UK Government

Scottish Government
Riaghaltas na h-Alba

AI for Engineering Biology:

Transatlantic cooperation on data, models and standards

Report on event hosted 17–18 March 2026 by the UK government in partnership with the University of Edinburgh and supported by the Scottish Government

THE UNIVERSITY *of* EDINBURGH

EDINBURGH INNOVATIONS

Published 5 August 2026

Report written by Johannes Boucsein, MD, MSc, of Inclusive Innovation

Background

The UK government's *AI for Science Strategy* (AI4SS)[1] and the US government's *Genesis Mission*[2] both made AI-enabled science national priorities, shaping UK and US research trajectories for the next decade. The AI4SS named engineering biology[3] a priority domain, while a key emphasis of the Genesis Mission is on 'scaling the biotechnology revolution'.[4] Importantly, and providing a departure point for the discussions summarised in this report, both national strategies emphasise that progress depends on the availability of high-quality, interoperable, and trustworthy AI-ready data.

As part of a broader effort to strengthen transatlantic cooperation in AI-enabled engineering biology, the UK government and the University of Edinburgh organised a symposium on 17–18 March 2026. The event was designed to test the contention that both data generation and data utilisation need to evolve to realise AI's true potential in engineering biology. The symposium provided a collaborative environment for stakeholders to explore cross-cutting opportunities and challenges associated with AI-ready data, models, and standards in engineering biology through a transatlantic, cross-disciplinary lens.

The event convened more than 80 participants from academia, industry, funding bodies, and government institutions. With true cross-disciplinary expertise in both AI and engineering biology a rare and emerging skill-set, participants were selected based on their ability to speak with authority to at least some aspects of the challenges under discussion, focusing on three application areas within engineering biology: detecting, preventing, and treating disease; food systems; and biomanufacturing. The programme emphasised interactive formats and shared sense-making with the aim of developing a common strategic view, identifying catalysts for progress, determining where cooperation might have the greatest impact, and building momentum and relationships for sustained dialogue on transatlantic cooperation. Inclusive Innovation facilitated the event.[5]

The following report summarises the most prominent themes and topics that participants brought forward *during the event*. This was a cross-disciplinary, exploratory event. Report reviewers noted that some discussions extended beyond the core focus (on establishing the data foundations for AI-enabled *engineering biology*) into questions of broader relevance to biological sciences.[6] These aspects are retained in the report in the interests of accuracy and completeness alongside the material directly relevant to the symposium's core objectives.

Part 1 of the report provides a high-level summary of the key opportunities, barriers, and action points. **Part 2** underpins this summary with a detailed analysis of each of the three application spaces.

---

[1] https://www.ukri.org/publications/uk-research-and-innovation-strategy-2026-to-2031/

[2] https://www.energy.gov/undersecretaryforscience/genesis-mission/genesis-mission

[3] In the UK, the term engineering biology (EB) is more common than the term biotechnology and will be used throughout the report. The UK government defines engineering biology as the design, scaling and commercialisation of biology-derived products and services that can transform sectors or produce existing products more sustainably. It draws on the tools of synthetic biology to create the next wave of innovation in the bioeconomy.

[4] https://www.energy.gov/undersecretaryforscience/genesis-mission/scaling-biotechnology-revolution

[5] https://www.inclusiveinnovation.org/

[6] Reviewers were symposium participants. Participants were generally invited to review the draft report where they had participated in panels or presented at the event. The author is grateful for their corrections and clarifications.

About Inclusive Innovation

Inclusive Innovation is a global facilitation and collaboration consultancy that helps people work more effectively across disciplines, sectors, cultures, and geographies to address complex societal challenges. Drawing on expertise in collaborative leadership, transdisciplinary research, and participatory workshop design, Inclusive Innovation creates interactive processes that enable diverse groups to think together, build trust, and move from dialogue to action. The organisation's work spans scientific research, international development, public policy, and innovation ecosystems, supporting institutions around the world in tackling issues that no single discipline or organisation can solve alone.

For the UK government and the University of Edinburgh, Inclusive Innovation worked closely with the organising team to design and facilitate both this symposium and an accompanying forum held the following day, in which a group of experts further explored biomedical opportunity areas in which AI might accelerate EB research and development (R&D).

Following the events, Inclusive Innovation prepared this report to capture the key themes, insights, and opportunities that emerged from the symposium discussions. Rather than serving as a verbatim record of proceedings, the report synthesises participants' contributions into a coherent summary intended to support ongoing dialogue and inform future activities.

Table of Contents

Part 1: High-Level Summary

Opportunity / Research Frontier

Integrating AI and engineering biology (EB)

Participants agreed that AI-enabled methods have significant potential to accelerate biotechnological R&D by improving foundational understanding of biology, facilitating biological engineering, and supporting translation to marketable biological applications.

During these discussions, two schools of thought were evident. Some argued for priority to be placed on the assembly of AI-ready data that would support deeper understanding of fundamental biology, arguing that this is a critical step in developing AI tools for biological engineering. Some opposed this point of view, arguing that an engineering approach does not necessarily require this kind of fundamental understanding to work, and that priority should be placed on generating data from engineered biology, or with direct relevance to bioengineering challenges.

Across the breakouts participants argued that genuine integration of AI and EB was not merely a matter of applying new technical tools. The process of knowledge generation had to be reconceived as "a new research model in which AI, data generation and biological engineering are tightly coupled".

On a practical level, participants identified several avenues where a closer integration of EB and AI promised near- and mid-term opportunities:

1. Making gene sequences and gene circuits easier to design and engineer, and predicting effects of changes made to the genome;
2. Improving data quality, curation and annotation;
3. Linking existing fragmented datasets and detecting patterns across multimodal data to move beyond correlative observations and enable predictive cause-and-effect models (e.g. predicting protein function from protein sequence data);
4. Identifying the most impactful research questions and data gaps for prioritisation (active learning);
5. Increasing the efficiency of production processes, ultimately making EB-enabled products competitive on the market;
6. Enabling real-time adaptation of experiments and production processes through digital twins;
7. Reducing the load of trial and error wet-lab experiments by enabling *in silico* research.

In the breakout discussion on Detecting, Preventing, and Treating Disease, participants discussed several advances that successful integration of AI could bring about, though not all related directly to engineering biology. These included: better identification of biomarkers for pathologies; improved design of target molecules for cell-therapy efficacy and manufacturability; novel small-molecule therapeutics with greater target precision and longevity; and advances in stem-cell programming for regenerative medicine and biological sensors. Reviewers of this report felt that, although the discussion generated useful

insights into how AI might affect the development of therapeutics, the discussion would have benefitted from a tighter focus on EB and that follow-up discussions would be worthwhile.[7]

In the breakout discussion on Food Systems, integrating AI and EB was anticipated to support: crop trait prediction and, in the longer term, active cross-species transfer of traits into food crops; soil health (including through the use of digital twins); biomanufacturing of soil amendments and replacements for synthetic fertilisers; microbiome engineering (soil, gut, and industrial); pathogen detection; biodiversity monitoring; and progress towards future food and feed systems (cell-cultivated products, alternative proteins and the like). Through improved techno-economic assessment (TEA) and life cycle assessment (LCA), AI could also support the commercialisation of sustainable novel food products.

In the breakout discussion on Biomanufacturing, promising applications of AI included: the assessment of experimental design spaces and novel products and process titres, rates, and yields with the aim of increasing cost-effectiveness; the identification of potential risks of failure and likely solutions during scale-up; predicting performance at industrial scale based on suitable bench-scale prototypes; self-driving bioreactors, automated wet labs, and real-time adaptation enabled by digital twins and biological sensors; and metabolic pathway optimisation for yield improvement.

Transatlantic collaboration

Participants felt that transatlantic collaboration could accelerate both countries' trajectories because the respective strengths of the US and the UK in EB and AI are structurally complementary, and combining them would accelerate the pace of innovation beyond what either country could achieve alone.

Participants described the UK as offering strong foundational research, high-value biological databases, population-level National Health Service (NHS) data, a well-networked and collaborative research community, access to European partners and infrastructure, and an agile start-up and spin-out ecosystem. The US was seen as bringing compute infrastructure at unmatched scale and sophistication, the most advanced AI models, leading agri-businesses, funding and investment at scale, and a long track record of successful translation, commercialisation, and scale-up.

Existing research networks and working relationships could be leveraged to support cooperation. Participants pointed specifically to a track record of cross-funded projects between standards-forming bodies and biofoundries that could serve as a model for what comes next.

[7] Reviewers mentioned among others: enabling sustainable drug production, removing the dependence on toxic chemicals; engineering therapeutic pathways circumventing off-target effects and reducing high attrition rates in drug development; automated design and rapid testing of new therapeutic antibodies.

Challenges, Constraints, and Bottlenecks

Data needs to be generated at greater scale, reproducibility, and quality

Both training useful AI models and gaining deeper insight into biological function require large quantities of trustworthy data, and it needs to be appropriate to the specific research questions under examination. Participants agreed that there was a requirement to increase data complexity: examples aired were linking genomic data with data on cell output at multiple levels ("multi-omics"), generating data purposefully for defined questions, and moving from observational to intervention data ("perturbation data"), whether on natural or engineered cells. Training more robust AI models – for prediction or causal modelling alike – requires linking datasets and increasing their complexity.

Existing data, however, is often ill-suited to the training requirements of AI models: it is mostly observational and limited in scope, and the available multi-omic datasets remain fragmented, poorly integrated, and drawn from naturally occurring rather than engineered biology. It is also often of limited quality, with weaknesses including missing, erroneous, or incomplete metadata, high heterogeneity across sources, and poor reproducibility. Even if existing data were better leveraged through quality improvement, standardisation, and cross-linking, generating *more* data, more *diverse* data, and more *connected* data would remain essential. Consensus formed that this is only achievable through systematic data generation and collaboration at previously unseen scales.

Negative or failed data is inaccessible

The unrealised value of "failed data" – data on experiments that did not succeed and, importantly, why they failed – was noted across application spaces. Not leveraging this asset means duplicating research effort and repeating mistakes, creating inefficiency and wasting resources. Negative data is also key to making AI models more robust and enabling them to predict failures as well as successes. Participants attributed the problem to a lack of positive incentives for making negative data available: scientific journals show little interest in publishing it, and dedicated infrastructure beyond the traditional publication system remains limited. Commercial actors, meanwhile, recognise the value of such procedural knowledge and guard it tightly as something providing a competitive advantage.

Trusted data at scale requires standards that don't yet exist

Data quality and standards are crucial for the coordinated generation of targeted, trusted new data, as well as for annotating and improving existing data. Established standards, however, often do not address the specific needs of model training, nor are they flexible enough to keep pace with a rapidly evolving field – and new standards are not readily available. Crucially, participants noted, it is often not yet clear what "AI-ready data" looks like for EB, yet data generation continues regardless, which raises questions of long-term value:

> *"We need to make sure that, five years from now, we don't need to retrofit data that we're generating over the next five years." – a panellist*

As a bootstrap approach, AI was proposed as one means of raising the quality of existing data – though this, too, hinges on first defining quality standards and building the models to apply them. A number of

participants also stressed that purely AI-based approaches were unlikely to succeed, and that human involvement would be critical to improving existing data resources.

