



Case Study: AI's Large Looting Models? The Emerging Generative Biology Stack as the Next Frontier of Biopiracy

Jim Thomas - ETC Group

This case study is part of a research collaboration between ETC Group and IT for Change and is supported by the Center for Global Digital Justice.

Abstract

This report sheds light on an emergent field: Generative Biology (GenBio), the application of generative AI to redesign genomes, proteins, RNA, and metabolic pathways. GenBio is expanding at extraordinary speed, propelled by Big Tech and venture capital, with pharmaceuticals as its primary market and growing applications across agriculture, materials, and energy. Promoted as a transformative solution for health, food, and climate challenges, its greatest value may lie in serving the AI industry itself—generating data-heavy workloads and reputational benefits—while shifting social, ecological, and economic costs onto society.

The study highlights urgent data justice concerns: renewed bioprospecting to feed AI models, the conversion of biodiversity into proprietary digital assets, erosion of consent and benefit-sharing, and the acceleration of digital biocolonialism. GenBio also threatens to displace livelihoods, erode biocultural rights, strain biosafety regimes, and entrench philanthrocapitalist and military agendas.

Against this backdrop, the report identifies key entry points for civil society engagement, including negotiations under the Convention on Biological Diversity, the FAO Seed Treaty, privacy and data-protection enforcement, and national or regional policy on AI training data. Precaution, human rights, ecological and data justice must guide governance efforts if deeper enclosures of life based on AI and digital technologies are to be resisted.

Table of Contents

Abstract.....	2
Section I - Overview and Introduction.....	4
Aims and Methodology.....	5
History and background of GenBio.....	5
Section II: Towards a definition and taxonomy for AIxBio.....	9
Biological design and ‘generative biology’.....	10
AI tools and lab automation.....	12
AI in cyber-physical biological systems.....	12
Bio computation.....	13
Section III: Interrogating the AIxBio/GenBio ‘stack’.....	13
Overview of the stack.....	13
a. Data. The prime importance of the data supply chain – the return of bioprospecting and biopiracy.....	15
b. Wet/Dry translation: Synthesis and Sequencing as essential bridges and control points	17
c. Models – AI giants upsell compute and cloud services.....	17
d. Product Markets: Pharma dominates. Deepened disruption of natural products.....	19
e. Finance. Heavy costs, high borrowing and Trump era drive military pivot.....	21
Section IV: Insights.....	22
Starting points for social, environmental and policy considerations from a data justice and human rights perspective.....	22
1. Bioeconomy: Greater resource extraction and violations of rights divided by AI-enabled GenBio.....	24
2. Biopiracy: A renewed bio-prospecting rush.....	25
3. Biocultural rights: Livelihoods and cultures under threat.....	27
4. Biosafety: Risk assessment, governance and oversight under strain by AI.....	28
5. Health Justice: the meaning of chasing ‘blockbuster’ pharma compounds and philanthrocapitalists involvement.....	29
6. Militarization: Defense-focused GenBio threatens rights, transparency and equity.....	31
Opportunities for Civil Society action.....	31
UN Food and Agriculture Organization (FAO).....	32
Digital Privacy Actions.....	33
National, Regional and Global Policy on AI Training Data.....	34
Final Thoughts.....	34

Section I - Overview and Introduction

Generative Biology (GenBio) describes the application of generative artificial intelligence (AI that creates novel or synthetic digital material) to the field of genetic engineering: the redesign and creation of fundamental biological systems such as genomes, metabolic pathways, RNA, and proteins. In this study, we use GenBio as shorthand for this narrower technological domain, while *AIxBio* denotes the broader convergence of AI and biotechnology.

Artificial Intelligence (AI) today refers to computational systems that rely on vast datasets and intensive processing power to build models aimed at predicting real-world outcomes. While AI is promoted as transformative across sectors—including biotechnology—the immediate gains from Generative Biology (GenBio) may accrue less to biology itself than to the AI industry. GenBio opens up a lucrative, data-intensive market, while also allowing AI companies to rebrand themselves as innovators addressing global problems, regardless of whether their technologies offer real or viable solutions.

By suggesting how generative AI technology might aid the production of high-value drugs to cure disease, materials, foods and chemicals, the tech industry is enlarging its scope beyond more familiar uses of AI – generating new text, media, video and sound – reassuring investors of a future product pipeline in the ‘real economy’. Association with developments in GenBio – especially for medicine – may lend Tech Titans greater moral legitimacy. A flagship example is Google DeepMind, a research lab and subsidiary of Alphabet/Google, whose lead scientists won a 2024 Nobel Prize in Chemistry for *AlphaFold2*, a GenBio model that revolutionized protein structure prediction.¹

In April 2025, DeepMind CEO (and now Nobel laureate Demis Hassabis) made the audacious claim that GenBio models could mean an end to disease “maybe within the next decade or so”.² By extending generative AI beyond the digital and into the physical and biological realms, GenBio not only fuels technological hype but also opens up a new frontier of commercial exploitation—bringing with it serious risks for biosafety, food and health systems, and corporate and digital control over life itself. The use of new genetic tools, coupled with a corporate-driven agenda to convert biomass into bio-based products, is often described as the ‘bioeconomy’ vision. When the dynamics of

¹ Ewen Callaway, “Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures,” *Nature*, 9 Oct 2024: <https://www.nature.com/articles/d41586-024-03214-7>

² Demis Hassabis quoted in Derek Lowe, “The End of Disease,” In The Pipeline blog, *Science*, 21 April 2025: <https://www.science.org/content/blog-post/end-disease>

emerging AI are added to this agenda, they amplify and complicate long-standing controversies over the push toward a high-tech, corporate-dominated “bioeconomy”.³

Aims and Methodology

Because the field of GenBio is extremely new—evolving month by month— this report seeks to:

- Develop an initial taxonomy/foundational definitions of the sector where AI and genetic engineering are overlapping and converging, surveying the types of entities and activities shaping the GenBio ecosystem, along with the strategic direction the field may take over the next 3–5 years;
- Present a conceptual framework of the GenBio industry (referred to here as the GenBio ‘stack’) and offer early insights to help assess near-term, real-world impacts, as a domain that has received little critical scholarly attention;
- Identify emerging areas of social concern, with particular attention to data justice and the urgent need for stronger data governance mechanisms.
- Demonstrate data justice norms and principles that would need to be brought into consideration.

This research is based on an initial survey of commercial and scientific activities at the intersection of AI—particularly generative AI—and biotechnology. It draws on commercial market analyses, academic literature, industry association updates, and direct participation in AIxBio events and conferences. The researcher also conducted in-person interviews with key industry and scientific figures⁴ and attended several major international policy and industry gatherings relevant to this emerging field. While this report is supported by IT for Change and the Center for Global Digital Justice, it was developed in collaboration with a parallel investigation into the market power and governance implications of the convergence of generative AI and biotechnology supported by the European Artificial Intelligence & Society Fund, and draws upon primary interviews originally undertaken in that project.

³ On the bioeconomy, see, for example, ETC Group, “The New Biomasters: synthetic biology and the next assault on biodiversity and livelihoods,” November 2010: <https://www.etcgroup.org/content/new-biomasters>

⁴ Interviewees included representatives from a GenBio startup (Basecamp Research), a leading global AI tech company (Google DeepMind), a leading venture capital investor in the space (NotBoring), a pioneer in Synthetic Biology (Stanford University) and a representative from an industry trade and media group (SynBio Beta).

History and background of GenBio

The year 2026 will mark two anniversaries with great symbolic significance for the field of biotechnology: fifty years since the founding of Genentech⁵, which pioneered the commercialization of biotechnology, and sixty years since the creation of the first AI chatbot, Joseph Weizenbaum's *Eliza*⁶. Synthetic biology pioneer Drew Endy has described the AI and computing sector as biotech's "big sister"⁷—both industries emerging from the same elite institutions and technocratic mindset, backed by the same venture capital networks, and rooted in the same geographic hubs. Yet while data companies ascended smoothly to dominate the global economy, biotech—cast as the Cinderella sister—faced powerful headwinds of public opposition and systemic challenges as it attempted to reengineer food, health, and energy.

By the early 2000s, as the first wave of biotech products stumbled amid public resistance and technical limitations, a new movement in the biosciences—synthetic biology—set out to reframe genetic engineering by purposefully adopting the language and logic of the computing industry. Promising greater precision and control, it embraced a "design-build-test" cycle and promoted the idea of standardized biological "parts" modeled on electronic components. Under this so-called digital biology paradigm, techniques like gene editing, DNA synthesis, and the 'rational design' of cells and genomes were rebranded as modular and programmable.

Around 2010-2012, genomic industrialist Craig Venter claimed to have 'edited' the DNA of a living organism to function as a kind of biological 'compiler', capable of 'uploading' and 'running genetic programs' within a living 'operating system'⁸. At University of California-Berkeley, Jay Keasling had redesigned yeast metabolic pathways relying on methods (and metaphors) borrowed from electronic circuit design⁹. Emerging synthetic biology unicorns such as Ginkgo Bioworks and Amyris began to boast of using large-scale computation and early AI to 'reprogram lifeforms. At the same time, firms

⁵ See the story of Genentech's origin story, as told by Genentech: <https://www.gene.com/stories/proof-of-concept>

⁶ For the story of the first chatbot – and the regrets of its inventor, see Ben Tarnoff, "Weizenbaum's nightmares: how the inventor of the first chatbot turned against AI," *The Guardian*, 25 July 2023: <https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai>

⁷ From panel discussion at SynBioBeta Conference 2025, "Unbound Biology: The Next Era of (Bio)Computing," San Jose Convention Center, 6 May 2025.

