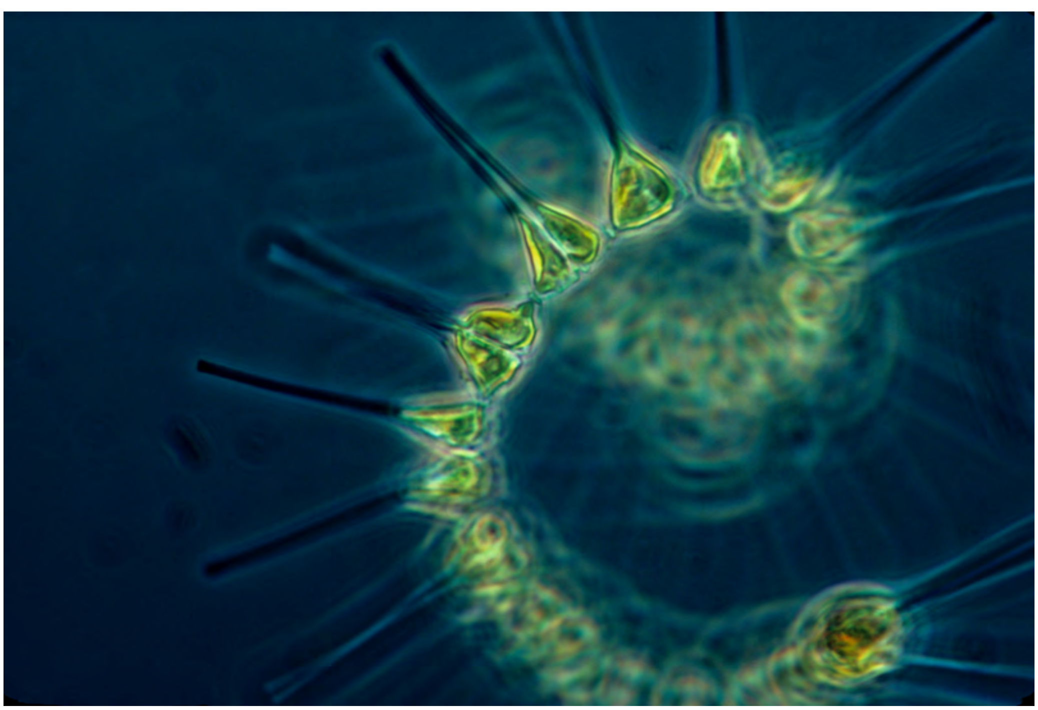
Roundtable participants drew a distinction between “AI-guided experimental design” and “AI-oriented experimental design,” expressing a clear preference for the latter. Whereas the first approach places the methodological onus on AI, the second approach integrates human expertise with AI-compatible workflows. To this end, one participant described using AI during the initial phase of data analysis before designing human-led experiments to test these hypotheses, ultimately feeding the results back into AI.

Alternatively, several experts shared experiences using robotic execution platforms like LILA AI to facilitate standardized and cost-effective experiments, allowing AI to perform longitudinal evaluations and dimensionality reduction. One participant highlighted: “Having automatic experiments allows you to open the door to all kinds of things. It’s not so much AI experimental design, but it’s really an experimental design allowing you to do AI down the road. And I think for me, that’s really crucial; that’s really a game changer.” In experiments that inherently have multiple versions, scientists can leverage these AI-ready data streams to benefit from techniques like reinforcement learning.

Hybrid Approaches

Throughout the convenings, participants from theoretical and experimental backgrounds alike expressed enthusiasm for hybrid modeling approaches. Such approaches combine prior mechanistic knowledge—such as conservation laws and physical constraints—with AI methods, enabling scientists to solve an aspect of a larger problem and refine their theories using computationally-learned quantities.

For example, participants explained that the construction of virtual cells is currently divided into two broad paradigms: purely predictive AI models lacking mechanistic insight, or manual recon-



NOAA / Unsplash

structions of every single interaction in the cell cycle. One expert posited that neither version is working well, instead advocating for the injection of more mechanistic insight into AI to predict actual cell behavior. They added that the most powerful development would be a complete cell simulation that allows for data generation and seamless integration with AI. By parameterizing coarse-grained models with experimental data, researchers can combine the strengths of statistical and mechanistic models.

One "gold standard" example discussed during the roundtable centered on the Pollard Lab at the University of California, San Francisco. This lab developed a machine learning architecture trained on high-throughput Hi-C genomic measurements to learn how primary DNA sequences dictate 3D chromosomal folding inside the cell. Using the model, the group was able to ask questions that would be impossible to test with natural data alone. The model enabled researchers to make sequence-level predictions about 3D genome folding that could be compared with experimental measurements. In doing so, the lab established a benchmark for closed-loop computational biology.

Distinguishing Across Models

Participants noted that a persistent challenge in biological modeling is non-identifiability. While a given model may generate accurate predictions, another model with slightly different parameters may produce the same results. Although AI systems alone are unlikely to provide true mechanistic understanding, these computational methods can help researchers to determine which models are indistinguishable using the available data and to design experiments that break this parameter degeneracy.

Dimensional Reduction

On a technical level, participants highlighted dimensionality reduction tools as useful for both extracting patterns from high-throughput data and for visualizing system dynamics. As one researcher observed, drawing on ideas from these tools allows them to better "structure [their] scientific thinking" in a human way. They noted that these architectures are particularly effective for processing nonlinear patterns across complex biological spaces.

D. High-Value Applications for Research and Decision-Making

Leveraging the Interface Between Food, Nutrition, and Microbes

Participants concurred that the highest-value applications of biological inference will involve understanding and managing how organisms adapt and interact with one another, whether host-microbe interactions or human immune responses to vaccines. Specifically, experts zeroed in on the interface between food, nutrition, and microbiomes as yielding the most societal benefit. If leveraged appropriately, systems knowledge can be used to produce food in ways that are sustainable, cost-effective, palatable, and nutrient-dense. These objectives have already been linked to the U.N. Sustainable Development Goals, with scientists worldwide issuing "a microbial food call-to-action."

Using AI Insights to Study Biology

Tools like AlphaFold are useful for prediction but can also be studied in their own right as complex systems that display emergent phenomena. One researcher described moving away from applying AI directly to their biological systems work, finding that concepts underlying AI were more useful than the software itself for understanding how biological function emerges. For example, by thinking of nature as tuning "parameters" across genomes and amino acid sequences, they were able to better conceptualize the many-to-many mappings through which different parameter combinations yield the same biological solution, or through which identical parameters yield different outcomes. Other participants agreed with this perspective, adding that concepts from AI such as optimization landscapes and learning dynamics are also helpful in informing biological investigation.

Addressing Brain Health, Collective Decision-Making, and Neuromorphic Computing

Participants also identified the brain as a critical frontier, placing it as the most complex biological system that researchers study. Within the space of neurologic disease, for example, AI could help scientists to conduct normal-versus-disease comparisons using inputs like blood samples and to analyze the progression of disease over time.

More broadly, AI could also help researchers study decision-making, cognition, and higher-order processes such as theory of mind, consciousness, and self-awareness. While AI facilitates increasingly high-throughput connectomics, scientists must still evaluate whether they possess the right theoretical concepts to interpret the resulting neural maps. Ultimately, these insights could inform collective decision-making. As participants observed, the major crises facing humanity—including war, disease, hunger, and climate change—have known solutions but require collective action to make true progress. In other words, the primary bottleneck is not a lack of scientific knowledge but the inability to incentivize human behavior change and collective decision-making.

Finally, studying the brain and the underlying mechanisms that govern cellular competition at scale could also inform novel computational architectures, such as task-specific neuromorphic computing.

Synthetic Biology and Protein Function

In thinking about the future of biology, participants converged around the long-term goal of accurately programming living systems through "synthetic biology." While current models can design shape to some extent, scientists still struggle to program protein function. Experts qualified this goal, adding that achieving such control will require substantially more data than was used to predict protein structure. Specifically, researchers will need to characterize which units a given protein binds to—and which units it does *not* bind to—across a range of environments and contexts.

Genome-Scale Metabolic Modeling

AI could make a transformative difference in genome-scale metabolic modeling, wherein researchers reconstruct an organism's complete network of metabolic reactions from its genome. Currently, this process tends to be performed through manual, step-by-step reconstructions that incur significant information loss at each stage. Participants acknowledged that a combination of AI models and machine learning could offer a far more comprehensive, automated pathway to reconstruct metabolic networks.

Scale, Evolution, and Emergent Behavior

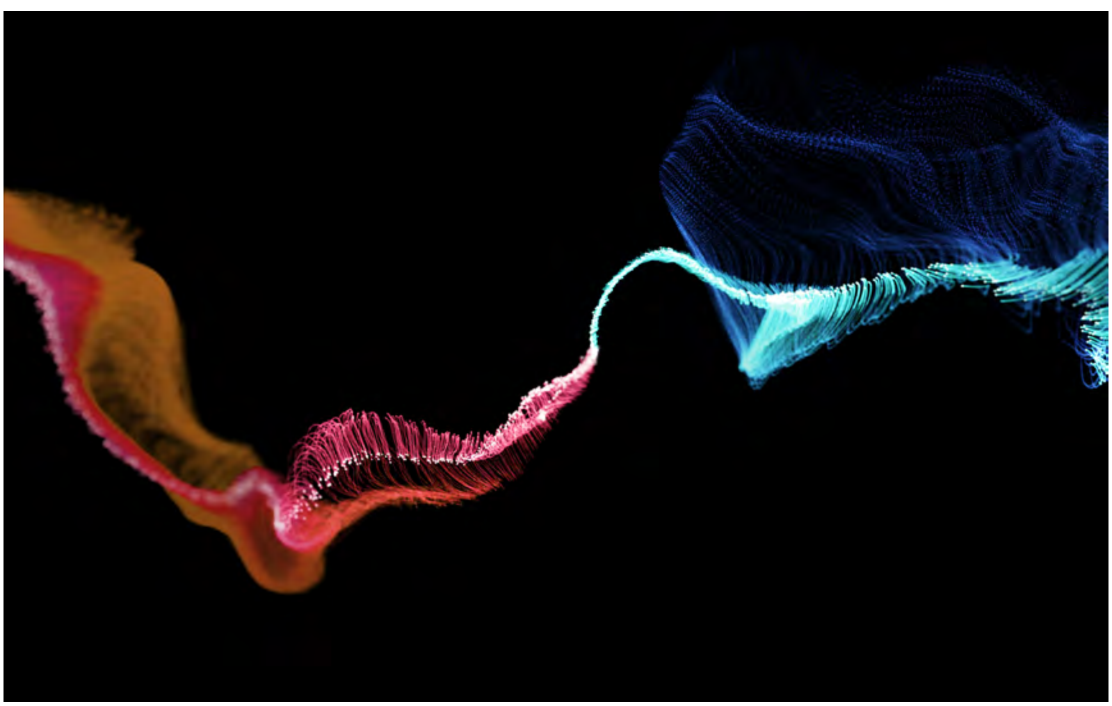
Roundtable experts agreed that complex biological systems are defined by cross-scale phenomena spanning molecular, cellular, organismal, and population levels, coupled with dynamic interaction, adaptation, and timescale differences. As a result, it can be difficult for researchers to understand what is the substrate on which they are or should be learning. Moreover, even phenomena occurring over very short timescales can drive impacts on an evolutionary scale. Moving forward, there is untapped potential for AI to assist in understanding how microscopic details yield emergent collective functions, as well as in working backward to decipher the origin of life and how the current chemistry of metabolism emerged.

Participants drew out cancer as one of the most clinically important multi-scale applications because it features collective cell behavior, spatial competition, and evolutionary processes.

Designing Adaptive Systems

In addition to prediction, a vital objective is developing optimal control frameworks for managing dynamic biological systems. Drawing from game theory, researchers represented at the roundtable seek to model the range of strategies that individual phenotypes might deploy under stress.

At a global level, participants noted that predicting the next pandemic is impossible. As one expert recommended, scientists and decision-makers should instead draw on biology as the inspiration for adaptive systems design. They explained, "I think of the vertebrate immune system as having evolved to deal with the fact that the next pandemic is [going to] come from is an impossible prob-



Richard Horvath / Unsplash

lem, and therefore you need a system which is hierarchical and designed to deal with the fact that you're certain there are going to be new pathogens attacking your body, but there's no way to know what they're going to be, and so you need an adaptive system. So, how can we develop an adaptive system that can respond to it? That's a place where AI, I could imagine, could be very useful."