Standardised, precise measurement tools were also described as the foundation of the design–build–test–learn (DBTL) cycle underlying EB R&D. Several interconnected gaps present themselves here: the absence of platform measurement standards; incompatible file types; limited software and equipment connectivity; inconsistent terminology, nomenclature, and units; inadequate provenance frameworks; and poor benchmarking.

Reviewers pointed to a further aspect not addressed during the event: beyond data standards, AI models require reporting standards too – including, for instance, reporting the training data alongside the model. Several model registries adhering to model-reporting standards are available (Croissant[8] and DOME-ML[9] were named explicitly) and offer a valuable opportunity to streamline the EB research process.

Lack of database capability and inadequate data governance models prevent collaboration and scale

Participants agreed that both data generation at scale and closer coordination between EB stakeholders require a mature database infrastructure supported by robust data governance – neither of which sufficiently exists at present. Such an ecosystem would need to provide access to vast quantities of data and metadata while also supporting the next generation of automated research through agent-ready access.[10] It would also have to be flexible enough to keep pace with the field, anticipating newly emerging kinds of data, such as sensor readings from production processes.

Participants pointed to the current lack of structured, equitable access pathways: because established FAIR[11] principles are often not enforced, large quantities of data go unpublished, while what is published is scattered across public databases, proprietary repositories, and journals. As a result, data that is abundant in theory remains hard to find and often impossible to access in practice.

Intellectual property frameworks are outdated

Participants cautioned that existing intellectual property (IP) frameworks do not adequately address value capture along the chain from data creation to model: once data is used to train an AI model, its commercial value accrues to the model. This is a key issue for maintaining existing databases and a major barrier to generating and investing in new data.

> *"We want people to make data for us to train models on, but how does the value created by the model then propagate back along the chain?" – a participant*

Key data and knowledge, along with the most advanced AI models, are often held by commercial actors. Slow, complicated legal processes – particularly around background/foreground IP arrangements –

[8] https://mlcommons.org/working-groups/data/croissant/

[9] https://dome-ml.org/

[10] access by automated software agents

[11] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

combined with a lack of frameworks for pre-competitive data sharing and weak incentives, discourage collaboration with academia, often producing "knowledge monopolies" in the form of proprietary models.

Addressing these issues goes hand in hand with resolving the current "lack of clarity around licences and terms of use" for sensitive health data – a barrier to collaboration both with commercial actors and across the Atlantic. The latter is further hampered by non-uniform regulation of AI and EB between the two countries.

Translation from science to marketable applications often fails

Translation from EB R&D into marketable products remains limited, because what works in the lab often fails at scale: cells and microbes behave differently in production-scale environments than in the research-scale wet lab, production becomes inefficient, and it is often unclear whether the root cause lies with the cell design or the production process.

Much of R&D relies on platforms chosen for their tractability, and knowledge often transfers poorly from cell line to human, plant cell to whole plant, and wet lab to industrial-scale incubator. Building more representative models of biological function requires population-scale, cross-species data; improving scale-up efficiency demands comprehensive real-time process data (inputs and outputs, state, waste-stream) and the sensors to measure it; and bringing novel foods to market requires earlier, more detailed TEAs. Robust, standardised assessments that predict real-world performance from bench-scale designs are also lacking.

Participants noted that in novel therapeutics, promising products often fail at the last, costly stages of clinical trials because of immunogenicity and unexpected side effects. AI could help address this, though reviewers noted the point applies across efforts to prevent and treat disease and is not specific to EB. In the food sector, many products that work at bench scale are not cost-effective or sustainable enough for industrial-scale production. For both health and non-health applications, the cost and time of development are high, as is the risk of failure even after significant expenditure. Consensus formed that the EB space at large needs funding and support that not only tolerates but encourages ambitious research at the frontiers of feasibility – even at the risk of failure.

Reviewers noted that in other disciplines facing similar scale-up and translation challenges, measurement and data standards have proven pivotal: beneficial across all Technology Readiness Levels (TRLs)[12], they become necessary for translation and scale-up. EB was seen as unique in its dependence on the interoperability of diverse processes and systems during scaling – which is precisely where AI and data quality become indispensable.

Grant structures do not reflect the scale and duration required for transformative collaborative work

The existing research funding structure was described as largely outdated and ill-suited to the long-term, coordinated, highly collaborative approach now needed. Current funding is geared towards per-project grants and individual research groups, providing insufficient support for scale or collaboration, let alone a joint mission. It comes with short timelines that fail to match the complexity of frontier research – much less the multi-decade, systematic knowledge generation the field requires. Funding is also limited in

[12] Technology Readiness Levels are a type of measurement system used to assess the maturity level of a particular technology.

scope, providing inadequate support for essential enablers: data standards, sustained and curated datasets, foundational models without immediate application, methods and wet-lab tools, Responsible Research and Innovation (RRI), and essential collaboration with metrology institutes, industry, and regulators. Finally, current structures are mostly built around within-sector grants and do not adequately support interdisciplinary work.

On transatlantic collaboration, participants raised parallel concerns: too few joint programmes, and short funding cycles that do not match the slow work of finding partners, developing ideas, and establishing working relationships. Grant application processes were described as cumbersome and sometimes redundant on both sides, requiring applicants to satisfy several institutions and differing review styles – turning applications into "double jeopardy". Against a backdrop of political uncertainty, this often poses too great a risk to be worth pursuing.

Currently available wet-lab technology has significant limits

Participants repeatedly observed that, beyond more and higher-quality data, the field also needs significantly improved wet-lab tools: genome-editing technologies such as CRISPR/Cas still offer limited precision and carry significant off-target effects; delivery of genomic payloads into human or plant cells suffers from similar drawbacks; and biosensors capable of real-time measurement in industrial-scale environments simply do not yet exist. One reviewer felt that stable, predictable, controllable genetic circuits were also a critical gap, though this was not recorded in the discussion notes.

The scale of data generation envisaged also demands a degree of automation to be efficient. The aspiration of fully automated, AI-enabled laboratories and real-time experimental adaptation via digital twins was voiced repeatedly – with the caveat that the technology is not yet sufficiently refined. Such capabilities are unlikely to be developed by EB alone and would require close integration of biology with chemistry, materials science, engineering, and computer science.

Several participants cautioned, however, that biology is inherently highly variable, and better technology may not overcome this constraint. Training AI to handle this noise, and developing better tools and approaches to assess uncertainty, may be as important as perfecting methods and instruments.

Integrating EB and AI requires coordination and a dual-skilled workforce

The scale of data generation needed can only be achieved through systematic efforts that link institutions and countries. The required governance and coordination were described as attainable only through large, integrated programmes; the level of management and coordination capacity needed, as one participant noted, does not currently exist in any single institution. Successful integration also depends on experts fluent in both disciplines, yet these remain rare: training pipelines are typically discipline-specific and offer limited opportunity to develop interdisciplinary skills or a practical understanding of AI's applications and limitations in research. Mobility barriers (visas, administration, tax, healthcare, pay gaps) further obstruct recruitment of the talent that does exist.

Responsible Research & Innovation requires practically applicable frameworks

The responsible use of AI in EB research requires a robust ethics ecosystem spanning regulatory institutions, education, and self-governance. Existing AI ethics frameworks remain abstract and high-

level, and are not fit to inform concrete research practice. This challenge could not be addressed simply by importing ethical frameworks from other disciplines such as the life sciences.

The impact of AI also reaches well beyond technical research: the pace at which the technology is evolving challenges existing decision-making processes, institutions, and governance frameworks. Panellists warned that this could lead to coordination errors and unexpected and undesirable consequences.

> *"Conversations in this space are often framed around datasets, technical capabilities and accelerating innovation. But we are also doing something else: we are reorganising entire research and innovation ecologies [...] How might we expansively re-imagine what collaboration could look like [...] before we end up building worlds that we did not intend?" – a panellist*

Established frameworks for stakeholder inclusion — particularly of vulnerable groups — do not adequately respond to the fundamental changes AI imposes. Panellists described it as vital to work across diverse perspectives and expertise to surface and address the environmental and social impacts of AI, including its effects on supply chains, energy consumption, and labour.

AI remains a frontier technology

Despite enthusiasm about the convergence of AI and EB across application spaces, many participants expressed hesitation about the measurable impact AI has so far had on research. While flagship models such as AlphaFold have helped predict protein structure from sequence – a long-standing challenge – they were not yet seen to have significantly advanced the causal understanding of biology, nor to have produced tangible commercial applications. Some participants pointed to the complexity of cell biology and cautioned that knowledge from one cell or organism may not generalise, arguing for narrower, specialised models aimed at clearly defined problems over the ultimate aspiration of universal cell models.

Some views were more critical still, describing AI as the current buzzword – its capabilities unclear, and offering little beyond traditional machine learning and statistical modelling. On this view, it may be more worthwhile to generate good data that answers the right questions than to treat AI as a panacea. As a case in point, participants in the detecting, preventing, and treating diseases application space admitted to "some magical AI thinking" while "lacking concrete ways forward for those cell or disease models". In food systems, some cautioned that "food is not tech" – that it is deeply interwoven with psychology, taste, culture, and tradition, and that focusing on AI at the expense of these risks building sophisticated technology that never leaves the lab. During drafting, reviewers noted the likely value of further cross-disciplinary discussion to identify where AI would have greatest utility in accelerating the ability to engineer biology, and to focus AI-ready data generation efforts accordingly.

What needs to change?

Fundamentally change data generation

Consensus formed that data generation must be scaled up and coordinated across research groups and institutions, and between countries. Participants stressed the need not only to generate more data, but to generate more comparable data, through shared standards, rich metadata, and comparable processes. Reviewers with expertise in standards and metrology noted that such comparability is achieved through inter-laboratory comparisons, which ensure datasets are consistent and shareable, an established practice in international metrology that remains overlooked by academia and funders.

At a strategic level, three diverging pathways were set out: an implementation-focused, challenge-driven approach; foundational research into the underlying cell biology; and active-learning-enabled, agnostic data generation at scale for better prediction, regardless of causal understanding. Participants did not converge on a single strategy, and reviewers felt this area would benefit from further discussion.

Build the infrastructure to leverage negative data

Realising the value of negative data requires building the infrastructure to host, share, and access data on failed experiments. Scientific journals, commercial actors, and the research community alike will need to be involved in building the incentives to share it.

Define data requirements and make quality a first-order priority

The first step is to define what "AI-ready" means for the EB space, then develop the data standards, measurements, and quality criteria to match. This requires parallel efforts in technical research, consensus-building, and the harmonisation of research protocols and practices. Established working relationships between metrology and standards-forming bodies in the UK and US offer a strong foundation for collaborative development and international alignment. Above all, it requires a change in mindset: data requirements and AI-readiness must become first-order priorities in the planning of R&D. Making data requirements and sharing practices preconditions of funding could be one way to drive progress.

