BLACK BOX BIOTECH

Integration of Artificial Intelligence with Synthetic Biology: Addressing the risks, hype and inequities underpinning Generative Biology

SEPTEMBER 2024

AFRICAN CENTRE FOR BIODIVERSITY

etc GROUP

TWN Third World Network

The African Centre for Biodiversity (ACB) is committed to dismantling inequalities and resisting corporate industrial expansion in Africa's food and agriculture systems.

AFRICAN CENTRE FOR BIODIVERSITY

www.acbio.org.za
PO Box 29170, Melville 2109, Johannesburg, South Africa.
Tel: +27 (0)11 486-1156

Researched and written by Jim Thomas, Scan the Horizon, formerly with the ETC Group
External review by Maywa Montenegro de Wit, Assistant Professor, Environmental Studies Department, University of California, Santa Cruz, which entailed substantial inputs, and also by Dan McQuillan, Lecturer in Creative and Social Computing, Goldsmiths College, UK
Editorial guidance and oversight: ACB executive director Mariam Mayet
Copy edit by Liz Sparg
Design and layout: Xealos Design Consultants

Acknowledgments
We would like to thank ETC Group and Third World Network for reviewing the text and collaborating with the ACB on the issues raised in this briefing. The ACB gratefully acknowledges the financial support of several donors though the views expressed may not necessarily reflect the views of our donors.

Table of Contents

Abbreviations 4

Overview 5

The issue 6
Policy implications 7
How we got here: Background to the Convention on Biological Diversity (CBD) 7
The basics: What are 'synthetic biology' (SynBio) and 'artificial intelligence' (AI)? 12

How AI and SynBio converge 14

1. Biodesign: Generative biology 15
Moving genetic engineering into a 'black box' (and out of CBD oversight) 18
What parts of genetic systems can now be biodesigned by AI? 20
2. Bioproduction: Boosting biotechnology 21
3. Biodigital systems: The example of 'robot-ready' crops 22
4. Biocomputation: The example of brain organoids 22

The generative biology rush: Technology titans move into biology (and biotech/biodiversity policy) 23

Implications for the CBD: Five urgent questions 26

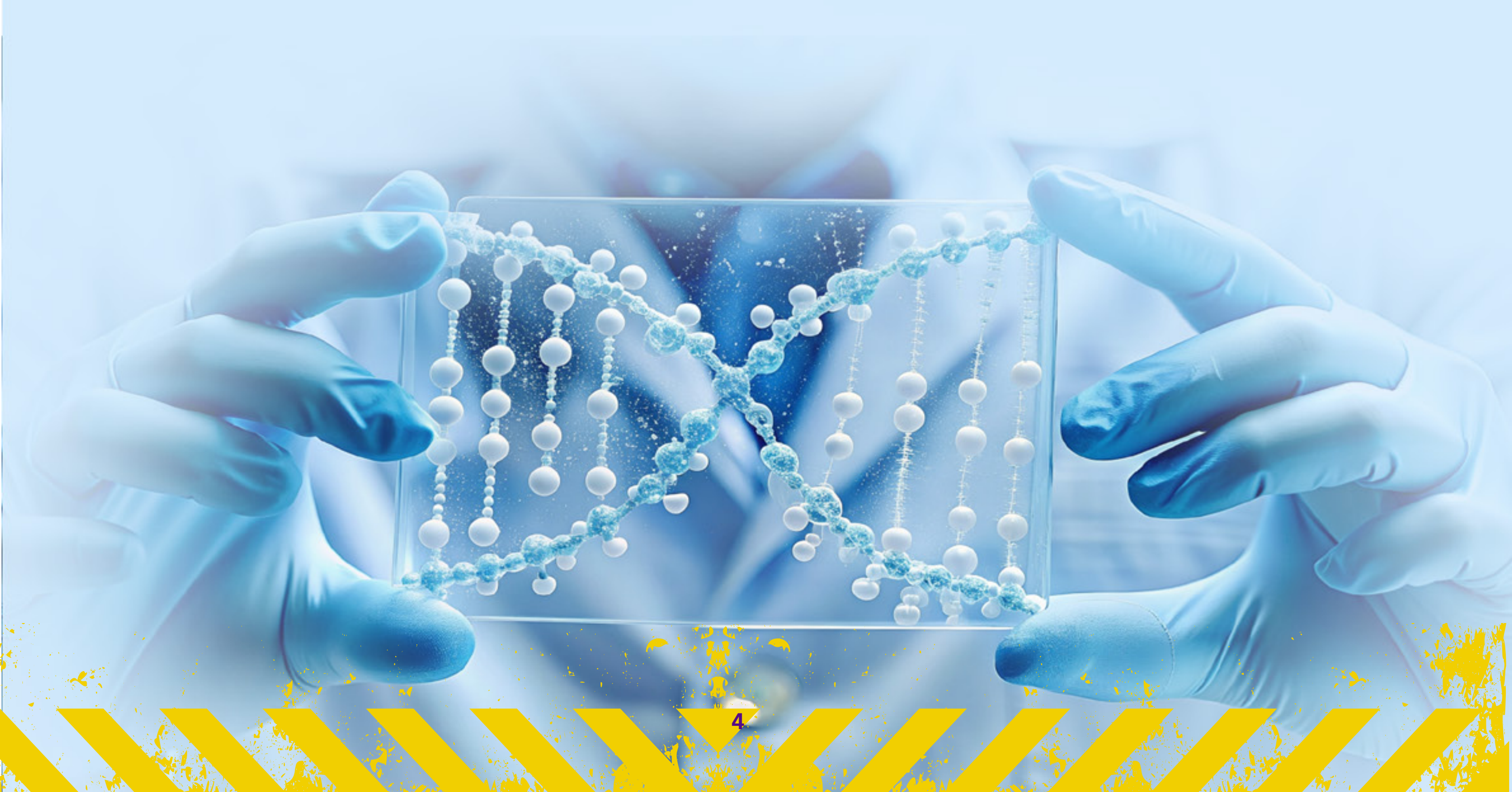
1. Does AI generative biology undermine access and benefit-sharing arrangements of the Nagoya Protocol and the governance of DSI? 26
2. Does AI generative biology undermine biosafety arrangements of the Cartagena Protocol on Biosafety? 27
3. Does AI generative biology pose biosecurity/bioweapons risks? 29
4. Will the integration of AI with SynBio improve or worsen health and sustainability? 29
5. What are the implications of AI/SynBio integration for traditional knowledge and practices? 31

A way forward for the CBD 32

Glossary of key terms 34

Abbreviations

AHTEG	Ad Hoc Technical Expert Group
AI	Artificial intelligence
CBD	Convention on Biological Diversity
COP	Conference of the Parties
DSI	Digital sequence information
GMOs	Genetically modified organisms
GURTs	Genetic use restriction technology
LLMs	Large language models
mAHTEG	Multidisciplinary Ad Hoc Technical Expert Group (on Synthetic Biology) SBSTTA
	Subsidiary Body on Scientific, Technical, and Technological Advice
SynBio	Synthetic biology



Overview

The United Nations Convention on Biological Diversity (CBD) and its protocols are the premier global instrument for oversight of modern biotechnology and have successfully adapted as new developments in the field have emerged. Today's 'generative' artificial intelligence (AI) tools, better known for text chatbots such as ChatGPT, are now being applied to generate novel digital sequences for genetically modified organisms (GMOs) and proteins. These models, developed by large digital technology firms, are trained on vast quantities of digital DNA or protein sequences, find patterns, and apply them to create novel digital sequences. This new industry, dubbed 'generative biology' by its advocates, is accompanied by promises that such AI 'biodesign' tools can deliver an array of technofixes for a more sustainable world. The claims now being made for generative biology echo the speculations made for previous cycles of GMOs and first-generation AI systems. Each fell short of initial commercial hype as new problems emerged.

Beyond the hype, the field of generative biology represents a bold grab on the world's digital sequence information (DSI) on genetic resources. Whether or not reliable bioproducts ever emerge, we are already witnessing significant investment flowing to these developments and powerful digital players pumping the hype cycle to generate fascination, hope, and investment in generative biology. With legitimacy granted by Silicon Valley funders, AI firms will likely attempt to significantly change governance conditions for modern biotechnology – claiming the CBD's core approach of defending precaution and equity is now outmoded in an age of AI. The CBD must be ready to separate fact from fiction. It must redouble its ground-truthing effort of horizon scanning, technology assessment, and monitoring by commissioning a sensible process to understand the implications of SynBio's integration with AI. CBD Parties should examine, and strengthen, the Convention's oversight arrangements for biotechnology in the face of a rapidly changing technological landscape.

COP15, UN Biodiversity - © Wikimedia Commons

The issue

Applying AI tools in genetic engineering and biotechnology involves important shifts that are more carefully examined in this briefing:

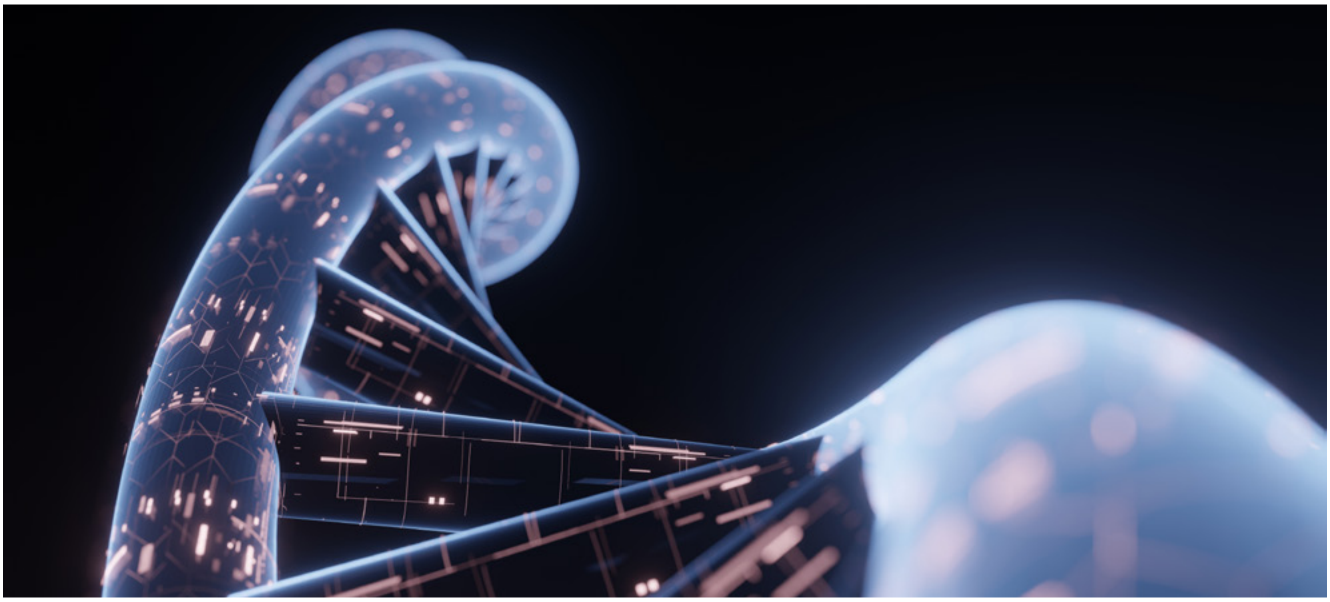
a) 'Black box' biotechnology: Using AI to digitally design genetic systems moves the process of genetic engineering into an unknowable algorithmic 'black box' where individual design decisions can be neither traced nor explained. This opaque 'black box' character of AI biodesign, which is described in greater detail below is inherent to generative AI. In the context of biotechnology it challenges current biosafety assessment capabilities, undermines monitoring requirements, and removes the traceability required to ensure fair and equitable benefit sharing from the use of genetic resources or to support systems of liability and redress. Already difficult arrangements over tracing the origin of DSI are about to become wickedly challenging or impossible as a result of using AI tools.