E. Persistent Challenges, Limitations, and Risks

Interpretable AI

Multiple roundtable experts underscored the need to move away from "black box" results and towards interpretable AI outputs. Participants identified two broad ways to think about interpretable AI: first, as dimensionality reduction to extract the "equations of motion" of a complex system in a reduced space; and second, as probabilistic or mathematical models that combine physical, mechanistic, and biological constraints alongside deep learning modules. They noted that machine learning tools frequently perform data compression without maintaining mass balance or physical properties and also described "interpretable AI" as an ambiguous term whose meaning depends on the context. In their view, the ability to generate an explanation matters less than the risk of over-explaining noise.

Despite the relative flexibility of current computational packages, experts noted that current limitations still constrain the ability to incorporate this second definition of interpretable AI. "I think that's a difficulty we're all struggling with, which is that all these packages [like PyTorch] allow us to do very good, efficient work, but at the cost of not being able to actually put the biological or physical insight that we may want to put in these models. There's probably exciting opportunities in writing a more general computational framework that will allow us to include these more mechanistic aspects directly into the model, along with the deep part," one researcher shared.

Although AI tools do not automatically lend themselves to mechanistic insights, other participants added that it is possible to extract such insights from models not originally built with physics in mind. To better understand underlying mechanisms, researchers can systematically compare deep learning predictions against physical models of increasing complexity. As one participant explained: "Once you have the ability to compare the amount of variance in a dataset that's explained by a physical model and then explained by a deep learning model, [that comparison] gives you a sense of what is being explained by the deep learning model that wasn't captured by the physics-based model, which can then point to new physics, even though the model itself wasn't built to ask a physical question in the first place." Participants also highlighted the use of forward simulations that feature a generative process combined with AI as a productive approach to learning underlying mechanisms and supporting interpretability.

Over-Interpretation and Reproducibility

According to participants, interpretation in the life sciences is often shaped by cultural tendencies as much as by the underlying evidence. Authors and peer reviewers may be drawn to striking computational visualizations, increasing the risk of overinterpretation. The growing volume of AI-assisted manuscripts also risks outpacing the capacity of peer review, weakening an important quality-control mechanism. AI systems, participants noted, frequently overlook online supplementary materials, where critical experimental details are often reported.

Additionally, replication is not always feasible in highly specialized fields, where few independent groups have the expertise or resources to reproduce complex experiments. Some long-term predictions—such as computational models forecasting Alzheimer's disease progression over decades—are also difficult to validate in the moment. As a result, scientific progress depends in part on trust within the research community. One participant argued that, although individual findings cannot always be independently verified, the durability of broader scientific principles and the success of subsequent generations of research provide a meaningful measure of progress. AI could also strengthen reproducibility by enabling more systematic efforts to evaluate and replicate published findings.

Participants suggested community norms that could help sustain scientific integrity, such as systematically annotating which claims have been validated by follow-up studies. Additionally, assigning confidence metrics based on the quality and type of experimental validation could improve interpretation.

Experts went on to link over-interpretation to the field's replication crisis, arguing that AI workflows can introduce hidden assumptions and unrecorded hyperparameters that compound existing problems with reproducibility. Current platforms fail to provide clear measures of certainty associated with observations, nor do most AI methods communicate how many hypotheses were tested behind the scenes. Participants further warned that machine learning can reinforce confirmation bias, pointing to single-cell RNA sequencing workflows that use UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction. Although researchers can construct biological narratives based on visual UMAP clusters, this "understanding" is "very likely to be misleading and potentially dangerous in the long run" because it is "fundamentally different from a physical understanding of the biological processes."

Binary Approaches to Pathogenicity and Immunology

Participants expressed concern that algorithms like AlphaMissense over-predict pathogenicity due to an over-reliance on binary processes. One expert explained: "The reality is that most proteins have multiple functions, and disruption of one of those functions can have a very different disease outcome. So if we had a better mechanistic understanding of not just whether something's bad, but what aspect of function is disrupted, that gives us an ability to better predict disease progression, as well as devote our efforts towards developing therapies that are more personalized to restore the precise function that was disrupted."

Some participants cautioned that in immunology, defining what needs to be predicted remains difficult because binary categories like "safe/unsafe" or "immunogenic/nonimmunogenic" vary continuously across environmental contexts and human lifespans.

Loss of Theoretical Focus

Roundtable discussants expressed concern that, with the proliferation of generative AI assistants, academic journals are being inundated with purely theoretical papers that offer ideas while lacking long-term dialogue with physical experimentation. At the same time, drawing from physics traditions, other participants countered there is value in publishing standalone theory papers without immediate testing, allowing validation to follow as a separate effort. Theory-driven models remain essential because the space of biological possibilities is far too vast for data fitting alone. One participant noted: "Where AI, I think, could really come in is telling us what are the atomic units of such theories, what is the basis we're going to live in, and how that basis connects to data."

Risk to Scientific Training and Exchange

Participants observed that AI has supercharged the work of researchers with strong domain-specific scientific training but weaker coding skills. However, they worried that rapid adoption of AI could threaten the traditional scientific training pipeline and result in the next generation of scientists losing foundational knowledge regarding the ins and outs of dynamical systems. These concerns culminated in a core question: "How do we train people to have that sort of grounded knowledge in an age where it's easier to have surface knowledge?" For instance, participants suggested that computational biologists should be trained as active developers who understand the internal mechanics of their tools, rather than as passive users of black box technologies.

Further, roundtable attendees warned that reliance on AI assistants risks siloing researchers from human exchange, causing them to lose the habit of seeking insights—and pushback—from colleagues in their own and neighboring departments or at academic conferences. Since human cogni-

Yoksel Zok / Unsplash

tion is high-dimensional and varied, replacing human discourse with text-constrained AI tools risks creating intellectual echo chambers.

Participants identified three central strengths of human-led scientific reasoning:

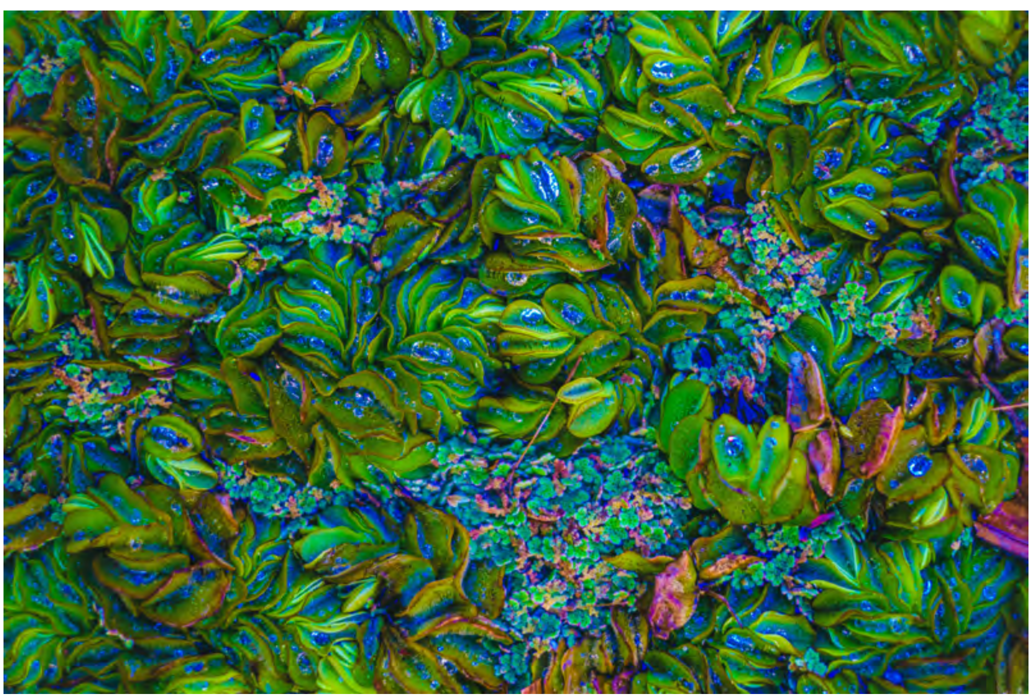
1. Critique of evidence and assumptions, bearing in mind the weaknesses and history of one's field;
2. Unique capacity for creativity; and
3. Social, political, and cultural perspectives.

One expert concluded: "I don't want to lose awareness of that and hand it over to AI, a computational tool. I want us to be agents over the relationship between our scientific thinking and our social milieu."

Limited System Ensembles and Constraints on Observed Variation

Participants argued that understanding of biological function is severely constrained by data availability and by constraints on natural phenotypic variation—such as developmental boundaries in embryos or microenvironmental degrees of freedom that are worth studying in their own right. One expert illustrated how this limitation pertains to their research, explaining that "we know from statistical physics that when you want to understand some collective behavior, you really need ensembles. Ensembles of systems with the same function, and we don't have that. We're restricted in a lot of ways to ensembles that nature produces [like] families of proteins. But those are very, very small compared to what we actually need to interpret something that's really quite noisy and complex."

Another researcher noted that overcoming the limited sampling space provided by evolution is the very argument for more comprehensive data collection. "I think that it is actually quite possible to make huge strides with high-throughput experiments that are quantitative, conducted in different



Soheb Zaidi / Unsplash

backgrounds—backgrounds that are as diverse as possible—[complemented] with analysis of the diversity of protein sequences that are out there. [We want] perturbation of all those wild types with multiple experiments. Particularly if they're quantitative [the data] can give us the decomposition into causal factors that we're looking for," they explained.

Privacy and Biosecurity Concerns

Deploying AI to facilitate a global census of all eight billion people, alongside a comprehensive survey of ecological networks, would provide unprecedented insights into health across scales. Participants warned, however, that both individuals and larger institutions should not be trusted with this level of private data. Even if anonymized, corporations excel at reconstructing identities, while academic teams lack the computational resources to provide an effective check on corporate data monopolies.

Researchers also highlighted severe biosecurity risks associated with generative AI models. For example, while commercial systems initially refused one participant's queries regarding smallpox synthesis, they were later able to train the model to yield detailed instructions for producing mouse-pox. The same participant noted that these safety regulations involve tradeoffs. While regulations are necessary to prevent harmful compounds from entering the world, restricting model outputs also prevents students from accessing detailed data as they learn.

F. Recommendations for Funders

Participants strongly agreed that funding for AI in biological systems work should be "orthogonal to the current vector of investment," which is heavily shaped by short-term market incentives such as pharmaceutical development. Grantmakers should focus on "putting the money towards efforts that are focused on the solution and the outcome rather than building from the foundation up," one researcher noted. Others added that as academic funding has shifted toward pure AI, teams working on theoretical and hybrid modeling approaches have found it increasingly difficult to secure support.

1. **Support a Diverse Portfolio of Open Research Problems:** Rather than honing in on one big question, funders should allow for varied experimentation. Across the board, participants expressed agreement with one attendee's framing: "There are points in fields where thinking about grand challenges is useful, and there are points where it's not. And I think this is a point where it's not. We are at a stage where all these problems at the intersection of AI and biology and the development of new methods in AI [are unfolding] so rapidly. There are so many open frontiers, and I feel the appropriate way for people to work is to work on lots and lots of diverse problems in lots of different ways, and start to see where the roadblocks appear. Right now, we're at a point where there's open road in all kinds of different directions but we know there are going to be challenges that emerge as we start to move along these roads in different ways."

2. **Enable the Transition from Prediction to Programmability:** Philanthropic capital should move in the direction of functionally programming biological systems, not just making predic-

tions. As one participant envisioned, the goal should be answering: "Could we make proteins in a single shot that have the function that we want? I think that that would have transformative applications from developing new therapeutics to creating enzymes that we could use across industry and environmental remediation."

3. **Invest in Applying the Concepts Underlying AI to Biology:** Funders should uplift research that draws on the concepts underlying AI—such as optimization dynamics, parameter tuning, and learning architectures—to reframe fundamental biological problems. Researchers from computational and biological backgrounds should also be encouraged to explore what AI can learn from biology, particularly drawing from neuroscience and evolutionary learning. "I really do think that, you know, investing more in applying concepts underlying AI to biology might be a very good use of funds for a foundation, rather than just doing what everybody else is doing, which is 'Let's use AI to solve all our problems,'" expressed one expert.

4. **Build High-Quality, AI-Ready Databases of Biological Dynamics:** Participants noted that data acquisition requires a high level of creativity, referencing the idea to use ice cores as evidence in climate change modeling. Building on the success of the Protein Data Bank in enabling structural prediction, funders should invest in the creation of centralized, AI-ready databases that capture protein dynamics, kinetic constants, catalysis rates, and thermodynamic properties. Such data is expected to enable key immunological insights, such as predicting antibody-antigen binding or T-cell receptor-antigen binding.

   - *Clinical & Longitudinal Data:* In clinical medicine, prioritize high-resolution longitudinal samples—specifically bodily fluids, single-cell profiles, and transcriptomics—with built-in quality control procedures.