Build accessible infrastructure and governance for data sharing

Data is increasingly the main currency of knowledge generation; using it effectively depends on infrastructure for safe storage, exchange, and processing, and on fair, efficient access. This means investing both in new cloud infrastructure and in the means to use it efficiently; retaining and sustaining data capacity through curated datasets; and developing modern data-sharing and IP frameworks to govern data use – while resolving the unsettled question of how value is distributed along the data-generation-to-model-development chain.

Greater effort should also go towards bridging the gap between academia and commercial actors, and exploring new modes of collaboration. Participants proposed more data and knowledge sharing at pre-competitive stages, involving commercial stakeholders in developing norms and standards, and greater investment in public–private partnerships.

Fundamentally change funding structure

Funding must be structured to support large-scale, multi-decade data generation. Time horizons should lengthen – ideally to five years or more – moving beyond the traditional three- or four-year project towards large-scale, collaborative, systematic data generation across institutions and countries. Transatlantic collaboration mirrors these needs and, in addition, requires aligning grant application and review processes, alongside funding opportunities dedicated explicitly to UK–US collaboration.

Strategically, participants sketched two approaches: large, centralised funds for major collaborative projects, or a more agile model of smaller grants for working groups tackling different aspects of one topic, linked and coordinated under a single umbrella.

Translating EB into marketable products often requires ambitious technical projects that are ill-suited to venture-capital incentives yet too engineering-heavy for academia. To address this, participants urged new funding models drawing on a mix of applied grants and contract research, modelled on the FRO[13] and BBN[14] structures used successfully in other sectors.

Finally, funders need to reconsider what counts as fundable: research into scale-up and translation, data standards and norms, and collaboration and governance are key enablers for the whole space, and should be recognised and funded alongside data management and foundational AI and EB research.

Make the workforce pipeline interdisciplinary and AI-ready

Talent pipelines and education must adapt to the rapid, fundamental changes in how science is done: this means both prioritising interdisciplinary work over technical silos and building the skills for AI-enabled research. Participants outlined several possible pathways: interdisciplinary programmes, PhD-level training, and application-oriented, apprenticeship-level routes developed in close collaboration with industry.

Develop and implement actionable RRI frameworks

Moving the space forward requires actionable ethics frameworks for the intersection of EB and AI, underpinned by the institutions, training, and self-regulatory mechanisms needed to implement them effectively. It also requires new decision-making processes that match the accelerating pace of progress and address stakeholder needs adequately. Here too, participants suggested that embedding RRI practices in research standards and funding requirements could be an effective way to promote adherence.

[13] focused research organisation

[14] Bolt, Beranek & Newman - an American research and development company based in Cambridge, Massachusetts that developed the predecessor of today's Internet

What can different actors do?

Data and model owners

Participants addressed several distinct holders of data and AI models. Of the NHS and other healthcare providers, they asked for easier access to data through streamlined IRB/REC[15] processes, common approval pathways, and the provision of compute infrastructure, so that sensitive patient data could remain with the hospitals and reduce the data-protection overhead. Recognising the commercial sector's importance and technological lead, participants asked for greater access to its high-quality data and the open-source release of its advanced biological models.

Scientists

Participants reflected on publication and incentive practices, hoping the scientific community involved at the intersection of AI and EB could take a broader view of success than one focused on publication metrics, and start recognising the less glamorous but necessary work of methods development, standards, and data generation. Negative data needed to be shared. They noted the need to develop cross-disciplinary skills, to fund roles dedicated to data collection, and to adapt to a new paradigm of research conducted with AI.

Regulators and policymakers

Participants asked regulators to keep pace with the speed at which the field evolves, and for consistent, stable policy to provide a basis for long-term commitment and investment in research and scale-up. They repeatedly called for proactive development of regulatory frameworks covering EB-enabled therapeutics, genetically modified organisms, and novel foods – frameworks that would guide translation into marketable products and therapeutics. Avoiding cross-departmental silos and greater alignment between the UK and US policy and regulatory environments were both named as practical near-term improvements.

Governments

The most repeated ask of governments – addressed to the UK, the US, and the two jointly – was stability: longer policy commitments, genuine long-term strategic planning, and frameworks kept steady over a horizon long enough to build, repurpose, and maintain the infrastructure, capability, and collaborative networks the science requires, particularly in metrology and standards. The US government specifically was asked to use its influence to support progress and collaboration. Participants asked the UK government for a "higher risk appetite" and a rethink of its incentive structures for transformative technology, including greater support for public–private partnerships.

Funders

Participants asked funders, above all, to change the structure of funding and broaden the scope of what is funded: longer-term schemes, more coordinated funding across agencies, explicit bilateral mechanisms, and support for the "unsexy" enablers of EB – data standards, model design, data-

[15] Institutional Review Board / Research Ethics Committee

infrastructure sustainability, training, and research into RRI and governance. Crucially, data curation and database maintenance need long-term funding once established; too often, infrastructure is built and then abandoned. Participants asked funders to use funding as a lever on data behaviour, enforcing data standards and data-sharing practices through grant terms and conditions. Novel mechanisms – such as "data credits" awarded for quality data generated at accredited core labs – deserve exploration.

Public institutions and philanthropy

These stakeholders were positioned as infrastructure providers and honest brokers. On the infrastructure side, the asks were concrete: provide resources and support – including robust computational infrastructure, secure databases, and access to biofoundries and scale-up facilities beyond the reach of individual research groups. The collaborative function matters as much as the technical one, and the expectation is proactive engagement over passive availability: coordinating across disciplines, seeking partnerships, working towards curated FAIR datasets, and committing to deliver for the common good rather than optimising for institutional visibility.

Part 2: Detailed Summary of Agenda and Breakout Group Discussions

How the Event Worked

Day one

Day one of the symposium opened with welcomes and introductory presentations setting out the purpose and aims of the convening. The morning and first afternoon session proceeded in plenary, with invited speakers addressing foundational topics critical to advancing AI-enabled EB, and panel discussions expanding on the material. Themes included AI-ready data and data standards, metrology, and Responsible Research and Innovation. In the afternoon, following a brief thematic introduction to the breakout themes by subject experts, participants moved into pre-assigned small groups in three breakout discussions. They remained in these breakout rooms for the rest of the day. Across the breakout rooms, all groups worked on the same four tasks, applied to their respective topic area.

- **Task one:** In small groups, collect and discuss pressing opportunities and research questions in EB and AI relevant to your application area.
- **Task two:** Assess the collected opportunities and research questions for their value if solved, and for the feasibility of solving them. Arrange them on a two-dimensional value/doability grid, and document any challenges or constraints that limit progress.
- **Task three:** How might AI help tackle these opportunities and research questions, and shift their position on the value/doability grid — that is, how might AI increase either the feasibility of an opportunity or its value if solved?
- **Task four:** Which data would you need to address the identified research questions and/or train the models discussed?

Day two

Day two opened with a report-back in which the breakout group leads gave a brief overview of the discussions and topics that had surfaced in each breakout discussion. This was followed by three case study presentations on successful AI application in EB R&D. Participants then returned to their breakouts to continue small-group work, working in the same room but with newly composed groups.

- **Task one:** For EB and AI in your application space, collect the respective strengths and advantages of the US and the UK, and identify opportunities for collaboration and mutual benefit.
- **Task two:** Given these opportunities, what is preventing collaboration?
- **Task three:** Collectively identify and agree a set of stakeholders key to the application space, then address each in turn, formulating the asks, hopes, and needs directed at them.

The convening closed with a plenary recap and an outlook on what might come next.

Breakout Discussion: Detecting, Preventing, and Treating Disease

Introduction

Before the breakout sessions, two leading EB experts working in this application space introduced the area and provided a broad overview of AI's potential to unlock a set of interlocking EB opportunities in the therapeutics pipeline.

The starting point was nucleic acids. Deep learning, it was argued, could improve both interpretation and design of RNA, relevant to therapeutics and point-of-care diagnostics, many of which depended on nucleic-acid platforms. The same logic extended to genome design: AI models could generate candidate genomes and predict their properties, optimising sequences in silico before moving to the wet lab. Further downstream, AI-designed vectors could help direct a genetic therapy's payload to the right location at the right time, with the further aim of improving specificity and control while reducing toxicity and off-target effects.

A second strand concerned biomarkers and cell therapies, particularly in oncology. AI was proposed for discovering biomarkers and designing target molecules, and for extending cell therapies from their present success in haematological cancers into solid tumours, including by improving the control and survivability of the designed cells themselves. Underlying much of this was a data problem: decades of multi-omic datasets existed but remained poorly integrated. AI, it was hoped, could link these datasets into a single framework, helping to draw out patterns and candidate targets.

Small-molecule drug design was another area of focus, with AI expected to improve precision of targeting and resistance to degradation. Reliably manufacturing these molecules at scale across diverse hosts remained an unsolved challenge. Stem cell reprogramming, for regenerative medicine and biological sensing, was raised as a further opportunity, with a caveat: mechanistic understanding remained limited. Finally, the presentation gestured towards AI-guided autonomous laboratories, in which AI and robotics might eventually automate the scientific method itself.

Against this range of opportunity, the challenges raised were few but important: training the AI models required to tackle these opportunities depended first and foremost on high-quality data, but much of the published biological data had quality issues: not standardised, lacking metadata, and failing to reproduce. The underlying complexity of biology, particularly in nucleic-acid sequences, was flagged as another issue that could not easily be solved by more computation alone. Finally, therapies destined for patients required extreme precision, demonstrated safety, and control in the form of off-switches – properties that were currently not well understood and often treated as an afterthought.

With this framing, participants moved to the breakout sessions. What follows focuses on the topics that emerged in participant conversations within the application-space sessions across the two days.

Opportunity / Research Frontier

Transformational potential aired for both applied and discovery-led science

Across the health discussion, participants located the live scientific frontier in writing and precisely controlling DNA for specific purposes. They largely echoed the long-horizon opportunities set out by the experts who had introduced the breakout session: cell and gene therapies, computation-assisted protein and antibody design, and biomarkers for cancer prevention and neurodegenerative disease. Across every table, the ambition was expressed in the same terms: precise, controllable intervention – "the ultimate goal of precision cell-engineering". Participants also mirrored the long-term aspiration of AI-enabled mechanistic cell models that would allow in silico experiments and greatly increase research efficiency.

One table described the necessary next research steps as a progression from single-gene, cis-regulatory, observational understanding of DNA towards complex, trans-regulatory, multi-gene writing of DNA. Participants shared the longer-term aspiration the application-space leads had set out – AI-enabled cell models permitting in silico design and largely bypassing the wet lab – but judged it too distant: the foundational understanding of cell biology was not yet advanced enough, the requisite data had not been generated, and its generation remained out of reach.