b) Boosted biotechnology: The industry claims that letting AI bots design lifeforms, proteins, and genetic parts (generative biology) both automates and speeds up the design of new organisms - which may never have existed before in nature. This approach is an attempt to increase the efficiency of biotechnology production systems and, if effective, may significantly increase the overall number of new lifeforms and proteins being produced, unleashing a deluge of such novel biotechnology entities entering the market or ecosystems. This increases the need for human capacity to regulate, monitor, and undertake biosafety assessment. It also raises serious biosecurity threats.

c) Biodigital risks: A feature of current technological development is the rise of cyberphysical systems – more complex real-world systems that are controlled by digital technologies, including AI, sensors, and automation. These cyberphysical systems increasingly include SynBio organisms, components, and products – blurring the line between biological and digital realms (known as biodigital convergence). These hybrid 'biodigital' systems, composed in part of genetically engineered organisms or components, will be increasingly designed, managed, or monitored by AI – taking humans out of the loop and giving rise to novel risks and vulnerabilities. Proponents hope that in the future, AI systems themselves may be based on using genetically engineered biocomputers. If this ever becomes feasible, it will bring additional biosafety risks and bioethical questions into the heart of digital industries.

Policy implications

Governments are already scrambling to try to catch up with the side effects, errors, and governance conundrums created by first-generation 'generative AI' programs – such as ChatGPT – while discovering the overreach of claims initially made by AI developers. The CBD, with three decades of experience tracking global biotechnology policy, is uniquely placed to assess the now-emerging field of 'generative biology' and offer sensible advice before AI risks become irrevocably entangled with genetic engineering risks. At the sixteenth Conference of the Parties (COP 16) of the CBD in October and November 2024, governments will have the option to commission a 'deeper dive' assessment to better understand the array of policy challenges arising from the rapid integration of AI with SynBio – and to propose how to address those challenges promptly in the frame of precaution and justice. Agreements made at COP 16 on DSI also need to robustly ensure that the digital AI giants now amassing DSI to train generative biology models are firmly covered by requirements concerning commercial utilisation of DSI.



Synthetic biology - **© kkssr, Shutterstock**

How we got here: Background to the Convention on Biological Diversity (CBD)

For three decades the Parties to the CBD have taken bold and precautionary action to keep the Convention and its protocols up to date with fast-moving developments in modern biotechnology. The CBD was initially brought into being as the first commercial GMO crop that was being released in developed countries.[1] Subsequent timely decisions of the CBD have tracked and addressed new developments in biotechnology as they have arisen – including managing the transboundary movement and global trade in GMOs, genetic use restriction technology (GURTS or 'terminator seeds'), the need to prevent biopiracy, the rise of SynBio and synthetically designed organisms, the development of gene drives (a technology that drives mutations through populations in the wild) and the unregulated use of DSI (see Figure 1). Addressing AI integration with SynBio is the next logical step on the CBD's journey.

1 Miller, S.K. 1994. "Genetic first upsets food lobby", *New Scientist*, 28 May, https://www.newscientist.com/article/mg14219270-700-genetic-first-upsets-food-lobby/

FIGURE 1

How the CBD has responded to key **developments in biotechnology** over 30 years

G – Key governance developments under CBD and protocols

T – Key technology developments

1987 (T) First GMO, ice-minus bacterium *(Pseudomonas syringae)* is approved for release into the environment (USA)

1988 (G) Ad Hoc Working Group of Experts on Biological Diversity starts to negotiate text of CBD – encompassing regulation of modern biotechnology

1992 (G) Establishment of the CBD, including agreement to develop an international biosafety protocol

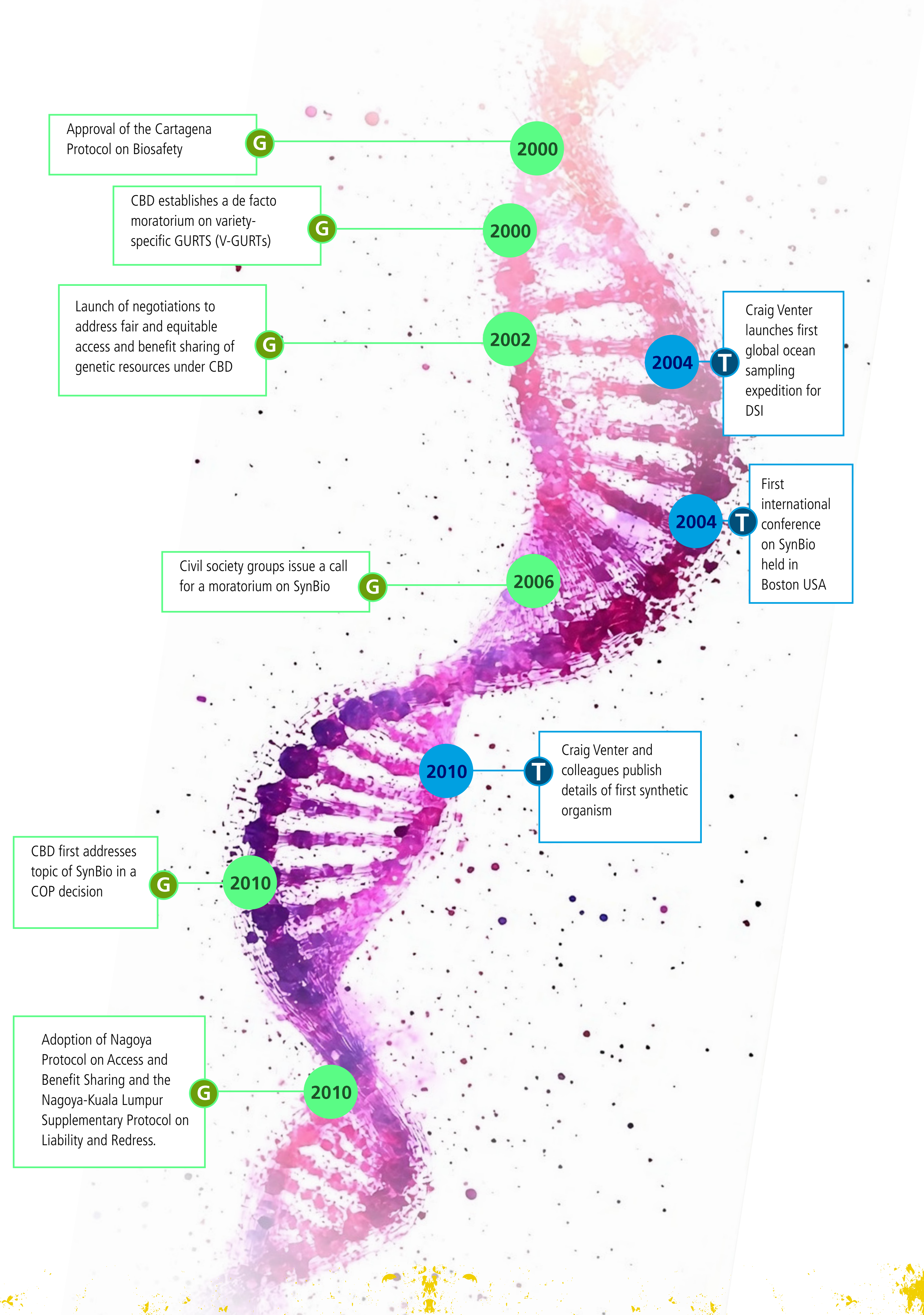
1994 (T) First commercial release of a GM crop by US company, Calgene ('Flavr Savr' CGN-89564-2 tomato)

1996 (T) Monsanto herbicide tolerant soybeans become first commercially grown and traded commodity GM crop

1996 (G) Establishment of CBD Working Group on Biosafety to regulate global transboundary movement of GMOs

1998 (G) Civil society uncovers research on sterile GMO 'terminator seeds' (GURTs)

1999 (G) Civil society publishes evidence of extensive biopiracy



2000
G
Approval of the Cartagena Protocol on Biosafety
2000
G
CBD establishes a de facto moratorium on variety-specific GURTS (V-GURTs)
2002
G
Launch of negotiations to address fair and equitable access and benefit sharing of genetic resources under CBD
2004
T
Craig Venter launches first global ocean sampling expedition for DSI
2004
T
First international conference on SynBio held in Boston USA
2006
G
Civil society groups issue a call for a moratorium on SynBio
2010
T
Craig Venter and colleagues publish details of first synthetic organism
2010
G
CBD first addresses topic of SynBio in a COP decision
2010
G
Adoption of Nagoya Protocol on Access and Benefit Sharing and the Nagoya-Kuala Lumpur Supplementary Protocol on Liability and Redress.

