From Virtual Cells to Programmable Humans: Advancing Digital Biology Through Hybrid AI Systems

Anonymous authors
Paper under double-blind review

Abstract

The convergence of artificial intelligence and systems biology is giving rise to a new paradigm in biomedical research—AI-powered virtual biological systems. From single-cell simulations to organ-level models and ultimately programmable virtual humans, this digital continuum holds transformative potential for disease modeling, personalized medicine, and therapeutic discovery. In this review, we critically examine the state of the art in AI-driven simulations, including the numerical foundations, multiscale integration strategies, and the emerging class of hybrid models that bridge mechanistic and data-driven approaches. We explore the challenges of validation, uncertainty quantification, and regulatory alignment across simulation scales, with particular focus on the development of simulation accountability frameworks such as SIM-CARDs. Ethical and privacy concerns, including algorithmic bias and data sovereignty in patient-specific models, are also addressed, alongside concrete proposals for governance and federated simulation workflows. Special attention is given to the technical complexity of multiscale modeling, including the integration of mechanistic solvers with neural architectures and the computational resources required for real-time, clinically actionable simulations. We conclude with a translational roadmap for virtual biology that projects validated virtual cells for drug screening by 2030, multi-organ simulations by 2040, and the emergence of programmable virtual humans by 2055. By unifying high-fidelity numerical models with explainable AI, and aligning simulation design with ethical, regulatory, and clinical needs, the field of digital biology is positioned to unlock scalable and trustworthy biomedical innovation.

Keywords: AI-driven simulation, virtual cells, programmable humans, digital biology, multiscale modeling, hybrid models, biomedical simulation, explainability, ethics, regulatory frameworks, uncertainty quantification

1 Introduction

Artificial Intelligence (AI) has rapidly emerged as a transformative tool in biomedical research, enabling the development of dynamic, scalable, and personalized computational models of biological systems. These AI-powered models—spanning virtual cells, organs, and entire physiological systems—promise a revolution in how we simulate, understand, and manipulate biology at scale. This shift from descriptive biology to computational simulation is not merely theoretical; it is already being operationalized in fields such as pharmacokinetics, systems biology, and personalized medicine Greene et al. (2019); Jimenez-Luna et al. (2022). Central to this transformation is the integration of machine learning (ML), mechanistic modeling, and multiscale data into hybrid frameworks capable of simulating complex, nonlinear, and stochastic biological behaviors Rackauckas et al. (2020); Chmielecki et al. (2023).

Despite these advances, the application of AI in modeling biological systems introduces a host of new challenges and responsibilities. Biological systems exhibit extraordinary het-

erogeneity, dynamic regulation, and context-dependent responses. Capturing this complexity requires combining mechanistic understanding—often encoded through systems of ordinary differential equations (ODEs), agent-based models, or partial differential equations (PDEs)—with data-driven AI models capable of learning unknown patterns from large-scale, multimodal datasets Karr et al. (2012); Rajkomar et al. (2019). These hybrid architectures, often called "gray-box" models, offer interpretability and flexibility but introduce new concerns about reliability, generalizability, and computational tractability.

As the field progresses toward simulating entire organs and ultimately programmable digital humans, ethical, regulatory, and computational issues rise to the forefront. For example, regulatory pathways for AI-based simulations are still in flux, with agencies like the FDA releasing evolving guidance for Good Machine Learning Practice (GMLP) FDA (2021a), while the EU AI Act and MDR establish complementary but more stringent criteria for risk and transparency Commission (2021; 2023). These frameworks are not yet fully adapted to the needs of dynamic simulation systems, which require continual updates, data privacy safeguards, and transparent documentation of modeling assumptions and limitations.

This review aims to critically evaluate the current landscape of AI-powered virtual modeling in biology, focusing on: (1) the modeling techniques used at the cellular, organ, and system levels; (2) the integration of mechanistic and machine learning approaches; (3) validation strategies and benchmarking efforts; (4) regulatory and ethical implications; and (5) a realistic roadmap for the future. We pay special attention to the convergence of computational biology, AI, and digital health, outlining the necessary infrastructure, governance frameworks, and cross-disciplinary collaboration needed to safely translate these technologies into clinical practice.