In practice, when asked to name the most pressing or promising EB challenges in health-related research to tackle first, two of the five flagship proposals focused on discovery-led questions relating to a better understanding of underlying cell biology or observational research, while three focused on applied bioengineering. This highlighted some disagreement in the group about where to draw the boundary around the field, and whether discovery-led questions were appropriate targets for collaboration under a banner of EB. It remained unresolved whether this was a signal about the maturity of the field, indicating it was still at a pre-engineering stage, or a sampling artefact arising from the over-representation of discovery scientists in the room.

The strengths of the UK and the US are systematic and complementary

Collaboration between the UK and the US was seen as a strong opportunity: the two countries broadly shared values and priorities, and their strengths complemented one another in a systematic way.

The UK was seen to hold a wealth of existing data (in NHS data and the UK Biobank), a well-networked and agile research community offering the tools and infrastructure for collaboration, and a leading position in Responsible Research and Innovation. The US offered high-performance compute infrastructure and a "strong tech and AI sector" that hosted many of the most advanced disease foundation models. It also provided deeper funding and venture capital, along with expertise in scale-up, giving it an advantage in the translation and commercial exploitation of science. The UK favoured a systems approach and had a well-networked research community; the US followed a mission-oriented, "get-things-done" philosophy: combining the two would play to the strengths of each. Underpinning this was a track record of research collaboration and earlier cross-funded projects between standards-forming bodies and biofoundries, which could serve as models for what might come next.

Challenges, Constraints, and Bottlenecks

To build AI models and advance foundational knowledge, a new scale of data generation is required

Participants described the current frontier in EB as moving from merely reading DNA to understanding its function and – further still – editing it precisely for desired outputs, controlling gene expression and timing, and building genetic circuits. The mainly observational data presently available was judged sufficient neither to answer these questions nor to train more capable predictive models. Advancing the space would require a shift in the kind of data generated: participants were unanimous in naming "integrated, multimodal, time-resolved" perturbation data, combined with multi-omic readouts, as the necessary next step. Experts who reviewed this report emphasised that, where the goal is to develop better, more effective engineered cells, the requirement is to generate data about engineered cells rather than naturally occurring biology.

Population scale (across many people or many cells) and cross-organism scope (across species or cell types) were treated as distinct axes rather than interchangeable ones. Reviewers noted that this would be broadly relevant to preventing and treating disease, but would only advance the development of AI for engineering biology specifically where it had relevance to the behaviour of engineered biological material. For target cells, one plausible trajectory ran from established HEK cell lines to primary cells, then to complex tissue, and finally to organoid models.

Generating data of this kind, at the scale and dimensionality required, was regarded as blocked by several major obstacles. It demanded research infrastructure that did not currently exist, underpinned by funding beyond the reach of "per-project grants": rather than funding single postdoctoral researchers and "some consumables for three years", moving forward required equipping research consortia with multi-million pound/dollar budgets and timeframes of five years or longer. Coordinating this type of interconnected collaborative work was described as attainable only through large, integrated programmes. It was also described as hinging on automation technology not yet sufficiently refined.

Active-learning algorithms were described as an elegant and readily deployable means of identifying the most consequential data gaps, making the process faster and more cost-effective. Participants cautioned, however, that the long experimental cycles characteristic of biology might undermine a highly iterative strategy, and argued instead for broad, parallel initial rounds of experimentation, followed by learning from the results.

At the level of research strategy, disagreement arose over what to prioritise. Some participants held that a deeper understanding of the underlying cell biology was the pertinent goal, and argued for question-driven data generation; others maintained that reasonably reliable prediction might be "good enough" to develop safe therapeutics, and argued for large-scale, agnostic data production alongside the production of data that supported a better understanding of engineered cell behaviour; yet others proposed focusing on moonshot implementation challenges that met a concrete need.

Lab knowledge often fails to translate to therapeutics

Participants described a strong translation gap between discovery research and therapeutics, due in large part to a lack of generalisability of knowledge. Current research relied on platforms chosen for their

tractability – cell lines, 2D culture, and animal models – that often provided poor representations of the intended target (complex tissue, in vivo, humans), and this frequently led to failure in human trials. Here, participants suggested, lay one of the most promising applications for advanced AI: generalised predictive cell models could accelerate the search for, and implementation of, EB solutions. In therapeutics research specifically, more robust models could speed up pipelines, predict effects and off-target effects in advance, and select promising candidates for clinical trials – increasing efficiency and heading off failure at the last and most costly stages.

Given the complexity inherent to biology, however, participants voiced doubt over whether universal cell models were obtainable at all, or whether smaller, cell-specific models offered a more attainable approach. A near-term win in drug development might be to train AI to predict failure modes earlier in the pipeline, reducing the proportion of late-stage failures caused by immunogenicity, ADME properties[16], and toxicity, and with them the cost and the risk to patients. To that end, participants suggested that historical trial data and negative data were underused, particularly for analysing why trials failed. A reviewer noted that this was an important discussion but had little direct relevance to EB.

Low data quality leads to weak AI models

> *"AI doesn't exist without data. The models are rubbish if the data is rubbish. […] If you can't rely on the data, it's all a house of cards." – a participant*

For the most part, data quality and the role of AI were discussed as interlinked. AI drew rather more of the group's attention, with data quality treated largely as a precondition for robust models. Within this framing, AI was broadly understood to have two applications: as a means of improving the quality of existing data and driving higher-quality data generation; and "AI for biology" proper, in the form of specific biological models, digital twins, and – in the long term – generalised cell models.

On the one hand, AI was thought capable of raising the quality of existing data and rendering it AI-ready – filling gaps in metadata, assessing discarded data for latent value, and improving the generation of new data. Yet despite "feeling eminently tractable", this AI-for-better-data approach remained out of reach: first, the cost of building the necessary models exceeded what academia could bear, given the expense of (manually) pre-processing large volumes of data. Second, the term "AI-ready data" itself remained poorly defined, leaving unclear what end state a data-quality-oriented AI ought to be working towards.

In mild tension with this caution, one participant questioned whether data quality needed to be treated as such a constraint on model training, drawing an analogy with large language models, which had reached considerable sophistication despite being trained on what they termed "internet garbage". Others promptly rebutted, arguing that the models required for EB applications in human health were considerably different and that the margin for error was extremely low.

On the other hand, AI was thought capable of improving the robustness of models and of "giving more prediction from less data" – a claim immediately contested on the grounds that such models remained fundamentally predictive and required real-world validation, which reintroduced the need to generate data. Another participant observed that "AI has already extracted much of the value available from easy,

[16] ADME principles describe pharmacokinetics through four key stages: absorption, distribution, metabolism, and excretion. These processes dictate how a drug enters the body, moves through tissues, changes chemically, and leaves the system.

pre-existing data", arguing that AI was not the panacea some had hoped for, and that there was no way around generating new experimental data.

AI-EB integration is still limited, and experts fluent in both fields are rare

A frequently stated concern was that knowledge and methods in both AI and EB were evolving rapidly and in parallel. Leaving the integration of the two disciplines to chance was described as a failing strategy; rather, it required dedicated effort and deliberate planning. One table went further, arguing that the process of knowledge generation had to be rethought altogether – building a "new research model in which AI, data generation and biological engineering are tightly coupled".

Two obstacles were identified: a lack of funding for specifically interdisciplinary projects, and a shortage of researchers literate in both disciplines. As a structural basis for EB-AI-integrated work, the model achieved in a collaboration between the US National Institute of Standards and Technology (NIST) and The Align Foundation[17] was named as an interesting approach, though participants cautioned that it did not scale well. The way forward, they suggested, was for funders to "recognise the importance of integrating both disciplines" and to prioritise funding at the interface.

Tighter integration also needed people fluent in both disciplines, and these were currently rare. As a case in point, one table noted that even in that room few individuals knew enough about both fields, and acknowledged having done "some magical AI thinking" while "lacking concrete ways forward for those cell or disease models". On one side, training pipelines were seen as too narrow, producing either AI or EB specialists but offering no track that combined the two and built the necessary interdisciplinary skillset. On the other, significant barriers to mobility (visas, administration, pay gaps) obstructed the recruitment of existing talent where it was most needed.

Better molecular tools are a precondition for real-world application

Across tables, participants noted that a major obstacle to advancing EB research and translating it into safe therapeutics was the lack of reliable and precise molecular tools: genome-editing technologies such as CRISPR were seen as too crude, unable to make precise modifications without generating significant off-target effects. Similarly, tools for the control of gene expression and technologies used to deliver genetic material into cells needed refinement. While AI might help improve some of these tools, for example through better antigen recognition in viral vectors, others were fundamentally technological problems that EB could not solve on its own, and that drew in chemistry and materials science.

UK–US collaboration is blocked by funding structure and volume

By far the most grounded theme, recurring across days and tables, was that although science and values were broadly aligned between the UK and the US, funding blocked collaboration.

Both the volume and the architecture of funding were identified as key barriers. The absence of joint programmes and grant opportunities for transatlantic collaboration forced initiatives onto ad-hoc funding and, in consequence, into instability and inconsistency. Participants pointed to previously existing opportunities such as the BBSRC–NSF[18] scheme, and to a history of successful NSF/UKRI cross-funded

[17] https://alignbio.org/

[18] https://www.ukri.org/opportunity/work-with-us-researchers-bbsrc-nsf-bio-lead-agency-2024

projects, and criticised the fact that at present there was no targeted funding and little opportunity to work together, "despite lots of interest from the community".

Short funding cycles both failed to match the scale of the research required and significantly constrained collaboration in their own right. On the one hand, they sat awkwardly against the slow but necessary work of developing project ideas, finding partners, establishing working relationships, and settling legal and IP contracts. Even with a perfect setup in place, the application itself often meant dealing with several institutions across both countries, cumbersome mechanisms, and redundancy on both sides; in some cases, this turned applications into "double jeopardy" and posed too great a risk to be worth pursuing.

The discussion around funding for transatlantic collaboration formed part of a broader discussion in this breakout space about a need to fund projects at bigger scale and with higher tolerance for risk.

A lack of clear IP frameworks creates friction for collaboration

IP was a smaller but still consequential theme, coming up across both days and on all tables. The common thread was that a "lack of clarity around licences and terms of use" – together with unresolved value capture along the axis from data creation to models – obstructed collaboration between the UK and US, across research groups, and in dealings with commercial actors.

Participants noted the absence of a "functional and fair" framework for distributing value between those who generate data and those who train and own the resulting AI models, which ultimately discouraged data generation: "We want people to make data for us to train models on, but how does the value created by the model then propagate back along the chain?" Value itself was not clearly defined, extending beyond commercial benefit to include academic recognition, licensing, and even policy change. Commercial actors were seen to guard access to what were often the most comprehensive datasets and most advanced models, unwilling to share them given their competitive worth. Public–private partnerships built on 12-month exclusivity contracts were suggested as one promising route. Another, directed at collaboration within academia, was to mandate open-source data and models, or public ownership of data, as a condition of funding. Overall, though, participants noted self-critically that researchers themselves would have to adopt "more realistic and flexible" positions on IP, data licensing, and public–private partnerships.