- **2014** (G): Establishment of first CBD Ad Hoc Technical Expert Group (AHTEG) on SynBio
- **2014** (T): First working engineered gene drive system (in fruit flies) reported
- **2015** (G): CBD AHTEG on SynBio first raises concerns of inequitable use of DSI and ecological risks of gene drive systems.
- **2016** (G): CBD establishes an operational definition of SynBio
- **2016** (G): CBD decides to undertake work on DSI
- **2016** (G): CBD first COP decision regarding engineered gene drives
- **2017** (G): CBD AHTEG proposes horizon scanning, assessment and monitoring of new developments in SynBio – also first raises use of AI as a new and emerging trend
- **2018** (T): DeepMind's AlphaFold 1 wins Critical Assessment of Structural Prediction (CASP) competition – first major use of AI in SynBio
- **2019** (G): CBD COP agrees on need for horizon scanning, assessment and monitoring process for SynBio
- **2022** (T): Scientists warn in *Science* of rapid development and risks of engineered self-spreading vaccines
- **2022** (G): CDB establishes process on horizon scanning, assessment and monitoring of new developments in SynBio, including the modified Ad Hoc Technical Expert Group (mAHTEG).
- **2022** (T): Public release of ChatGPT2 catalyses new investment rush in generative AI
- **2023** (G): mAHTEG on Synthetic Biology identifies integration of AI and SynBio (as well as development of self-spreading vaccines) as urgent key trends for CBD to address through deeper dive assessment

AI gene editing, - **© Rawpixel.com, Shutterstock**

In light of the fast-paced development of new biotechnologies in the field of SynBio, Parties at CBD COP 14 in 2018 agreed on the need for a broad and regular process of horizon scanning, assessment, and monitoring that could identify key trends and developments in SynBio early enough to take governance action.[2] Such a process was then formally launched at COP 15 in 2022 to help the CBD stay relevant to developments occurring in the field of SynBio.[3] An open online forum – followed by a series of virtual and in-person meetings of the mAHTEG on Synthetic Biology – identified five key trends for which an initial assessment was undertaken. This included the topic, 'Integration of Artificial Intelligence with Synthetic Biology'. In its report to SBSTTA 26,[4] the experts of the mAHTEG urged the Parties to the CBD to act on the topic of AI and SynBio integration. The mAHTEG advised that "the accelerated development of artificial intelligence and machine learning in the field of synthetic biology may have significant adverse impacts on the objectives, principles and provisions of the Convention and that those potential impacts need further evaluation".[5]

The mAHTEG proposed:[6]

- The CBD initiates a policy formulation process.
- "A request to the multidisciplinary Ad Hoc Technical Expert Group to undertake a further assessment leading to a report addressing, inter alia, potential impacts on biosafety, the sustainable use of biodiversity, equitable access and benefit-sharing, social, economic and cultural aspects, impacts on traditional knowledge and practices, and other relevant matters".
- The CBD secretariat develops a technical series publication addressing AI and participates in UN system level activities on AI.
- Parties "Consider the development of effective and equitable governance arrangements for artificial intelligence data sets, foundation models, algorithmic biodesign tools, automated science tools and the use of synthetic biology organisms, components and products in cyberphysical systems".

Unfortunately, at CBD SBSTTA 26, constructive discussions on SynBio were severely hampered by a small group of biotech-industry aligned countries which declared that they wished to discontinue the process of horizon scanning, assessment, and monitoring that had previously been agreed on.[7] Despite the obstructions by this small group, heavily bracketed terms of reference were nonetheless drafted that included a request for the mAHTEG to:

2 UN CBD Decision CBD/COP/DEC/14/19 – 30 November 2018: "Agrees that broad and regular horizon scanning, monitoring and assessing of the most recent technological developments is needed for reviewing new information regarding the potential positive and potential negative impacts of synthetic biology vis-à-vis the three objectives of the Convention and those of the Cartagena Protocol and Nagoya Protocol".

3 UN CBD Decision CBD/COP/DEC/15/31 paragraph 4

4 The 26th meeting of the Subsidiary Body on Scientific, Technical, and Technological Advice (SBSTTA) took place in Nairobi, Kenya, from May 13 to 18, 2024. This body operates under the CBD and provides the COP with advice on scientific, technical, and technological matters related to the implementation of the Convention

5 Report of the multidisciplinary Ad Hoc Technical Expert Group on Synthetic Biology contained in CBD/SBSTTA/26/4 – Annexe 1

6 Ibid

7 Sirinathsinghji, Eva. 2024. "SynBio discussions held to ransom", SBSTTA 26 Daily Eco report of the CBDA, 18 May, https://cbd-alliance.org/sites/default/files/2024-05/ECO-6-SBSTTA-26.pdf

"Continue the in-depth assessment, including on [their] potential [positive and negative] impacts [in the light of][on] the objectives of the Convention [ecological, socioeconomic – including potential threats to livelihoods, the sustainable use of biodiversity – and ethical and cultural considerations, and taking into special consideration indigenous peoples and local communities, women and youth] of:
(i) The potential impacts of the integration of artificial intelligence and machine learning into synthetic biology".[8]

At COP 16 of the CBD, Parties have the opportunity to ensure that such an assessment of AI/SynBio integration moves ahead, in line with the precautionary approach, and that it is not further blocked by industrial interests.

The basics: What are 'synthetic biology' (SynBio) and 'artificial intelligence' (AI)?

Synthetic biology (SynBio) encompasses several new and emerging developments in modern biotechnology (also known as genetic engineering). The first GMOs were made in a laboratory by using a bacterium (*Agrobacterium tumefaciens*) and restriction enzymes to transfer DNA from one organism into the genome of another (a technique known as recombinant DNA, or rDNA). However, biotechnologists increasingly rely on a family of 'gene editing' technologies, such as CRISPR/Cas, that cut and modify host DNA using RNA-guided enzymes combined with different pathways of cellular repair. Gene editing may be used to delete a single base to deactivate the function of a gene (a "knockout"); it may also be used to insert small sequences or whole new genes, including from "transgenic" donors, to correct or gain function. Like previous genetic engineering techniques, these gene-editing techniques are currently heralded for unprecedented gains in speed, cheapness, and precision. Yet any of these technical features can exacerbate existing environmental harms and social inequalities. Moreover, the history of biotechnology shows that each generation of breakthroughs is routinely proclaimed as superior to the "error-prone" technology it displaces but then in short order displays its own errors. Gene editing for example continues to face technical challenges, including unintended, off-target effects.[9]

While gene editing and rDNA for agriculture evolved largely at the intersection of molecular biology and plant/animal breeding, synthetic biology has theoretical and methodological roots in engineering. Synthetic biology applies engineering principles to develop new biological parts, devices, and systems or to redesign existing systems found in nature. Synthetic biologists – who span fields of computer science, systems biology, engineering, and biophysics, among others – increasingly design new DNA codes using a computer and manufacture synthetic DNA strands from chemicals.

8 UN CBD SBSTTA decision 26/4 – CBD/SBSTTA/REC/26/4, available at https://www.cbd.int/doc/recommendations/sbstta-26/sbstta-26-rec-04-en.pdf

9 "Unintended changes induced by CRISPR/CAS cause novel risks", *Testbiotech*, last modified 4 December 2019, https://www.testbiotech.org/en/news/unintended-changes-induced-crisprcas-cause-novel-risks/

They also seek new ways to design, build, and use genetic molecules other than DNA,[10] such as RNA, and proteins such as enzymes, alongside genetic regulatory elements. As per their training, synthetic biologists tend to treat living genetic molecules as if they were programmable software code for organisms, endeavouring to rewrite that 'code' for industrial purposes. This reductive approach makes sense within the constraints and conditions of a model in a lab. However, life does not, in fact, operate as a computer and the machine metaphor often fails in complex, real-world systems.[11]

Artificial intelligence (AI) is a set of technologies based primarily on machine learning and deep learning, and used for data analytics, predictions and forecasting, object categorisation, natural language processing, recommendations, data retrieval, and more.

Traditional AI, also known as rules-based or deterministic AI, relies on pre-programmed rules and algorithms to perform specific tasks. This type of AI has been around for decades, evolving with advances in computing power. In contrast to the open-ended data sets used in generative AI, traditional AI systems typically use structured and curated data sets and models suited to particular purposes. As such, these AI systems are often used in finance, manufacturing, and healthcare to solve bounded problems and perform repetitive tasks.

'Discriminative AI' software can learn its own rules. It takes large unstructured sets of data and self-sorts them into meaningful groupings or identifies images or patterns. To do this, the AI software is first 'trained' on other data and is helped to make associations until it develops its own models (with structured weightings) for making distinctions. This training process, inspired by neural networks in the brain, is called 'machine learning'.

Generative AI differs in key respects. Like traditional AI and discriminative AI, generative AI trains on large data sets. But whereas traditional AI is limited to rule-bound prediction, and discriminative AI sorts existing data, generative AI models use their training datasets to generate novel data examples. Generative AI is thus capable of creating novel outputs from a vector space that it has generated through learned data. For example, whereas a discriminative AI system is trained to distinguish between different images of animals, a generative AI might be asked to look at relationships within its data to generate a new picture of a "cat" based on the materials it has seen and the model it has developed. Generative AI often uses very large amounts of training data harvested from the internet to synthesize images, generate text, or transfer a style.

Since late 2019, when ChatGPT-2 was released to the public, the vast majority of generative AI in use has been based on transformer models (large language models) and diffusion models (see glossary). For example, the popular generative AI program ChatGPT is a transformer-based large language model (LLM) that draws on the massive training dataset of language scraped off the internet, to stitch together credible-seeming text. Programs like Stable Diffusion, DALL-E, and Midjourney do the

10 Schmidt, M. 2010. "Xenobiology: A new form of life as the ultimate biosafety tool", *Bioessays*. 32(4):322-31. doi: 10.1002/bies.200900147. PMID: 20217844; PMCID: PMC2909387. https://pubmed.ncbi.nlm.nih.gov/20217844/

11 Boudry, M., & Pigliucci, M. 2013. "The mismeasure of machine: Synthetic biology and the trouble with engineering metaphors", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, Volume 44, Issue 4, Part B

same but for images (much to the chagrin of artists and designers whose creative works are being appropriated). These programs use their internal rules to predict what might be the correct image or string of words to satisfy the user's natural language query ('prompt'). That prediction may look credible but on closer analysis may be wrong (and frequently is – see box on AI hallucinations, bias, and data poisoning).

Computer scientists Arvind Narayanan and Sayash Kapoor explain that the fallibility in large language models like ChatGPT is due in part to the fact that AI is incapable of ascertaining causality or context, both of which give language meaning.[12] The models are simply making statistically probable linkages, operating as what experts have described as a "stochastic parrot."[13] Unlike AI in science fiction, these systems do not actually "learn" in ways comparable to human/animal brain learning; AI is not deliberative, conscious, or despite its name, "intelligent". It is, however, a powerful prediction engine based on a very large amount of data.

Thus far, Generative AI is proving to be more hype than substance. Its test cases like ChatGPT are demonstrating that simply adding greater amounts of data doesn't equate to greater accuracy or predictive power. Like all models, these systems are hampered by the limitations and biases of the data on which they are trained (again, see the section below on AI hallucinations, bias, and data poisoning) but also by flawed assumptions underlying the technology.