⁸ Daniel G. Gibson, "Programming biological operating systems: genome design, assembly and activation," *Nature Methods*, 11(5):521-6, May 2014. doi: 10.1038/nmeth.2894.

⁹ J. D. Keasling, "Synthetic biology and the development of tools for metabolic engineering," *Metabolic Engineering*, 2012 May;14(3):189-95. doi: 10.1016/j.ymben.2012.01.004.

like Zymergen¹⁰ and Transcriptic mimicked the commercial model of cloud computing to establish fully automated, AI-assisted ‘cloud laboratories’—offering remote biotech experimentation through a web interface.

With much of synthetic biology borrowing heavily from the digital business playbook, it was perhaps inevitable that the explosion of developments in generative AI would also seep into biotech. Google’s DeepMind subsidiary, based in London, turned to neural networks to tackle the so-called “protein folding problem.” While most discriminative AI models of the 2010s were trained on large datasets of text or images, DeepMind’s AlphaFold was trained on publicly available digital sequences of proteins, with the aim of predicting how a given amino acid chain might fold into a three-dimensional protein structure. In 2018 and again in 2020, AlphaFold competed in—and won—the international CASP (Critical Assessment of Structure Prediction) protein-folding competition. At the close of CASP 2020, CASP founder John Moult declared the protein-folding problem solved.¹¹ (In reality, AlphaFold still struggles with certain classes of proteins and offers no deeper understanding of the underlying ‘rules’ that govern protein folding¹².)

In 2022, commercial large language models (LLMs) burst into public consciousness amid a wave of generative AI hype and speculative investment—driven by platforms like OpenAI’s ChatGPT and Google’s Bard (later rebranded as Gemini). Eager to tap into the billions flowing into AI, synthetic biology CEOs began claiming that the ‘transformer’ architecture¹³ behind LLMs could also be applied to the ‘language of biology,’ enabling the generation of novel proteins and DNA sequences. Some even framed the potential as comparable to “splitting the atom.”¹⁴ Startups like Generate Biosciences and Ginkgo Bioworks began promoting a shift into a new “generative biology” phase, in which

¹⁰ See <https://en.wikipedia.org/wiki/Zymergen> and <https://www.ycombinator.com/companies/transcriptic>

¹¹ Will Douglas Heaven, “DeepMind’s protein-folding AI has solved a 50-year-old grand challenge of biology,” *MIT Technology Review*, 30 November 2020: <https://www.technologyreview.com/2020/11/30/1012712/deepmind-protein-folding-ai-solved-biology-science-drugs-disease/>

¹² https://www.reddit.com/r/labrats/comments/1b1l68p/people_are_overestimating_alphafold_and_its_a/

¹³ A transformer model is a neural network architecture that can automatically transform one type of input into another type of output. Transformer models are particularly adept at determining context and meaning by establishing relationships in sequential data which in turn is the basis of large language models (LLM’s) such as Open AI’s ChatGPT.

¹⁴ Jason Kelly of Ginkgo Bioworks, speaking on No Priors Podcast, 28 September 2023, episode 34: “DNA as Code – Cell Programming and AI”, YouTube: <https://www.youtube.com/watch?v=snt-fMsCDVI>.

chatbots—not bioengineers—would design DNA, proteins, and even entire genomes from an AI prompt, offering what they described as “text-to-protein”¹⁵ interfaces.

In early 2024, Mustafa Suleyman—one of the co-founders of Google DeepMind—co-authored (with Michael Bhaskar) a high-profile book titled *The Coming Wave*¹⁶ that predicted humanity was nearing its most historically-significant “inflection point” due to the convergence of artificial intelligence and synthetic biology. Shortly thereafter, Suleyman was appointed CEO of Microsoft’s Artificial Intelligence division¹⁷. By the end of 2024, nearly all the world’s major AI corporations—Microsoft, Google, Amazon, NVIDIA, Alibaba, and Meta—had become active players in GenBio.¹⁸ That same year Google DeepMind’s leadership was co-awarded the Nobel Prize in Chemistry for AlphaFold2,¹⁹ further cementing the field’s rising status and the growing entanglement of Big Tech with the life sciences.

Considering the field didn’t exist just three years earlier, the pace at which GenBio has expanded is dizzying. A 2024 survey by Future Markets²⁰ identified 97 companies active in generative biology, collectively raising over \$1 billion in investment annually since 2021—though new names appear almost every week. Three-quarters of these companies are focused on applying GenBio to drug discovery, and nearly every major pharmaceutical corporation—including Roche, GSK, Merck, Bayer, Pfizer, and Sanofi Aventis—has partnered with at least one GenBio startup. Interest is also growing across agribusiness, materials, and energy sectors. Future Markets predicts GenBio could become a \$50 billion industry by 2035, serving an end-user market valued at up to \$140 billion.²¹

¹⁵ Shelly Fan, ‘ChatGPT for Biology: A New AI Whips Up Designer Proteins With Only a Text Prompt,’ Singularity Hub, 27 May 2025: <https://singularityhub.com/2025/05/27/chatgpt-for-biology-a-new-ai-whips-up-designer-proteins-with-only-a-prompt/>

¹⁶ Mustafa Suleyman and Michael Bhaskar, *The Coming Wave: Technology, Power and the 21st Century’s Greatest Dilemma*, Crown Publishing Group, Random House, 2023.

¹⁷ <https://www.theguardian.com/technology/2024/mar/20/microsoft-hires-deepmind-co-founder-ai-division-mustafa-suleyman>

¹⁸ Jim Thomas, “DSI, AI and Technology Titans” CBD Alliance Eco article no 70 (7) - 29th October 2024. <https://www.cbd-alliance.org/en/2024/dsi-ai-and-technology-titans>

¹⁹ Ewen Callaway. “Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures” 9th Oct 2024 Nature 634, 525-526 (2024).

²⁰ Future Markets, "[Global Market for Generative Biology 2024-2035.](#)" 2024.

²¹ Future Markets, "[Global Market for Generative Biology 2024-2035.](#)" 2024.

Investor excitement over the potential profit from allowing AI systems to reprogram the foundations of life is increasingly matched by growing concern from civil society and international policymakers. Both generative AI and genetic engineering are already deeply contested domains, and the emergence of a new commercial sector at their intersection—touching health, food, the environment, and corporate consolidation—amplifies a host of unresolved social, economic, moral, legal, safety, and power-related questions. Rather than resolving existing controversies, GenBio intensifies and compounds them.

AI-driven synthetic biology has, for example, already surfaced as a topic in the negotiations of the UN Convention on Biodiversity (more details below) and that convergence also raises security concerns that automated GenBio capabilities collapse the barriers to a rogue or hostile actor generating novel toxins, bioweapons or pandemic-creating pathogens. As such, states and militaries have a heightened interest to put themselves in the front row of development and industrialization of GenBio – increasing the hype and bringing defense budgets to the table. Amidst the market froth and conflicting hopes and fears, a clear and grounded picture of the actual state of GenBio (and AIxBio more broadly) is essential for coming policy processes.

Section II: Towards a definition and taxonomy for AIxBio

Definitions and delineations remain far from settled among stakeholders. The language of *AIxBio*—as used by trade group SynBioBeta and various industry commentators²²—encompasses a broad and evolving range where artificial intelligence, machine learning, and advanced computation are reshaping the practice of biotechnology. This AI–Bio interface can be grouped into four broad areas:

- 1. Biological design (generative biology), including**
 - a. Protein engineering and**
 - b. Generative biology for crops and agriculture**
- 2. AI tools and lab automation**
- 3. AI in cyber-physical biological systems²³**

²²

See <https://www.synbiobeta.com/read/inside-aixbio-demo-day-2024-the-top-startups-merging-ai-and-biotechnology>

²³ A Cyber-Physical Biological System connects computing and machinery to interact with living things; for more, see Sustainability Directory article, 01 May 2025: <https://fashion.sustainability-directory.com/term/cyber-physical-biological-systems/>

4. Biocomputation

Biological design and ‘generative biology’

GenBio refers to the use of AI tools to intentionally modify the genetics of living organisms—such as redesigning metabolic pathways (the cell’s chemical processes) or creating new DNA and RNA sequences. Future Markets²⁴ defines GenBio in its industry survey as “an emerging field that combines computational techniques, artificial intelligence, and biological data to design, optimize, and create novel biological systems, materials, and molecules.” At its core, GenBio seeks to use generative AI models to expand the design of biological molecules. These computational frameworks are trained to ‘learn’ patterns from large datasets—such as DNA databases—and generate new, plausible examples with similar properties. Such models can be applied across various forms of biological data, including DNA and RNA sequences, protein structures, gene expression profiles, and metabolic networks.

Protein engineering. In practice, most current generative biology activity is concentrated on redesigning and generating new proteins—a focus shaped largely by the perceived success of AlphaFold. Proteins are complex molecules composed of chains of amino acids that fold into specific three-dimensional structures. They are ubiquitous in biological systems and play a wide range of roles, from food proteins and enzymes (which catalyze or regulate chemical reactions) to key components of DNA, RNA, and other genetically active molecules. Often described as nature’s nano-machines, proteins underpin much of the dynamic activity within living systems—making them a prime target for computational redesign, despite ongoing uncertainties around function, interaction, and behavior in real-world contexts.

The ability to generate entirely novel, never-before-seen proteins has generated significant excitement, driven by the idea that such proteins could exhibit new, useful properties and biological functions.