IP and data-sharing restrictions were also identified as barriers to transatlantic collaboration. Sharing valuable but sensitive NHS data was a particular challenge, alongside "fundamental differences" in how IP was handled. Additionally, far-reaching rules enforced by UK universities' technology transfer offices, in particular, often deterred prospective US partners.

What needs to change?

Funding needs to increase and change fundamentally in structure

Participants agreed that data generation at scale was the way forward – whether by a challenge-driven or an agnostic route – and that it would require a coordinated effort across institutions, tightly coupling AI and EB research. This hinged chiefly on funding.

The current funding landscape, participants across all tables indicated, was not equipped to support data generation on the scale required: too thin in volume, too narrow in scope, too fragmented across institutions, and built for the short timeframes of traditional, per-grant research. The changes called for followed directly: raising funding volume in the UK to match US levels, and reworking the structure of funding schemes – extending timelines to match the pace research actually takes ("ideally five years or more"), harmonising application processes across funding bodies, and moving from single-group grants towards support for large-scale, coordinated research. Retaining promising spin-offs in the UK was seen to require investment in scale-up, access to venture capital, and support for commercial exploitation.

Misaligned funding was likewise seen as the principal obstacle to UK–US collaboration, and the remedies mirrored those for data generation at scale: funding dedicated specifically to transatlantic work, aligned application processes between the two countries, reduced workload and lower risk for applicants, and longer funding cycles that allowed reliable, long-term collaboration.

Developing trusted, safe data frameworks; simplifying and aligning IP

Noting the "huge institutional overhead" of building certified trusted research environments one data provider at a time, participants considered national databases necessary – both physical infrastructure able to host large volumes of sensitive data securely, and a trusted framework governing access in a structured and fair manner.

The second major barrier to collaboration identified, between institutions and across national borders, was the state of IP. Participants considered it necessary to simplify and harmonise IP frameworks between the US and the UK, to create standardised legal pathways for sharing data with international partners and commercial actors, and to invest more in public–private partnerships. Making data-sharing a precondition of funding, or mandating open-source data, were considered as further options.

One intervention was proposed for the long-term: preparing regulatory standards for therapeutic products that are EB- and AI-native, and finding ways to handle the potential lack of mechanistic understanding of novel therapeutics that AI-enabled research might produce.

What can different actors do?

Funders

Participants asked funders to prioritise changes to the structure of funding over its quantity: longer-term schemes, more coordinated funding across agencies, and explicit bilateral mechanisms. Joint NSF/UKRI calls were named specifically, as was funding dedicated to the scoping of collaboration. A second cluster asked funders to use funding as a lever on data behaviour, enforcing data standards and data-sharing practices through grant terms and conditions. A tension surfaced within the community over what should be funded: some argued for a challenge-driven approach and "moonshot" projects, others for the opposite – favouring a genuinely integrative approach that focused on developing tools for AI, EB, and AI-EB as strong, transferable foundations across challenges.

Governments

Participants' most repeated ask of governments was stability: keeping priorities and frameworks steady over a horizon long enough to build the infrastructure, frameworks, and collaborative networks the science required. In practical terms, participants asked for better coordination of existing compute and data infrastructure, joint funding schemes, and the development of shared data standards. The US specifically was asked to use its influence to support progress and collaboration, adopting a collaborative stance. Of the UK, participants asked for a "higher risk appetite" and a reconsideration of its incentive structures for transformative technology, including greater support for public–private partnerships.

Regulators

Participants asked regulators to keep pace with the speed at which the field evolved. Working towards regulatory alignment between countries, and developing standards and guidance on data formats, would ease both scientific collaboration and translation into commercial products. One request stood out, because it seemed disconnected from the rest: "Do something about the publication system and paywalls."

Data, infrastructure, and model owners

Without explicitly naming individual actors, participants addressed holders of data, research infrastructure, and AI models, asking for easier and faster access to compute and to publicly funded biofoundries. Of the NHS and other healthcare providers, they asked for easier access to data through streamlined ethics approval processes, common approval paths, and the provision of compute infrastructure so that sensitive patient data could remain with the hospitals, reducing the data-protection overhead. Recognising the importance and the technological lead of the commercial sector, participants asked for free access to its high-quality data and the open-source release of its advanced biological models.

Scientists

Turning to their own community, participants reflected on publication and incentive practices, hoping to move away from primary focus on publication metrics towards recognising the less glamorous but necessary work of methods and tools development, standards, and data generation. They noted the need to develop cross-disciplinary skills, to fund roles dedicated to data collection, and to adapt to a new paradigm of research conducted with AI. Social scientists, specifically, asked to be embedded in both research projects and policy.

Breakout Discussion: Food Systems

Introduction

"How are we going to feed 10 billion people by 2050?" – A panellist

Ahead of the breakout sessions, the two application-space leads gave a brief introduction to the current challenges and opportunities in food systems where EB could provide solutions. They set out a clear problem statement: the global food system was increasingly coming under pressure from climate change while being one of its largest drivers at the same time. It produced about a third of global greenhouse gas emissions and was the primary driver of deforestation and biodiversity loss, thus exacerbating the conditions that fuelled threats to its existence. At the same time, the global population was expected to increase to about 10 billion people by 2050. Feeding this growing population sustainably, securely, efficiently, and safely would require rapid innovation.

AI offered a potential catalyst for progress. In crop development, AI was increasingly used to predict phenotype from genotype, allowing faster breeding and selection of climate-resilient crops. AI could also integrate biological, environmental, and economic data and lead to more robust predictions of productivity, supply-chain risks, and environmental impact. In novel food production from precision fermentation and cell cultivation, AI was increasingly being used to predict and optimise production processes and scale-up, as well as to inform techno-economic assessments (TEA).

Application of AI in the food systems application space, however, remained fragmented. This prompted the presenters to ask whether the real opportunity lay in the integration of different ways of leveraging AI from R&D to market placement, enabling true end-to-end biological design. On a technological level, this could be the key to developing closed-loop bioengineering platforms, greatly accelerating research and design processes. For research itself, it could open up a whole new level of complexity in designing crops, microbes, and cells, optimising for performance and downstream processing right from the start. And from a systems perspective, true integration across scales could help design more resilient systems spanning the entire range from molecular to ecosystem level.

Despite all these technological opportunities, a caveat applied: ambition for EB did not stop at delivering healthy, nutritious, cost-efficient food products, but extended to positive impact on global food systems at large. The presenters argued EB first and foremost needed to consider its impact on future generations and *do no harm* – whether that concerned planetary boundaries or the policies it lobbied for.

Continues over page

Opportunity / Research Frontier

Fundamental and applied frontiers

By and large, the discussion mirrored the opportunities for food systems offered by EB and AI that the two presenters had laid out, and broadly delineated two research frontiers: on the one hand, unravelling the interaction between genotype, environment, and phenotype and shaping it for purpose; and on the other, turning novel food breakthroughs into competitive, marketable products through better biomanufacturing and techno-economic integration. Participants did not consider these two areas disjoint but overlapping, and one table proposed a flagship project that aimed to integrate both: closed food production systems for challenging environments (deserts, favelas, deep-space travel to Mars) that would require crops designed for extreme conditions, circular systems, complex real-time monitoring, and extreme process efficiency.

Sustainability was a third important theme, repeatedly brought up across all tables. Participants expressed a feeling of urgency in the face of the increasing challenges brought about by climate change and population growth.

> *"So many crops, so many problems, so many people – so little time!" – a participant*

Participants positioned AI in a long tradition of crop breeding, stressing that other forms of machine learning had been used successfully for crop selection for a long time. In this view, EB was the new player that enabled a move beyond prediction-driven crop selection into active crop design grounded in causal understanding. The decisive opportunity that deep-learning-enabled AI opened up was the capacity to integrate large amounts of heterogeneous data across different data types, offering a level of complexity and inference previously unattainable.

At the genotype-environment-phenotype intersection, research questions focused on first understanding gene function and then designing plant genomes with the aim of engineering more resilient and more effective crops, reducing the need for land, fertiliser, and pesticides. They also focused on moving beyond the plant itself and understanding and manipulating the complex interactions with the different elements involved in crop production, including root structures, soil conditions, and the microbiome. On a practical level, participants aspired to move desirable traits across plants (e.g. C4 crops) and non-plant traits into plants (for example, bacterial nitrogen fixation); to engineer plant metabolism and organelles for greater efficiency; and to engineer microbiomes that supported plant growth and protected against pests.

Regarding translation, biomanufacturing, and scale-up, the opportunities of AI centred on assessing project doability, identifying bottlenecks in and potential for improving biomanufacturing production processes, and enabling deeper, more robust risk assessments before investment in scale-up. AI was also seen to offer the potential to improve TEAs, evaluate market signals, and align regulation processes and research. While both frontiers – the genotype-environment-phenotype axis and research translation – were considered in discussion, the first of the two received considerably more attention. It is not clear whether this is a true signal of maturity (i.e. whether engineering biology in the food systems space *is* largely pre-biomanufacturing) or a result of selection bias among event participants.

Transatlantic collaboration

Participants described collaboration between the UK and the US in the food systems space as a large opportunity. The two countries had different strengths that complemented one another in a systematic way, and combining them could have a decisive impact.

The UK's key strength was seen in its well-developed academic research ecosystem: the country offered deep expertise in EB, data standards research, and sequential data generation, as well as food systems and systems biology. The UK hosted advanced facilities for plant science and phenotypic assessment, as well as several world-leading seed banks (including pre-green-revolution genomic checkpoints such as the *Watkins Landrace Wheat Collection*[19] and the *Millennium Seed Bank*[20]). Equally important, the UK offered an agile collaboration culture supported by mature infrastructure (for example *CARMA*[21] and *NAPIC*[22]) and a strong network of European and Commonwealth collaborations in plant, microbial, and cell-cultivated food science. Programmes such as *C-SPIRIT*[23] offered a blueprint for synergistic working.

Participants also mentioned the regulatory and policy environment as a particular strength of the UK. The Precision Breeding Act[24] allowed fast deployment of crops; the FSA operated proactively with standardised and agile processes; and the broader policy stance – strategic, pragmatic, and environmentally aware – created conditions supporting research. Finally, the UK offered a strong cultural connection to food sources and a circular economy, as well as broad environmental awareness and serious environmental policy.

The US, in turn, offered expertise in bioinformatics and systems biology, as well as a track record of applied research at world-leading institutions in sustainable food production. The country led in computer and AI infrastructure and offered a mature level of cloud adoption. The unique environment of Silicon Valley, large-scale data generation facilities, and nascent self-driven autonomous labs reinforced this position.