How **AI** and **SynBio converge**

The 'Big Tech' View:

> **"Limits are now being breached. We are approaching an inflection point with the arrival of these higher order technologies, the most profound in history. The coming wave of technology is built primarily on two general-purpose technologies capable of operating at the grandest and most granular levels alike: artificial intelligence and synthetic biology."**
>
> **– Mustafa Suleyman, CEO of Microsoft AI, Founder of Infection AI, and Co-founder of DeepMind (now owned by Google).[14]**

12 https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil

13 https://dl.acm.org/doi/10.1145/3442188.3445922

14 Suleyman, M. *The coming wave: Technology, power and the 21st century's greatest dilemma*. 2023. Crown Publishing Group, Random House, p.55.

There are at least four ways in which AI and SynBio are becoming integrated, as described in greater detail below:

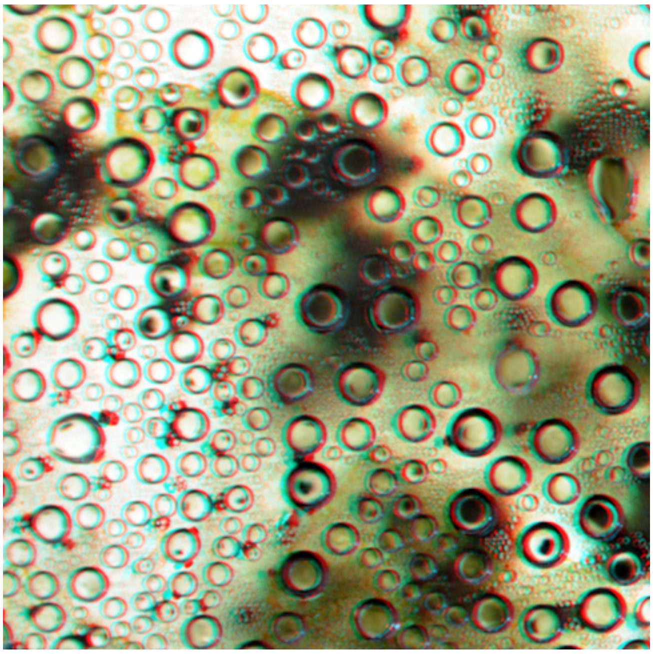
Biodesign (generative biology): AI models that generate new genetic or protein sequences. The design is digital, but then protein engineering and biochemistry are used to turn that into synthetic biological molecules.

Bioproduction: The use of AI to improve the hardware and efficiency of processes used in industrial biotechnology production (such as fermentation in vats or cell culture in petri dishes).

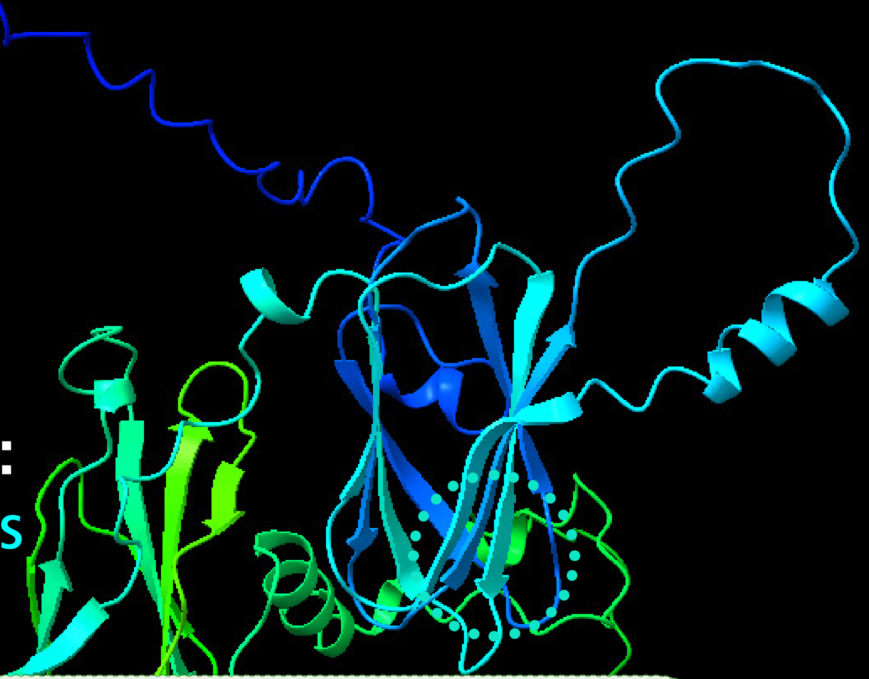
Biodigital (cyberphysical) systems: Technical systems that mix AI and SynBio components (e.g. using engineered crops or RNA sprays in digital agriculture, where decision-making is carried out by algorithms).

Biocomputation: The use of engineered DNA, cells, bacteria or other biological parts as computational hardware for AI processes.

1. Biodesign: Generative biology


Biodesign - **© wim hoppenbrouwers, Flickr**

Just as generative AI models for text and images can appear to create novel pieces of writing or new pictures by selecting and rearranging from the massive sets of data they were trained on, so AI models trained on large amounts of biological information can appear to generate credible-seeming new arrangements for biological molecules – including new DNA sequences and protein sequences. The idea goes that, rather than train an AI LLM on the texts of languages such as English or French, the same model can instead be trained on what synthetic biologists simplistically describe as biological 'languages' – such as DNA code or the sequence of amino acids that make up proteins. The AI software then develops its own model to sort and 'understand' those biological codes, which, in turn, can be prompted to generate 'new' biological language – such as novel DNA codes for genomes or new arrangements of amino acids for novel proteins. All of these processes depend on first accepting a highly reductive (and incomplete) understanding of how genetic systems function.



Google's AlphaFold: Generative biology's 'ChatGPT moment'

The breakthrough 'proof of concept' for letting AI models interpret biology and carry out SynBio was an AI model called Alphafold, developed by Google's AI subsidiary DeepMind. Alphafold was initially trained on data describing 170,000 proteins from a public repository of protein sequences and structures.[15]

Its goal was to discern how proteins arranged in certain sequences of amino acids go on to fold into specific functional structures. At the start of this century the scientific labour required to explain the link between the code of any single protein and how it then subsequently folded into a shape that could take up the length of an entire PhD. In 2018, Alphafold successfully predicted the folding structures of 13 proteins from their protein code, and by 2021, Google was claiming that Alphafold 2 had predicted the folding arrangements for all known protein codes (almost 200 million proteins)[16] – although others observed that this was an overclaim and not backed up by lab science.[17]

Wowed by these results (and Google's press releases), scientific publications proclaimed that the Alphafold AI had 'solved' the problem of protein folding. In March 2024, Google DeepMind released Alphafold 3, which, they claim, can not only predict any protein structure but also "the structure and interactions of all life's molecules with unprecedented accuracy."[18] Alphafold 3 is claimed (again by Google in their press releases) to be able to predict the structure of any DNA, RNA, ligands, proteins, and more, and also to predict how these biomolecules might interact with each other at the cellular level. Having trained a model supposedly able to make such powerful predictions about the basic biomolecules of life, generative AI companies now are treating prediction as fact and claiming they can reliably generate new, never-before-seen proteins, DNA, RNA, and other structures for genetic engineering, using AI-directed generative biodesign. Thus, "generative biology" has been established as a new industry. In particular there is a rush of new AI 'protein design' companies inspired by Alphafold (see below).

15 "Method of the Year 2021: Protein structure prediction", *Nat Methods* 19, 1 (2022). https://doi.org/10.1038/s41592-021-01380-4

16 Callaway, Ewen. 2022. "'The entire protein universe': AI predicts shape of nearly every known protein". *Nature*. 608 (7921): 15–16

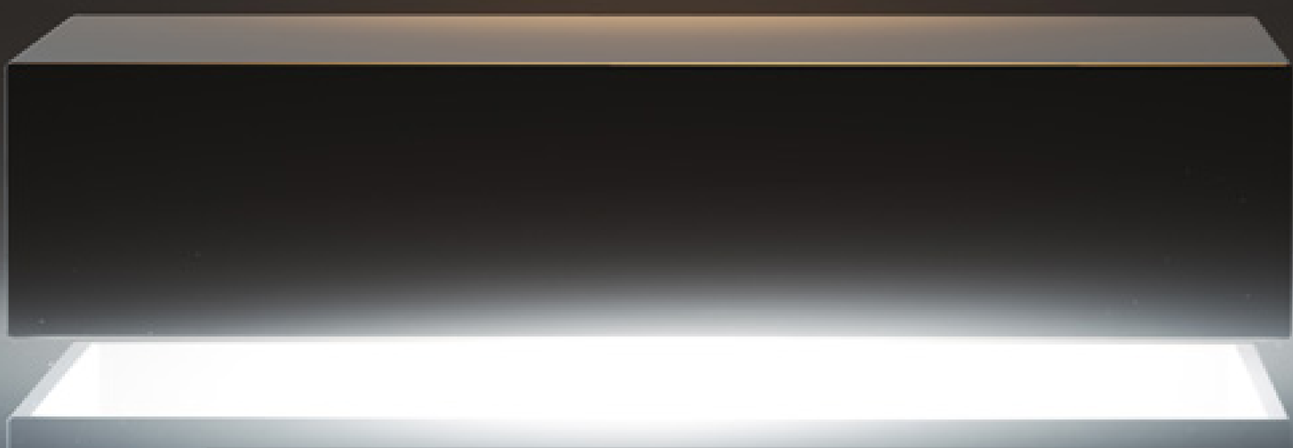
17 Terwilliger TC, Liebschner D, Croll TI, et al. 2024. "AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination", *Nat Methods*. 21(1):110-116. doi: 10.1038/s41592-023-02087-4. Epub 2023 Nov 30. PMID: 38036854; PMCID: PMC10776388. Includes the observation: "We note that as AlphaFold prediction does not take into account the presence of ligands, ions, covalent modifications or environmental conditions, it cannot be expected to correctly represent the many details of protein structures that depend on these factors".

18 Google DeepMind Lab, AlphaFold Team, "AlphaFold 3 Predicts the Structure and Interactions of All of Life's Molecules", May 8, 2024, https://blog.google/technology/ai/google-deepmind-isomorphic-alphafold-3-ai-model/#life-molecules.