Novel proteins are now being engineered for a range of applications—including as alternative food proteins, industrial enzymes, novel materials, and even as genetic molecules. The GenBio company Profluent²⁵, for example, uses generative AI models to create new gene-editing proteins similar to the CRISPR-Cas9 system, aiming to accelerate the pace of genetic engineering. Other companies are developing proteins

²⁴ Future Markets, "[Global Market for Generative Biology 2024-2035](#)" (2024)

²⁵ See <https://www.profluent.bio/>

designed to regulate gene expression—affecting whether a gene’s activity is switched on or off—or to function directly as therapeutic agents.

Following the initial focus on proteins, increasing attention is now being directed toward re-writing the hereditary genetic molecules—DNA and RNA—of viruses and organisms using AI. A particularly prominent development in this space is *Evo2*, released in February 2025.²⁶ Developed by the Arc Institute at Stanford University, in collaboration with NVIDIA, *Evo2* is described as a multimodal “foundation model” for biology. (A foundation model is a general-purpose model—like ChatGPT—that can operate across multiple domains of knowledge, as opposed to more specialized systems.) While DeepMind’s AlphaFold was trained to predict protein folding, Arc claims that *Evo2*, trained on 9 trillion DNA nucleotides, can predict genome changes, proteins, RNA, and binding domains. It is reportedly capable of generating or redesigning entire genomes for bacteria, viruses, and similar organisms.

Just as AlphaFold was backed by AI giant Google, *Evo2* is associated with NVIDIA, which supplies most of the Graphics Processing Units (GPUs, the hardware) used for generative AI). NVIDIA had previously developed its *GenSLM* model, capable of generating DNA sequences for microbes and viruses²⁷. Other specialized RNA-focused models have also emerged, including those developed by startups Atomic AI²⁸ and Inceptive Bio²⁹, as well as by Chinese tech giant Alibaba³⁰. Interest in RNA has grown in the wake of the success of mRNA vaccines (e.g., for COVID-19). Meanwhile, tech titan Amazon is collaborating with EvolutionaryScale³¹—an AI life sciences company founded by scientists formerly employed at Meta (Facebook).

Generative Biology for crops and agriculture. The same technologies are now being applied to crops and plants. A comprehensive report by *Save Our Seeds* surveys

²⁶ Arc Institute news release, “AI can now model and design the genetic code for all domains of life with Evo 2,” 19 February 2025: <https://arcinstitute.org/news/evo2>

²⁷ Isha Salian, “Gen AI for the Genome: LLM Predicts Characteristics of COVID Variants,” NVIDIA blog, 13 November 2023: <https://blogs.nvidia.com/blog/generative-ai-covid-genome-sequences/>

²⁸ <https://atomic.ai/>

²⁹ <https://inceptive.life/>

³⁰ “Alibaba Cloud’s AI Technology Sparks Breakthrough in RNA Virus Discovery” Oct 12 2024 - Online at https://www.alibabacloud.com/blog/alibaba-cloud%E2%80%99s-ai-technology-sparks-breakthrough-in-rna-virus-discovery_601654

³¹ Matt Wood, “Revolutionizing Generative Biology with AWS and EvolutionaryScale,” AWS news, 25 June 2024: <https://aws.amazon.com/blogs/industries/revolutionizing-generative-biology-with-aws-and-evolutionaryscale/>

the range of approaches and plant-specific AI models currently used by GMO crop scientists.³² One example is *AgroNGT*, a model developed by Google and InstaDeep and deployed by Syngenta³³, which was trained on 10 million genome sequences from 48 plant species. Genetic engineers are using generative biology tools to modify crops such as poplar, potato, and maize—enabling increasingly complex edits at multiple sites within plant genomes³⁴. Some researchers believe that AI-designed genome edits can be made at few enough sites to sidestep European regulatory thresholds³⁵. Meanwhile, AI-biodesigned microbes are also being developed for use in soils or as next-generation pesticides.

AI tools and lab automation

AI is also being integrated into biotechnology through machine-vision tools that accelerate and automate routine laboratory tasks. Large language models, trained on a vast corpora of scientific literature, are being positioned as creative aides for researchers—such as Google’s “co-scientist” application, which claims to help generate testable hypotheses accompanied by “a summary of relevant published literature and a possible experimental approach.”³⁶ The frontier is now advancing into *Agentic AI* (AI “agents” that autonomously perform digital tasks on behalf of users) and *Embodied AI* (robotic or mechanical systems guided by AI-driven decision-making). Both forms are beginning to enter genetic engineering laboratories, further automating experimentation and tightening the integration between digital systems and biological manipulation.

AI in cyber-physical biological systems

Genetic engineering is increasingly being used to design biological components—such as biosensors—as part of larger engineered systems that rely on AI and algorithmic

³² Benno Vogel, “When chatbots breed plant varieties,” *Save Our Seeds*, January 2025: https://www.saveourseeds.org/wp-content/uploads/EN-with-summary_-_online.pdf

³³ Instadeep news release, “Syngenta and InstaDeep collaborate to accelerate crops seeds trait research using AI Large Language Models,” 18 June 2024: <https://instadeep.com/2024/06/syngenta-and-instadeep-collaborate-to-accelerate-crops-seeds-trait-research-using-ai-large-language-models/>

³⁴ Benno Vogel, “When chatbots breed plant varieties,” *Save Our Seeds*, January 2025: https://www.saveourseeds.org/wp-content/uploads/EN-with-summary_-_online.pdf

³⁵ Test Biotech, “AI-designed insecticidal NGT 1 plants,” May 2025: <https://www.testbiotech.org/en/publikation/ai-designed-insecticidal-ngt-1-plants/>

³⁶ Google Blog, “We’re launching a new AI system for scientists,” 19 February 2025: <https://blog.google/feed/google-research-ai-co-scientist/>

decision-making to function. For example, the synthetic biology company InnerPlant has developed soybeans that emit specific light frequencies when under stress.³⁷ These signals are then interpreted by an AI system, which triggers automated responses such as additional irrigation or agrochemical spraying. In effect, such crops are being engineered to be “robot-ready”—designed from the outset to integrate seamlessly with AI-driven agricultural systems.

Bio computation

Biotechnologists are also exploring ways to implement computation—including AI circuits—directly within biological cells, potentially replacing traditional silicon chips. One example is the company Cortical Labs, which has engineered brain cells (known as organoids) to interface with computer chips and perform computational tasks.³⁸ These “brain cell computers” have been trained to carry out image recognition, a basic form of artificial intelligence. Although Cortical already offers a commercially available organoid-based computer³⁹ and is reportedly in discussions with Amazon, the broader field of biocomputation remains years away from widespread commercial deployment.

Section III: Interrogating the AIxBio/GenBio ‘stack’

Overview of the stack

One way to grasp the emerging shape of the GenBio industrial enterprise—and to understand its underlying business plan—is to map out the different parts of the industrial ecosystem required to make this new approach function. In digital industries, the full combination of hardware, software, network connections, and applications is sometimes referred to as a *stack*. Generative AI scholars are beginning to identify key components of the ‘AI stack’: for example, generative AI requires training data, foundational models, and specialized models. It also depends on infrastructure, including high-powered ‘compute’ (clusters of GPUs housed in data centres), as well as the necessary systems for cooling, electricity, and network connections to move data to and from AI servers.

³⁷ See InnerPlant website, “Products,.” <https://innerplant.com/products/>

³⁸ Adithi Iyer, “Is OI the New AI? Questions Surrounding ‘Brainware,’” The Petrie-Flom Center, Harvard Law School, 13 March 2024: <https://petrieflom.law.harvard.edu/2024/03/13/is-oi-the-new-ai-questions-surrounding-brainware/>

³⁹ Jess Kinghorn, “The world's first 'body in a box' biological computer costs \$35,000 and looks both cool as hell plus creepy as heck,” *PC Gamer*, 06 March 2025: <https://www.pcgamer.com/hardware/the-worlds-first-body-in-a-box-biological-computer-costs-usd35-000-and-looks-as-cool-as-it-is-creepy/>

Because GenBio relies on generative AI, the ‘stack’ for generative biology includes all these elements but is further complicated by the need for a biological *wetware* layer: lab equipment, reagents, and validation and safety protocols required for engineering the biological world. Data collection is also more complex, since living materials must first be accessed, extracted, and digitized before being used in AI systems.

Figure A presents a general overview of the full ‘stack’ required to carry out generative biology in the real world. It outlines a process that begins with the acquisition of genetic resources (genes, proteins, RNA, etc.) as training data, moves through the AI-enabled biodesign phase, and culminates in the physical construction, testing, and potential commercial deployment of biotech products and processes. The schema in this report was developed through iterative conversations and feedback from scientists and industry practitioners, but it remains our own initial attempt to model the system.

The diagram reveals that the ‘flow’ of work and material through the GenBio stack begins with acquiring genetic resources and other biological data, which are digitized and used to train AI systems. This is followed by the AI-driven biodesign phase—the core of the project—and then by physical bioengineering in laboratories, where candidate designs are built and tested for potential commercial applications such as drugs, crops, food ingredients, and materials. The synthetic biology industry operates on a “design-build-test” cycle—borrowed from software and chip design—where biological constructs are repeatedly iterated to get new organisms or proteins to ‘optimize’ and ‘work’ for industrial purposes.

This recursive loop becomes even more central in AI-generated biodesign, where human researchers cannot simply inspect AI-generated code and intuit whether it will function. Validation occurs only once the code is implemented in living systems and tested, though testing will inherently be open to interpretation and limited, temporally and physically. Mapping out the full stack also highlights several critical pinch points—bottlenecks and sites of control—in the GenBio pipeline.