   - *Metadata Curation:* For microbiome and field sequencing, invest in data curation infrastructure to address the pervasive issue of poor-quality metadata in public repositories, ensuring samples are optimized for machine learning.

5. **Target Funding for Domain-Specific Foundation Models:** Funders should support open-access, domain-specific foundation models tailored to distinct categories of biological systems—such as protein dynamics, genome sequences, and microbial communities—that integrate mechanistic rules to guide downstream experimentation. While data is limited for specific sub-tasks, aggregating standardized data across entire biological domains reveals vast amounts of underutilized data ready for foundation model pre-training.

G. Conclusion: Toward an Explanatory Paradigm in the Life Sciences

The integration of artificial intelligence into biological research marks a shift in how researchers study, model, and manipulate living systems. Machine learning can compress high-dimensional data, navigate vast sequence spaces, and automate complex experimental loops. Yet, predictive accuracy alone does not constitute a scientific explanation.

Participants argued that biological discovery will depend on sustained interaction among computational models, biophysical principles, and wet-lab experimentation. Achieving this vision will require embedding physical mechanisms in neural architectures, generating quantitative functional datasets, preserving human judgment and critical reasoning, and funding a diverse range of research problems. Together, these approaches can help ensure that advances in computational power contribute to a deeper understanding of biological systems.

Timothy Dykes / Unsplash

4. Neural Systems, Cognition, & Human Behavior

A. Participants

We aim to synthesize and share perspectives from the discussions as a whole rather than to attribute any quotations or viewpoints to specific individuals. Participants with whom we engaged in roundtable discussions are listed below (alphabetically by last name):

- **Anton Arkhipov, Ph.D.** – Investigator and Lead of Bio-Realistic Modeling Team, *Allen Institute*
- **David Freedman, Ph.D.** – Professor and Chair of Neurobiology, *The University of Chicago*
- **Kris Ganjam** – Director of Generative AI, *Allen Institute*
- **Mehrdad Jazayeri, Ph.D.** – Professor and Director of Education, Brain and Cognitive Sciences, *Massachusetts Institute of Technology*; Investigator, *Howard Hughes Medical Institute*
- **Nancy Kanwisher, Ph.D.** – Professor of Cognitive Neuroscience, Brain and Cognitive Sciences, *Massachusetts Institute of Technology*
- **Saul Kato, Ph.D.** – Associate Professor of Neurology, *University of California, San Francisco*
- **Nikolaus Kriegeskorte, Ph.D.** – Professor of Psychology and Neuroscience and Director of Cognitive Imaging, *Columbia University*
- **Eve Marder, Ph.D.** – University Professor of Neuroscience, *Brandeis University*; Recipient, *National Medal of Science*
- **Russell Poldrack, Ph.D.** – Professor of Psychology and Associate Director of Stanford Data Science, *Stanford University*
- **Anil Seth, Ph.D.** – Professor of Cognitive and Computational Neuroscience and Director of the Sussex Centre for Consciousness Science, *University of Sussex*
- **Michael Shadlen, M.D., Ph.D.** – Professor of Neuroscience, *Columbia University*; Investigator, *Howard Hughes Medical Institute*
- **Andreas Tolias, Ph.D.** – Professor of Ophthalmology, *Stanford University*
- **Anthony Zador, M.D., Ph.D.** – Professor of Biology and Chair of Neuroscience, *Cold Spring Harbor Laboratory*

B. Introduction

Since the emergence of [cybernetics] in the mid-20th century, scholars have actively debated the relationships between engineered computational architectures and biological cognitive systems. Historically, this knowledge transfer largely operated as a one-way path, with computer scientists drawing inspiration from biological neural networks to design artificial algorithms. However, roundtable experts identified that this dynamic has shifted markedly in recent years, with researchers of cognition and behavior now asking how engineered AI models can inform humanity's fundamental understanding of biological brains.

While a two-way dialogue holds deep promise for both fields, participants shared that methodological and disciplinary divides continue to hold back progress. "These two revolutions [the modern neural networks revolution and the AI revolution]—while they've granted all our wishes in a way—have disrupted our science, and we don't have a mature methodology for using these wonderful new tools and connecting the bigger, better data to the bigger, better models that we now have," summarized one participant.

This chapter seeks to map constructive pathways for "NeuroAI," the emerging field at the intersection of neuroscience and artificial intelligence. Specifically, this chapter focuses on how AI can help model complex neural and cognitive processes that give rise to observable, naturalistic behavior; where current computational and biological models remain limited when capturing the multi-scale dynamics of the mind; how insights from biological brains can inform the design of interpretable, resilient, and human-aligned intelligent systems; and what methodologies are needed to bridge the cultural and technical divides separating computational engineers from experimental neuroscientists.

At the center of this exploration is a tension between the predictive utility of simplified models and the divergences they display from the biological reality of brains. Although modern artificial neural networks excel at processing high-dimensional informational streams, participants cautioned that treating models as direct proxies for living brains poses serious epistemic risks.

<u>Core Takeaways</u>

- **The two-way street of "NeuroAI":** While historical AI development drew heavily on abstracted neural concepts, modern convergence allows brain science and machine learning to explore shared algorithmic spaces using unified software tools.
- **AI as a "test zero" in cognitive science:** AI models can serve as a computational "test zero" for cognitive theories—demonstrating that a proposed theoretical framework can successfully perform a complex perception or motor task before evaluating its biological fit.
- **Comparative transparency and model organisms:** Neural networks provide a uniquely transparent and manipulable testbed, serving as a first step for refining theories and experimental designs that can be later applied to biological brains.
- **Bridging the space between cure and understanding:** In clinical brain health and psychiatric disorders, AI may accelerate diagnostic prediction and therapeutic interventions before researchers achieve complete mechanistic understanding, buying vital time for deeper scientific inquiry.
- **Acknowledging the limits of computer-brain analogies:** Researchers should guard against conflation between computational models and biological reality. Relatedly, scientists have a responsibility to use AI to amplify human intuition and cross-disciplinary reasoning rather than outsource critical thinking.

C. Areas of Excitement and Progress

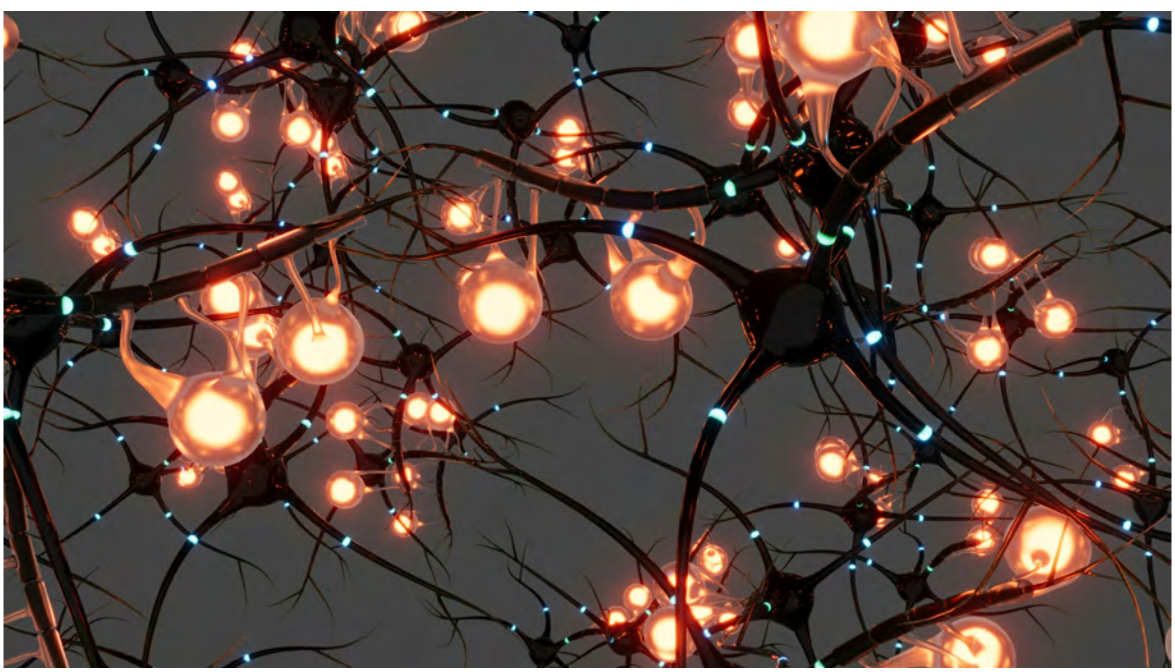
"NeuroAI" and the Convergence of Disciplines

Participants described excitement over "NeuroAI" as an emerging field sustained by a positive conceptual and experimental feedback loop between neuroscience and artificial intelligence. Reflecting on this history, one expert observed, "The history of AI is the history of taking ideas from neuroscience, squinting, abstracting, and then applying them to AI. And then, to some extent, we take those distillations, and in some cases, they reflect back to neuroscience." They added that "what made [this initial cross-pollination possible] was the fact that the same people who were making progress in what at that time was called computational neuroscience were trying to build better neural models"—referencing foundational figures such as Geoffrey Hinton and Yann LeCun.

As NeuroAI continues to gain traction, participants emphasized that the field cannot rely solely on rare individuals who happen to straddle both disciplines; rather, progress will require investment in shared institutional spaces alongside shared tools. While some contemporary engineering perspectives suggest that raw computational scale reduces the need for biological inspiration—drawing on the analogy that engineers today can design aircraft without necessarily studying bird flight—participants argued for a continued grounding in neuroscience and for collaborative infrastructures that promote two-way knowledge transfers. "It's not that we take the algorithm from AI and it turns out to be a great theory, but it's the convergence of tools that makes it possible to explore those together across engineering and brain science," one researcher commented.

New Modeling Possibilities

This convergence is creating new opportunities to construct computational models of cognition. Roundtable participants expressed considerable enthusiasm for computational architectures capable of simulating and testing theories of human cognition at scale. "If you have a theory about how a cognitive process works, it's not a real viable theory if you can't implement it in a computer simula-



Bhautik Patel / Unsplash

tion and [have] it actually perform that cognitive process, right? So, theories of cognition have to be AI models, almost by definition, that [is] sort of Test Zero," one attendee posed. As another participant articulated, "the use of AI models to explore fundamental questions of how you build brains, and what you need to build in architecturally, and what kind of priors, and what you can get with almost no priors and just lots of experience [is] cracking wide open huge questions—the most fundamental questions in cognition and neuroscience."

Looking further ahead, AI may eventually enable accurate, full-scale emulations of the human brain at the resolution of individual neurons and synapses. Participants noted that the process of developing such a high-resolution model is valuable in itself, as constructing a computational model forces researchers to re-examine existing theoretical assumptions and identify entirely new data requirements.

Finally, researchers expressed optimism that AI will not only facilitate neural network models of brain computation, but also the modelling of other components like the mind and cognition in naturalistic environments. The ability to train interpretable models of complex behaviors—ranging from interpreting sensory stimuli and deliberating over decision spaces to generating precise motor responses and executing task sets—provides unique advantages over animal experimentation alone.

Comparative Transparency and Hybrid Frameworks

Although deep learning models are often characterized as "black boxes," participants noted a comparative transparency when evaluated against biological animal brains. Artificial networks can serve as fully observable "computational model organisms," allowing researchers to record, measure, and manipulate individual units, weights, and activation vectors across the entire network. This level of visibility provides a powerful proving ground for mechanistic interpretability. In one participant's framing, these model organisms provide a "testbed," allowing researchers to ask "What are the strategies that work to actually understand an organism that's not even as complicated as us, but that can do really interesting stuff?" before moving forward experimentally. Notably, *in silico* testing is not a replacement for live biology. Rather, validating strategies in AI models allows scientists to de-risk hypotheses and design more targeted, high-value experiments when returning to the wet lab.

To ground these models in biological reality, experts advocated for hybrid approaches. One scientist suggested: "The way to do it is, yes, to have the machine learning component there. All the biological parameters for which you have some measurements [can also be used] to constrain the values of those parameters as a part of the optimization process. That way, [you're] building models that both reflect the data, the structure, the architecture of the circuit that you're trying to learn something about but [also using] modern machine learning tools to get some function out of it."

Accelerating Big Data Analysis and the Practice of Science

Realizing these modeling ambitions depends on equally dramatic advances in data collection and integration. The growing collection of neuroscientific data, including data gathered under naturalistic conditions, is yielding rich, multi-scale datasets that range from single-cell transcriptomics to continuous behavioral recordings. Machine learning provides powerful methods for integrating and analyzing these high-dimensional data streams, enabling scientists to generate hypotheses and test them in biological systems.

AI is also reshaping the day-to-day practice of science. Beyond data processing, researchers described using AI to coordinate collaborations, streamline code, mine scientific literature, and "red team" manuscripts. As one participant shared, "this whole ability for the AI to help coordinate the scientific enterprise is really what excites me, by really being that glue for the collective intelligence of how we input ideas and make connections."