A word of **caution**

While Alphafold impressed many as an example of AI-driven big science, established laboratory-based protein scientists are already raising a red flag to not be blinded by the hyped capabilities of generative biology.[19] Like all AI systems, Alphafold doesn't solve protein folding per se – it merely offers likely predictions. Protein scientists still need to do the laboratory work to check the accuracy of those digital predictions and in several cases and classes of protein, the predictions are proven wrong or misleading. This echoes how the initial commercial hype of generative AI in text (ChatGPT) or images (Dall-E or Midjourney) airbrushed out the real problems and limits of generative AI technology. It also echoes how early claims about the precision of both rDNA and later gene editing were wildly overstated. There is reasonable concern that, in the commercial rush now underway to use generative AI models to generate novel living genetic systems, policymakers may be ignoring mistakes and dangers. Unlike digital text or digital graphics, digitally designed novel lifeforms and proteins may have a very direct and unpredictable biological impact on the living world – especially if they are designed incorrectly, based on wrong assumptions, and released irresponsibly.

19 See Reddit thread, "People are overestimating Alphafold and it's a problem", https://www.reddit.com/r/labrats/comments/1b1l68p/people_are_overestimating_alphafold_and_its_a/

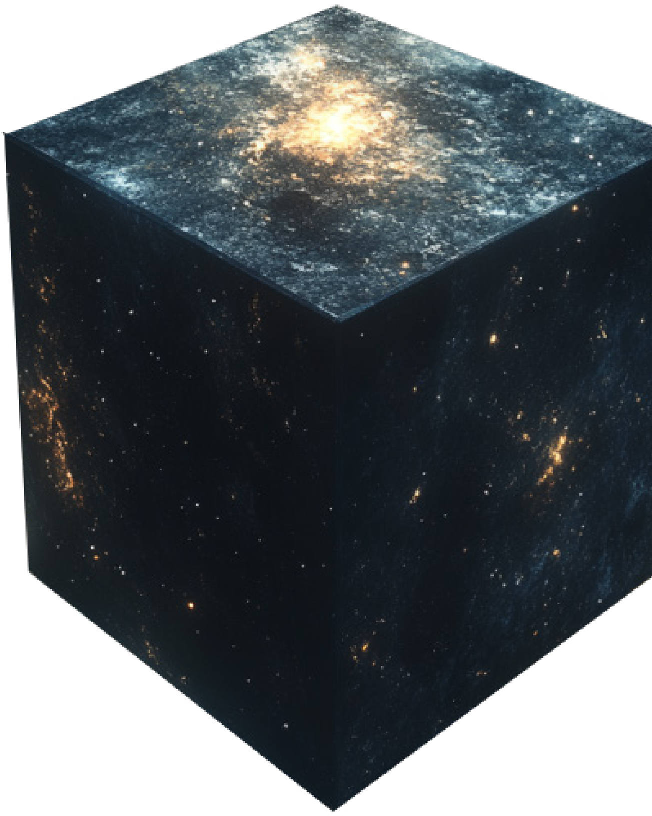


Moving genetic engineering into a 'black box' (and out of CBD oversight)

An important metaphor that has developed from critical scholarship on AI is the concept and risks of the 'black box'.[20] The 'black box problem' refers to the opaque way an AI system evolves its own model during 'training' in a manner of applying statistical weights to certain connections it observes. The huge number of operations undertaken renders the reasoning for this self-generated model untransparent and unknowable to a human observer – even a technical expert. Subsequently, when a generative AI system generates a particular image or makes a particular prediction as output, it is not currently possible to interpret or explain why it generated that outcome – e.g. why it chose a particular word, image, or genetic sequence in its output.

For genetic engineering, moving biodesign decisions to be obscured in the 'black box' marks a fundamental and consequential shift in how genetic engineering interventions are carried out – cutting humans out of the loop in design decisions. Previously, a human genetic engineer would select or alter an existing digital or actual genetic sequence or protein sequences for a particular explainable reason or a conjecture they could substantiate. They would then experimentally try them out to come to a result, usually documenting the process. Now, an automated AI system applies an internal self-generated model that cannot be explained, based on statistical predictions from a repetitive analysis of thousands or millions of sequences. The computer then selects combinations of genetic sequences drawn from its training data but does not identify the source and it strings them into a new sequence. This is how genetic biodesign decisions are moved into an unknowable 'black box'.

20 See for example "AI's Mysterious Black Box Problem Explained", University of Michigan-Dearborn, accessed July 25, 2024, https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained.



Of course, creating a black box process is a decision that the designers of an AI system choose to make and gamble with. In that sense, they should become directly culpable for any outcomes from their invention.

AI systems that carry out black-box genetic engineering have at least three important implications that affect governance:

1. **Biosafety** – If the AI system that designs a genetic engineering intervention cannot explain either why it selected a genetic sequence or even what starting material that sequence was drawn from, it becomes difficult to undertake an independent biosafety analysis of that design. The uncertainties associated with genetic engineering increase greatly as does the inability to spot problems, errors, and mistakes.

2. **Access and benefit sharing** – If the designer of a genetic engineering intervention does not disclose what digital sequences the design was drawn from or learned from (let alone where the original natural source came from), it is not possible to share benefits back to the original holder of that genetic resource or for that original holder to exercise rights over access. Attempting to ensure appropriate access and fair and equitable benefit sharing thus becomes impossible.

3. **Liability and redress** – If the designer of a genetic intervention is an automated probabilistic computer system that generates its own opaque model to make decisions, it may not be clear who is legally responsible when something goes wrong in the outcome. There is no human check or accountability over the genetic engineering process. Even though developers of AI systems know they are building black box systems that may be riddled with errors, some courts have found that AI developers providing software for self-driving cars are not liable for deadly car crashes, where a human is technically a backup driver.[21] However, the law is less clear with fully autonomous cars.[22]

These three areas ('Biosafety', 'Access and benefit sharing' and 'Liability and redress') describe the purposes of the three CBD protocols (Cartagena Protocol, the Nagoya Protocol and the Nagoya-Kuala Lumpur Protocol) that support governance of biotechnology and its impact on the living world. Additionally, the CBD attempts to address socio-economic considerations that will also arise when AI systems move into biotechnology. By moving genetic engineering into the black box, the integration of AI with SynBio directly challenges the core governance approach that the CBD has established over the field of modern biotechnology.

21 David Shepardson and Heather Somerville, "Uber Not Criminally Liable in Fatal 2018 Arizona Self-Driving Crash: Prosecutor", Reuters, March 5, 2019, https://www.reuters.com/article/business/uber-not-criminally-liable-in-fatal-2018-arizona-self-driving-crash-prosecutor-idUSKCN1QM2P3/.

22 Jenna Greene, Column, "Driverless Car Problems Are Outpacing Liability Laws", Reuters, December 11, 2023, https://www.reuters.com/legal/transactional/column-driverless-car-problems-are-outpacing-liability-laws-2023-12-11/.

What parts of genetic systems can now be biodesigned by AI?

DNA and RNA – AI LLMs trained on all of the world's existing DSI can generate novel DNA or RNA sequences for SynBio applications.

> We took all the DNA data that is available – both DNA and RNA data for viruses and bacteria that are known to us – about a hundred and ten million such genomes. We learnt a language model over that and we can now ask it to generate new genomes.
>
> **– Anima Anandkumar, NVIDIA Corporation describing NVIDIA's GenSLM model**[23]

Proteins – Building on Alphafold's success at predicting protein structures, dozens of companies now claim their AI platforms can generate designs for never-before-seen proteins with novel commercial characteristics. Some are even working on text-to-protein generators.

> The new idea is, can I make a foundation model that … speaks 'protein' just like GPT4 speaks English?
>
> **– Jason Kelly, CEO Gingko Bioworks (partnered with Google Ai)**[24]

Gene editors – In the past decade, the key tools for genetic engineering have been gene-editing proteins such as CRISPR-Cas 9, originally found in natural bacteria. AI companies such as Profluent are now using generative AI to generate novel artificial gene-editing proteins that they claim will cut, splice, and replace DNA and RNA more efficiently than CRISPR.[25]

Epigenetics – Engineers at the University of Toronto associated with startup TBG Therapeutics have developed an AI system called ZFdesign that generates zinc fingers (ZF). These act as transcription factors – epigenetic elements that activate or deactivate different parts of DNA. The researchers feed data from billions of interactions between ZF proteins and DNA into a machine-learning model, which then generates engineered zinc fingers that bind to a given DNA sequence.[26] Generative AI is also being used for histone modification [27]– another epigenetic mechanism.

23 Kristen Ye, "The AI Podcast - Anima Anandkumar on Using Generative AI to Tackle Global Challenges – Ep203", NVIDIA, accessed July 25, 2024, https://blogs.nvidia.com/blog/anima-anandkumar-generative-ai/.

24 Jason Kelly speaking on No Priors Podcast ep34: "DNA as Code – Cell Programming and AI ", YouTube, uploaded by No Priors Podcast , https://www.youtube.com/watch?v=snt-fMsCDVI.

25 Callaway E. 2024. "'ChatGPT for CRISPR' creates new gene-editing tools', *Nature* 629(8011):272. doi: 10.1038/d41586-024-01243-w. PMID: 38684833.

26 Ichikawa DM, Abdin O, Alerasool N, et al. 2023. "A universal deep-learning model for zinc finger design enables transcription factor reprogr

27 DaSilva LF, Senan S, Patel ZM, et al. 2024. "DNA-diffusion: Leveraging generative models for controlling chromatin accessibility and gene expression via synthetic regulatory elements, bioRxiv [Preprint]. 1:2024.02.01.578352. doi: 10.1101/2024.02.01.578352. PMID: 38352499; PMCID: PMC10862870.

Bioproduction - **© Greg Emmerich, Flickr, CC BY-SA 2.0**

2. Bioproduction: Boosting biotechnology

AI systems can also redesign the hardware, infrastructure, and workflow of industrial biotechnology – such as laboratory processes, fermentation facilities, and cell cultures. They can also enable robots to carry out parts of laboratory work, for example by using machine vision in place of human scientists. Growing a genetically engineered cell culture or fermenting engineered bacteria or yeast in a bioreactor (a large steel vat) is complicated and prone to contamination and other problems. It is claimed that by using AI to monitor and automate processes, bioproduction systems can become more efficient or able to identify, prevent, or combat contaminants or imbalances.[28] If AI-managed bioproduction can increase production speed or quantity, this may make genetically engineered products that previously could not be produced in quantities beyond the laboratory, industrially viable. This, in turn, requires that governments increase their human capacity for monitoring, risk assessment, technology assessment, regulation, cleanup up, and recall. The same capabilities can also be used by malicious actors to produce toxins or dual-use biowarfare agents in large quantities.

28 Amerging Technologies, "How AI is Revolutionizing the Fermentation Industry", March 15, 2024, https://www.linkedin.com/pulse/how-ai-revolutionizing-fermentation-industry-amergingtech-1dsec/.