2 AI-Driven Modeling of Virtual Cells

The concept of a virtual cell—a fully in silico representation of cellular processes—lies at the heart of digital biology. AI-driven modeling of virtual cells enables the integration of high-throughput omics data, mechanistic understanding, and empirical rules to create dynamic, predictive simulations of single-cell behavior under various conditions Waltemath et al. (2016); Karr et al. (2012). These models have already found applications in drug discovery, toxicity screening, and synthetic biology Chowdhury et al. (2022).

The foundation of virtual cell modeling can be broadly categorized into three methodological pillars: mechanistic models, data-driven models, and hybrid approaches. Mechanistic models, such as ODE-based systems biology representations, capture well-known biochemical pathways including gene regulatory networks, signaling cascades, and metabolic fluxes. These models rely on known kinetic parameters, yet their scalability is often hindered by parameter uncertainty and incomplete biological knowledge Klamt (2009). In contrast, data-driven models leverage machine learning—particularly deep learning architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers—to learn high-dimensional mappings between inputs (e.g., drug concentrations) and outputs (e.g., apoptosis rates) directly from data Angermueller et al. (2016); Yang et al. (2022).

Hybrid models, or gray-box systems, seek to reconcile these approaches. For instance, Universal Differential Equations (UDEs) embed neural networks within ODEs to model unknown components while retaining mechanistic interpretability Rackauckas et al. (2020). A notable application is the use of hybrid modeling for predicting single-cell transcriptomic dynamics across time, where neural networks infer regulatory effects not captured by static pathway maps Chen (2020). These approaches show promise but pose computational challenges—particularly in parameter estimation, model stiffness, and convergence during training.

Validation of virtual cell models requires stringent evaluation strategies. Cross-validation against single-cell RNA-seq datasets and perturbation experiments are often employed, alongside comparison with CRISPR screening data Luecken (2022). However, a lack of gold-standard datasets for benchmarking limits consensus on model generalizability. To

address this, initiatives such as the DREAM Challenges and Human Cell Atlas have begun establishing standard tasks for virtual cell simulation and prediction Norman (2019). Additionally, uncertainty quantification methods—such as Bayesian neural networks and ensemble learning—are increasingly adopted to provide confidence estimates for model predictions Gal & Ghahramani (2016).

A representative case study is the application of a virtual hepatocyte model for predicting drug-induced liver injury (DILI). A hybrid model incorporating pharmacokinetic parameters, intracellular signaling dynamics, and transcriptomic responses was used to predict hepatotoxicity profiles of novel compounds. Validation was performed against both in vitro assays and known toxicological benchmarks, achieving a balanced accuracy of 82% Ghallab et al. (2022). Nonetheless, the model revealed limitations in capturing rare failure modes associated with mitochondrial dysfunction, emphasizing the need for robust outlier modeling.

Despite encouraging progress, modeling the cellular level remains bottlenecked by three core issues: (1) data sparsity and bias in high-dimensional omics; (2) the difficulty of incorporating stochasticity in gene expression; and (3) limitations in causal inference from observational datasets. Future directions must prioritize causal representation learning, integration of time-resolved data, and modular architectures that allow for reuse across cell types and conditions Schölkopf (2021); Luecken (2022). In sum, virtual cell modeling has transitioned from conceptual aspiration to practical implementation, though significant challenges remain in ensuring biological plausibility, computational efficiency, and clinical applicability.

3 AI-Powered Modeling of Virtual Organs

Modeling biological organs with high fidelity remains a central challenge in computational medicine. Virtual organs aim to replicate the behavior of complex tissues such as the heart, liver, and lungs using AI-enhanced simulations. These models integrate structural, functional, and sometimes patient-specific data to reproduce physiological dynamics. While mechanistic modeling has been traditionally dominant in this space, the growing interest in AI-driven and hybrid approaches is redefining possibilities. Deep neural networks, including recurrent neural networks (RNNs), graph neural networks (GNNs), and convolutional architectures, have been applied to organ-scale dynamics, often in combination with physics-informed constraints to respect underlying bio-mechanics Karniadakis et al. (2021); Raissi et al. (2019). These models are capable of approximating PDE solutions, capturing flow dynamics, and learning spatiotemporal behaviors without fully solving physical systems numerically.

Heart modeling offers a key testbed. Models of cardiac electrophysiology and hemodynamics rely heavily on multi-physics solvers, but physics-informed neural networks (PINNs) are increasingly employed to improve simulation speed and parameter fitting Sun et al. (2023). For example, Karniadakis et al. demonstrated PINNs in modeling arrhythmogenic substrates, achieving accurate voltage propagation estimates while circumventing stiff PDE solvers Karniadakis et al. (2021). Similarly, surrogate AI models now approximate ventricular pressure-volume loops with high precision for real-time decision support Fritz et al. (2022).