Experimental Design and Piloting

Another contribution of AI lies in helping researchers decide what experiments to perform next. New research and development (R&D) platforms—such as those developed by the nonprofit FutureHouse and now commercialized by its spinoff Edison Scientific—demonstrate how machine learning is reshaping experimental workflows. By integrating AI from the beginning, researchers can concentrate on the most promising experiments and adapt their process as the experiment progresses to optimize over an objective function (i.e., minimizing uncertainty about answers to specific questions).

As an example, one researcher specializing in brain imaging shared that their lab now routinely pilots functional magnetic resonance imaging (fMRI) experiments on AI models, with these pilots almost always aligning with observed reality. While the principal investigator is firm about continuing to collect live brain data, this initial computational phase allows the team to use time and other resources more efficiently.

Together, participants viewed these developments as shifting AI from a tool for analysis toward a platform for scientific discovery.

D. High-Value Research Applications

Participants identified several research domains where these emerging capabilities could have the greatest impact.

Subcellular Mechanics and Dendritic Computation

Participants advised that to advance beyond high-level approximations, the field of neuroscience must make a concerted effort to model fine-grained subcellular processing, particularly within dendrites. They agreed that modeling a single synapse alongside its surrounding dendritic architecture represents a critical, missing building block in current computational frameworks. Progress will require methods that combine real-time holographic imaging with electrical recording and the simultaneous perturbation of specific somas and subcellular components.

Enabling Data Integration

Researchers pointed to a persistent data integration problem within neuroscience, noting that these challenges exist across animal models, levels of explanation, and data modalities. As one expert detailed, “One of the things that strikes me about neuroscience is that a lot of people collect data, but a lot of it is still relatively individualized, not systematized very well. It’s very hard to share data between labs, with some exceptions” like UK Biobank. Given this bottleneck, participants viewed AI as a tool for bringing data together into searchable registries. A key prerequisite in building such registries will be identifying the minimal metadata standards necessary to enable cross-laboratory data integration and ensuring that scientists adhere to these standards during all future research.

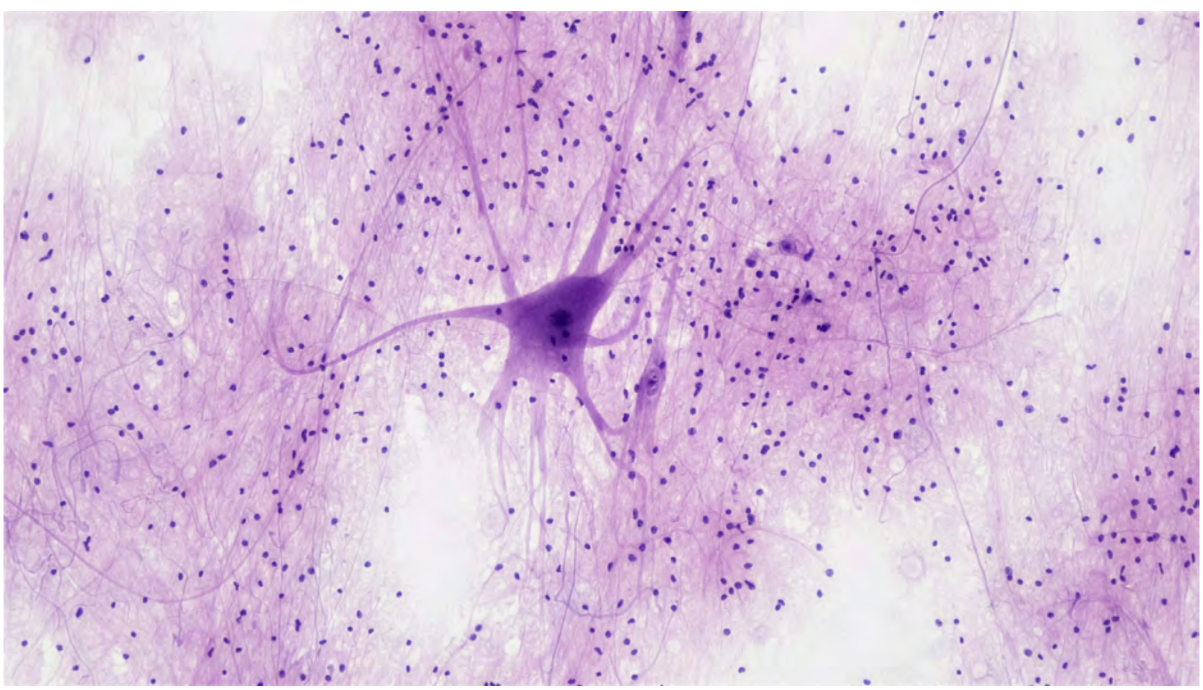
Defining Minimally Viable and Generally Intelligent Systems

Constructing AI models enables scientists to evaluate how much biological detail must be explicitly programmed into a system versus how much structure emerges spontaneously. One participant noted that this approach reflects British neurobiologist David Marr’s foundational formulation: the nature of the computations that underlie perception depends more on the problem being solved than on the specific hardware implementation. The failure of a specific artificial model to replicate biological behavior, then, can highlight domains where biological architecture matters most.

While some participants cautioned that AI systems have the advantage of being trained on Internet-scale data, others cited studies in which models trained on human-scale linguistic inputs were still able to produce intelligible language. These data-limited settings reflect common benchmarks of artificial general intelligence (AGI), such as the framework proposed by DeepMind co-founder Demis Hassabis. According to Hassabis, AGI should be capable of generating transformative scientific theories, akin to those of Newton or Einstein, using only the literature published before their actual discoveries.

Brain Repair and Stability

A key biological mystery identified by participants centers on how the brain maintains operational stability and long-term memory representations across decades, despite ongoing molecular turnover,



Bioscience Image Library by Fayette Reynolds / Unsplash synaptic remodeling, and processes of cellular repair. Machine learning models may help decipher the dynamic principles governing functionality by providing a window into small-scale evolution.

Discovering Real-World Perceptual Constraints

AI architectures can also reveal which features of biological intelligence arise from constraints imposed by the physical world. One participant called attention to a study in which investigators trained an artificial neural network on realistic auditory input, yielding an architecture that spontaneously replicated numerous human psychophysical phenomena in sound localization. However, when trained on sound inputs created in a synthetic environment without echoes—these being one of the primary factors complicating natural sound localization—the model failed to produce any of those human-like psychophysical phenomena. As the authors of the paper note, their results demonstrate "how artificial neural networks can reveal the real-world constraints that shape perception."

Participants noted that these insights relate to Moravec's Paradox, which posits that tasks that are easy for humans are difficult for AI agents and vice versa. Uncovering which computational problems are intrinsically complex versus straightforward can reframe inquiries regarding machine consciousness. For instance, attendees observed that many tasks that current AI systems struggle to perform overlap directly with functions associated with conscious thought in human beings, such as confidence judgments, rapid one-shot learning, and out-of-distribution generalization. Therefore, studying these computational limitations allows researchers to clarify the specific functional utilities of consciousness in biological organisms. Interaction with the physical world, autonomous navigation, energy-efficient operation, and alignment with real-world physical environments were identified by researchers as among the most challenging hurdles in machine learning. These challenges translate directly to the field of robotics, where the current generation of machines struggles less with motor skills and more with building an internal representation of physical reality that captures three-dimensional visual geometry, mass, friction, and object dynamics.

Advanced Behavioral Annotation and Novel Computing Architectures

Building on advances in perception, participants envisioned using AI to help infer cognition. While tools like DeepLabCut have greatly contributed to physical pose tracking and external behavioral observation, current technologies remain limited to recording observable mechanics. Participants noted that a promising next step lies in expanding these frameworks to construct AI annotation models capable of decoding complex, internal cognitive states—such as holding information in working memory or executing a real-time change of mind. Deploying machine learning architectures within interactive and recordable environments like video game tasks presents a powerful opportunity to enrich existing technologies.

Attendees also saw opportunities to combine neuroscience and computer science to design new brain-inspired computing architectures that feature greater energy efficiency, processing power, and operational flexibility.

Clinical Translation

Beyond basic neuroscience research, participants described important opportunities for clinical translation. In clinical health, psychiatric and neurodegenerative disorders present complex challenges because they can be polygenic, heterogeneous, and deeply modulated by environmental variables. "We don't have the data to understand how…these emergent, dysfunctional states come from both the background of several mutations working together, several genetic variants, plus something else coming from the environment or the experience of the organism," one participant explained.

As data collection and integration proceed, machine learning may accelerate clinical treatments for these conditions, likely in advance of complete mechanistic understanding. One researcher reflected: "Traditionally, what I think we mean by understanding something is having a model in our heads that is compact enough that we can manipulate that model. It's a sort of cartoon understanding of the world. And traditionally, to make progress in science, we need understanding. I think there's this weird decoupling that we're seeing—that we're still sort of struggling with—which is [that] we might be able to cure things, fix things, make progress, make predictions without actually understanding how the system is making the prediction."

While a majority of participants admitted that this prediction-based approach is not ideal, others added the caveat that clinical interventions informed by these approaches could deliver near-term benefits while researchers pursue deeper mechanistic understanding. For instance, pattern recognition and prediction may allow scientists to identify the most effective techniques for cellular repair and neurogenesis. Researchers expressed optimism that integrating AI with transcriptomics and synthetic biology will allow them to screen new therapies faster, push more control into cells, and perform safer perturbation experiments.

Neurobiological Alignment

In addition to these immediate applications, participants highlighted several scientific frontiers that remain unresolved. Neurobiological alignment emerged as one promising direction for AI research. One expert argued that current AI systems primarily imitate the outputs of human intelligence, learning from artifacts such as text and video, rather than the brain's underlying mechanisms. They proposed that "a grand challenge would be to create an AI system that is not just aligned at the behavioral level, but at the neurobiological level," then see if and how it performs differently.

As an early step in that direction, researchers have compared object recognition behavior of deep artificial neural networks with both the monkey and human visual hierarchies. Although scientists have found some level of explanatory power, substantial variance remains unexplained. Participants suggested that incorporating recurrent connections, stochasticity, and canonical microcircuits could improve the models' biological fidelity.

Learning From Limited Resources

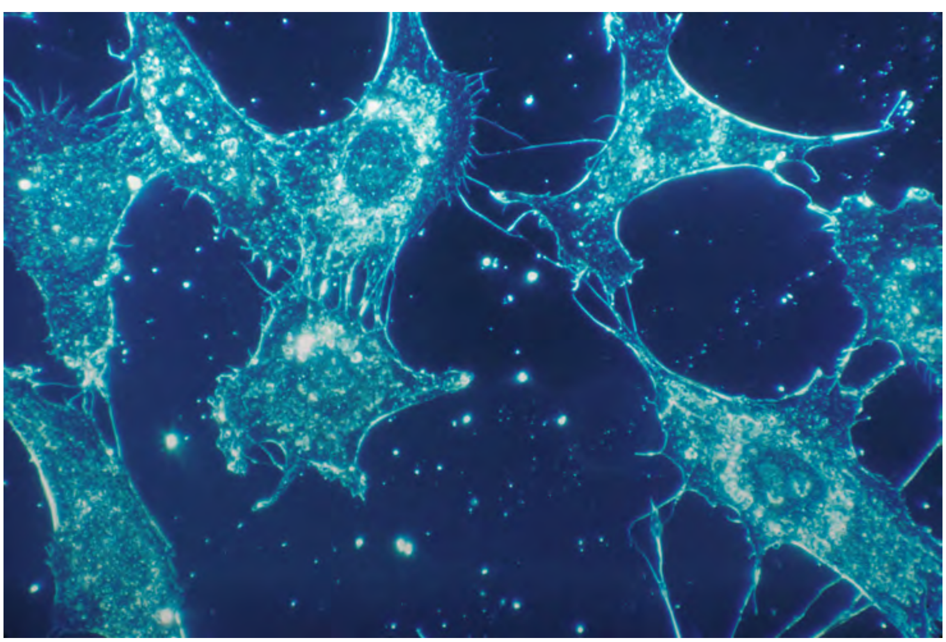
Biological intelligence operates with remarkably limited computational resources and energy. This efficiency contrasts sharply with current large language models, which have parameter counts and training datasets that exceed the sensory data a human acquires throughout development. How humans achieve comparable capabilities from far less experience remains an open scientific question. As one participant noted, "We can get away with far fewer examples in training, and probably doing something more sophisticated than brute-force reinforcement learning, which has been one of the major ways that we've been training the engineered networks." Participants suggested that part of the explanation lies in innate inductive biases shaped by evolution. However, because genetic code carries limited storage capacity, those biases are thought to be highly abstract.