3. Biodigital systems: The example of 'robot-ready' crops

A feature of current technological development is the rise of cyberphysical systems – more complex real-world systems that are controlled by digital technologies, including AI, sensors, and automation. These cyberphysical systems increasingly include SynBio organisms, components, and products, blurring the line between biological and digital realms (known as biodigital convergence). An example is the use of genetically engineered crops and RNAi sprays in digital agriculture systems managed by AI. For example, SynBio company InnerPlant has bioengineered soy and other crops to emit a fluorescent protein when stressed (e.g., in need of water or attacked by pests). This fluorescence is tuned to be detected by sensors mounted on AI-enabled farm machinery, which can, in response, apply water or pesticides.[29] In effect, the genetically engineered plants have been biodesigned to be 'robot-ready' – that is, supposedly to uniquely communicate with AI-operated farm machinery. Together, the bioengineered crops and the AI farm machinery enable a bigger cyberphysical system of digital agriculture to change farming – one that locks in pesticide and big data use.

4. Biocomputation: The example of brain organoids

AI models are currently based on large-scale computation by silicon chips in racks of servers in large temperature-controlled warehouses, known as data centres. These consume enormous amounts of energy, water, and chemical coolant on an ongoing basis. AI hardware developers are actively seeking low-energy alternatives to silicon-based computation and one future direction may be harnessing SynBio to make living cells or DNA that carries out computation – known as bio-computation. While this dream of biocomputing is an old one, the massive energy and resource costs of current silicon-based AI are making it a renewed focus. In one proof of principle, an Australian company, Cortical Labs, has built working AI computational circuits out of brain organoids (cultured brain cells). They call this 'Organoid Intelligence' (OI) or 'Brainoware' and bioengineers have trained these simple, living AI circuits to play computer games or recognise speech.[30] If synthetic biological parts become common in computation this would bring genetic engineering risks and the task of managing biological pollution into a whole new field of industry.

Agri AI robots - **© Steve Long, Flickr**

29 "Products", InnerPlant, accessed July 25, 2024, https://innerplant.com/products/./

30 Adhiti Iyer ,13th March 2024, Is OI the New AI? Questions Surrounding 'Brainoware' Blog - https://blog.petrieflom.law.harvard.edu/2024/03/13/is-oi-the-new-ai-questions-surrounding-brainoware/

The **generative biology** rush: Technology titans move into biology (and **biotech/ biodiversity policy**)

One of the most striking features of AI-driven SynBio is that much of the work is being led by the largest technology companies in the world. Most of these new AI/SynBio leaders are digital titans, with no previous experience in biotechnology or stewarding biodiversity but extensive experience in implementing monopolistic business models and skirting regulations. They are striking joint agreements or acquiring smaller biotechnology startups. These data firms are acting from common metaphors in SynBio that claim that living systems can be simplistically manipulated, like biological machines that run on code and can be programmed and debugged. However, in reality, the 'wet' world of biology, ecosystems, and life processes is often nothing like computing or coding. They entail quite novel, living risks. Digital technology titans coming from a Silicon Valley culture of 'move fast and break things'[31] may be badly equipped to handle bio-risks.

Biochemistry - **© Day Of Victory Studio, Shutterstock**

31 The Silcon Valley mantra of 'move fast and break things' is often attributed to Mark Zuckerberg of Meta/Facebook - see https://www.snopes.com/fact-check/move-fast-break-things-facebook-motto/

Technology giants at play:

Google/Alphabet – Google DeepMind (AI research laboratory) developed the high profile Alphafold program. Google has a joint venture with leading SynBio company Gingko Bioworks to generate novel proteins.[32] They have also created their own biotechnology company called Isomorphic Labs, which is using AI to generate new drug compounds for major pharmaceutical companies.[33]

Microsoft – The CEO of Microsoft AI, Mustafa Suleyman, recently published a high-profile book (called *The Coming Wave*) on how the convergence of SynBio with AI will transform society (and create new risks).[34] His firm is developing several generative AI tools for SynBio, including a generative medical platform called BioGPT.[35]

Amazon – In June 2024, the world's largest provider of data cloud services announced it was collaborating with a company called EvolutionaryScale to host ESM3 – a generative biology AI platform trained on "billions of protein sequences spanning 3.8 billion years of evolution".[36] According to Amazon, ESM3 can understand complex biological data from various sources and generate entirely new proteins that have never existed in nature. Meanwhile, the Bezos Earth Fund (associated with Amazon founder Jeff Bezos and his girlfriend) has launched a 100 million dollar 'AI for Climate and Nature' program that focuses on using generative AI for alternative proteins and other materials.[37] Amazon is reportedly also interested in brain organoid biocomputation.[38]

NVIDIA – The world's largest AI chipmaker is also out front in generative biology. Their GenSLM AI platform has been trained on hundreds of thousands of genomes to generate novel microbial and viral genomes. They have particularly been developing candidate COVID sequences for vaccines and appear to have successfully predicted some new Covid variants.[39] Nvidia is also collaborating with Amazon's ESM3 platform.[40]

32 Jake Wintermute, "Google and Gingko: Foundry-scale Data meets AI" - https://www.ginkgobioworks.com/2023/08/29/google-and-ginkgo-foundry-scale-data-meets-ai/

33 https://www.isomorphiclabs.com

34 Suleyman, M. *The coming wave: Technology, power and the 21st century's greatest dilemma*. 2023. Crown Publishing Group, Random House, p.55

35 Newton,W. 2023. "What is BioGPT and what does it mean for healthcare?" 9 Feb, https://www.clinicaltrialsarena.com/news/biogpt-healthcare/

36 Wood, M. 2024. "Revolutionizing generative biology with AWS and EvolutionaryScale" 25 June, https://aws.amazon.com/blogs/industries/revolutionizing-generative-biology-with-aws-and-evolutionaryscale/

37 "AI for Climate and Nature", Bezos Earth Fund, accessed July 25, 2024, https://www.bezosearthfund.org/ai-climate-nature

38 Lee, Z. 2023. "This AI startup wants to be the next NVIDIA by building brain cell-powered computers", Forbes, June 21, https://www.forbes.com/sites/zinnialee/2023/06/21/cortical-labs-brain-computer/

39 Salian, I. 2023. "Gen AI for the genome: LLM predicts characteristics of COVID variants", Nvidia Blog, 13 Nov, https://blogs.nvidia.com/blog/generative-ai-covid-genome-sequences/

40 Nagel, B. 2024. "NVIDIA and AMS-backed bioresearch startup launch 'ESM3' model'", 26 June, https://pureai.com/Articles/2024/06/26/Bioresearch-ESM3-Model.aspx

Meta (Facebook) – In 2022, Meta unveiled its ESMfold platform as a rival to Google's Alphafold. Meta claimed that ESMfold housed more than 600 million protein structures and was 60 times faster than Alphafold.[41] However, the project was shut down in 2023 amidst large staff layoffs at Meta.[42]

Salesforce – This leading US data cloud company has developed ProGEN, an AI LLM for generating novel proteins. ProGEN was trained by feeding the amino acid sequences of 280 million different proteins into a machine-learning model.[43] As a proof-of-concept, Salesforce then tuned the model by priming it with 56,000 sequences from just one class of protein: lysozymes (used for food ingredients).[44]

Alibaba – In 2023, scientists at the leading Chinese technology giant, Alibaba, published results from its LucaProt AI platform, which was trained to identify RNA viruses. According to the researchers, LucaProt identified 161,979 potential RNA virus species and 180 RNA virus supergroups. They asserted,

This study marks the beginning of a new era of virus discovery, providing computational tools that will help expand our understanding of the global RNA virosphere and of virus evolution.[45]

41 Rosenbush, S. 2023. "Scientists at DeepMind, Meta press fusion of AI, biology", *Wall Street Journal* – 22 March, https://www.wsj.com/articles/scientists-at-deepmind-and-meta-press-fusion-of-ai-biology-4b92af6f

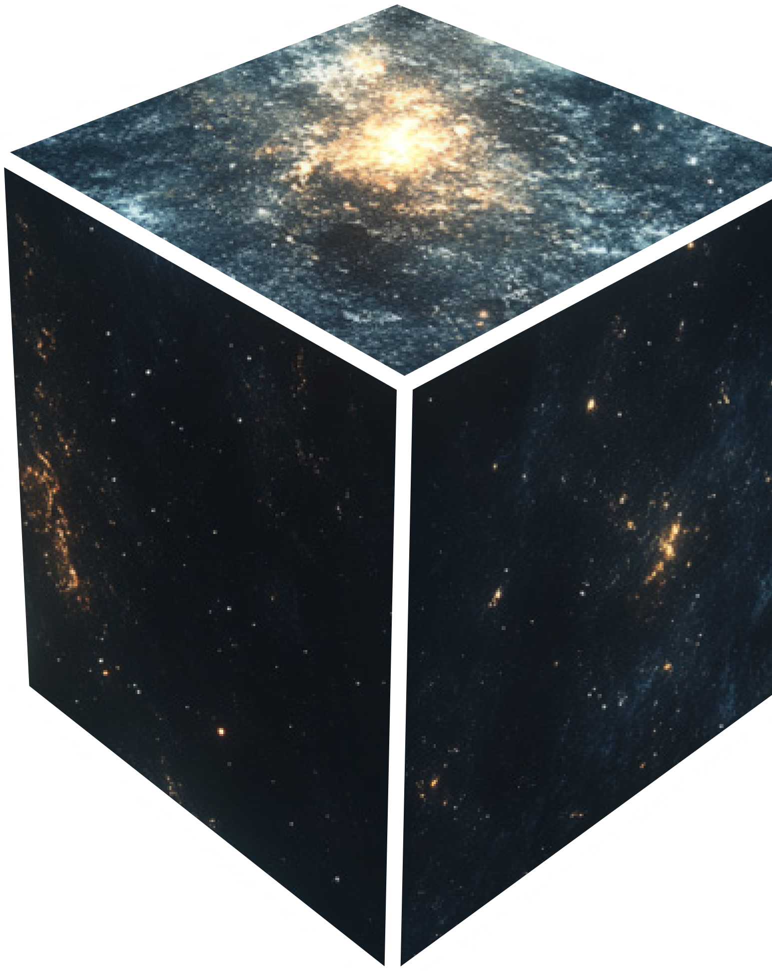
42 Wodecki Jr, B. 2023. "Meta lays off team behind revolutionary protein-folding model', *AI Business*, 8 Aug.