In the liver domain, virtual simulations have enabled drug-induced liver injury (DILI) prediction and hepatotoxicity assessment. Hybrid models, integrating mechanistic compartmental pharmacokinetics (PK) with deep generative models, provide both biological interpretability and predictive capacity Yu et al. (2021); Dahlquist et al. (2022). For example, liver lobule simulations now incorporate cell signaling dynamics with histopathological images to capture zonal heterogeneity, allowing finer-grained toxicity predictions Elbadawi et al. (2021). These models often require massive multiscale datasets and are prone to overfitting without proper regularization strategies and uncertainty quantification Zhang et al. (2023).

Despite progress, several challenges persist. First, the multiscale nature of organs introduces scale-bridging complexity—from ion channels to tissue mechanics. Second, integrating di-

verse data modalities (e.g., genomics, histology, hemodynamics) remains an open frontier Viceconti et al. (2021). Lastly, benchmarking and validation are critical. Studies often use clinical imaging or biomarker trajectories for validation, but generalizability remains limited due to inter-patient variability Lambin et al. (2021). To address this, some efforts propose simulation testbeds using public datasets like MIMIC-IV and the UK Biobank, facilitating reproducibility and comparison Goldberger et al. (2000); UK Biobank (2022).

Emerging standards such as the Functional Mockup Interface (FMI), CellML, and SBML are also enabling model interoperability across toolchains, which is vital for co-simulation of cardiac-lung interactions or liver-gut axes Ghoniem et al. (2015); Hucka et al. (2003). A concrete example is the integration of finite element simulations with GNN surrogates for cardiac tissue stress prediction, yielding a 10x speedup without compromising fidelity Fritz et al. (2022). These case studies demonstrate the growing maturity of AI-powered organ models and highlight the importance of rigorous multiscale validation pipelines.

4 Integration of Mechanistic and Data-Driven Models

The integration of mechanistic and data-driven models represents a pivotal advancement in biomedical simulations, bridging first-principles biological knowledge with the pattern-extracting power of modern AI. Mechanistic models offer explainability and are grounded in physical and physiological laws, often formulated as systems of ordinary or partial differential equations (ODEs or PDEs). In contrast, data-driven models, particularly deep learning approaches, excel in flexibility and scalability, especially when experimental data is abundant. By combining these two paradigms, hybrid models can yield predictive tools that are both interpretable and accurate across scales.

Recent developments in physics-informed machine learning have made such integration feasible at scale. For example, Physics-Informed Neural Networks (PINNs) can solve differential equations while respecting underlying biophysical laws Karniadakis et al. (2021). These models have been applied in cardiac electrophysiology, where traditional finite element solvers are complemented by neural networks trained on experimental data Fritz et al. (2022). The use of surrogate models accelerates simulations significantly while retaining physical fidelity, thus allowing for real-time decision support in clinical contexts.

However, hybrid modeling introduces new challenges. One critical issue lies in the alignment of scales and resolutions between mechanistic and data-driven components. Mechanistic models may operate at millisecond timescales or micrometer resolutions, whereas machine learning models often function at aggregated levels. Mismatches can lead to spurious predictions unless carefully aligned. Furthermore, hybrid models require strategies for efficient parameter mapping, transfer learning across domains, and management of conflicting data priors. Advanced data assimilation techniques, such as ensemble Kalman filters or variational inference, are increasingly used to harmonize these models Xu et al. (2022).

A practical case study is the integration of a mechanistic liver model, which includes zonation-based enzyme distribution, with a generative adversarial network (GAN) trained on biopsy images for simulating drug-induced liver injury (DILI). This hybrid approach enabled prediction of both systemic pharmacokinetics and localized tissue-level hepatotoxicity, outperforming either model alone in preclinical settings Elbadawi et al. (2021). In another example, finite element cardiac models have been coupled with convolutional neural networks to form bidirectional feedback loops between mechanical stress computation and real-time echo imaging interpretation Fritz et al. (2022); Sahli Costabal et al. (2021).

From a computational standpoint, integrating mechanistic and data-driven models increases the need for interoperability standards. Tools like the Systems Biology Markup Language (SBML), CellML, and the Functional Mockup Interface (FMI) facilitate modular composition and co-simulation of models developed in different frameworks Ghoniem et al. (2015); Hucka et al. (2003). However, model exchange is still hampered by a lack of standardization for metadata and ontology alignment.