Understanding these biases is also tied to the open question of how memory and learning—especially one-shot learning—operate at the cellular circuit level. For example, although biological brains do not appear to use standard backpropagation, recent empirical work suggests hints of backpropagation-like mechanisms in brain circuits. Participants also noted a novel finding indicating that multiple distinct learning rules can coexist.

Complementing the direct applications outlined above, participants identified **several fundamental questions in the field** that remain unanswered or underexplored:

- The mechanisms underlying mental time travel;
- Defining the computational or biological nature of a thought;
- Decoding the exact principles governing perception;
- Evaluating whether biological brains perform computation in a formal sense;
- Identifying the wiring rules that allow a biological system to develop intelligence; and
- Understanding how AI agents evaluate the capabilities of other AI systems in multi-agent reinforcement learning environments to develop social intelligence.

Together, these questions define long-term research priorities for NeuroAI rather than problems likely to be solved by incremental scaling alone.



National Cancer Institute / Unsplash

E. Persistent Challenges, Limitations, and Risks

Despite these advances, participants emphasized that significant conceptual, technical, and institutional barriers remain.

Bridging Disciplinary Divides and Neuroscientific Scales

Despite the move toward NeuroAI, participants shared that meaningful integration between computational and biological research remains difficult. They noted that, historically, biologists have questioned purely computational approaches, while students trained in computational methods are often unfamiliar with biological realities. As one researcher observed, "The problem still exists that you need to integrate these things within individuals."

A lack of unified modeling tools capable of bridging scales—from molecular and biophysical mechanisms to neural circuits, behavior, and cross-species evolutionary comparisons—remains a major obstacle. Attendees noted that simulating the brain from first principles requires resolving billions of neurons and trillions of synapses over time scales spanning milliseconds to years, creating immense computational bottlenecks.

Identifying the Appropriate Level of Observation and Scientific Scope

Participants argued that unlike genome sequencing, where base-pair resolution provides a clear observational target, neuroscientists lack consensus regarding the necessary level of detail required for accurate brain simulations. As one expert explained, "Where simulation has been powerful and successful in physical sciences is when we really feel like we have the unit interactions down and that there's uniformity in the system, so that you can use finite element models for mechanics and for fluid dynamics…. We don't have that in the brain, so I think this is a big challenge for any kind of modelling."

Additionally, researchers discussed how increasingly capable AI systems may reshape the scientific scope of neuroscience. Using the example of interactions among multiple AI agents, a participant suggested that the field should study intelligent systems more broadly. "I think we do need to expand the remit of neuroscience. It is all about intelligent systems [and] cognition. Studying behavior of AIs is really starting to become pretty important, [including] how to align their behaviors, how they model us, and how we model them," the attendee advocated.

Functional Decomposition

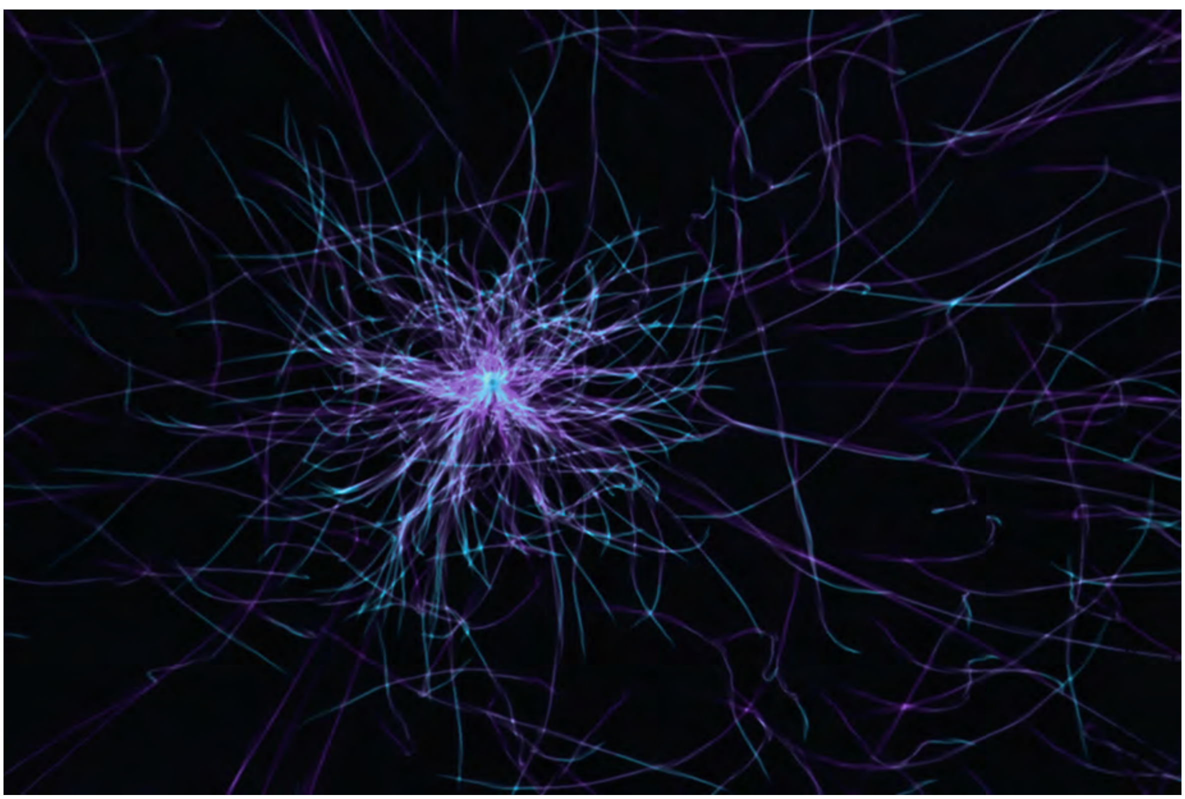
A central debate in neuroscience centers on the extent to which human cognitive architecture is organized into distinct, modular components versus a deeply integrated system. Some participants pointed to developmental psychology and functional neuroimaging as evidence of distinct parts. Further, they noted that artificial neural networks can spontaneously reproduce these functional alignments with minimal prompting, suggesting there may be deep computational constraints driving why both biological and artificial systems develop shared structural organization. Others emphasized how little is understood about what specific brain regions actually do, adding that without knowing a component's precise functional objective, assessing how its underlying architectures contribute becomes exceptionally difficult.

Reflecting on this tension, one researcher challenged the assumption of clean modularity: "I don't know if you can say that we've figured out if the mind is well separated into modular units and layers. If [so], there's a whole lot of mush and overlap. And maybe that's the power. Maybe that's what we're learning from the fact that you can do these feed-forward inference passes on these models to get them to do wonderful things without much recurrent dynamics. Maybe that's a hint that the mind is really working together a lot more than in separable subcomputational units," they reasoned.

Their explanation illuminates how the lack of clean functional separation in biological systems directly parallels the interpretability challenges facing modern AI. Large language models, deep convolutional networks, and complex parameterized systems present a similar opacity, proving "sort of challengingly non-dissociable for people who are studying them as artifacts of cognition themselves or [of] problem solving or computational goals."

Risk of Overgeneralization

Speaking to overgeneralization from the perspective of AI itself, participants emphasized that a core component of scientific reasoning is distinguishing system-specific idiosyncrasies from genuinely general principles. In one participant's view, "an intelligent experimentalist [is] one who knows which of their findings are idiosyncratic to their system and which are emblematic of general principles." Given this definition, they highlighted that a major challenge for AI systems going forward will be to have an internal understanding of which instances are general and which are particular.



Sandip Kalal / Unsplash

Experts also warned against human-led overgeneralization, particularly by equating biological brains too closely with simplified computational models. Researchers cautioned that a major fallacy is "reifying" the metaphor of the brain as a computer as though it were wholly true. They argued that this confusion licenses researchers to ignore biological details that are essential for understanding neural computation.

In identifying how this metaphor breaks down, participants noted that many AI models assume stationary environments, whereas biological reality is inherently non-stationary. As one expert explained, using past predictions to directly predict the future assumes a certain level of stationarity. This assumption, they argued, remains underacknowledged in predictive AI modeling. Other key differences mentioned by participants included energy efficiency, backpropagation techniques, and generalization out of distribution.

Scientific Pipeline and Societal Risks

Participants cautioned that an over-reliance on AI assistants could risk eroding the scientific training pipeline, isolating students and even later-career researchers from constructive exchanges with colleagues. They also noted that researchers can become so distracted by studying AI mechanics that they lose important foundational knowledge and the ability to ask deep questions about biological systems. As one expert summarized, "I think that the meta-issue [for] the whole of AI is if we use it to amplify our thinking, it can be good; if we use it to outsource our thinking, it generally is very bad."

Regarding the broader societal risks of AI, participants elevated immediate dangers—specifically calling out misinformation, job loss, and corporate profiling using behavior tracking—over sensationalized fears of superintelligence taking over humanity. "We don't need to worry about the creation of a new potential bad actor. It's the pairing of bad actors with tools of massive power amplification" that we should worry about, one attendee reflected.

F. Recommendations for Funders

Participants expressed appreciation for long-standing neuroscience initiatives funded by the BRAIN Initiative, Howard Hughes Medical Institute, and Wellcome Trust. Looking ahead, researchers stated the importance of sustained support for these transformative projects while actively counteracting incentives that encourage researchers to chase short-term technological trends over careful integration between experimental data and biophysical models.

1. **Prioritize New Data Collection:** Resources should target high-resolution, chronic data acquisition tools that bridge human and animal research.
    - *MRI-Compatible Human Chronic Neuropixels:* Enable the engineering and clinical translation of MRI-compatible chronic Neuropixel arrays that enable simultaneous 3D functional imaging, multi-site electrical recording, and focal micro-stimulation. With this technology, neuroscientists would be able to characterize the evolution of representa-

tions within the brain and map structural connectivity—all with shafts that are smaller than the current standard intracranial electrodes.

- *Harmonized Multi-modal Human Datasets:* Support deep data collection across small cohorts—integrating fMRI, intracranial recordings, physiological feeds, and naturalistic behavioral tracking—to construct individual-specific brain foundation models. Participants predicted that within five years, these efforts could enable clinicians to predict optimal drug or brain stimulation parameters for depressed patients with minimal baseline imaging.

- *Naturalistic and Physiological Contexts:* Expand data collection beyond confined laboratory environments to capture real-world interactions and systemic body-brain physiological interactions.

2. **Support model development through integrated and comparative approaches:** Participants expected to see high return on investment through projects that integrate large-scale experimental data acquisition with a variety of computational approaches, including AI architectures, foundation models, and biologically realistic circuit models. Specifically, they proposed building a deep understanding of how the circuits of the brain carry out their functions.

    - *Integrated Circuit Mapping:* Combine connectomics, slice physiology, *in vivo* recordings, and computational modeling into unified pipelines. For one expert, achievable progress in five years would include having foundation models and biorealistic models of a particular part of the brain that can reproduce *in vivo* activity while respecting the substrate and the mechanisms.

    - *Comparative Model Organism Research:* Fund comparative research across simpler model organisms (such as worms, flies, or mice) to identify cognitive features like distributed oscillator systems that are conserved at the human level.

Shawn Day / Unsplash

3. **Provide embedded, cross-disciplinary training experiences and centers:** Funders should bridge the cultural divide separating computational engineers and experimental biologists across settings.

    - *Cross-Disciplinary Embeddings:* Establish fellowship programs, modeled on initiatives such as the Eric and Wendy Schmidt AI in Science Fellowship at the University of Chicago, that embed computer scientists in wet-lab environments and experimental scientists in computational settings. As one expert noted, "Despite all that has happened with the integration of neuroscience, neurobiology, computational work, and AI, I think the most vast gap still exists between the computational people and the experimental people." Another attendee encouraged funders and tech companies to directly consult neuroscience experts when working on topics like consciousness.

    - *Public-Benefit University Centers:* Establish and fund independent, university-based multi-disciplinary centers dedicated to AI for public benefit. These centers should aim to bring neuroscience expertise into discussions of human cognition and AI governance while ensuring that commercial interests do not shape those discussions alone.

4. **Invest in an "AlphaFold for Neuroscience":** Funders should support unified computational initiatives that model complex neurodevelopmental and functional mappings. As one participant detailed, "If the data were to be available, I think one version of AlphaFold for neuroscience would be, 'How do you go from early developmental ingredients to the outcome, to the structure that comes at the end?'" A parallel objective introduced by participants would involve building frameworks that bridge structure to function, constraining the space of functional possibilities based on structural connectivity.