43 Madani, A. n.d. "ProGen: Using AI to generate proteins", Salesforce AI Research, accessed 25 July 2024, https://blog.salesforceairesearch.com/progen/

44 Univ of California. 2023. "AI technology generates original proteins from scratch', *Physics.org*, 26 Jan, https://phys.org/news/2023-01-ai-technology-generates-proteins.html

45 Xin Hou, Yong He, Pan Fang, et al. 2023. "Artificial intelligence redefines RNA virus discovery", bioRxiv; doi: https://doi.org/10.1101/2023.04.18.537342

Implications for the CBD: **Five urgent questions**



Amidst the current rush of big technology companies, investment, and hype around generative biology, and the use of AI to boost biotechnology production, the Parties to the CBD urgently need to find answers to at least the following five questions:

1. Does AI generative biology undermine access and benefit-sharing arrangements of the Nagoya Protocol and the governance of DSI?

Current arrangements for access and benefit sharing from utilisation of genetic resources depend on tracing a clear line from the original accessing of a genetic resource (e.g. taking a seed or tissue sample) to the final commercial utilisation of that knowledge (e.g. in a gene-edited product). That may involve tracing the digitisation of sequence information (e.g. DNA codes, protein sequences) and their storage in databases of DSI. However, how AI-driven generative biology works hides all of this in an algorithmic black box, breaking any link between acts of access and acts of utilisation. In this way, the opportunity is lost for benefit-sharing or for communities to exercise a veto over end uses.

It is clear, by the large investments flowing into so-called generative biology, that technology giants believe there are huge profits to be made from generating novel proteins, DNA sequences, viruses, and the like. To generate these biomolecules, technology titans are grabbing (for free) all of the DSI currently available in databases and using it as training data. Yet when it comes to benefit-sharing they will claim that it is not possible to identify which specific original genetic resources caused their models to suggest a particular output. Because of the black-box nature of generative AI, they will argue governments, regulators or even experts can't trace the link between access and utilisation. This will have implications for the current negotiations on DSI and strengthen the argument for a multilateral mechanism and global fund. Big technology firms running generative AI models based on DSI should be clearly identified as commercially utilising DSI and should pay into the fund that has been established.

This very serious challenge to the current paradigm for equitable sharing of benefits echoes similar fights that are now emerging across the domain of AI policy. Artists, writers, journalists, photographers, and designers have launched legal cases against AI companies who trained their models on their images and writings scraped from the internet (i.e., for free).[46] They have even written software to identify specific artworks in training data sets.[47] Technology companies claim they can evade copyright payments since their systems are only 'trained on' these inputs and the new outputs they generate are 'novel' and different. They will likely make the same arguments to the original rights holders of genetic resources. It may be necessary for governments to insist a tax or license be levied on AI companies for any use of DSI in training datasets. There may be technical ways of forcing AI companies to maintain traceability and to force generative biology to become explainable, rather than based on a black box. At least one generative biology startup, UK-based Basecamp Research, already claims that their AI model is Nagoya-ready.[48]

2. Does AI generative biology undermine biosafety arrangements of the Cartagena Protocol on Biosafety?

Decisions of the CBD and Cartagena Protocol compel Parties to undertake biosafety evaluations of living modified organisms and to assess the risks of products, components, and organisms developed through SynBio. This is in line with the precautionary approach. To assess the biological safety of a novel organism or product, assessors need to know certain types of information: they need to know what genetic resources were added and the organism that they came from (donor organism) as well as to understand why those sequences were selected. When genetic biodesign is carried out in an AI black box, biosafety assessors may be denied this, and other important information, needed to carry out safety and other (e.g. socio-economic) assessments.

The recent history of generative AI in other domains also raises serious red flags that AI-designed organisms or proteins may not be reliable, may have unexpected (and unexplainable) biodesign decisions built into them, and be designed with problematic biases (see box on AI hallucinations, bias and data poisoning). Combining the risks and uncertainties associated with genetic engineering with the emerging risks and uncertainties accompanying generative AI creates a perfect recipe for 'black swan' type disasters (high-impact, unusual risks that cannot be foreseen).[49]

A broader concern for regulators is what has been called 'the pacing problem'. As automated design and building of engineered synthetic organisms, viruses, and proteins becomes easier and less reliant on human laboratory work, regulators may expect to see a rising deluge of applications for GMOs and novel materials that will have to be assessed and monitored. Some military scholars are already warning against any move to hand biosafety assessment of AI-generated products and organisms to automated AI biosafety assessors – arguing for the importance of keeping human expertise firmly in the loop.[50]

46 A round-up of AI legal copyright cases can be found here: "Case tracker: Artificial intelligence, copyrights and class actions", BakerHostetler, accessed 25 July, 2024, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/

47 See https://haveibeentrained.com

48 See https://www.basecamp-research.com/biotech

49 See https://en.wikipedia.org/wiki/Black_swan_theory

50 Vindman, C., et al. n.d. *The convergence of AI and synthetic biology: The looming deluge*, preprint, accessed 25 July 2024, https://arxiv.org/pdf/2404.18973

AI hallucinations, bias, and data poisoning

Although AI companies present their systems as being competent to generate text, images, and genomes of living organisms, the first couple of years of generative AI systems have thrown up very significant problems with the quality and accuracy of outputs. A part of these problems follows from the fact that an AI system is only as good as its training dataset, which must be carefully curated, cleaned up, and managed, to get meaningful and fair results. Otherwise 'garbage in' creates 'garbage out'. Data is not 'pure and immaculate'[51] but can be messy, easily misclassified, missing, or heavily biased. More significantly though, Generative AI systems are not designed to create or verify 'truth' but to only to create credible-seeming outputs. In this way, they may never reach the standard of good science.

One of the most common output problems from generative AI is known popularly as 'hallucinations'. This describes a generative AI model producing a text or image that appears to make sense but on closer investigation is full of errors, lies, falsehood, and improbable elements. AI-generated images will, for example, commonly show people with six fingers, scramble text or blur features. AI-generated text may wrongly report facts, mishandle numbers, or invent references. In 2023, analysts estimated that AI chatbots hallucinate 27% of the time with almost half (46%) of their responses containing factual errors. A digital image or text full of such hallucinatory errors (as is common) can be annoying or even funny, but if a living organism or protein is created with significant hallucinatory errors, it could prove dangerous.

Scholars however have warned that the term 'hallucinations' may be a misleading metaphor since in fact, the fundamental architecture of generative AI systems means that these systems are not interested in producing truthful outputs. The errors are not mistakes but baked into how generative AI works. Scholars have even suggested that generative Ai systems should be scientifically classified as "bullshit machines" since they "are simply not designed to accurately represent the way the world is, but rather to give the impression that this is what they're doing."[52]

One additional source of errors is a bias built into the training data. For example, AI text systems have been found to reflect social biases against ethnic and racial stereotypes that are found in the training data, or to deepen gender and racial biases in hiring or legal decisions. For this reason, generative AI companies hire many thousands of human annotators to 'clean up' their outputs.[53] In a biodiversity context, the selection of biological data from only certain commercial crops or drawn from limited ecosystems and geographies could build in problematic biases. It has also been shown that commercial or politically motivated entities can deliberately 'poison' datasets (add or remove specific data) to drive biased outcomes or to make options invisible.[54] Exerting strong governance over training datasets therefore becomes an extremely important but difficult policy necessity.

51 Bronson, K. 2022. The immaculate conception of data: Agribusiness, activists, and their shared politics of the future. Montréal: McGill-Queen's University Press

52 Hicks, M.T., Humphries, J. & Slater, J. ChatGPT is bullshit. Ethics Inf Technol 26, 38 (2024). https://doi.org/10.1007/s10676-024-09775-5

53 Dzieza, J. 2023. "AI is a lot of work", *The Verge*, 20 June

54 See "Data poisoning: The essential guide", *Nightfall AI*, accessed July 25, 2024, https://www.nightfall.ai/ai-security-101/data-poisoning

3. Does AI generative biology pose biosecurity/ bioweapons risks?

One of the risks of AI generative biology that some governments are already acting to address is the potential for such systems to generate new pathogens, toxins, or other outputs that could be weaponised. AI generative biology models could be used to find deadly proteins or harm crops or to redesign viruses and bacteria to spread more easily or attack certain species. At least one science paper describing an AI protein generator describes how that system can be easily used to artificially generate analogs of natural poisons such as snake venom.[55]

4. Will the integration of AI with SynBio improve or worsen health and sustainability?

Part of the reason that large AI corporations are racing into generative biology is that they hope to publicly project the vision for a potentially limitless range of speculative technofixes, thereby improving the social acceptability of AI. Generative biology may theoretically be used to create new alternatives to animal protein (Alt-proteins), which its proponents argue could reduce the climate intensity of processed food. It is imagined that AI proteins might be developed for materials that will sequester carbon dioxide and that new generative biology-designed pesticides or viruses might eradicate malarial parasites. Pharmaceutical companies are particularly hopeful that they can use generative biology tools to create new drugs or to replace natural product compounds with new generative biology-designed compounds that they can patent and sell.

For now, all of these promises are speculative and presented as overly simplistic technofixes. Common sense requires that governments carry out ground-truthing to separate hype from promise but also technology assessments to understand the real and likely societal impacts and consequences. The history of synthetic chemistry and more recently of GMOs has shown that bold promises by industry need to be subject to societal scrutiny and the precautionary principle, in case they turn out to be false solutions, either not working or triggering more intense problems.

In the case of integrating AI with SynBio, policymakers need to be attentive to the real environmental costs that are already emerging from generative AI.[56] Kate Crawford, a leading AI scholar and research professor at the University of Southern California, suggests that looking at specific AI models misses the point and that, "If we really look at this big picture, it's a gigantic planetary system." To make AI work, one needs a huge amount of data, as well as huge amounts of computation. Currently, this computation takes the form of NVIDIA chips (H100 and A100) manufactured in Taiwan. Computation takes energy, and the energy required to run these models is staggering – predictions range between 5 to 50 times more energy-intensive than traditional search functions. The heat generated by this processing needs to be cooled, meaning copious amounts of water. Researchers from the University of California Riverside and the University of Texas Arlington estimated that about 700,000 litres of water could have been used to cool the machines that trained ChatGPT-3 at Microsoft's data facilities (Li et al. 2023). ChatGPT-4 is even more energy-intensive, and both Microsoft and Google report that their energy consumption has spiked in recent years — up 50% in five years in Google's case (Malmo, 2024) — largely due to generative AI.

