Roundtable Session 1 - Table 3 - Can AI Transform CGT Process Development?

Facilitator: Krzysztof Krawczyk, *F. Hoffmann-La Roche AG*

Scribe: Rafaella Buzatu, *Leiden University*

Abstract:

While the term "AI" dominates today's headlines, the underlying concepts have been operating in process development under various names, such as multivariate data analysis, machine learning, and expert systems, for many decades.

Today, however, the paradigm is shifting from merely monitoring a process to actively designing it. Generative AI can navigate complex parameter spaces that traditional Design of Experiments (DoE) cannot easily solve, while digital twins combine mechanistic equations with machine learning to simulate cell growth, optimize transfection efficiency, and predict purification outcomes. Simultaneously, Large Language Models (LLMs) can ingest large numbers of unstructured documents - legacy batch records, deviation reports, shift logs, equipment manuals, and scientific literature, to facilitate root-cause analysis.

Yet, applying it in cell and gene therapy (CGT) process development presents severe operational hurdles. AI models thrive on large, standardized datasets that are rarely available in low-run, high-variability CGT processes. Meanwhile, "black-box" predictions and hallucination risks complicate defining robust control strategies. Finally, in a sector where "the process equals the product," integrating third-party AI introduces critical data confidentiality and IP risks. The recent controversy where private mathematical drafts on the Navier-Stokes problem allegedly influenced frontier model training serves as a stark cautionary tale for biopharma.

This roundtable discussion will explore whether AI can deliver transformational benefits to CGT process development today despite these challenges.

Discussion Questions:

1. How can teams train reliable models when constrained by small data sets and high donor heterogeneity?
2. Is the concept of a hybrid digital twin a viable option in CGT process development?

3. How do we establish a defendable design space using non-explainable ("black box") algorithms?
4. How can biopharma leverage external AI capabilities while safeguarding proprietary process trade secrets?

Notes:

Disclaimer: In the spirit of this Table discussion, the following summary has been generated using ChatGPT-6.0 based on notes taken by the scribe.

The discussion explored the opportunities and limitations of applying AI to cell and gene therapy (CGT) process development, particularly in the context of limited datasets, high donor variability, process control, digital twins, and data confidentiality.

A central theme was that the main bottleneck for AI implementation in CGT is not necessarily the technology itself, but the availability, quality, and accessibility of relevant data.

1. Data availability and sharing remain major challenges

CGT process development often involves small datasets, limited manufacturing runs, and substantial donor-to-donor variability. Although autologous therapies may eventually generate considerable amounts of patient-specific manufacturing data, the availability of meaningful readouts and associated patient information remains a challenge.

Participants discussed the potential of leveraging data across products, platforms, and organizations. However, limited platform knowledge, inconsistent data structures, and reluctance to share proprietary information currently restrict these opportunities. A shift towards greater data collaboration may be necessary to fully exploit AI capabilities.

Importantly, the issue is not always a lack of data, but also the inability to efficiently access, integrate, and interpret existing information. Bioreactor measurements, laboratory notebooks, failed experiments, historical batch records, and meeting minutes represent potentially valuable but underutilized sources of knowledge.

2. Knowledge management may be the most immediate application of AI

One of the most promising near-term applications identified was the use of AI, particularly large language models (LLMs), to extract and organize knowledge from unstructured information.

For example, AI could identify previously unsuccessful experimental conditions, recover knowledge from historical documentation, and prevent unnecessary repetition of failed

experiments. This could be particularly valuable for rare disease therapies, where development resources and experimental opportunities are limited.

AI could also support Quality by Design (QbD) approaches by helping identify potentially relevant process parameters, organizing existing evidence, and supporting more consistent critical quality attribute (CQA) assessments, which currently involve considerable manual and subjective evaluation.

3. AI-assisted design spaces and control strategies require human oversight

The discussion questioned whether AI could reliably support the establishment of design spaces, acceptable process parameter ranges, and control strategies when experimental data are limited.

Although AI may help explore complex parameter interactions beyond those investigated through conventional Design of Experiments (DoE), its predictions remain dependent on the underlying data and require experimental validation.

The lack of explainability of certain models, alongside the hallucination risks associated with generative AI, raises concerns about their use in regulatory decision-making. Participants emphasized that AI should support, rather than replace, expert scientific judgment, with outputs critically evaluated and validated before implementation.

A more realistic near-term approach may involve applying AI to specific, well-defined tasks, such as molecular binding predictions, formulation optimization, or structured CQA assessment, rather than attempting to automate entire process development workflows.

4. Digital twins offer potential, particularly when developed incrementally

Hybrid digital twins, combining mechanistic understanding with data-driven predictive models, were considered a promising long-term direction.

Such models could be built progressively as experimental and manufacturing data become available, allowing predictions of untested conditions, identification of influential process parameters, and recommendations for process adjustments.

However, participants questioned whether meaningful digital twins could be established before sufficient experimental evidence exists. While AI may provide initial guidance, reliable predictive models require experimental data for development and validation.

The potential value of digital twins may therefore extend beyond initial process design into continued process verification, post-commercialization process improvement, and knowledge transfer to future generations of products.

5. Data confidentiality and intellectual property require new approaches

The potential use of external AI platforms raises significant concerns regarding proprietary manufacturing information and intellectual property.

At the same time, restricting access to data limits the development of sufficiently robust and generalizable AI models. Participants recognized a fundamental tension between protecting commercially sensitive process knowledge and enabling the data sharing necessary for effective AI development.

Industry-wide collaboration, potentially through manufacturing organizations with access to multiple products and processes, was discussed as one possible way forward, provided appropriate confidentiality safeguards can be established.

Overall conclusion

The discussion highlighted that AI has substantial potential to improve CGT process development, but its application must be aligned with the maturity and availability of data.

In the near term, AI may deliver the greatest value through knowledge management, data integration, evidence synthesis, and decision support, rather than fully autonomous process design. Over time, as experimental and manufacturing datasets accumulate, more advanced predictive applications, including hybrid digital twins and AI-assisted control strategies, may become increasingly feasible.

Ultimately, realizing this potential will require improved data infrastructure, greater collaboration and knowledge sharing, appropriate protection of proprietary information, and continued human oversight and experimental validation.

Key takeaway: The opportunity is not simply to use AI to design better processes, but to use AI to better capture, connect, and learn from existing knowledge—creating the foundation for increasingly predictive and data-driven CGT process development.