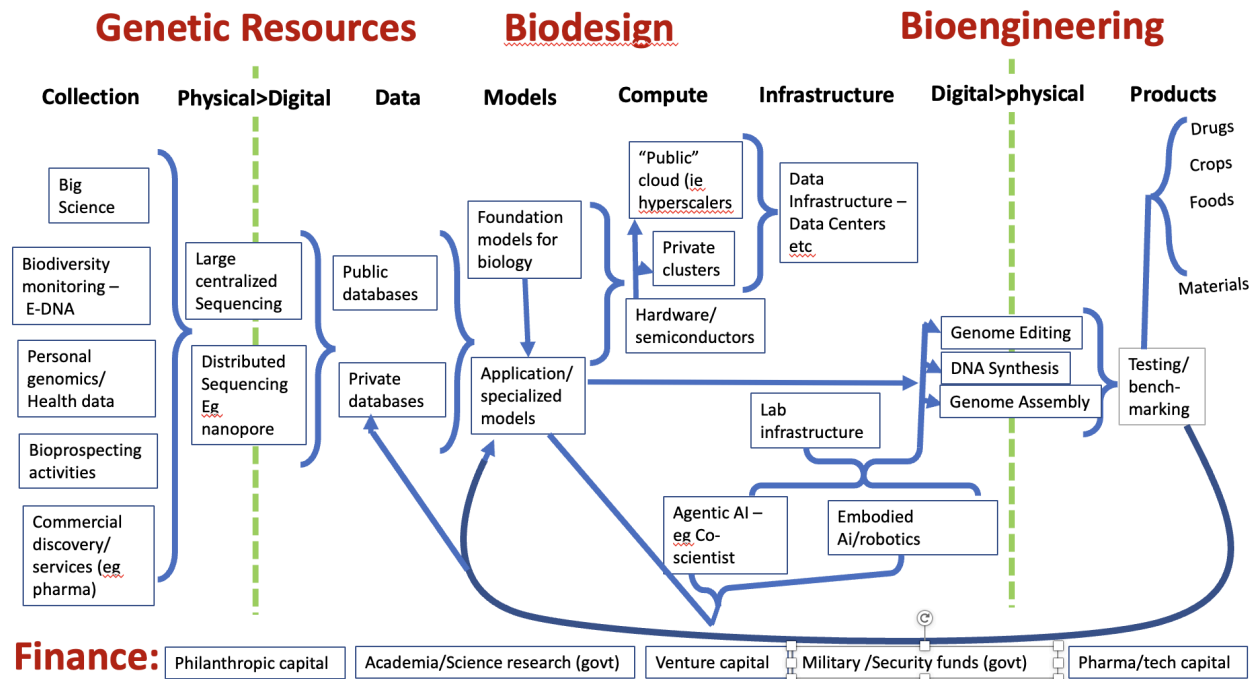


Fig A: Mapping the GenBio 'stack'

a. Data. The prime importance of the data supply chain – the return of bioprospecting and biopiracy

Across all generative AI sectors, tech companies are expressing an urgent need to access more—and better—data to train their models. Faced with persistent issues such as high error rates and so-called “hallucinations,” many firms are betting that simply increasing the volume of training data and the number of model parameters will lead to improved accuracy. Others are beginning to acknowledge that the *quality* of training data is equally, if not more, important—requiring careful selection, curation, and annotation to enhance model performance. In response, a whole sub-industry dedicated to AI *data labelling* has emerged, highlighting how central the politics and labor of data are to the future of generative AI.

This need for better data is even more pronounced in the case of GenBio models, where the biological “possibility space” being mapped is vastly larger—and far less explored—than, for example, human language. Yet the digital genetic sequences used to train GenBio models to date are mostly drawn from a small number of academic databases, originally compiled for research—not for training AI. These datasets are rarely well-annotated, often lack essential contextual information, and are heavily biased in terms of what was collected and where. Most sequences come from a limited set of

commercially valuable species, largely sourced from the Global North. One company, Basecamp Research (described more fully below), has bluntly described these datasets as “small, skewed, and stagnant.”⁴⁰

As a result, generative biology developers are intensely focused on gaining access to more—and better—biological data to feed their AI stacks. At a recent panel of GenBio entrepreneurs, when asked what they would prioritize if given funding, nearly all responded that increasing the pool of training data was their absolute top priority.⁴¹

To access more training data, GenBio firms must first acquire and sequence more biological material. This has triggered a renewed wave of large-scale bioprospecting—organized expeditions and projects targeting biodiverse and often remote regions to collect biological samples from across nature. A similar rush occurred in the 1990s during an earlier phase of genetic engineering investment and was met with fierce resistance from Indigenous and local communities, who denounced it as biopiracy.⁴² Today, a new biopiracy surge is underway—this time to feed AI systems. Samples are not only being collected but also cleaned, curated, and annotated specifically to train generative biology models.

At the forefront of this AI-driven bioprospecting is Basecamp Research, which is conducting DNA sampling in national parks and paying communities in the Global South to collect biodiversity from their own territories. Basecamp positions itself as a GenBio data infrastructure company, claiming its key commercial advantage lies in organizing the data supply chain. The company’s BaseData database is marketed as the world’s largest repository of microbial digital genetic sequences⁴³, each linked with geolocation and contextual metadata designed to improve AI model performance. Just as sequencing and synthesis companies are expected to play outsized roles in the GenBio

⁴⁰ Oliver Vince, Phoebe Oldach, Valerio Pereno et al., “Breaking Through Biology’s Data Wall: Expanding the Known Tree of Life by Over 10x using a Global Biodiscovery Pipeline,” Basecamp Research, [2025], p. 8: <https://basecamp-research.com/wp-content/uploads/2025/06/Breaking-Through-Biologys-Data-Wall-Expanding-Known-Tree-of-Life-by-10x.pdf>

⁴¹ Panel on “Bigger Data vs. Better Models – Finding the Right Scale for Bio-AI,” SynBioBeta Conference 2025, San Jose Conference Center, 07 May 2025.

⁴² See for Example RAFI Update, “Biopiracy update: The Inequitable Sharing of Benefits” Sept/Oct1997 - <https://etcgroup.org/sites/www.etcgroup.org/files/publication/437/01/raficom56biopiracyupdate.pdf>

⁴³ Oliver Vince, Phoebe Oldach, Valerio Pereno et al., “Breaking Through Biology’s Data Wall: Expanding the Known Tree of Life by Over 10x using a Global Biodiscovery Pipeline,” Basecamp Research, [2025], p. 1: <https://basecamp-research.com/wp-content/uploads/2025/06/Breaking-Through-Biologys-Data-Wall-Expanding-Known-Tree-of-Life-by-10x.pdf>

economy, Basecamp argues that whoever controls the data supply chain could become a multi-trillion-dollar player.

b. Wet/Dry translation: Synthesis and Sequencing as essential bridges and control points

One important aspect that distinguishes the GenBio stack from other AI stacks is the movement of genetic information between two fundamentally different realms: the living, “wet” world of cells, genes, and atoms, and the “dry” virtual world of bits and electronic circuits where AI operates. At one conference presentation, a GenBio entrepreneur evocatively described this interface as occurring “on the beach”—the meeting point where the wet and dry worlds converge.⁴⁴

The technologies of genetic sequencing (digitally reading genetic code from physical biological specimens) and gene synthesis (physically constructing genetic molecules like DNA from digital sequences) are central to generative biology. Sequencing and synthesis form the essential bridges between the biological and computational realms—making the entire GenBio process possible. Because the AI models at the core of this field depend entirely on the quality and scale of their training data, there is an inherent need to sequence vast quantities of high-quality biological material to feed into model training (and continual retraining). At the same time, since generative biology is inherently experimental—and AI outputs can only be validated through feedback from the real, “wet” biological world—there is an equally critical need for substantial synthesis capacity to test and verify the models’ predictions.

Whereas leading players in the generative AI space are often defined by their access to massive computational resources, the GenBio stack requires more than just compute. In this context, large-scale access to *computation*, *sequencing*, and *synthesis* are all essential for the workflow to function effectively. This has important implications for future market power within the GenBio sector. Those who control these key translational technologies—bridging the digital and biological domains—will hold disproportionate influence over the entire AIxBio stack.

c. Models – AI giants upsell compute and cloud services

At the heart of the GenBio stack are the generative AI models themselves, which attempt to predict—or “generate”—new biological sequences to engineer into living organisms or to produce novel proteins. As noted earlier, these models are built on the *transformer* architecture, which uses a mechanism called “attention” to assess the

⁴⁴ Jacob Uszkoreit speaking on a panel on ‘Hyperscale Biology: Designing Intelligence in Molecules,’ SynBioBeta Conference 2025, San Jose Conference Center, 06 May 2025.

relative importance of each element in a sequence and assign weights accordingly. The multi-trillion-dollar data and AI corporations that dominate the generative AI space through transformer-based models are now turning significant attention to generative biology, viewing it as a potentially lucrative application of their model-building technologies. For example, Jacob Uszkoreit, one of the co-inventors of the transformer architecture, now leads his own heavily funded generative biology startup, Inceptio Bio, which has raised \$120 million.⁴⁵

Alongside Google's AlphaFold and EvolutionaryScale's ESMFold (originally developed at Meta and now an independent company backed by Amazon), a key 'foundation model' in generative biology is the Arc Institute's *Evo2*, developed in collaboration with NVIDIA. Arc itself is funded by prominent tech CEOs, including the founders of Stripe and the cryptocurrency Ethereum⁴⁶. NVIDIA, which dominates the AI hardware market by supplying most of the processing GPUs used for AI computation, also provides grants and direct investments to researchers and companies working in generative biology—actively helping to build out the field.⁴⁷