The sheer scale of the US market – and the agricultural industry in particular – created sustained demand for innovation and offered the venture capital required for translation and scale-up. The size was seen as complemented by an entrepreneurial culture and mature scale-up infrastructure. Grant funding, too, operated at a size and duration that sustained genuinely ambitious goals. A permissive regulatory environment and the diversity of natural environments across the country were further advantages to the food systems application space.

Combining the strong foundational research, collaborative culture, leading databases, and established international networks of the UK with the US's scale-up and implementation expertise, venture capital, and world-leading agri-businesses promised to unfold a synergy that created an impact neither country could have on its own.

[19] The A E Watkins Landrace Wheat Collection is a collection of heritage varieties of wheat from around the world offering a greater genetic diversity than those found in modern varieties

[20] The Millennium Seed Bank is the world's largest international conservation project for wild plant seeds, holding over 2.5 billion seeds from more than 40,000 plant species worldwide

[21] The Cellular Agriculture Manufacturing Hub - https://carmahub.co.uk/

[22] The National Alternative Proteins Innovation Centre - https://napic.ac.uk/

[23] Centre for Sustainable Plant Innovation and Resilience - https://c-spirit.org/

[24] UK legislation, 2023: https://www.legislation.gov.uk/ukpga/2023/6/contents

Challenges, Constraints, and Bottlenecks

Key data does not exist, is inaccessible, or hard to leverage

Both frontiers – understanding genome function and successfully translating research – came with a data-shaped challenge. They differed in how central the data gap was and in the kind of data required.

In general, biological research was currently based on observational data and modelling correlation; while a wealth of descriptive data on plant and animal genomes existed, gene function was often not clear – nor how best to modify it. Both understanding genome function and modelling and manipulating the interaction between genome and environment required moving from observational data to intervention data, and linking data across data types. Participants cautioned that better leveraging existing data was not enough, and that moving forward required the large-scale generation of new data.

Several data gaps were identified. First and foremost, participants pointed to a lack of data linking changes in genome to changes in function and phenotype. On a cellular level, single-cell transcriptomics including proteomics and metabolomics, linked with genome perturbation data, were required. On a larger scale, field data connecting molecular changes to whole-plant and field-level outcomes were described as missing. Similarly, datasets comprehensively combining soil, climate, bacterial, satellite, and genomic data were currently not available.

In a slight disagreement, one participant objected that the subfield of predictive breeding already had large amounts of data, and that integrating them better was an attainable near-term win that would enable more robust prediction models and have a measurable impact – even without creating the new data required for causal understanding of genome function.

With sequencing becoming increasingly cheap, the larger challenge was generating phenotype and environmental data at scale: that "it requires time to grow a plant or a chicken" is a constraint that AI cannot overcome. As a consequence, iterative data generation strategies might not lend themselves to EB, and parallel approaches might be more efficient: waiting for the results of one study before starting the next was simply not time-efficient; rather, studies should be conducted in parallel, informing each other along the way.

For precision-fermentation foods or cell-cultivated foods, participants pointed out, understanding genome function was a lesser immediate concern, and the real challenge lay in making things work at scale. Consequently, the binding data constraint was located downstream of the biology and included process data, techno-economic data, and scale-up economics. Participants cautioned that strain performance data, metabolomics data, thermochemical data, and cost data were currently not readily available. While the data gap existed, however, it was arguably smaller and partly a data access and usability problem. Participants discussed that a lot of high-quality data existed, particularly on novel foods, but that it was challenging to harvest, because it was buried in journal articles or held by commercial actors. The existing data was also often heterogeneous, did not share common standards, and lacked metadata – all barriers to leveraging it efficiently.

Funding structure and volume don't match the challenge

There was a consensus that funding structure and funding volume did not match the needs of the space, blocking both scientific advance and research translation. On the one hand, short funding cycles did not correspond to the long-term efforts required to generate the necessary data at scale; they eroded momentum and fragmented expertise. On the other, the lack of recurring funding opportunities and generally diminishing public funding compounded this. Taken together, this forced research into a perpetual cycle of ad-hoc, short-term proposals and prevented the long-term efforts desperately needed to tackle pressing problems.

At the same time, there were not enough programmes that funded collaboration, both across disciplines and internationally. Mirroring other application spaces, dedicated mechanisms for large-scale UK–US collaboration were largely absent.

Scope mattered too: participants criticised that biomanufacturing scale-up was often dismissed as "not research" by funding agencies, regardless of the application space. UK venture capital investment lagged the US considerably, and the outputs of UK research had a tendency to end up benefiting investors and institutions outside the UK. To move forward, foundational research into biology, research into enablers such as AI and data quality, and applied translational work all needed support.

Food system improvements require more than R&D

More than in any other application space, participants stressed that food systems could not be seen through a technological lens alone. A technically successful product had to clear a number of non-technical gates before it reached a plate, and could fail at any one.

> *"Remember that food is not tech. But don't forget that tech is a valuable enabler."*
> – a participant

The first gate was social permission to operate. Participants noted that many societies regarded biotech as deeply problematic, a legacy of the controversies around the first genetically modified organisms. While the scientific case might be clear to practitioners, closing the gap with the public was critical. Participants cautioned against "repeating the mistakes of the genetically modified organism (GMO) scandal", and described engaging with the public, addressing social media rumours, and countering misinformation as essential.

The second gate was access to a market controlled by a small number of gatekeepers. Food producers and distributors dominated the supply chain and effectively determined what reached the market; translation and scale-up hinged on their collaboration. Some participants expressed frustration that these actors favoured traditional foods and at times actively blocked novel ones. Winning them over was seen as crucial, not least because they held valuable insight into consumer demand, logistics, and industrial-scale production – knowledge academia otherwise struggled to access.

The third gate was market demand. Participants pointed to novel proteins that could already be scaled up but generated few market signals and attracted little venture capital. Techno-economic assessments and market analysis were often afterthoughts instead of essential pillars of research design.

A further gate was the resilience of the food system itself, with R&D having to anticipate what the system might need in the years to come: participants stressed the need, above all, to develop crops and food products that were robust against climate change and resource-efficient, addressing the intensifying competition for land and water. Understanding how crops interacted with a changing environment was central to developing resilient supply chains.

The final gate was the plate itself: even where genuine demand existed, novel foods had to be optimised as much for taste and texture as for nutritional value, acknowledging that food was deeply rooted in culture and eating traditions. Designing three-dimensional, appetising food structures was a challenge repeatedly stated.

IP friction discourages commercial actors from collaborating

For food systems R&D, commercial actors played a decisive role. Participants noted that large agri-businesses held some of the most valuable datasets – elite crop varieties, industrial breeding and fermentation data, alternative protein data – and were key drivers of AI innovation. Despite a perceived willingness to collaborate, participants identified ownership, IP structure, and international mistrust around regulations for data sharing as major challenges. In particular, complicated background/foreground arrangements and investors targeting specific outputs regularly caused friction; as a result, food-system-specific pre-IP data sharing barely existed – a gap that could potentially be addressed by borrowing from other sectors that had navigated this challenge more successfully.

A particularity of the field of novel proteins was that research in this area had largely been driven by commercial actors, leading to large parts of the data and knowledge being proprietary. Key companies, however, were now struggling to survive, and several had already wound down their business – and with them went the data and procedural know-how. Conserving this knowledge in the public realm was described as a time-critical matter that could potentially have a huge impact on the sector.

Regulatory pathways are unclear and lack alignment

Around regulation, participants expressed considerably heterogeneous views. On the one hand, some pointed out that standards and regulatory pathways for novel foods and food-systems-related applications were still immature when compared with those for health and human biological applications. They asked for clearer regulatory frameworks and flexible pathways as a way of supporting translation.

On the other hand, the Precision Breeding Act, as well as "agile and pragmatic" regulatory bodies, was highlighted as a UK strength, and "permissive regulation" in the food sector was described as an advantage of both the UK and the US. In this view, regulatory misalignment between the countries was seen as the bigger issue, complicating the approval of novel products for both markets. Phytosanitary regulation regarding material transfer, and national security controls preventing sharing (such as ITAR[25], export controls, and NSIA[26] provisions), were named as concrete regulatory barriers to transatlantic R&D collaboration.

[25] International Traffic in Arms Regulations
[26] The UK National Security and Investment Act

AI has yet to demonstrate its impact

Some suggested that AI was extremely helpful for addressing "large, fuzzy" problems without clear modelling approaches, integrating different data types of heterogeneous quality, and modelling biology in latent rather than physical spaces. The most important advantage of deep learning models, however, could be their ability to learn: to make sense of hugely complex data spaces and identify the most impactful questions to solve and data gaps to fill (active learning), thus helping to guide research strategy in novel and partly uncharted territory. Some participants suggested that they were not yet convinced that deep learning AI offered concrete advantages over previous forms of statistical modelling. While AI offered better tools for modelling data, in essence it was often "just Monte Carlo times five", and many AI problems could be solved by more traditional mathematical approaches as well. One participant noted dryly that AI was useful for attracting funding, given its current buzzword status.

What needs to change?

Move from observational data to intervention data, generated at scale

Consensus formed that the field had to move beyond generating observational data and describing correlation, transitioning towards the coordinated large-scale generation of new intervention data for causal understanding. Participants surfaced several priority gaps: multi-omics and perturbation datasets linking genomic changes to cellular function; field-level data connecting molecular changes to whole-plant outcomes; and integrated datasets spanning soil, climate, bacterial, satellite, and genomic sources.

Because generating phenotype and environmental data was inherently constrained by time, generation strategies should favour parallel over iterative approaches. Where near-term wins were available, as in predictive breeding, they should be seized: better integrating existing data could yield more robust prediction models and measurable impact.

For novel foods – precision fermentation and cell-cultivated products – the need was to generate and surface process, techno-economic, strain-performance, metabolomic, thermochemical, and cost data. Some of this data already existed but was inaccessible, being buried in journal articles or held privately, so the task was as much about ensuring systematic capture and fair access to both existing and newly generated data.

Redesign funding structure, volume, and scope to fit the challenge

Participants called for funding with longer time horizons, and for recurring, reliable funder support that allowed research structures to mature and momentum to build. Funding had to expand in scope in two ways: firstly, there was a need for funding dedicated explicitly to collaboration – across disciplines and internationally; secondly, funders needed to broaden their scope to support research on translation, scale-up, and increasing efficiency as well as foundational biological science. Investing in scale-up was seen as key to realising the true economic value of UK research output.

Engage beyond academia

Realising the aspiration to change food systems for the better, address climate change, and deliver socio-economic impact demanded that the research community step out of the lab and engage more with

society and industry. This meant investing in public engagement, and developing a shared language and a shared vision. It meant making techno-economic assessments and genuine market demand a pillar of research design, not an afterthought, and reaching out to commercial actors and winning champions among agri-businesses, food producers, and distributors.

Friction discouraging collaboration with commercial actors needed to be reduced through the provision of solid, pragmatic, and flexible IP frameworks. Participants proposed easier background/foreground arrangements, as well as building the legal frameworks and infrastructure for pre-competitive data sharing, potentially borrowing from sectors that had navigated this challenge more successfully. Most urgently, the proprietary data and procedural know-how of currently failing novel protein companies had to be conserved in the public realm before it was lost with the companies – a time-critical opportunity of potentially large impact.