G. Conclusion: Toward a Convergent Neuro-Computational Science

The growing convergence between artificial intelligence and neuroscience represents one of the most promising intellectual frontiers of the twenty-first century. Machine learning offers tools for modeling cognitive architectures, analyzing high-dimensional neural data, and designing experiments. Biological brains, in turn, demonstrate forms of adaptable, energy-efficient, multi-scale intelligence that engineered systems have not yet fully reproduced.

Advancing this reciprocal research agenda will require mature scientific methods grounded in data sharing, cross-disciplinary training, and sustained exchange between computational and biological researchers.

5. Genomics & Human Variation

A. Participants

We aim to synthesize and share perspectives from the discussions as a whole rather than to attribute any quotations or viewpoints to specific individuals. Participants with whom we engaged in roundtable discussions are listed below (alphabetically by last name):

- **Joshua Akey, Ph.D.** – Professor of Ecology and Evolutionary Biology, *Princeton University*
- **Žiga Avsec, Ph.D.** – Genomics Research Scientist, *Google DeepMind*
- **Andrew Carroll, Ph.D.** – Genomics Product Lead, *Google AI*
- **Kushal Dey, Ph.D.** – Associate Professor of Computational and Systems Biology, *Memorial Sloan Kettering Cancer Center*
- **Erez Lieberman Aiden, Ph.D.** – Professor and Department Chair of Biochemistry and Molecular Biology, *University of Texas, Medical Branch*; Professor of Biosciences, *Rice University*
- **Xihong Lin, Ph.D.** – Professor of Biostatistics and Coordinating Director of Program in Quantitative Genomics, *Harvard T.H. Chan School of Public Health*
- **Quaid Morris, Ph.D.** – Professor of Computational and Systems Biology, *Memorial Sloan Kettering Cancer Center*
- **Ben Raphael, Ph.D.** – Professor of Computer Science, *Princeton University*
- **Steve Reilly, Ph.D.** – Associate Professor of Genetics, *Yale School of Medicine*
- **Fritz Roth, Ph.D.** – Professor and Chair of Computational Systems Biology, *University of Pittsburgh School of Medicine*
- **Jay Shendure, M.D., Ph.D.** – Professor of Genome Sciences and Director of Brotman Baty Institute for Precision Medicine, *University of Washington*
- **Cas Simons, Ph.D.** – Rare Disease Program Lead, *Centre for Population Genomics*
- **Sanjay Srivatsan, Ph.D.** – Assistant Professor, *Fred Hutchinson Cancer Center*
- **Lincoln Stein, M.D., Ph.D.** – Acting Scientific Director and Head of Adaptive Oncology, *Ontario Institute for Cancer Research*

B. Introduction

The effort to map and sequence the human genome was once framed as uncovering the definitive "blueprint" of life. Yet, as functional genomics has expanded, researchers now confront a new frontier: reading sequence data is fundamentally different from understanding its operational rules. Today, the central objective of applying artificial intelligence to genomics and molecular biology is to move the field from observational cataloging toward true functional prediction of how genetic variation shapes biological outcomes. Roundtable participants identified this predictive capacity as the first step toward new, mechanism-driven interventions that improve human healthspans and lifespans.

This chapter explores how artificial intelligence can uncover structure in genomic and multiomic data at scale; how machine learning may reveal relationships between variation and biological function; and how high-dimensional data integration may reshape knowledge about human variation.

Answering these questions requires greater clarity about what constitutes biological interpretation. Throughout the convenings, participants debated whether scientific interpretation centers on drawing static schematic diagrams or on accurately predicting the downstream effects of physical perturbations. Several participants noted that developing a tractable perturbation theory is far more achievable than attempting to solve a complete, multi-scale biological system at once. They also emphasized the need to evaluate distributions of variant effects across specific genomic, environmental, and developmental contexts rather than studying mutations only in isolation. This approach would require moving beyond tests of isolated variants in single, steady-state cells and toward evaluating them across diverse genetic and environmental backgrounds.

From an evolutionary perspective, participants framed the ability to distinguish what is conserved from what varies across species as an important benchmark of how well researchers understand living systems. Despite substantial advances, current models still struggle to predict how genomic variants affect gene expression across diverse cellular and biological contexts. Participants also outlined a longer-term goal for the next century of genomics: determining whether a computational model could predict what kind of organism would emerge given a genome, a particular cellular environment, and the fundamental laws of physics.

Core Takeaways

- **From isolated variants to context distributions:** The field should move past localized single-variant predictions to map distributions of variant effects across diverse genetic, environmental, and developmental contexts.
- **Machine learning accelerates multiomic integration:** AI can bridge historically siloed domains, combining genomic variants, spatial transcriptomics, medical imaging, and longitudinal health records into unified predictive frameworks.
- **Causal mechanisms vs. statistical plausibility:** Black-box predictors excel at identifying pathogenic variants or transcription factor binding but fail to explain underlying mechanisms such as identifying target genes or specific cell types.
- **Generative models require rigorous statistical safeguards:** Synthetic multiome and proteome datasets offer powerful workarounds for severe data gaps, provided downstream statistical methods account for potential model errors.
- **Strategic philanthropy should prioritize experiments over algorithms:** Funders should under-index on standalone algorithms and invest heavily in high-throughput perturbation assays, open data infrastructure, wild-type species research, and disease-specific multiomic datasets.

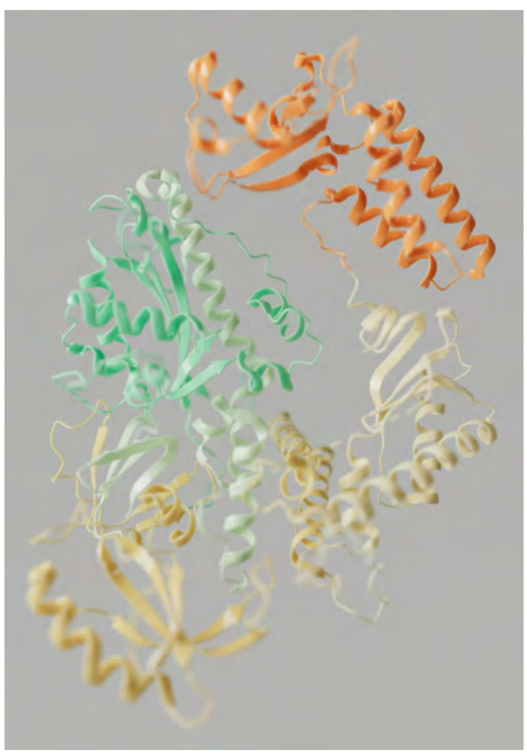
C. Areas of Excitement and Progress

Pattern Identification Across Omics and Regulatory Spaces

Participants highlighted several areas in which machine learning is advancing pattern identification across large biological datasets. For example, in imaging genomics, they pointed to algorithms that derive complex imaging phenotypes from large-scale repositories such as the UK Biobank, in turn revealing subtle structural patterns that more traditional analyses may fail to detect. In transcriptional regulation, contributors noted that machine learning models can now predict how increasing or decreasing the expression level of a gene alters broader transcriptional patterns. Experts also pointed to large-scale multiplex assays as one way to help resolve variants of uncertain significance (VUS) in Mendelian disorders.

However, experts emphasized that key mechanistic gaps remain. Models linking noncoding DNA variation to function can prioritize candidate variants, yet struggle to determine downstream biological consequences or predict the direction of their effects on cellular function. Similarly, tools such as DeepMind's AlphaMissense can predict variant pathogenicity but do not explain the underlying biological mechanisms. As one participant summarized, "Across the full spectrum of missense variants in clinical genetics, we can do reasonably well at predicting this not entirely useful thing [due to its binary definition] called pathogenicity, but we don't know why."

Participants contrasted promoter regulation, where variants are relatively localized with enhancer variants, which tend to be more complex, context-dependent, and specific to particular tissues and cell states. Despite progress in mapping enhancers, they emphasized that the current picture only covers a small fraction of the cell types and states present throughout human development and that linking specific enhancers to their regulated genes remains a challenge. One researcher noted, "I think that a really important future goal in biology is being able to not only understand and predict



Planet Volumes / Unsplash

regulatory features throughout the genome, but understand what particular genes they interact with over what developmental time points and tissues that those enhancers talk to those genes in."

Even within local regulatory regions, individual genes exhibit diverse modes of function, with independent noncoding variants driving distinct tissue-specific or shared expression changes. This creates complex regulatory architectures. Linking these variants to downstream pathways and mechanisms becomes even more challenging in genome-wide association study (GWAS) data, where tens of thousands of implicated haplotypes contain unknown causal variants. Beyond detecting patterns in GWAS data, researchers must ask how to link GWAS associations to their target genes and how cis- and trans-regulatory networks interact to shape phenotypes across scales, from a single enhancer to hundreds of kilobases of genetic material.

In terms of binding, attendees explained that models are quite skilled at predicting where various types of DNA binding proteins bind in the genome—often analyzed prior to checking pathogenicity—but that, as with the previous research areas, knowledge about how these proteins interact across cell types, protein combinations, and larger regulatory networks remains limited.

Working Around Data Gaps

Attendees also discussed AI as a potential strategy for addressing experimental data gaps arising from the sheer number of base pairs, tissues, cell types, and contexts under consideration. They characterized experimental data as a matrix with many holes, suggesting that generative models could produce synthetic multiome and proteome data to address missing observations. This approach, they explained, could be particularly useful for populations underrepresented in whole-genome sequencing datasets.

However, participants emphasized that downstream statistical methods must account for potential errors introduced by generative models, with one contributor arguing that even if the underlying model is wrong, the right statistics can still make the downstream interpretation valid. They also noted that many generative and foundation models are not well-rounded. For instance, some models currently perform well on some omics prediction tasks while continuing to struggle with accurate gene expression prediction.

At the human scale, data gaps are also driven by the fact that current GWAS datasets substantially overrepresent certain populations, particularly those of European ancestry. While there is no true substitute for collecting more representative data, participants suggested that AI could help decode basic genomic rules—which operate universally across ancestries—and subsequently advance knowledge about small populations lacking dedicated GWAS studies. The same was discussed for rare disease studies. As one participant noted, "I think that there's both a fear of concentration of benefits, but also a democratization, if applied well."

Acceleration of Big Data Aggregation and Analysis

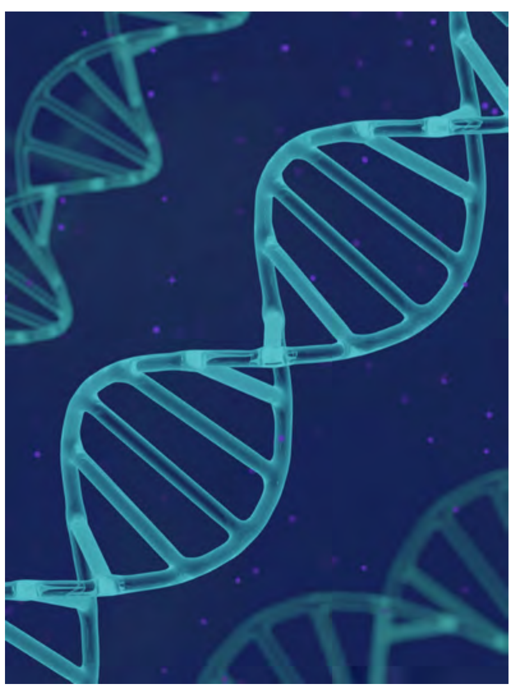
Machine learning algorithms can support the aggregation and re-analysis of disparate data modalities within new paradigms, uniting variation across humans and evolutionary organisms. Participants noted that the study of human disease variation has been historically siloed from functional genomics, the sub-field that aims to connect variants at the cellular level to disease outcomes. Purpose-built, supervised models can integrate genomic variants with medical imaging, lipidomics, and proteomics to predict whether particular variants are oncogenic. In rare disease research, experts described how natural language models can draw on published literature to help compile current gene-disease associations for clinical interpretation. The next extension involves directly linking patient medical records to case studies in the literature—while erecting appropriate privacy behaviors—using language models to evaluate phenotypic similarities that may not be captured by relying solely on [Human Phenotype Ontology](#) terms. Along these lines, machine learning might be integrated with environmental exposure data to examine how an individual's life history may impact their disease progression.

AI can also help researchers incorporate new data modalities. As an example, one contributor spotlighted that "we're in the stage where we're now generating all this telomere-to-telomere [T2T] data, we're adding all the structural variation to the human genome…. How do we update all of our basic pipelines to account for structural variation? How do we include this new data modality? That's a great application of AI. We know how we want [these things] to work, we just [have] to add this new modality to them."

Participants noted that these computational approaches generate inferences about population history, patterns of selection, and conservation over both short and long evolutionary timescales, adding that AI can help meet the substantial computational demands created by long-read sequencing.