When asked about NVIDIA's commercial interest in generative biology, Anthony Costa, the company's Head of Digital Biology, stated that NVIDIA has no intention of producing biotech products or operating molecular biology labs. Yet he described the sector as potentially holding "infinite value" in the coming decades.⁴⁸ Since generative biology fundamentally depends on the computational processing of massive and ever-expanding datasets, NVIDIA stands to benefit directly from the field's growth—by selling more GPUs to power its development. Similar interests appear to be driving the involvement of companies like Google, Microsoft, and Amazon—whose investment in EvolutionaryScale aligns with their dominant positions in cloud computing through Google Cloud, Azure, and AWS respectively. Academic generative biologists told us that their computational needs are already so high that they are pursuing agreements with these commercial cloud providers, as university and even pharmaceutical computing clusters are no longer sufficient.⁴⁹ The level of compute—and therefore the energy, water, and other physical resources—demanded by the AlxBio stack appears to

⁴⁵ Katie Tarasov, "After leaving Google, Jakob Uszkoreit started Inceptio to apply AI to drug development" CNBC, 12 July 2024: <https://www.cnbc.com/2024/07/12/inceptio-ceo-jakob-uszkoreit-says-ai-will-transform-pharmaceuticals.html>

⁴⁶ See Arc Institute webpage: <https://arcinstitute.org/about>

⁴⁷ Personal communication with scientist/investor Elliot Hershberg (Not Boring)

⁴⁸ Anthony Costa, Director of Digital Biology at NVIDIA speaking at 'AI & Drug Discovery' session, SynBioBeta 2025, San Jose Conference Center, 06 -May 2025 .

⁴⁹ Thematic conversations AlxBio stream at Spirit of Asilomar Conference, Monterey, CA, 23 Feb 2025.

be especially high yet remains largely unmeasured. Basecamp Research, a UK-based generative biology company, told us they operate the largest private NVIDIA GPU cluster in the country⁵⁰. Researchers also noted that a shift toward so-called *physics-based models* could significantly increase resource consumption even further. Unlike LLM's, Physics-based models attempt to digitally simulate the underlying physics of all the processes in a cell - requiring far more computation.

In addition to providing cloud infrastructure, Google and Microsoft are also marketing AI tools for laboratory and scientific use, such as Google's *co-scientist* application—leveraging their core software expertise but raising concerns among some biotech scientists about the growing removal of human judgment from research processes. As with generative AI more broadly, the GenBio space involves two types of models: large, general-purpose *foundation* or *multimodal* models (like Evo2 or ESMFold), and more specialized *application-layer* models focused on specific tasks—such as designing proteins, enzymes, or RNA. Many GenBio startups are working on these application-specific models, aiming to generate novel compounds, molecules, or engineered organisms.

d. Product Markets: Pharma dominates. Deepened disruption of natural products.

In conversations with GenBio founders, CEOs, investors, and observers, every interviewee agreed that the ultimate source of commercial value in generative biology lies in the new products companies hope to bring to market at the end of the pipeline. Overwhelmingly, these envisioned products are pharmaceutical drugs or other medical compounds. Several interviewees emphasized that the high cost and complexity of using AI for biodesign means the goal is to land a blockbuster drug—one with returns large enough to justify the investment. Pharmaceuticals are expensive to sell but relatively cheap to manufacture, with most of their value tied up in intellectual property, licensing and marketing.

Drug discovery, however, has been a field marked by diminishing returns in recent decades, despite ever-growing investments in computational tools and research infrastructure—a dynamic that one interviewee referred to as “Eroom’s Law” (Moore’s Law in reverse)⁵¹. Generative biology firms are now positioning themselves as the solution to this stagnation, hoping their AI-powered models will finally reverse the trend.

⁵⁰ Personal Communication with Basecamp Research.

⁵¹ Personal Communication with scientist/investor Elliot Hershberg (Not Boring).

The biotech sector already follows a well-established dynamic in which small drug discovery startups pursue target compounds and then partner with large pharmaceutical companies to test, license, and commercialize them. Pharma firms act much like equity investors—backing or betting on smaller companies, and potentially acquiring them if their products show promise. At least one generative biology company—Isomorphic Labs, a spin-off from DeepMind and another Google/Alphabet subsidiary—is currently testing a drug in-house⁵². However, there is little evidence so far that other major data and AI firms are willing to invest directly in building wetlabs or pharmaceutical testing infrastructure themselves. That said, DeepMind’s move into this space may foreshadow a longer-term push by Big Tech to enter the pharmaceutical sector directly—much as they have previously moved into transportation through self-driving car initiatives.

Beyond pharmaceuticals, another emerging target area for generative biology is so-called natural products such as cosmetics, flavours, and food ingredients. Major consumer goods firms—including International Flavours and Fragrances (IFF), Unilever, and Procter & Gamble—are already partnering with GenBio companies to develop AI-designed alternatives to their natural counterparts. The first AI-designed GenBio consumer product to reach the market is a *Reb M* stevia sweetener, modified from the natural compound using an AI-designed enzyme developed by Seattle, Washington-based Arzeda. (The founder of Arzeda shared the 2024 Nobel Prize in Chemistry with DeepMind’s protein scientists.) Arzeda’s engineered sweetener is now incorporated into commercial products, including under the *Splenda* brand⁵³.

ETC Group has reported extensively over the past two decades on synthetic biology’s threats to small-scale producers of natural products like vanilla, stevia, and saffron via market disruptions.⁵⁴ The introduction of generative biology tools appears to be accelerating—and heightening—this threat. In a recent public statement, even some AlxBio practitioners acknowledged the potential harm to these communities⁵⁵, while also admitting they had not seriously considered the issue.

⁵² William Gavin, “Google-backed, AI-developed drugs are headed to trial by 2026, DeepMind CEO says,” *Quartz*, 23 January 2025: <https://qz.com/google-ai-designed-drugs-deepmind-isomorphic-insilico-1851745806>

⁵³ Personal communication with Arzeda.

⁵⁴ For example, see ETC Group, *Synthetic Biology, Biodiversity & Farmers*, 2016: https://www.etcgroup.org/sites/www.etcgroup.org/files/files/etc_synbiocasestudies_2016.pdf

⁵⁵ Yana Bromberg, Russ Altman, Michael Imperiale et al. *et al.*, “Artificial Intelligence and the Future of Biotechnology,” Rice University: <https://doi.org/10.25611/1233-X161>.— This ‘Entreaty’ from the Spirit of Asilomar Conference 2025 acknowledges that “AI advances in biotechnology development also carry other potential ramifications, such as workforce skilling challenges and needs, and the potential to destabilize developed and developing bioeconomies, with potentially adverse impacts on Indigenous peoples and marginalized groups in the Global South and elsewhere.”

e. Finance. Heavy costs, high borrowing and Trump era drive military pivot

The broader field of generative artificial intelligence is already proving to be an extraordinarily expensive investment space, with capital expenditures on infrastructure—such as data centres and GPU clusters—reaching levels not seen since the railway-building booms of the late 1800s.⁵⁶ But building out a GenBio stack is even more capital intensive. In addition to the demands for computation, energy, and digital infrastructure, generative biology firms must also invest heavily in wet labs, bioprospecting, sequencing, synthesis, and securing partnerships to scale up manufacturing and bring physical biotech products to market.

At the same time, the economics of private startups and tech capital—particularly in biotechnology—are becoming increasingly strained due to broader structural shifts. High interest rates, new U.S. tariffs on physical materials, and significant cuts to public science and research funding (including under the Trump administration⁵⁷) are all influencing the strategic decisions of biotech firms. The rapid pivot of many biotech companies toward AI reflects, in part, the outsized flow of venture capital and private equity into the AI sector, where funding remains more readily available. The strong orientation toward pharmaceuticals and partnerships with Big Tech also signals a pragmatic turn toward the few remaining sources of significant capital.

Most significantly, the GenBio sector—mirroring broader trends across the AI and tech industries—is increasingly courting military and security institutions. In the U.S., biotech policy leaders are adopting an assertive, security-focused rhetoric that leverages anti-China sentiment, particularly among Republican lawmakers and the Trump administration.⁵⁸ Chinese advancements in generative biology are being framed as an existential threat to U.S. industrial and economic dominance. This geopolitical narrative is further fueled by fears that GenBio tools in the hands of adversaries could be used to develop novel pathogens or toxins.

⁵⁶ Edward Chancellor, “Victorian rail mania has lessons for AI investors,” *Reuters*, 12 July 2024: <https://www.reuters.com/breakingviews/victorian-rail-mania-has-lessons-ai-investors-2024-07-12/>

⁵⁷ Adam Feuerstein, “Biotech is in a dark place,” *STAT+ News*, 13 Feb 2025: <https://www.statnews.com/2025/02/13/biotech-industry-sentiment-crispr-medicines-fda/>

⁵⁸ Allyson Park, “U.S., China Engage In Race to Militarize Biotechnology” *National Defense Magazine* 8/1/2025 - <https://www.nationaldefensemagazine.org/articles/2025/8/1/us-china-engage-in-race-to-militarize-biotechnology>

In response, the US's National Security Commission on Emerging Biotechnology and others recommend building a robust GenBio infrastructure—funded through national defense budgets—as a matter of strategic necessity. A 2025 report by the Commission proposes a large-scale federal program to sample genetic material from public lands for AI training datasets, alongside the creation of a national backbone of sequencing and synthesis facilities to support a domestic GenBio ecosystem.⁵⁹

The result of this appeal to military funding is a deepening entanglement between the emerging GenBio industry and U.S. national security ideology. This raises serious concerns about how human rights, equity and global justice considerations will be addressed—if they are acknowledged at all.