Clarify and align regulation

As regulation for novel foods remained immature relative to that for health applications, clearer frameworks and flexible pathways would support translation. Where regulation was already a strength – and here participants cited the Precision Breeding Act, pragmatic regulatory bodies, permissive food-sector rules in both countries – resolving misalignment between the UK and US was key. This also included phytosanitary and national security measures currently complicating transatlantic collaboration.

What can different actors do?

Funders

Participants' asks of funders were clear: create long-term strategic programmes that address the scale of the research problems; and have greater willingness to de-risk innovation aiming to take products across the "valley of death". Some participants accused funders of being too willing to abandon frontier projects which encounter difficulties.

Funders could also use their position to drive data quality: applying stricter requirements on data sharing and standards, tied to funding conditions, would shift behaviour faster than exhortation. Novel mechanisms – such as "data credits" awarded for quality data generated at accredited core labs – deserved exploration.

Finally, participants asked funders to reconsider what counted as fundable: both research into data standards and research into scale-up and translation were seen as crucial enablers of EB. This also included long-term commitment to research infrastructure, data curation, and database maintenance. Withholding funding from these foundations eroded the field's agency.

Policymakers

Participants made their most pointed requests of policymakers. The most endorsed point was the need for clear, consistent policy. Another widely supported ask was developing a regulatory framework to support emerging technologies, including updating existing policy on genetically modified organisms. Participants also repeatedly raised a need to work towards a level playing field for novel foods,

comparable to that for traditional foods, arguing that creating market demand or stepping in at points of market failure could help kick-start the new bio-economy.

To a significant degree participants' asks of policymakers mirrored those directed at funders: longer policy commitments, genuine long-term strategic planning, bolder and less risk-averse positions, and – this stood out – real commitment on climate and environment. Avoiding cross-departmental silos, and achieving greater alignment between UK and US policy and regulatory environments, were both named as desirable near-term improvements.

Regulators

Regulators were asked to engage with academia and to collaborate: providing clear regulatory guidance, streamlining procedures, aligning internationally, cooperating to identify where current regulations failed and taking a flexible approach where they did. Participants called for evidence-based decision-making and robust ethics guidance as necessary foundations for a level playing field between novel and traditional foods.

Publicly funded institutes

The role of large, publicly funded institutes was perceived as one of providing infrastructure and being honest brokers. As such, these institutes should build and maintain the required research infrastructure at a scale not obtainable by individual research groups or academic institutions. This included robust computational infrastructure, modern and secure databases, and scale-up facilities. It also entailed developing and enforcing governance frameworks that secured fair and equitable access to both data and infrastructure.

This collaborative function mattered as much as the technical one, and participants called for proactive engagement over passive availability. This meant coordinating across disciplines, seeking partnerships, and committing to delivering for the common good rather than optimising for institutional visibility.

Innovators and commercial research

Stakeholders in commercial research were seen to enjoy advantages of scale not available to publicly funded research. They were asked to engage in pre-competitive collaboration with academia and to share data, pipelines, foundation models, and the lessons learned from collaborative approaches. They were also invited to think long-term and build for legacy rather than immediate return.

Growers, producers, and food distributors

Commercial growers, producers and distributors hold knowledge that researchers lack: first-hand understanding of problem spaces and practical knowledge of bottlenecks at commercial scale. Participants argued that for fruitful collaboration, these actors should "participate actively in scientific discussions", provide data on customer needs and public opinion, identify problems clearly, and prioritise what science should deliver. Acting as an informed customer for innovation, including within B2B and supply-chain contexts, could give these stakeholders genuine influence over the direction of research. As the essential functional link between science and the public, these stakeholders were asked to be open to novel foods and technologies and to champion them.

Researchers

Participants demanded more genuine collaboration in research across disciplines, across comfort zones, and with investment in people who make cross-disciplinary work possible at a functional level. Researchers should think early and concretely about what data will be needed midway through a project; focus on what is actually worth capturing; should not fabricate problems to solve; should not over-promise; and should invest more time in critically examining results before publishing, rather than optimising for output volume. The encouragement was to think bigger and be open to new possibilities, integrating AI and EB into research without waiting for certainty.

Educators

Training programmes needed to go beyond the PhD and postdoc pipeline to address gaps at technician level and bridge the AI skills gap. They needed to be updated frequently to reflect how quickly the field moved. Educational institutions should collaborate closely with researchers and industry on curriculum development, cultivate interdisciplinary thinking, and avoid siloed programmes that train narrow specialists for problems that require broad coalitions.

Participants also called on educators to take a role in public engagement. They were asked to develop accessible language for end users such as farmers, and to actively promote critical thinking around contested concepts like GMO, rather than leaving public discourse to louder and less careful voices.

> *“Educate for the future, not for now”* – a participant

Breakout Discussion: Biomanufacturing

Introduction

"We need more Engineering in EngBio" – a participant

In the biomanufacturing breakout participants focused their discussion on scaling up from the wet lab bench to industrial production: optimising production processes, maximising yield, and increasing cost-effectiveness in a competitive market. The tangible, well-defined challenges participants formulated centred on production and scale-up – measuring feedstock, genetic heterogeneity, and output data in large fermentation tanks in real time; predicting the behaviour of novel designs at industrial scale; identifying the right candidate designs for scale-up; and reducing inputs, capital expenditure, and production cost to improve cost-effectiveness at scale.

Data collection and data standards were examined through a production-and-process lens, with a clear instrumental role: to standardise production, increase efficiency, and build stronger predictive AI models that would in turn enable better scale-up and more efficient production. While some of the biomanufacturing group's challenges required genuinely new data – real-time process data, flow and state data, waste data – many of the data bottlenecks were seen as problems of access and sharing. Together, limited sharing of existing pre-competitive data, lack of collection infrastructure, and weak incentives prevented data sharing and collaboration. 'Failed data' was singled out as a key asset, on the view that learning from failed experiments would reduce duplicated effort and improve efficiency.

Little of the discussion related to advancing understanding of fundamental biological mechanisms, with participants demonstrating a clear focus on process innovation and commercialisation. One table suggested that biomanufacturing "should be thought of more as underpinning/cross-cutting" across application spaces than as an application space in its own right.

Opportunity / Research Frontier

Across every small group in this application space, participants described the same goal: to build products that worked at scale and were competitive in a commercial market: "in the end we want to develop sellable products". Current EB projects mostly failed to scale from promising wet lab research designs to industrial production, and failed to be cost-effective.

"We failed to learn how to turn into a real bioeconomy" – a participant

The frontier in this space had two broad strands: increasing the efficiency of biomanufacturing processes, and improving the process of scaling up and commercialisation. The first strand involved reducing input and optimising yields, recycling feedstock or using alternative feedstock, and generating real-time data from within industrial-scale fermentation tanks to improve monitoring and quality. Participants also discussed opportunities to reduce resource requirements and compress timescales by replacing wet lab components of research designs with AI and other digital tools. The second strand involved "predicting large-scale performance at a smaller scale", learning to identify promising designs early, and working out how to improve scale-up processes.

Alongside this, participants suggested there was a considerable opportunity for transatlantic cooperation. Both the US and the UK had made AI and EB current priorities, promising to leverage their mature EB ecosystems and world-class talent pipelines. Existing scientist-driven synergies, shared leadership in data standards, and a shared vision were a good basis for collaboration, and leveraging the two countries' complementary strengths could move the space forward and secure global leadership.

The UK was seen to hold a mission-focused research infrastructure supporting translation, together with a growing pilot/scale-up ecosystem and the "highest concentration of biofoundries in Europe". The US held a cutting-edge synthetic biology capability, a highly skilled workforce, and the experience of decades, synthetic biology having originated in the country. The funding landscape of both countries was seen as strong, though valued for different aspects. The US was credited with a "rich and diverse funding landscape", the most venture funding, mature philanthropic funding, "a very strong scale-up investment structure", and a strong not-for-profit landscape enabling public–private partnerships. The UK's advantage lay in the simplicity of funding mechanisms for EB and AI, and in "strong and direct" government support for AI, research infrastructure, and bioprocess scale-up, with the Biotechnology and Biological Sciences Research Council (BBSRC) and the Engineering Biology Infrastructure Programme (EBIP) named directly. The UK also offered "tighter connectivity" and "strong cross-domain collaboration", including with Europe, all of which facilitated in-person collaboration.

Challenges, Constraints, and Bottlenecks

Wet-lab success does not sufficiently predict performance at scale

Designs that seemed promising in the wet lab too often failed to reproduce at scale. This was attributed in part to microbes behaving differently in larger, less controllable environments and when exposed to different feedstocks. One table, however, expanded the question, asking "Is it the cell design or the process?", and pointed at a key barrier: it was often not clear *why* promising designs failed at scale.

Being able to predict how a biomanufacturing design would perform at industrial scale was seen as a high-value challenge and was raised by all three small groups in the room. One table framed it as increasing the predictability and reproducibility of multi-dimensional systems in order to enable efficiency, and another called for more predictability in the scale-up process. Two approaches were proposed: developing "an evaluation framework for robustness" that would allow the field to move beyond trial and error; and embedding a strategy for scale-up in the design process right from the start.

Participants accepted biological variation as an inherent challenge to predictability and industrial exploitation, one table observing that biology was not designed to be stable but to evolve. One technical avenue discussed was the deployment of single-cell technologies able to measure heterogeneity at every scale – lab, seed-train, and commercial stage – judged a high-value but low-doability challenge.

Genuine data gaps exist in several domains

Participants identified two types of data that did not yet exist in sufficient quantity: process data required to study and improve biomanufacturing processes and scale-up, and intervention data that would allow them to better understand underlying biological processes.

One table stated a need for "huge amounts of data, metadata, and analytical capacity", including domain-specific dynamic process data at large scale and a domain-specific LCM[27] vector database. Another focused exclusively on process data: multimodal data; real-time data covering genotypic variation assessment and metabolic outputs; waste-product data rather than final-product data alone, together with data on process impurities; and data on process integration and process engineering. That table also noted an under-specification, namely that the key attributes to capture in bioprocesses had yet to be defined. Generating the data was considered a high-value challenge but achievable only on a longer time horizon of more than five years. On the one hand, collecting data from the production process required continuous and better measurement capabilities and advanced biological sensing in real time, with the necessary sensors either not yet existing or being too expensive to deploy at scale. On the other, running experiments with the sole aim of filling data gaps was considered extremely costly in the first place.

Against the engineering-first framing, one participant argued repeatedly that scale-up failed because the underlying biology was not well understood, and that without such understanding the field would not be able to synthesise working chromosomes. Understanding the underlying cell biology, in this view, required omics data showing how cell behaviour changed at different scales and conditions – though the same participant conceded that this might instead be a data quality problem, there being a great deal of omics data but very poor metadata, curation, and connection.