To overcome these challenges, there is a growing interest in developing modular platforms that allow pluggable AI components into simulation pipelines. For instance, frameworks like

OpenCOR and Sim4Life now support hybrid modeling through plugin architectures. Additionally, formal causal reasoning frameworks, such as Structural Causal Models (SCMs), are being explored to constrain machine learning outputs with biologically plausible interventions Pearl (2009); Runge et al. (2019).

Hybrid models must also contend with uncertainty propagation. Uncertainty can stem from both mechanistic parameters (e.g., diffusion coefficients) and data-driven predictions (e.g., black-box neural nets). Techniques like Bayesian deep learning, Monte Carlo dropout, and interval analysis are used to quantify and propagate this uncertainty Zhang et al. (2023). Properly quantified, this uncertainty can guide experimental design and clinical decision-making.

As the field matures, it is likely that model transparency, interpretability, and traceability will become prerequisites for clinical acceptance. Standards such as TRIPOD-AI, MINIMAR, and CONSORT-AI are being adapted to assess not only pure ML systems but also hybrid pipelines Liu et al. (2020); Rivera et al. (2021). These developments signal a shift toward simulation systems that are not only predictive but also accountable and auditable across their entire life cycle.

5 Projected Timelines and Translational Roadmap

The progression from AI-powered virtual cells to programmable virtual humans in biomedical research represents a paradigm shift, demanding not only technological readiness but also regulatory adaptation, interdisciplinary collaboration, and ethical foresight. To navigate this complex trajectory, we propose a phased roadmap spanning from foundational development to clinical and societal integration, with projected timelines and defined milestones.

Phase I (2025–2030): Virtual Cell Prototypes. This phase emphasizes the construction and validation of virtual cell models capable of mimicking molecular signaling, metabolic flux, and cellular differentiation processes. AI approaches, particularly deep generative models, are now able to learn the latent dynamics of cellular phenotypes from high-dimensional omics data Ma et al. (2018); Yuan & Cai (2022). During this phase, progress will hinge on the availability of single-cell multi-omics datasets, better spatial transcriptomics resolution, and robust experimental benchmarks such as perturbation assays Cao (2020); Lohoff et al. (2022). Readiness milestones include reaching predictive accuracy benchmarks in CRISPR screening simulations, benchmarking against known metabolic outputs, and standardized evaluation on datasets such as Human Cell Atlas Regev (2017).

Phase II (2030–2040): Virtual Organs and Tissue Simulations. Building on validated cellular models, Phase II will target multi-cellular and tissue-level simulations with integrated vasculature, immune response, and organ-specific microenvironments. Techniques such as multiscale modeling, graph neural networks, and hybrid finite element-AI pipelines are projected to dominate this decade Viceconti et al. (2021); Sahli Costabal et al. (2021). Validation milestones will include predictive accuracy of tissue response to drugs, in silico organ perfusion simulations, and virtual biopsy comparison to histopathological gold standards. Initiatives such as the VPH Institute and Human Physiome Project provide scalable blueprints for these models Hunter & Borg (2003). Key bottlenecks include simulation stability, data heterogeneity, and computational load. Contingency plans involve leveraging federated learning to access distributed organ-level datasets without violating privacy constraints Rieke et al. (2020).

Phase III (2040–2055): Programmable Virtual Humans. This final phase envisions virtual human models integrating cell, tissue, and organ-level simulations governed by parameterizable AI modules. These models are not merely digital twins, but programmable entities capable of intervention modeling, treatment simulation, and hypothetical testing. Key technologies will include cross-modal transformers, structural causal modeling, and generative simulation Lu et al. (2023). Milestones include achieving FDA-approved virtual clinical trials for narrow use cases (e.g., cardiotoxicity), real-time biophysical feedback integration with wearable data, and transparency-compliant AI decision logs. Benchmarks

like UK Biobank, MIMIC-IV, and OpenSim simulations will form the basis of this validation effort Johnson et al. (2016); Allen et al. (2015). Regulatory frameworks, such as FDA GMLP and the EU AI Act, must be expanded to include simulation drift metrics and SIM-CARD accountability frameworks. Crosswalks to ISO standards (ISO 14971, 62304) will guide lifecycle safety and risk tracking FDA (2021b); ISO (2019; 2006).