Section IV: Insights

Starting points for social, environmental and policy considerations from a data justice and human rights perspective

The assumption that generative biology can be governed in just, safe, and equitable ways overlooks the deeper structural harms embedded in its very development. GenBio depends on transforming living materials into data, enclosing them in opaque AI systems, and converting them into proprietary assets. Rather than focusing solely on regulatory frameworks for managing biological and digital components, national and international policy must critically evaluate—and, we argue, ultimately reject—the large-scale deployment of GenBio technologies. A precautionary approach grounded in human rights, ecological justice, and democratic control must take precedence over efforts to retrofit legitimacy to an extractive and destabilizing enterprise.

Nevertheless, governance frameworks remain necessary. A just regime must be capable of tracking the transformations between biological and digital domains and of addressing how interventions in the digital realm reshape power relations, economies, and rights in the living world. This requires legal and social tools not only to regulate risks but also to defend livelihoods, protect cultures, and uphold diverse ways of living.

At the core of the GenBio stack are raw materials that are simultaneously biological and digital—molecular descriptions of DNA, RNA, proteins, and associated contextual information such as ecosystems, geographies, and biodiversity. The capture of these

⁵⁹ [National Security Commission on Emerging Biotechnology](#) (NSCEB), “[Charting the Future of Biotechnology: An Action Plan for American Security and Prosperity](#).” April 2025. For example, “Congress should authorize and fund the Department of Interior to create a Sequencing Public Lands Initiative to collect new data from U.S. public lands that researchers can use to drive innovation,” p. 38- 41.

resources parallels other forms of digital extraction: social data harvested by platforms, health data by hospitals and devices, or environmental monitoring data by sensors. In each case, questions of justice arise not only at the point of extraction but also in downstream impacts, when data is repurposed in ways that affect ecosystems, economies, and communities.

This dynamic depends on the creation of a global infrastructure for the extraction and surveillance of biodata from plants, soils, waters, and human and animal tissues. Establishing such infrastructure risks reviving colonial practices and power relations, particularly in societies that have long stewarded biological resources.

Examples of these large-scale projects include:

- *The Earth Biogenome Project*, a global consortium launched in 2018 aiming to sequence, catalog, and characterize the genomes of all of Earth’s eukaryotic biodiversity within ten years.⁶⁰
- *The Earth Microbiome Project*, begun in 2010, a collaboration of over 500 researchers cataloguing microbial diversity.⁶¹
- *BaseCamp Research’s BaseData*, a private database of genetic sequences from environmental samples collected through ongoing expeditions in biodiversity hotspots.⁶²

While often presented as serving the “public good,” the same digital sequence information harvested by these projects is already moving into AI training pipelines or proprietary databases. In practice, this converts what are ostensibly common goods into private assets—enclosing biodiversity in ways that benefit investors while sidelining the communities who steward these resources.

With the growing emphasis on training multimodal “foundation models” on millions or billions of parameters, commercial pressure to scale up biodata harvesting is intensifying—deepening historical injustices and creating new forms of *digital biocolonialism*.

Framing GenBio within the lens of *digital biocolonialism* highlights that the issue is not only one of consent or benefit-sharing, but of geopolitical and economic domination. Just as earlier forms of colonialism appropriated land and resources, today’s enclosures

⁶⁰ <https://www.earthbiogenome.org/>

⁶¹ <https://earthmicrobiome.org/>

⁶² As described in Oliver Vince, Phoebe Oldach, Valerio Pereno et al., “Breaking Through Biology’s Data Wall: Expanding the Known Tree of Life by Over 10x using a Global Biodiscovery Pipeline,” Basecamp Research, [2025]: <https://basecamp-research.com/wp-content/uploads/2025/06/Breaking-Through-Biologys-Data-Wall-Expanding-Known-Tree-of-Life-by-10x.pdf>

operate through the capture and privatization of genetic sequences, biodiversity information, and health data. The South once again risks becoming a frontier for extraction—this time of biodata—while wealth, power, and control accumulate elsewhere.

Consent frameworks are also increasingly inadequate. Many genetic sequences now used to train GenBio models were collected before generative AI existed, meaning that prior informed consent cannot reasonably be assumed. In practice, consent is often ignored or reduced to a purely formalistic process, stripping it of substantive meaning. This gap underscores the urgency of reassessing benefit-sharing obligations and establishing governance mechanisms that respond to AI-enabled biological design.

From a data justice perspective, these developments reflect a broader pattern of dispossession: communities' data is extracted without consent, repurposed for private or state interests, and embedded into opaque systems that consolidate elite agendas. Recognizing this dynamic is essential if civil society is to intervene effectively in governance arenas and push for precaution and recognition of collective rights.

The following topics are useful starting points for further investigation and policy attention:

1. **Bioeconomy: Greater resource extraction and violations of rights divided by AI-enabled GenBio**

Computation by data centers already accounts for up to 2% of global energy use—a figure projected to rise to 3–4% by 2030 due to the growth of generative AI.⁶³ Water consumption for cooling and electricity generation is similarly intensive, placing additional pressure on deserts, fragile ecosystems, and agricultural regions. Meanwhile, the surge in demand for chips, networking components, and devices is fueling an explosion in mineral extraction, with serious ecological, geopolitical and human rights consequences.

While tech companies are investing hundreds of billions of dollars to expand AI infrastructure and computing capacity, the primary beneficiaries of this boom include energy corporations (particularly fossil fuel companies), mining interests, and private water utilities. The primary casualties may be the human right to water (as outlined in

⁶³Anon., “AI Is Poised to drive 160 percent increase in Data Center Power Demand”, Goldman Sachs, 14 May 2024: <https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand>

ICESCR, Article 11)⁶⁴, the right to a clean, healthy, and sustainable environment⁶⁵, and the resilience of Earth’s life-support systems.

Our investigation suggests that much of the current hype around GenBio is being driven—at least in part—by market-making efforts from major AI and data infrastructure firms such as Google, NVIDIA, and Microsoft, which stand to profit directly from a computationally intensive sector. These same firms also benefit reputationally from unproven panacea narratives, suggesting that GenBio products—such as biomaterials, alternative proteins, or new pharmaceuticals—can reduce the climate- or resource-intensity of industry, while solving or curing all manner of ills.

Civil society and policymakers must scrutinize the real-world resource demands of GenBio and interrogate the sustainability and health claims being made. Particularly relevant here is *Jevons Paradox*,⁶⁶ which shows that increases in industrial efficiency often lead to increased, rather than reduced, extraction and energy use.

2. Biopiracy: A renewed bio-prospecting rush

The generative biology industry’s existential dependence on large, high-quality training datasets is fueling a new bioprospecting rush. Companies like Basecamp Research are striking deals with communities, national parks, and governments to access soils, seeds, waters, flora, and fauna.⁶⁷ Yet it remains unclear whether local communities fully understand what they are handing over—or for what purposes their biological heritage will be used. This raises urgent concerns about consent, transparency, and exploitation.

⁶⁴

<https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights>; *United Nations Committee on Economic, Social and Cultural Rights (CESCR). General comment no. 15 (2002). The right to water (arts. 11 and 12 of the International Covenant on Economic, Social and Cultural Rights)*, E/C.12/2002/11, 2003.

⁶⁵ *United Nations General Assembly. The human right to a clean, healthy and sustainable environment : resolution / adopted by the General Assembly. A/RES/76/300, 2022.*

⁶⁶ Greg Rosalsky, “Why the AI world is suddenly obsessed with a 160-year-old economics paradox,” *Planet Money Newsletter*, 04 February 2025: <https://www.npr.org/sections/planet-money/2025/02/04/g-s1-46018/ai-deepseek-economics-jevons-paradox>

⁶⁷ Oliver Vince et al, “Breaking Through Biology’s Data Wall: Expanding the Known Tree of Life by Over 10x using a Global Biodiscovery Pipeline”:

<https://basecamp-research.com/wp-content/uploads/2025/06/Breaking-Through-Biologys-Data-Wall-Expanding-Known-Tree-of-Life-by-10x.pdf>

The growing demand for DSI is likely to accelerate biological data grabs that amount to a new wave of digital biopiracy. Conservation initiatives tied to the *Kunming-Montreal Global Biodiversity Framework* are already engaged in large-scale monitoring and extraction of environmental DNA (eDNA) and biodiversity data⁶⁸—datasets that can easily be absorbed into GenBio models without the knowledge or consent of the communities connected to source ecosystems. Particularly concerning is the \$256 million acquisition of 23andMe’s private DSI database by pharmaceutical firm Regeneron, now holding more than 15 million individuals’ DNA and billions of associated health datapoints⁶⁹. These sequences could be ingested into AI models for uses never imagined or agreed to by those who provided samples.

Such practices raise not only questions of access and benefit sharing under the *Convention on Biological Diversity and its Nagoya Protocol*, but also deeper data justice concerns. Once biological information is digitized, it is severed from its origins, obscuring community connections and converting shared heritage into proprietary assets. Once fed into AI models, these sequences become irretrievable and untraceable, making restitution or benefit-sharing nearly impossible.

Two harms are especially important to distinguish:

- Biopiracy and Enclosure: the appropriation of genetic resources and knowledge historically stewarded by communities, disrupting cultural and governance systems that cannot be monetized or replaced.
- Livelihood Displacement: the introduction of GenBio products that undercut or replace traditional production, threatening peasant and Indigenous livelihoods.