Pairing AI with Emerging Domain-Specific Approaches

Participating researchers identified spatial genomics, transcriptomics, and proteomics platforms as promising new avenues for computational modeling. These methods measure transcript and protein expression levels at single-cell resolution while preserving tissue architecture, allowing researchers to observe cell-cell communication and spatial organization. As one participant highlighted, "I think



TSD Studio / Unsplash

this is incredible technology that has great potential for training machine learning systems to identify patterns in cell-cell interactions, and to work out the programs for normal and abnormal development."

Aiding Experimental Workflows and Scientific Communication

Participants described machine learning as contributing to lower instrument costs, improved data quality, and new assay designs, opening up new possibilities for researchers to test hundreds of hypotheses across multiomic datasets. AI could also make scientists more productive by helping automate large computational workflows and software-based informatics tasks. Specific applications include scaling up investigations of "thousands of Mendelian hits that we've been sitting on for decades." Physical procedures such as liquid handling that require precise interaction with the laboratory environment were described as candidates for future AI facilitation, with one attendee hoping to see systems with sufficient "embodiment" for these applications.

Participants also discussed how large language models could "raise the bar of scientific communication" by summarizing literature, reformatting datasets, creating graphs, and generating publication-ready presentations. Researchers could also use these agents to assist with annotating the literature and reimplementing published methods to address reproducibility issues in the field.

D. High-Value Research Applications

Evolutionary History and Comparative Genomics

Machine learning offers powerful tools to model complex, long-term evolutionary demographics by integrating contemporary human genomes with ancient DNA samples. Sequencing also provides ways to study recent evolutionary dynamics, such as viral infections that affect the pregnant body.

Participants identified comparative genomics across species as a critical lens for understanding fundamental biology. They noted that cell-type-specific regulatory programs can be highly conserved across species, making it valuable to study human biology alongside other species. Participants warned against the trap of "human exceptionalism," which leads many labs to focus exclusively on human cell lines while disregarding other useful models. As one expert explained, "Even if we had all the [human] data, we need to go to other organisms to understand the machinery, right? Because all those variants exist, but we don't have all the haplotypes. So every haplotype is going to be a little bit unique, every protein might look a little bit unique from person to person. So you really want to have a grid model. You need to extrapolate out beyond it to understand the full shape of it, and for that you want really different sequences."

Although obtaining cross-species longevity data remains difficult, contributors pointed to two structured sources that offer immediate promise: large livestock populations optimized for specific traits and zoo animals across the world that are tracked through daily-resolution veterinary care. Other examples include the open-science Dog Aging Project, which features information on over 50,000 companion dogs.

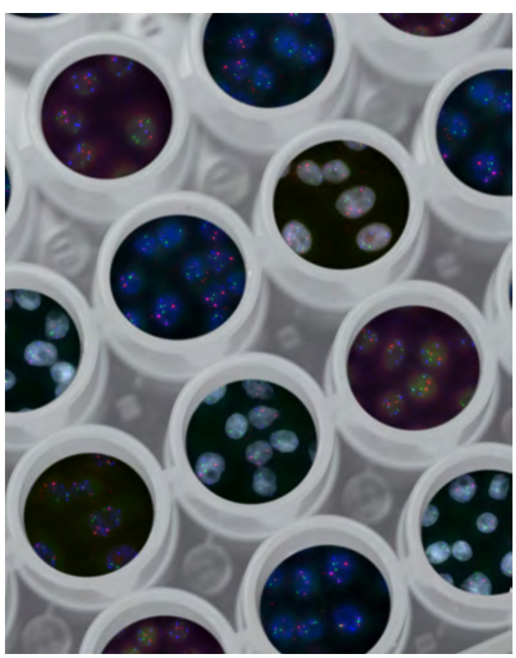
Polygenic Risk Scores

Participants discussed how machine learning might offer pathways to refine Polygenic Risk Scores (PRSs) by discovering underlying mechanisms, context-specific variant effects, and target populations. They noted that newer, AI-powered models more effectively capture high-signal-to-noise variants near gene promoters but struggle with distal, weak-effect variants operating in context-specific cell states that may not be represented in steady-state tissues. One expert cautioned that only "a limited handful" of disease-associated variants can currently be mapped well by these models. The expert also described "a lack of alignment" between model-based expression predictions, including those from AlphaGenome, and findings from expression quantitative trait locus (eQTL) studies. Although the expert felt that using AI models to inform individual risk scores remains "a little far" from clinical application, they saw promise as datasets expand and model architectures improve.

Attendees further highlighted the need to examine interaction effects between inherited genetic variation and environmental factors like diet to better understand the extent to which disease risk is genetically determined.

Precision Medicine and Targeted Cell State Interventions

Within precision medicine, machine learning could help move clinical genetics beyond simplistic binary pathogenic classifications and toward continuous probability distributions of medically relevant traits—in turn informing personalized care by clinicians who would need to be trained in this technology. On the clinical diagnosis side, researchers are now working to systematically correlate high-accuracy long-read sequencing, methylation profiling, and detailed phenotypes across cohorts to push rare disease diagnosis rates toward 100%. One participant detailed: "Given a genome sequence, [we want] to come up with a probability that this person will experience each of the possible traits that's worth tracking for medical reasons. The clinical world is doing this crazy simple thing of saying, 'Is a variant pathogenic or not?'.... We want a probability because then we can start using decision theory and say, given that probability, what is the expected utility of applying this



National Cancer Institute / Unsplash

treatment, or that treatment, or that treatment, or no treatment?" A further suggestion involved combining genome sequencing with immune-system measurements to improve predictions of individual treatment responses, a use case that may also help with candidate identification in clinical trial design.

Translating probabilistic predictions into clinical practice presents challenges, however. Attendees noted that implementation would largely fall to healthcare companies, which they expected would lack training in these tools. Moreover, a key therapeutic challenge remains: designing interventions that drive cells into specific states, including cell death, differentiation, protein production, and regeneration. Participants discussed moving from passively reading genomes to actively writing novel haplotypes and synthetic functions, creating an iterative process between reading and writing biological sequences. However, one expert underscored the difficulty of "writing different haplotypes that have never existed and will likely never exist, or different proteins, different regulatory sequences, and doing things at a scale that we actually need to test in biological systems. That's one big technical, unsolved piece that would greatly accelerate hypothesis testing in the lab," they explained.

Antigen Mapping

Mapping sequence relationships between T-cell receptors (TCRs) and antibody sequences and their target antigens was, for participants, a revolutionary yet achievable five-year goal. Sequencing blood samples would provide researchers with insight into what the immune system is responding to—knowledge that is expected to enhance therapeutic development in cancer and neurodegenerative diseases. As one contributor commented, "I've been surprised that none of the big AI companies have actually gone in this direction. This seems way more transformative and more tractable than building the virtual cell, because this [project is] very specific. It's sequence-based; it uses protein language models, DNA language models, all this stuff. There's a fair amount of data, but more could be generated."

Other fundamental questions remain underexplored in the field, such as:

- Decoding the vast regulatory information in the 3′ UTR (three prime untranslated region) to predict mRNA stability, localization, and translation; and
- Developing methods to identify the specific cell types in which disease variants are active, an ongoing challenge in Alzheimer's research.

E. Persistent Challenges, Limitations, and Risks

Unrepresentative Data

A primary bottleneck in contemporary genomics is the historical overrepresentation of a small number of ancestries—predominantly populations of European descent—and a narrow range of disease profiles. Participants described this disparity as an urgent yet entirely addressable challenge that requires expanding empirical data collection among underrepresented global populations, including Indigenous communities. As one expert asserted, a central priority must be “the solvable problem of making sure our genomic databases have full global diversity represented.” They continued, “We’ve obviously had, historically, very limited to a handful of ancestries dominating these data sets, and again, it’s a very tractable problem.” Achieving global representation was presented as a fundamental requirement for scientific accuracy. Without comprehensive data on variation and phenotypes across all populations, underrepresented groups will continue to be excluded from the clinical benefits of modern precision medicine.

Contributors also highlighted severe empirical limitations in modeling dynamic, multi-scale biological systems. Since immune cells migrate throughout the body, mutate, and interact within spatial microenvironments, static *in vitro* models struggle to capture how cell states shape disease phenotypes. Participants noted that the focus on a small number of cell types and states, including within cancer research, represents a significant disconnect with [Waddington’s landscape] of development and underscores the need for richer, spatially resolved multi-scale datasets.

Limitations of Chromatin Structure Modelling and Assay Methodologies

The ability to map enhancer variants to target genes is complicated by limitations in models of three-dimensional chromatin structure, with one attendee describing current models as “primitive.” Focusing on model development in this space—including predicting where each piece of DNA is aligned with others and how the nucleoprotein complex is involved—is expected to enable understanding of transcriptional machinery in physical terms rather than relying solely on primary sequences. As one expert posited, “I think it’s going to be similar to the difference between looking at a primary amino acid structure and looking at a folded protein.” Additionally, existing functional assays suffer from methodological constraints. While deep mutational scanning in haploinsufficient cell lines measures net cellular fitness, experimental assays must expand to decompose this aggregate measure into specific components, such as protein stability and multi-protein binding affinities.

Planet Volumes / Unsplash

Balancing Empiricism and Reductionism

Roundtable experts highlighted a tension between approaches that directly reflect observed data versus ones that draw on simplified models of biology. For instance, machine learning models excel at reconstructing biological patterns through self-supervised masked-input tasks. However, this high predictive accuracy does not guarantee that a model has uncovered a true physical principle; it may simply come from learning a batch effect.

Several attendees expressed skepticism toward "virtual cell" efforts, noting that current transformer-based architectures—originally designed for human language and vision—possess inductive biases that may be fundamentally ill-suited for generalizing from physical sequence relationships. Cells rely on individual molecules dynamically interacting in space, whereas transformer models depend on fixed computational links. Moreover, current virtual cell models lack sufficient data and are largely trained on reference genomes rather than genomic-specific experimental data, creating a disconnect from the world they seek to represent. "Before we start believing things that are coming out of virtual cells, we have to really think about the underlying architectures of the models that are modeling the virtual cells. To get there, I think we need a lot of data," one participant reflected.

To prevent models from hallucinating false biology, contributors endorsed iterative, "experiment-in-the-loop" strategies wherein generative predictions and foundation models at-large are continuously benchmarked and validated against downstream wet-lab experiments. As one advocate for a bottom-up approach argued, researchers need to build models that can accurately mimic the most fundamental levels of biology—such as individual genes—before going into full genomes. Another participant posed, "What we really want [to know] is, how do you extract physical principles from AlphaFold, the physical principles of protein folding? Could we create new generations of AIs that have the physics baked into them and are thinking about the physics as they're thinking about the problem?"

Risks of Over-Fitting and Over-Interpretation

Unlike classical computational statistics, which start from assumptions of the underlying system through linear or differential equations, deep learning draws directly from the data. "Deep learning has given us the ability to discover patterns without a preconceived notion of what's going on under the covers," one participant expressed with excitement.

This dynamic latent-space mapping allows machine learning to identify subtle biological linkages that traditional statistical approximations overlook. As one expert explained, "The power of various ML models is that their internal representation of the spaces of points can be dynamically learned. Things that are close to each other can be represented in a close way. This lets you find associations where [a] little inexactness is going to obscure some link across a variety of problems." Through contrastive and self-supervised learning tricks, AI models can implicitly map relationships across multiomic assays without requiring explicit, manual labeling—a major advantage given the field's lack of massive, standardized datasets. The genomics data landscape was contrasted with language and vision models, which were trained on large corpora like ImageNet.

However, reliance on learned data distributions and generative models introduces notable risks of over-fitting and of over-interpreting "salacious," complex predictions. Participants warned against a growing "hype cycle" driven by "wrong-sized" or misguided projects that generate plausible-sounding hypotheses that cannot be falsified or validated in the real world. They noted that it is exceptionally easy to prompt generative models to yield positive, publishable results, creating a dangerous temptation to confuse statistical plausibility with biological causality—a temptation that will require shifts in academic culture and incentives.

Attendees noted that these risks are particularly salient when computational tools are "getting ahead of the appropriately cautious clinical treatment." While tools like AlphaMissense offer powerful predictions, one participant explained that the algorithm's approach to classifying variants as pathogenic or benign based solely on a single computational source conflicts with clinical standards, which require multiple independent evidence streams from actual patient cohorts.

Validating complex foundation models is further complicated when training sets consume every available data repository, leaving no datasets available for true out-of-sample benchmarking. As one participant observed, "I think everybody will appreciate how difficult it is to validate models that have thrown everything in the kitchen sink into building the model, because then you are searching for some independent, organized, well-trusted source of data that you can use to validate, but it was already used in training. What do you do?" To combat this challenge, contributors called for strict data provenance tracking for all new model iterations. Another expert proposed creating tightly controlled synthetic datasets to test whether agentic LLMs can reason and reconstruct known biology without relying on biased literature. "I'm fascinated by the idea that you could create *de novo* datasets where you understand everything that's going in as input to an agentic LLM that is then trying to reason, and seeing what it can learn. You have a lot more tight control over understanding exactly what happened and where," they suggested.