A critical path analysis (Table 1) maps interdependencies and highlights bottlenecks: (1) data harmonization across modalities, (2) scalable validation strategies, and (3) explainability for regulatory pathways. These are the three pillars upon which translational success rests. Our risk register includes potential failures due to data bias, lack of causal generalization, and under-specification of AI behavior in novel clinical contexts.

Table 1: Phased roadmap milestones and risk dependencies

Phase	**Milestones**	**Risks / Dependencies**
Phase I (2025–2030)	CRISPR simulation benchmark, metabolic modeling fidelity, Human Cell Atlas alignment	Lack of single-cell perturbation gold standards, variability in omics platforms
Phase II (2030–2040)	Organ perfusion modeling, in silico biopsy fidelity, immune-tissue interaction simulation	Heterogeneous clinical imaging standards, high HPC demand
Phase III (2040–2055)	Virtual trials FDA approval, real-time wearable-AI integration, SIM-CARD regulatory endorsement	Unclear regulatory validation pathways, uncertainty propagation, explainability bottlenecks

Ultimately, the translational path from AI-modeled cells to programmable virtual humans will depend not only on scientific innovation but also on interdisciplinary governance and ethical co-design. Transparent benchmarks, international regulatory alignment, and scenario-based validation will be the linchpins of adoption.

6 Governance, Regulation, and Ethical Implications

As AI-powered virtual biological models transition from research prototypes to translational clinical tools, the demand for robust governance and ethical stewardship intensifies. The trajectory toward programmable virtual humans traverses a spectrum of ethical, regulatory, and societal concerns, including data sovereignty, model explainability, and the mitigation of algorithmic bias.

A primary ethical challenge lies in the use of patient-specific data to train and personalize these models. Clinical datasets often contain sensitive information, and AI models trained on such data may unintentionally reveal personal traits or medical conditions if not properly anonymized. Privacy-preserving technologies—such as federated learning Rieke et al. (2020), differential privacy Dwork & Roth (2014), and secure multi-party computation Mohassel & Zhang (2017)—are increasingly being integrated into modeling pipelines to address these concerns. However, these techniques introduce trade-offs in model performance and development complexity, and more work is needed to standardize their implementation in biomedical AI workflows.

Another critical concern is algorithmic bias, which arises when training data fails to represent diverse patient populations or encodes historical inequities. Biased models risk perpetuating disparities in diagnosis or treatment outcomes. For instance, if a virtual heart model is trained predominantly on male data, it may underperform in female patients Obermeyer et al. (2019). Strategies such as subgroup-specific validation, fairness-aware learning, and post-hoc debiasing are necessary but insufficient without upstream data reform. Regulatory agencies such as the FDA and EMA are increasingly demanding transparent documentation of bias mitigation strategies during model approval Topol (2019); FDA (2021b).

To facilitate transparent regulation and clinical trust, we propose the concept of "SIM-CARDs" (Simulation Cards for Accountability, Regulation, and Disclosure), a documentation framework inspired by model cards Mitchell et al. (2019). Each SIM-CARD would detail simulation assumptions, data lineage, failure modes, validation status, and intended clinical scope. For example, a SIM-CARD for a virtual liver model would list drug metabolism pathways included, exclusion criteria for datasets, and performance metrics on population subgroups. These would serve both regulatory review and end-user interpretation needs. A draft template is illustrated in Table 2.

Table 2: SIM-CARD Template Example for Virtual Liver Model

Field	**Description**
Simulation Scope	Hepatic metabolism simulation for drug X, time frame: 24h
Data Lineage	Liver biopsy RNA-seq from GTEx v8; clinical PK data from Phase II trial
Assumptions	Homogeneity in hepatic zonation; linear pharmacokinetics
Validation Metrics	AUROC = 0.84 (internal), AUROC = 0.79 (external)
Failure Modes	Unreliable prediction in hepatic cirrhosis or pediatric patients
Regulatory Notes	Pre-submission filed with EMA; requires CE marking for deployment
Update Policy	Recalibration every 6 months or after model drift ¿5% detected

The ethical and legal implications of simulating human physiology extend further into potential misuse. While digital twins can aid in optimizing treatments, they could be repurposed for unregulated drug testing, insurance discrimination, or even bioweapon design. Proactive governance—via multidisciplinary oversight boards and anticipatory risk frameworks—will be essential. Lessons from genomics regulation and bioethics can inform early-stage guardrails for virtual human simulation Mittelstadt et al. (2016).