In this sense, digitization itself becomes a new form of enclosure—transforming living materials into proprietary data assets that communities can no longer access or govern. This mirrors broader dynamics in the digital economy, where social, health, and environmental data are extracted without consent and embedded in opaque systems that reinforce elite agendas.

As AI models scale to billions of parameters, the commercial pressure to accelerate biodata extraction is intensifying. By scaling up harvesting under opaque and extractive

⁶⁸ Emma Granqvist et al, “The transformative potential of eDNA-based biodiversity impact assessment”, *Current Opinion in Environmental Sustainability*, Volume 73, 2025, <https://doi.org/10.1016/j.cosust.2025.101517>.
<https://www.sciencedirect.com/science/article/pii/S1877343525000107>

⁶⁹ Rhys Northwood, “The Genetic Goldmine: Why Regeneron’s \$256M 23andMe Acquisition is a Biotech Masterstroke,” *Ainvest*, 19 May 2025: <https://www.ainvest.com/news/genetic-goldmine-regeneron-256m-23andme-acquisition-biotech-masterstroke-2505/>

terms, GenBio risks reviving colonial logics in digital form — a new phase of *digital biocolonialism*. Addressing these harms requires updated governance mechanisms, robust enforcement of prior informed consent, and recognition that digitization itself can constitute dispossession.

A resurgent wave of digital biopiracy may violate communities' rights to benefit from the protection of the moral and material interests resulting from their resources and knowledge—as enshrined in the International Covenant on Economic, Social and Cultural Rights (ICESCR, Article 15(c))⁷⁰ and clarified in CESCR General Comments 17⁷¹ and 25⁷².

While these dynamics of enclosure and digital biocolonialism highlight the structural injustices of GenBio, their impacts are most tangible in the daily lives of those who depend on biodiversity for survival.

3. Biocultural rights: Livelihoods and cultures under threat

An estimated 20 million small-scale farmers and agricultural workers—primarily in the Global South—depend on cultivating 200 to 250 botanical crops grown across roughly 250,000 hectares worldwide.⁷³ These crops supply natural ingredients to the flavour and fragrance industry, as well as to pharmaceutical companies, and represent a vital source of income and livelihood. In many cases, they are also central to traditional and Indigenous cultures, embedded in cosmologies and protected through customary laws and stewardship practices.

Generative biology poses a twofold threat. First, the development of synthetic substitutes risks displacing livelihoods. Genetic sequences taken from high-value crops can be used to engineer “nature-identical” replacements—undermining the value of the original crops for growers. A prominent example is stevia: the company Arzeda has used AI to design an enzyme that converts the widely available Rebaudioside A

⁷⁰

<https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights>

⁷¹ United Nations Committee on Economic, Social and Cultural Rights (CESCR), *General comment no. 17 (2005). The right of everyone to benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production*, E/C.12/GC/17, 2006.

⁷² United Nations Committee on Economic, Social and Cultural Rights (CESCR), *General comment No. 25 (2020) on science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights)*, E/C.12/GC/25, 2020.

⁷³ See, for example, ETC Group, *Synthetic Biology, Biodiversity & Farmers*, 2016: https://www.etcgroup.org/sites/www.etcgroup.org/files/files/etc_synbiocasestudies_2016.pdf

molecule into Rebaudioside M, marketed as a more palatable sweetener. Splenda now sells this enzyme-derived product as “natural,” directly competing with farmer-grown stevia originally cultivated by the Guaraní people of Paraguay. Even where synthetic products have not yet reached the market, the mere prospect of substitution is enough for major flavour and fragrance companies to negotiate lower prices, squeezing small producers further.

Second, these shifts represent a deeper violation of *biocultural rights*. Natural products like saffron, vanilla, and stevia are not merely commodities; they are woven into systems of knowledge, health, and spirituality. When AI-driven GenBio systems enclose these resources and transform them into proprietary digital assets, they undermine sovereignty over genetic resources and sever cultural relationships to biodiversity.

Consumers, too, face risks. Products derived from GenBio may be poorly labeled—or not labeled at all—despite the possibility of algorithmic errors producing unexpected side effects. This raises concerns about transparency, health, and informed choice.

From a rights perspective, GenBio may undermine:

- the rights of peasants and rural peoples (as recognized under UNDRIP),
- the rights of Indigenous Peoples (UNDRIP), and
- the rights of all individuals to health, safety, and informed choice (ICESCR and other human rights instruments).

By displacing livelihoods and eroding cultural heritage, generative biology deepens historical injustices and places fragile economies and ways of life at risk. Safeguarding biocultural rights requires not only strengthening benefit-sharing and consent mechanisms but also confronting how digital enclosures convert heritage into speculative assets, leaving producers, communities, and consumers exposed.

Alongside these livelihood and cultural risks, generative biology also raises profound concerns for biosafety. Even where communities are bypassed, states remain bound by international human rights law to regulate and monitor new technologies that may endanger life, health, or ecosystems.

4. Biosafety: Risk assessment, governance and oversight under strain by AI

If generative biology accelerates the design of new organisms, proteins, and molecules, national and international biosafety mechanisms could be stretched beyond their limits. The risk is amplified by the “black box” nature of many generative models, which cannot explain the rationale behind their genetic engineering decisions. Regulators may thus

be left effectively blindfolded, tasked with approving or rejecting products whose underlying logic is opaque.

Evidence is already emerging of how these dynamics enable regulatory evasion. A report by *Save Our Seeds*⁷⁴ found that generative biology techniques can be tuned to make fewer than 20 base pair changes—alterations small enough to potentially bypass scrutiny under existing European biosafety rules. In this way, a longstanding loophole could be widened into a technological highway for escaping oversight.

International human rights law is clear: states have duties to prevent corporations and individuals from violating rights, to regulate and monitor private actors, to require robust environmental and human rights impact assessments, and to act with precaution when activities may threaten the free exercise of rights. They must also guarantee access to information, public participation, and access to justice for all individuals potentially affected by GenBio innovations. These obligations are explicitly reinforced in *Advisory Opinion 23 of the Inter-American Court of Human Rights on Environment and Human Rights*⁷⁵, and in *General Comment 25 of the UN Committee on Economic, Social and Cultural Rights*.⁷⁶

Taken together, the convergence of opaque AI design, weak regulatory mechanisms, and accelerated biotech development poses a biosafety challenge of unprecedented scale. Unless precaution and accountability are embedded into GenBio governance, society risks approving products and processes without ever fully understanding their consequences.

5. Health Justice: the meaning of chasing ‘blockbuster’ pharma compounds and philanthrocapitalists involvement

The commercial model of “health” is heavily skewed toward reinforcing costly, privatized healthcare systems and delivering treatments for wealthy populations—such as high-margin obesity drugs—while continuing to neglect orphan diseases, socially rooted health conditions, and the structural determinants of ill health.

⁷⁴ Benno Vogel, “When chatbots breed plant varieties,” *Save Our Seeds*, January 2025: https://www.saveourseeds.org/wp-content/uploads/EN-with-summary_online.pdf

⁷⁵ Inter-American Court of Human Rights. Advisory Opinion OC-23717. The Environment and Human Rights.

⁷⁶ United Nations Committee on Economic, Social and Cultural Rights (CESCR), General comment No. 25 (2020) on science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights), E/C.12/GC/25, 2020.

The narrative that generative biology will cure (all!) disease functions as a powerful public relations tool for the AI industry, which faces growing public skepticism, concern over energy and resource use, and criticism of its oligarchic business models. Even if actual medical breakthroughs prove elusive, there remains enormous value for Big Tech in sustaining the illusion of progress. In this context, the AIxBio field is already attracting substantial support from tech-aligned *philanthrocapitalist* actors. Foundations such as the Gates Foundation (Microsoft)⁷⁷, Open Philanthropy⁷⁸, the Chan Zuckerberg Initiative (Meta)⁷⁹, the Ellison Foundation (Oracle)⁸⁰, and the Bezos Earth Fund (Amazon)⁸¹ are all actively supporting GenBio-related efforts. Their investments not only generate good publicity but also help maintain the value of the tech stocks upon which these foundations are built.

It must be recalled that international human rights frameworks require precaution in the face of uncertainty. As CESCR General Comment 25 (para. 56) and Inter-American Court of Human Rights Advisory Opinion 23 (para. 180) state: *In the absence of full scientific certainty, when an action or policy may lead to unacceptable harm to the public or the environment, actions must be taken to avoid or reduce that harm.*

“Unacceptable harm” includes impacts that are:

- (a) threatening to human life or health;
- (b) serious and effectively irreversible;
- (c) inequitable to present or future generations; or
- (d) imposed without adequate consideration of the human rights of those affected.

Technological and human rights impact assessments are vital tools to identify risks early in the process of scientific and commercial innovation. The growing role—and self-interest—of philanthrocapitalist entities in shaping the future of the GenBio field must be closely scrutinized and publicly debated.

⁷⁷ Eg BMGF has funded Exscientia a AIxBio drug company - see <https://www.gatesfoundation.org/about/committed-grants/2020/10/inv004656>

⁷⁸ Eg <https://www.openphilanthropy.org/grants/securebio-ai-biological-capabilities-dashboard/>

⁷⁹CZI Sets 4 Scientific Grand Challenges To Transform Human Health at the Intersection of AI and Biology - <https://chanzuckerberg.com/blog/ai-biology-grand-scientific-challenges/>

⁸⁰ <https://www.eit.org/news/professor-jason-chin-joins-eit-to-lead-its-new-generative-biology-institute>

⁸¹ <https://www.bezosearthfund.org/initiatives/ai-for-climate-and-nature>

6. Militarization: Defense-focused GenBio threatens rights, transparency and equity

Synthetic biology and the broader biotech industry have long-standing ties to military funding—links that are often obscured from public scrutiny. For more than a decade, the U.S. Defense Advanced Research Projects Agency (DARPA) has been one of the most significant public funders of synthetic biology research. What marks a shift in the current moment, however, is the overt rhetorical framing: previous appeals to ‘green’ innovation or ‘global health’ benefits are now giving way to a more aggressive narrative of safeguarding American dominance in the face of geopolitical threats.⁸² Generative biology is increasingly positioned as a tool of national security, echoing the sentiments of the U.S. MAGA movement and promoting a muscular techno-nationalism.