Data standards are the necessary basis for automation and AI models

Data standards were discussed as enablers of automation, collaboration, and better AI models. One table treated the definition of domain-specific standards as a high-value target achievable within one to three years. Standardisation of data collection and metadata across production steps would facilitate crowd-sourced data generation and data sharing, and would increase the value of open-access data. Ultimately this was hoped to create a virtuous cycle in which identifying gaps in data standards and metadata ontology promoted better submissions, leading to better modelling approaches, in turn leading to better data.

A second table discussed data standards in relation to the question of whether the current research pipeline produced the right type of data: there was a great deal of omics data, but the quality of the data, including metadata and curation, was very poor. Existing standards, with the Protein Data Bank cited as an example, were seen as diluted by application across biology fields with differing requirements; participants at that table suggested learning from systems biology standards, and concluded that, rather than working around existing structures, the field should develop guidelines and establish standards that served its purpose.

A third table touched on standards only once, as enablers of collaboration, while noting the difficulty of getting even two people to agree.

Key data is inaccessible for lack of incentive and outdated infrastructure

Participants suggested that a great deal of work and money did not result in accessible, shared data, creating inefficiencies and hindering progress. The issue was attributed both to a lack of collection systems and suitable infrastructure and to misaligned incentives in a highly competitive environment, one table asking where the data was after fifteen billion dollars of company spending over the past decade.

[27] Large Cellular Model

Another table likewise located the data problem in industry behaviour: pre-competitive data existed but went unshared because companies "don't speak to each other", "processes are proprietary", and sharing was "constrained by IP and secrecy".

Where infrastructure for data sharing existed, it often did not keep pace with the rapid development of the sector. Participants pointed at a lack of incentives to update, upgrade, or retrofit it to support new data capture and sensor integration. They also questioned whether established FAIR data-sharing principles remained sufficient for the volumes and complexity of data now generated: beyond publication-based sharing, more structured ways of finding and accessing data were needed to enable efficient progress – and to prepare the field for AI-enabled automated research and agentic access to data.

A specific data-sharing gap mentioned across all tables was negative data – data about experiments that did not go as planned – which was held to be extremely valuable in stopping the repetition of mistakes already made elsewhere. Participants discussed that such data was being produced but was not accessible, because it was not captured systematically and because there were no incentives to publish this type of procedural and meta-knowledge.

Limited investment in commercial scale-up drives high-value UK start-ups away

Despite the UK's lively and efficient start-up environment, commercial scale-up was seen as lagging behind: it was limited by costs, regulation, and access to facilities, and behind this lay a conceptual barrier, in that scale-up was not viewed as research. While foundational research on biomanufacturing was well funded in the UK and the EU, commercial scale-up lagged. Lacking financial support, start-ups often remained small or moved to the US, where they could attract more venture capital.

The absence of bilateral funding mechanisms remains the principal barrier to transatlantic cooperation

By far the most cited barrier to transatlantic collaboration was the lack of bilateral funding mechanisms and a consistent funding strategy. Participants criticised both the volume of funding for international streams and its nature, funding for international collaboration being perceived as transient and ad hoc; what was demanded instead were "specific, irrevocably committed funding streams for bilateral efforts". Existing funding schemes, one table added, should be aligned between the two countries.

The knowledge-generation pipeline lacks scale, infrastructure, and coordination

This was chiefly a theme of one table. Participants there argued that the field generated knowledge at the wrong scale: research organised around individual PhD and post-doctoral projects was too small and too isolated, duplicated effort, repeated mistakes, and did not produce the data needed to move forward – the pipeline being geared towards publication and career-building rather than generating data for models, increasing data quality, and translating research into marketable products.

The proposed direction was advancing towards large-scale, coordinated efforts in a challenge-driven approach. While the first steps in this direction were seen in better coordination, data sharing, and data governance, the long-term aspiration was to move towards a "government-funded scale-up facility (or

network of them)" and massive automated data generation using automated "dark labs / cloud labs". A second table offered an analogy for shared infrastructure of this kind, observing that astrophysicists do not build their own telescopes but buy time on them, and asking whether the same could be done for one large network of scale-up facilities generating data for models.

The funding implication was specific: not "large, chunky funding" but "smaller groups synced up knowing that they are all part of the same scheme". This in turn required coordination across funders, researchers, and government; existing institutions – UKRI specifically – were said not to seem to have the capacity for coordination of this type.

Biomanufacturing needs to decrease cost to become competitive

This theme was raised largely by one table. For biomanufactured products to become more competitive in the market, manufacturing costs needed to fall, specifically the input cost of energy, labour, raw materials, and capital expenditure. A particular challenge was the cost of generating synthetic DNA, described as high and as not having fallen in step with the decreasing cost of DNA sequencing. Reducing it was rated high value, because synthetic DNA played such a central role in EB – "DNA is the currency of engineering biology". The UK was said not to have sufficient in-country DNA synthesis capabilities, creating a dependence on external suppliers which participants cautioned strengthened the position of those suppliers, China being named specifically, and created IP issues. Generalising the point, the optimisation of supply chains was held to be key. Learning from the successful use of AI in other engineering disciplines could be transferable to biomanufacturing, leading to new ways of designing and engineering, reducing cost, and increasing competitiveness.

What needs to change?

Funding needs to support collaboration and research on enablers

Participants framed funding both as a matter of volume and as a matter of structure, and asked for three changes. Participants pointed to limited common funding sources accessible to the US and the UK, a lack of specific, irrevocably committed streams for bilateral efforts, an absent dedicated funding mechanism and strategy, and the unfinished task of aligning UK–US funding schemes. Funding agencies were asked for joint funding instruments "rather than coordinated funding", and for long-term funding extending beyond single three-year projects.

Participants also asked to change what counted as fundable: research into translation and scale-up should be recognised as 'real' science and resourced accordingly. Participants noted that the funding geography cut both ways, biomanufacturing research being too low a TRL for funding in the US but easier to support in the EU.

Investing in scale-up was a topic particular to the UK: participants asked that the UK dedicate capital to growth at scale of "£50 million to £150 million", matching US levels. Participants also wanted incentives reaching beyond the obvious sectors, observing that it was difficult to attract money for ventures that were not in pharma or health, and asking the government to direct funding at areas of market failure.

Develop standards that meet the needs of biomanufacturing

Standards were seen as crucial enablers of scaling and of the efficient use of AI. The recurring demand was that standards be built deliberately rather than inherited: rather than working around existing structures, the field should develop guidelines and establish standards that served its purpose. General-purpose conventions had been worn too thin across biology's sub-fields to serve the specific needs of manufacturing well. At the transatlantic level, participants pointed to regulatory misalignment and a lack of co-development of standards – compounded because the structure of data, curation, and even nomenclature differed between the two countries. Standards bodies should develop clear and precise measurements, data quality standards, and future-proofed data formats. While acknowledging that precise data requirements were not yet settled, it was seen as critical to define the key attributes to capture in bioprocesses first.

Build IP frameworks and infrastructure that enable data sharing and collaboration

Across all tables, participants held that IP frameworks and a lack of infrastructure for sharing and cooperation were a fundamental barrier blocking data access. The solution might lie in a "pre-competitive knowledge base that captures, curates, and manages data from different stakeholders, including academia, industry, and government".

Similarly, negative data remained an underused asset, with the result that people made the same mistakes in different places. Journals should publish negative results and offer dedicated negative data repositories, and funding agencies should fund negative data publishing. Journals should also insist on metadata and software deposit, and funding agencies should make data standards and sharing a condition of funding.

Similarly limiting was the lack of highly capital-intensive research infrastructure, or the lack of access to it. Participants sketched a network of dark labs or cloud labs and proposed a "buy-time-not-build" model comparable to other research areas that self-govern access to high-capital research infrastructure, such as telescopes in astronomy.

What can different actors do?

Standards and metrology bodies

This drew the longest list of asks: maintaining a prolonged UK–US lead; delivering bioprocess standards with defined, common measurements (endorsed); an agreed approach to uncertainty (endorsed); consensus and ontologies for automation; a measurement approach to LCA; agreed analytical methods; digital, sensor, and reference material calibration; guidance on sensors and integration across scale; bilateral structures; reduced adoption barriers; and future-proofed, flexible data formats. A single participant noted bluntly: "Speed".

Funding agencies

A deliberate distinction opened the cluster: joint funding instruments "rather than coordinated funding". Participants asked for long-term funding beyond single terms; central facilities and strategic infrastructure,

as well as the funding to use and access them; funding for collaborations and AI–EB expert teams; and support for skills, training and research programmes. They requested funding for both publication of negative data and foundational data generation, collective funding of metadata curation, data standards and sharing as a funding condition, and metascience on funding effectiveness.

Government and policymakers

Participants asked government to promote the bioeconomy under a national and international strategy, with long-term commitment matched by actual appropriations rather than stated priorities alone; to fund areas of market failure; to offer tax incentives or "at least not disincentives" for bio-enabled products; to support pro-AI and pro-EB advocacy; and to credit curation and slow development so that the whole team was valued.

Regulatory bodies

Participants called for pragmatism and a balanced approach to risk among regulators; for clear, implementable rules "only where needed"; for greater speed; for a new GMO framework, along with platform mechanisms and adoption pathways; for clear paths and responsibilities for biomanufactured products; and for framework-compliant guidelines for data-generating processes. One participant challenged the framing of risk, asking what safety and environmental impact were being compared to.

Scientific journals

Participants provided a dense cluster of overlapping asks: data structure obligations for peer-reviewed publication; mandatory metadata, with software deposit and citation; deposit into publicly accessible repositories, crediting data generators via a "d-index"; and publication of negative results and dedicated negative data repositories, with one note recalling "**μ**" as a prior attempt. Participants also wanted journals to pay for metadata checking and to credit all research products, and took aim at publisher economics: "5–10% profit margin not 70%".

Researchers and AI scientists

Of their own community, participants asked for re-skilling, upskilling, and cross-training of current groups; higher skills and productivity; AI–EB degree and training programmes; knowledge exchange on applying AI methods effectively; early education to spark interest; and building to match the productivity of countries with more than a decade's head start. One note pressed culture, asking for respect for the team – "we science not just me science". The sharper notes addressed researchers' conduct towards the public: "to listen", "don't gaslight the public", and "be less cocky intellectuals".

Commercial partners, demand-side firms, and infrastructure vendors

Specialty chemical firms were asked to release some existing bioprocess data into a pre-competitive knowledge base. Infrastructure vendors should treat academia differently from big pharma; collaborate; build modular, flexible, interoperable, and open systems, API-controllable and "digital-twin"-compatible, including self-serviceable machines for lay use; and provide standardised data and sensor data outputs, with metadata reachable inside proprietary software.

UK Government

Scottish Government
Riaghaltas na h-Alba

THE UNIVERSITY of EDINBURGH

EDINBURGH INNOVATIONS