55 Zulfiqar H, Guo Z, Ahmad RM, et al. 2024. "Deep-STP: A deep learning-based approach to predict snake toxin proteins by using word embeddings", Front Med (Lausanne). doi: 10.3389/fmed.2023.1291352. PMID: 38298505; PMCID: PMC10829051

56 Berreby, D. 2024. "As use of AI soars, so does the energy and water it requires", *Yale E360*, 26 February, https://e360.yale.edu/features/artificial-intelligence-climate-energy-emissions

AI Big data - **© frank60, Shutterstock**

Additionally, AI computation is carried out on thousands of electronic servers, whose fabrication requires extensive water use (for chips), industrial battery components, and the extraction of metals, rare earths, and other minerals from the Earth. The recent generative AI boom that was accelerated by the release of ChatGPT has sparked a global boom in building energy-hungry AI data centres. This, in turn, is reversing hard-won carbon reduction gains in the energy sector and diverting freshwater away from farmland and poorer communities, to cool AI chips. The largest AI companies, such as Microsoft, are now spearheading a renaissance of nuclear power facility construction, to keep their data centres running.[57] Even the attempt to supply renewable energy to data centres is driving a problematic expansion of critical mineral extraction. AI companies argue that AI will become more energy-efficient over time and that AI can also aid in efforts to tackle climate change. However, there is increasing concern that AI is exacerbating ecological threats at precisely the moment when scientists warn there is very little time to turn the global climate and biodiversity crises around. Computation and training on genomic DSI may be particularly energy and resource-intensive and should not be adopted lightly by Parties already struggling to protect biodiversity, water, food security, and net zero goals.

57 Hiller, J. 2023. "Microsoft targets nuclear to power AI operations", *Wall Street Journal*, 12 Dec, https://www.wsj.com/tech/ai/microsoft-targets-nuclear-to-power-ai-operations-e10ff798

5. What are the implications of AI/SynBio integration for traditional knowledge and practices?

The second aim of the CBD concerns the sustainable use of biological diversity. Bodies such as the Working Group on Article 8J play an essential role in protecting traditional knowledge and practices that underpin the conservation and sustainable use of biodiversity. Generative biology companies seeking to produce alternatives to natural products, including medicinal and food ingredients, may be directly undermining sustainable value chains that support millions of small farmers worldwide. The digital genetic resources that train AI models are often originally taken from the territories of indigenous peoples and local communities. These resources are used in their ceremonies, food systems, and ways of life and exist in close relationships with the communities. Rendering that knowledge and those relations as DSI to train corporate AI systems for producing commercial synthetic organisms and proteins extends the centuries of colonial theft and appropriation that Indigenous peoples and local communities have long suffered. Scholars have termed this new wave of digital appropriation of life 'data colonialism'.[58]

In the coming years, discussions over generative biology are likely to be joined by other urgent discussions over the implications of deploying AI systems in forestry, land use, conservation, marine management, restoration, species management, agriculture, and more. Indigenous peoples and local communities – as the original holders of biological and other resources that are being taken, trained on, and manipulated through AI – should be taking the lead in global policy decision-making around AI at the CBD and beyond.

Seed exchange - **© P.Kimeli_CCAFS, Flickr**

58 Couldry, N. & Mejias, U.A. 2020. "The costs of connection: How data are colonizing human life and appropriating it for capitalism", *Social Forces*, Volume 99, Issue 1, Page e6, https://doi.org/10.1093/sf/soz172

A way forward for the CBD

In the years to come, questions of how AI will transform conservation, sustainable use, and fair sharing of benefits can be expected to become a major topic of policy debate and action dominating all programmes of the Convention. For now, the topic of generative biology and integration of SynBio with AI is of urgent and rising importance for Parties to grapple with, separate hype from reality and get on top of.

Moving genetic engineering and production of novel biomolecules into the black box of generative AI may, in the end, fail to deliver the glittering promises currently being made by big technology companies. However, in the meantime, Parties need to recognise that this approach undermines all three key pillars that have been erected to ensure a just and precautionary development of biotechnology: biosafety, fair and equitable benefit sharing, and means for liability and redress. It also raises serious socio-economic considerations.

Bean farming - **© Alliance of Bioversity International and CIAT, Flickr**

At the same time, the application of generative AI tools is being promised to increase the volume and complexity of bioengineered organisms and products to levels that may overwhelm regulatory capacities. This comes at exactly the time when regulators most need to ensure more carefully that they have a 'human in the loop'. Hype and promises from digital giants newly taking command of the bio-economy will need to be investigated and sorted through – by human beings, not algorithms – particularly drawing on situated knowledge of Indigenous people and local communities. The CBD will need to maintain its historic commitment to the precautionary principle in the face of Silicon Valley's preferred approach of 'move fast and break things' (sometimes called 'the innovation principle').[59]

At COP 16 of the CBD, Parties have a very clear and straightforward opportunity to prevent the issue of generative biology from upturning and hollowing out the decades of good work that the CBD has led, to help establish good governance over biotechnology. Following the recommendations of the mAHTEG on Synthetic Biology, Parties should:

- Initiate a policy formulation process addressing the integration of AI with SynBio.
- Request the mAHTEG to undertake a further assessment of the integration of AI with SynBio, leading to a report that addresses, inter alia, potential impacts on biosafety; the sustainable use of biodiversity; equitable access and benefit-sharing; social, economic and cultural aspects; impacts on traditional knowledge and practices; and other relevant matters.
- Request the CBD Secretariat to develop a technical publication series addressing AI and to participate in UN system-level activities concerning AI governance.
- Urge Parties to consider the development of effective and equitable governance arrangements for AI data sets; foundation models; algorithmic biodesign tools; automated science tools; and the use of SynBio organisms, components and products in cyberphysical systems.
- Insist that big technology firms and generative biology companies that train their AI models on DSI are utilising DSI for commercial purposes, and therefore should be subject to the multilateral mechanism and the obligations to pay into the global fund on DSI.

59 Corporate Europe Observatory. 2018. "The 'innovation principle' trap: Industries behind risky products push for backdoor to bypass EU safety rules", 5 December, https://corporateeurope.org/en/environment/2018/12/innovation-principle-trap

Glossary of key terms

Access and benefit sharing – Refers to how genetic resources may be appropriately accessed with consent, and how the benefits that result from their use are fairly and equitably shared with the people or countries that provide them (providers).

Artificial intelligence (AI) – A set of technologies that are based primarily on machine learning and deep learning, and are used for data analytics, predictions and forecasting, object categorisation, natural language processing, recommendations, data retrieval, and more.

Biocomputation – Systems of biologically derived molecules or engineered living cells that perform computational processes in place of silicon-based computation.

Biodesign – The application of methods and approaches of creative design and engineering to the deliberate design of living materials and systems.

Bioproduction – The growth, development, and harvesting of industrial compounds (e.g. small molecules or proteins) in biotechnological facilities such as fermentation vats or cell culture. Also sometimes called 'bioprocessing'.

Biotechnology – The UN CBD defines biotechnology as "any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use". It commonly refers to 'modern' molecular biotechnologies such as gene manipulation and gene transfer, gene editing, DNA typing, and cloning of plants and animals.

'Black box' problem – The intrinsic lack of transparency or external interpretability of AI systems whose internal models are hidden from being understandable.

Cyberphysical systems – More complex real-world systems in which real-world activities or objects are controlled by digital technologies, including AI, sensors, and automation.

Deep learning – A form of machine learning that uses multilayered neural networks to process data in a way that is inspired by the human brain.

Diffusion model – A generative AI model that learns by adding random noise to destroy data and then reconstruct data. Used for example for Stable Diffusion (generative AI image maker).

Digital Sequence Information (DSI) on genetic resources – A policy term that refers broadly to genetic sequence data and other related digital data. This includes the details of an organism's DNA, RNA, protein sequences, and other molecular biological codes that can be expressed in digital form.

Discriminative AI – Focuses on classifying existing data. It does this by learning the decision boundary that separates different classes of data – e.g. image classification or identifying the taxonomy of species.

Explainable AI – Proposals to develop AI models that do not hide their workings in a 'black box' – but rather store their workings where it is possible to trace the reason and source for a particular output (still theoretical at this point).

Foundation models – Large-scale, broad, adaptable AI models with many parameters, trained on a broad gamut of datasets in a self-supervised manner. Can be further trained on more specific datasets for different applications.

Generative AI – Deep-learning models that can take raw data and generate statistically probable outputs when prompted. Generative models encode a simplified representation of their training data and draw from it to create a new work that's similar, but not identical, to the original data.

Generative biology – An emerging industrial term used by companies that apply generative AI models trained on genetic or biological sequence information to predict designs of new biological sequences (e.g. DNA or protein sequences).

Hallucination – A response generated by AI that contains false or misleading information presented as fact. Might include images or pieces of text that look credible but are impossible or erroneous.

'Human in the loop' – An emerging concept in AI discourse describing processes that step away from full machine automation by ensuring human agency remains over decision-making and outcomes.

Large language models (LLMs) – Machine learning models that ingest large amounts of text, from which they analyse, classify, and then generate credible-seeming language-based texts.

Liability and redress – Legal system established under the CBD, and in particular the Nagoya-Kuala Lumpur Supplementary Protocol on Liability and Redress, which set out a liability regime for damage caused to biodiversity (e.g. from living modified organisms) as well as procedures to ensure repayment or redress for damage.

Machine learning – The use of computer systems able to learn and adapt without following explicit instructions. Machine learning systems use algorithms and statistical models to analyse and draw inferences from patterns of data.

Modern Biotechnology – The UN CBD defines "modern biotechnology" to mean "the application of:

- In vitro nucleic acid techniques, including recombinant deoxyribonucleic acid (DNA) and direct injection of nucleic acid into cells or organelles, or
- Fusion of cells beyond the taxonomic family, that overcome natural physiological reproductive or recombination barriers and that are not techniques used in traditional breeding and selection".

The Precautionary Principle or Precautionary Approach – An approach in environmental law that rebalances the burden of proof in the case of risk. The 1992 Rio Declaration defines the precautionary approach as "where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation".

Protein large language models (Prot-LLMs) – LLMs trained on protein-related sequence data – including amino acid sequences, protein folding patterns, and other biological information – that attempt to predict protein structure, functions, and interactions.

Robot-ready crops – Crops genetically engineered to communicate information to robotic agricultural equipment (e.g. fluorescing when stressed by insect attack or drought). The crops are designed to support digital agriculture.

Synthetic biology – Defined by the UN CBD as "a further development and new dimension of modern biotechnology that combines science, technology, and engineering to facilitate and accelerate the understanding, design, redesign, manufacture and/or modification of genetic materials, living organisms, and biological systems".

Training data – The initial set of data used to train a machine learning algorithm or model.

Transformer-based model – An example of a generative AI model based on an encoder and decoder to analyse the relation of words (or similar) to other words in a piece of text. Used, for example, for the ChatGPT generative AI chatbot.