This realignment—linking the bioeconomy to military outcomes—raises serious concerns about how GenBio will be applied, posing new challenges to international humanitarian law. It risks entrenching inequities, legitimizing narratives of domination and superiority, and even reviving eugenics logic under the guise of national defense. As more GenBio infrastructure, data pipelines, and research funding are reclassified as “essential to national security,” the already limited transparency and accountability in this field may shrink further, narrowing space for democratic oversight, public debate, and dissent.

Opportunities for Civil Society action

GenBio is advancing faster than governance systems can respond, leaving communities exposed to new forms of dispossession. Yet existing international frameworks and emerging policy arenas do provide entry points to contest enclosure and demand accountability. Civil society can use these spaces to push for precaution, equity, and collective rights in the face of accelerating biodata extraction.

UN Convention of Biological Diversity (CBD)

More than 15 years of negotiations under the *Convention on Biological Diversity* (CBD) have established the foundational principle that Digital Sequence Information (DSI) should fall under the same broad Access and Benefit Sharing (ABS) obligations as physical biological material.⁸³ This protracted process culminated in October 2024 with the creation of the *Cali Fund*—a multilateral mechanism where commercial users of DSI

⁸² [National Security Commission on Emerging Biotechnology](#) (NSCEB), “[Charting the Future of Biotechnology: An Action Plan for American Security and Prosperity](#),” April 2025.

⁸³ Digital Sequence Information on Genetic Resources - <https://www.cbd.int/dsi-gr>

can contribute financially in lieu of negotiating individual ABS agreements with the original providers.

Crucially, the agreement establishing the Cali Fund⁸⁴ explicitly acknowledged the rise of AI and GenBio by identifying AI and software companies as *commercial users* of DSI—entities that may be required to contribute to the fund. A newly formed steering committee, which will meet twice annually and include civil society representation, has been tasked with overseeing the operation of the fund. This offers one potential avenue for advancing more just and accountable governance of data flows within the GenBio economy.

In parallel, the CBD also maintains an ongoing agenda on *Synthetic Biology*⁸⁵, focused on horizon scanning, technology assessment, and monitoring of emerging issues. This process provides another important platform where civil society can propose governance measures specifically targeting the extraction, use, and commodification of biodata and GenBio technologies—consistent with the Convention’s mandate to ensure “the fair and equitable sharing of benefits arising from the utilization of biodiversity.”

Finally, DSI governance is also anchored in *Target 13* of the Global Biodiversity Framework, which calls on Parties to: “*Increase the sharing of benefits from genetic resources, digital sequence information, and traditional knowledge.*” This target, and its associated action agenda, can serve as a mechanism for pushing governments and corporations to confront the inequities embedded in the GenBio data pipeline.

UN Food and Agriculture Organization (FAO)

Parallel opportunities for governance may also exist within bodies established under the UN Food and Agriculture Organization—particularly the *International Treaty on Plant Genetic Resources for Food and Agriculture* (commonly known as the *Seed Treaty*)⁸⁶—which has been actively grappling with the issue of Digital Sequence Information (DSI) for several years. The Seed Treaty is one of the few international legal instruments that explicitly recognizes Farmers’ Rights, including the right to save, use, exchange, and sell seeds, and to participate in decision-making about plant genetic resources. However, the rise of digital tools that extract genetic sequences from seeds and convert them into data—without access agreements or benefit-sharing—poses a serious threat to those rights. The use of DSI in AI-driven GenBio models risks

⁸⁴ CBD Decision CBD/COP/DEC/16/ 2 - <https://www.cbd.int/doc/decisions/cop-16/cop-16-dec-02-en.pdf>

⁸⁵ UN CBD page on Synthetic Biology see <https://www.cbd.int/synbio>

⁸⁶ <https://www.fao.org/plant-treaty/en>

decoupling biological innovation from the territories, communities, and knowledge systems that have nurtured and developed these resources over generations. As DSI flows freely into proprietary AI systems while farmers remain excluded from its benefits, there is a growing risk of digital enclosures replacing physical ones. Strengthening DSI governance within the Seed Treaty—guided by the principle of free, prior and informed consent and the fair and equitable sharing of benefits—is crucial to protecting the rights of peasants and Indigenous peoples and ensuring that emerging technologies do not undermine food sovereignty.

Digital Privacy Actions

The repurposing of Digital Sequence Information (DSI) for training AI models in GenBio represents a significant shift in the use of data that was often originally collected for public-interest scientific research or conservation purposes. This re-use raises serious legal and ethical concerns, particularly where consent was never given—or could not have been imagined—for such commercial and computational applications. There may be grounds for complaints or legal actions to be brought before national and regional privacy authorities regarding violations of the original terms under which biodata was collected. For example, in the European Union, the holding and processing of certain types of biodata may fall under the General Data Protection Regulation (GDPR), which explicitly covers *genetic data* as a special category of sensitive information (see Recital 34)⁸⁷. The use of such data outside its original scope—especially without proper transparency, consent, or benefit-sharing—should be scrutinized under existing privacy and data protection frameworks.

Civil society can demand transparency and accountability from public research institutions, conservation programs, and biotech firms regarding the secondary use of biodata—especially where this use may violate original terms of consent or ethical research standards. Organizations can file complaints or inquiries with national and regional data protection authorities (such as those enforcing the EU’s GDPR) to trigger regulatory reviews of genetic and biological data’s use in AI development and practice. Advocates can also push for new national and international guidelines that clarify the legal status of DSI and assert that its use in AI must be subject to human rights, privacy, and benefit-sharing obligations. Furthermore, civil society can work to build public understanding and grassroots capacity—particularly in biodiverse regions—on data sovereignty and digital biopiracy, empowering communities to assert control over how their biological heritage is being extracted, digitalized, and commercialized.

⁸⁷ Recital 34. EU GDPR - <https://www.privacy-regulation.eu/en/recital-34-GDPR.htm>

National, Regional and Global Policy on AI Training Data

While the repurposing of Digital Sequence Information (DSI) for generative biology models raises serious concerns around digital sovereignty and data rights, existing Access and Benefit Sharing (ABS) frameworks—particularly the CBD’s tacit recognition that AI applications of DSI should trigger contributions to the Cali Fund—offer a valuable precedent. This recognition can help catalyze a broader conversation about the need to properly compensate communities whose biological and environmental data are extracted, digitized, and fed into AI models without their prior knowledge or consent. While CBD negotiations on DSI have focused narrowly on genetic information, similar forms of digital data—such as data from soil, water, climate, and species observations—are increasingly being collected under the guise of biodiversity conservation. These environmental datasets are also valuable inputs for AI systems, including those used in agriculture, synthetic biology, and conservation technologies. There is growing scope, therefore, for the ABS framework and the Cali Fund to be extended or adapted to include other forms of ecological and community-derived data, helping ensure that communities retain agency over how – or if – their territories and knowledge systems are transformed into digital capital.

This moment offers an important opportunity for civil society to intervene in global governance spaces—such as the CBD and emerging digital rights forums—to demand stronger protections, binding transparency obligations that ensure free prior informed consent is provided by communities, and redistributive mechanisms. By linking digital justice with collective rights, civil society can play a pivotal role in exposing new forms of biopiracy and pushing for policies that centre prior informed consent in the age of AI. At the national level, civil society should advocate for national access and benefit sharing laws to explicitly include AI environmental data sets ensuring communities retain sovereignty over digitized ecological knowledge.

Final Thoughts

Generative Biology is unfolding with astonishing speed, propelled by massive investments from tech giants, venture capital, and aligned philanthrocapitalist foundations. Its promises—disease eradication, sustainable materials, revolutionized agriculture—dominate headlines, while the risks accumulate largely out of public view. Oversight, precaution, and democratic deliberation lag dangerously behind. We are witnessing the rise of a powerful new industrial sector before rules are even written, let alone enforced.

At its core, GenBio accelerates long-standing dynamics of dispossession: transforming living resources into digital data, enclosing them within opaque AI systems, and

converting them into proprietary assets. These processes threaten to deepen inequality, displace livelihoods, erode biocultural heritage, and overwhelm already fragile biosafety regimes. They also entrench corporate power and techno-nationalist agendas, narrowing the political space for communities to assert rights and for societies to shape technological trajectories in the public interest.

Yet this moment also presents a powerful opportunity for resistance and re-imagination. Civil society, Indigenous peoples, farmers, and local communities can assert collective governance over biological and digital resources, demand enforceable rules on data collection and use, and insist on precautionary limits. Legal interventions, public policy engagement, and grassroots mobilization all have a role to play in ensuring that the hype over the deployment of AI in the design of living organisms is signaled as a clear threat, and that this phenomena does not multiply the injustices of earlier biotech and digital enclosures.

Lessons learned from past struggles have shown that without community sovereignty, transparency, and binding safeguards, generative biology will exacerbate structural inequities into our food, health, and ecological systems, and deepen corporate control.