Moving forward, the scientific community must apply the exact same rigorous evidentiary standards to AI models as they do to human scientists. As one participant concluded, "We should apply the same standards of requiring evidence that we do to these models as we do for any other paper or any other scientist, and I think we should give them less of the benefit of the doubt that we give other people."

Developing and Integrating Robust Statistical Methods with AI Models

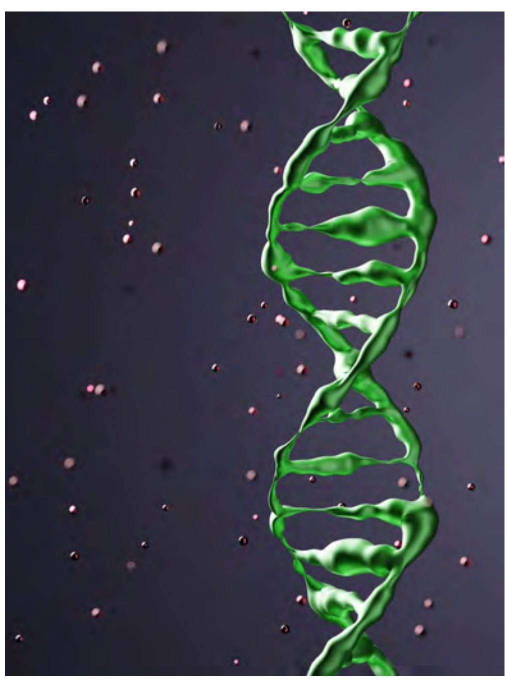
As generative AI and foundation models enter genomics, researchers are grappling with how to marry deep learning with classical statistical rigor. Traditional statistical frameworks in biology relied on a few, highly specific, asymptotic approximations and required explicitly pre-defined assumptions, such as linear or differential equations. In contrast, deep learning departs from these rigid assumptions by learning non-linear distributions directly from massive empirical datasets. As one participant remarked, "I don't think of AI as a statistical tool yet, but I believe people will come up with statistical bounds for these things. But maybe it gives us a way to process, synthesize, and accrue information in a different kind of way—and change the way we do experiments."

A primary advantage of generative architectures—such as transformers, diffusion models, or hybrid approaches—is their ability to model complex joint distributions that traditional statistical frameworks struggle to calculate. For example, when generating synthetic DNA sequences across hundreds of millions of genetic variants in common-variant GWAS, capturing the full linkage disequilibrium (LD) structure mathematically is extraordinarily difficult using classical approaches, whereas generative models learn these relationships directly.

However, because the underlying theoretical guarantees of deep learning are still known only in very simple settings, relying on non-linear latent representations introduces persistent risks of overfitting and poor generalization. To ensure scientific validity, downstream statistical safeguards are essential. As one participant summarized, “The model can be wrong, but if we do downstream statistical analysis right, that will ensure that downstream analysis or scientific discovery is trustworthy by accounting for the fact the generative model can be wrong.”

Integrating AI with Bayesian statistics offers promising avenues alongside clear trade-offs. Historically, setting “priors” in Bayesian modeling was strictly expert-driven. Machine learning can now scan vast biological literature to construct richer prior distributions automatically. Yet, because these AI-generated priors are less interpretable, transferring and generalizing models across different contexts become challenging.

This problem of mismatched priors is particularly acute when shifting from computational models to clinical practice. Current resources carry different forms of historical ascertainment and representation bias. ClinVar disproportionately reflects variants encountered and interpreted in clinical and research settings, while population resources such as gnomAD (Genome Aggregation Database) remain limited by the composition and ancestry representation of the cohorts available for sequencing: variants reported in ClinVar reflect patient cohorts who were specifically suspected of having a disease and referred for clinical sequencing, rather than general population distributions. Applying broad AI predictions built on these biased databases creates a mismatch between real-world clinical priors and training priors. Participants emphasized that the field must move beyond simplistic binary concepts of pathogenicity toward calculating true posterior probabilities of individualized penetrance and disease severity. As one expert detailed, “The whole problem with the concept of pathogenicity is, ‘this variant could cause disease in somebody, in some context.’ But that doesn’t really help the patient in front of you. What we want to know is: they’ve got a pathogenic variant, but



TSD Studio / Unsplash

are they actually going to get the disease? We are not yet set up to ask those questions in a broad AI-driven way, and that feels super promising to me. We need individualized estimates of pathogenicity, penetrance, and severity. We especially need better ways to understand low-penetrance variants and to counsel patients about rare variants that may or may not cause disease."

Finally, participants stressed that sequence-only foundation models trained on a single human reference genome are fundamentally limited in predicting molecular phenotypes and gene expression. Predicting how variants function across diverse cell types will require optimal transport-based transfer learning to bridge limited experimental data. However, because transfer models can also produce errors, rigorous statistical bounds remain necessary. Ultimately, researchers cautioned that while language models benefit from matching input and output formats, biology currently measures inputs well while relying on incomplete proxies—such as gene expression—for outputs. Moving beyond sequence-only models to capture the true, multi-scale output space of cellular phenotypes remains one of the most critical statistical and modeling frontiers in genomic science.

Erosion of Scientific Training Pathways and Experimental Substitution

Participants expressed concern that the rapid integration of AI tools is altering the social compact between principal investigators and graduate students. When principal investigators rely on AI to generate ideas, their incentives to train students may diminish, potentially weakening the development of scientific skills. One researcher observed: "the crux of the educational process is this apprenticeship that has trained generations of biologists. The apprenticeship comes from this sort of mutual need, right? I need you to help carry these sorts of things out and move them forward, because I have the knowledge and can educate you. I think that that social compact is actually breaking."

They also worried that some students approach research in the wrong order, searching for problems to which they can apply AI rather than beginning with fundamental scientific questions. To this end, a majority of participants warned against using experiments on machine-learning models as substitutes for experiments in biological systems.

Despite these critiques, contributors acknowledged that experienced researchers may have the contextual knowledge needed to use AI productively and that some students use language models thoughtfully as tutors or to integrate tested code from published papers into their own models.

Genetic Determinism, Biosecurity, and Information Risks

Participants highlighted risks related to the resurgence of genetic determinism, in which predictive models might be applied inappropriately to IVF embryo selection or to produce designer traits. They worried that democratized models may also allow individuals to sequence their own genomes and over-interpret variants, creating opportunities for healthtech companies to prey on unfounded fears.

Attendees noted that, within the biosecurity landscape, pathogen safety has shifted from a process of physical containment to one of information security. As DNA synthesis advances, controlling access to genomic sequence information is vital. One expert detailed: "It's not a matter anymore of making sure people don't have access to the physical pathogens. It's making sure people don't have access to the information that allows you to create pathogenic dangers to the human species. Once it becomes a matter of not physical security, but information security, AI has to play a really big role in keeping us safe."

Drawing analogies from nuclear development, participants anticipated that sensitive bioinformation would eventually propagate globally and suggested that red teaming initiatives could buy time to develop defensive technologies, global pathogen-surveillance networks, vaccines, and rapid therapeutic-response platforms.

F. Recommendations for Funders

Participants noted that investment in large language models (LLMs) has been guided by scaling laws, meaning funders have mapped increases in compute, data, and model size to expected performance gains. They suggested that establishing similar scaling metrics in genomics could help guide and motivate data collection, simultaneously urging funders to resist focusing on AI compute at the expense of generating diverse forms of biological data. In general, discussants called on funders to balance concentrated investments with support for diverse ideas and to frame broad engineering goals around outcomes that extend beyond phenotype prediction.

1. **Support Experimental Validation, Not Just Model Development:** Funders should invest in active learning strategies that directly couple generative AI predictions with high-throughput experimental perturbation assays. As one participant emphasized, "At the end of the day, or for a lot of these cases, we will need some sort of experimental validation, so investing in these high-throughput experimental models that can allow us to go in and perturb a system and to actually ask a question about a particular variant is absolutely essential. [There's] a lot of work in this space, but it is not yet scaling to the genome scale that we need to keep up with the predictions that these models are going to make. So, it's a real need for continued investment."

2. **Realign Incentives Around Model Organism and Wild-Type Research:** Funding should move toward wild species population biology, complementing existing inbred laboratory models. As one participant argued, scientists are under-indexing on the importance of studying variance in wild populations other than humans. To incentivize funding, contributors suggested pitching influential examples such as the transfer of knowledge about Gila monster venom to develop GLP-1 drugs. "If you actually want to make predictions about human physiology and intact living organisms, you will never be able to do it in culture, right? The path to doing it is by learning to do [make predictions] in model organisms and essentially learn a transfer function to project those learnings onto human systems, which are just another node in a phylogeny of mammals," one expert explained.

3. **Focus on the Dissemination of Information and Scientometric Tracking:** The spread of information can be as—if not more—influential for the acceleration of science than individual technologies. Funders should support the development of LLMs capable of keeping pace with the real-time publishing landscape and of drawing on GitHub repositories and supplemental paper materials. Grantmakers should also fund MCP (Model Context Protocol) servers to make

resources like [ENCODE (Encyclopedia of DNA Elements)] directly accessible to language models. At the same time, researchers should be taught how to structure their projects for seamless AI accessibility down the line. Participants further stressed the importance of addressing paywalls in the current for-profit publishing space. As one expert with a clinical genomics background noted, their sub-field has been particularly slow to embrace open-source publishing, in turn making it difficult to train LLMs on the knowledge already available. Drawing on the concept of scaling laws, funders should invest in scientometric tracking systems to measure how often biological datasets are utilized, as unmeasured data usage cannot be incentivized.

4. **Advance Deep, Disease-Specific Data Ecosystems:** Philanthropic capital should target specific disease systems (e.g., Alzheimer's tissue archives), integrating *in vitro* stress assays, *in vivo* mouse models, human clinical data, and spatial omics. As one expert detailed, "If I'm a funder, then I would probably go for a very focused, targeted system—like one particular disease system with a specific kind of tissue or organ of interest—and then try to generate data that gives different views of that system…. [Then we can] see, with all the depth that we have in terms of the data modalities, how much [we can] really push some of these methods." They added that such an approach is useful for parsing which limitations are due to modeling versus data.

   Key data acquisition initiatives include:

   - *Cross-Species Developmental Datasets:* Fund datasets across species, early development, and population-level individuals to integrate evolution and cell-type differences into models. A specific example proposed during the roundtables would center on serial sectioning of developing mouse embryos from single-cell to birth using spatial tran*scriptomics and proteomics to create 3D expression models.*

TSD Studio / Unsplash

- *Expanded Human Multiomics*: Expand biobanks (like the UK Biobank) by increasing population diversity and capturing comprehensive metabolomic, methylomic, and proteomic data.
- *Privacy-Preserving Health Datasets:* Structure electronic health records, histopathology, and imaging in ways that can be shared freely without personal health information risk.
- *Pedigree Data Capture:* Leverage health systems (such as the longstanding collaboration between Geisinger Health and Regeneron) to capture parent-child pedigree data, allowing AI to map rare genetic variation and infer haplotype phase from parent-child trios without the need for long-read sequencing.

5. **Conduct Field-Wide Conceptual Mapping and Bridge Research Communities:** Data generation should be preceded by field-wide roadmap exercises in which experts align around core strategic priorities. Funders should also establish grant mechanisms that integrate AI model generation with experimental validation, fostering active methodological collaboration between computational and experimental researchers. By coming together, these groups will be better able to identify which new datasets offer the greatest strategic utility for model accuracy. Ultimately, method development must answer one central question raised during the roundtables: "How can we advance trustworthy scientific discovery by integrating [the] generative model and the statistical method and also the domain science to accelerate discovery?" Participants also noted that industry-academic collaborations remain essential for training the next generation of scientists.

G. Conclusion: Toward a Causal Paradigm in Genomic Science

Applying artificial intelligence to genomics represents a fundamental shift in how scientists model, interpret, and care for the health of living systems. Machine learning offers powerful tools to compress multiomic data, discover non-linear regulatory patterns, and optimize experimental workflows. However, predictive performance should not be conflated with the long-term goals of mechanistic understanding and intervention design.

Ensuring that AI serves as a partner in biological discovery requires maintaining a rigorous dialogue between computational models, statistical safeguards, and wet-lab experimentation. By moving past analysis of isolated variant predictions, building open data infrastructure, embedding physical constraints into modeling architectures, and prioritizing context-dependent experimental validation, the scientific community can harness computational power to decode the foundational rules governing life.

SCIENCE & SOCIETY
aspen institute