7 Validation and Benchmarking

As AI-driven virtual biological systems move closer to clinical and translational deployment, rigorous validation and benchmarking become non-negotiable for ensuring trustworthiness, reproducibility, and regulatory approval. Virtual cells, organs, and programmable humans each present unique challenges in model assessment, requiring diverse validation metrics, datasets, and methodological rigor.

The validation of virtual cell models often involves correlating predicted cellular behaviors with high-resolution in vitro data, such as single-cell RNA-seq, proteomics, and electrophysiology. Standard metrics such as area under the curve (AUC), root mean square error (RMSE), and dynamic time warping (DTW) are used to compare predicted and observed cellular responses across stimuli or time-series perturbations Williams et al. (2021); Lopez et al. (2018). However, reliance solely on statistical metrics can obscure biological plausibility. Therefore, hybrid validation strategies incorporating knowledge-driven constraints and perturbation-based testing (e.g., CRISPR knockout simulations) are gaining traction Campbell et al. (2021); Kermany et al. (2018).

At the organ level, benchmarking requires simulation fidelity across multi-physics and multi-scale phenomena. For instance, virtual heart models must reproduce both electromechanical coupling and hemodynamic outputs Niederer et al. (2019); Wang et al. (2021). Benchmarks such as the Cardiac Electrophysiology Web Lab Noble et al. (2012) or VPH Benchmarking Initiatives provide structured frameworks for comparing virtual organ simulations across labs. Key metrics include ejection fraction, pressure-volume loops, or temporal waveform alignment with patient ECGs. Sensitivity analysis and uncertainty quantification using Bayesian or ensemble methods are increasingly mandated Ghosh et al. (2022); Wu et al. (2021).

Validation of programmable human models is inherently more complex, requiring evidence across a hierarchy of physiological levels and temporal scales. We propose a three-stage framework: (1) internal consistency (e.g., cross-modality coherence across tissues), (2) retrospective validation (comparison with longitudinal clinical records), and (3) prospective synthetic control trials or real-world evidence comparison Muehlematter et al. (2021); Karpatne et al. (2017). For example, a virtual diabetic patient model might be benchmarked on HbA1c trajectories from NHANES cohorts, calibrated using glucose-insulin homeostasis equations, and validated prospectively in synthetic arms Fralick et al. (2020).

To support reproducibility and transparency, a simulation accountability checklist should accompany each model. Inspired by datasheets for datasets Gebru et al. (2021), our "Simulation Accountability Sheets" include fields for assumptions, parameter sensitivity, failure modes, validation datasets, and update history. These templates align with FAIR principles and emerging regulatory guidelines such as the FDA's Good Machine Learning Practice (GMLP). We also propose the release of a minimal benchmark suite, comprising gold-standard reference problems across cell, organ, and human scales, using publicly available datasets such as the Human Cell Atlas, MIMIC-IV, and UK Biobank. A comparison table mapping tasks to datasets and evaluation metrics can serve as a roadmap for the community.

Robust benchmarking protocols will be vital for clinical readiness. Quantitative thresholds must be defined for specific validation gates, including calibration slopes within ± 0.05, coverage rates above 90%, and cross-site generalization with ¡10% performance degradation. Future efforts must also integrate model explainability into validation processes, using SHAP, LIME, or saliency maps to audit learned behaviors and ensure alignment with domain knowledge.

8 Causality, Explainability, and Clinical Trust

The successful integration of AI-driven virtual models into clinical workflows demands a robust framework for causal inference, explainability, and clinical interpretability. While predictive accuracy remains a key metric, the ability to understand causal relationships and provide transparent reasoning for model outputs is essential for trust, particularly in high-stakes applications such as diagnosis and treatment planning.

Causal modeling in virtual biology is increasingly framed through the lens of Structural Causal Models (SCMs), which encode domain knowledge as graphical models or structural equations to model interventions and counterfactuals Pearl (2009); Schölkopf (2022). In organ-level simulations—such as cardiac or renal systems—embedding SCMs allows for hypothesis testing, intervention simulations, and policy optimization Berner et al. (2021); Yoon et al. (2018). For instance, SCM-based models of cardiac arrhythmia can be used to simulate ablation effects or drug perturbations under multiple comorbidity scenarios. However, challenges remain in defining valid causal graphs from biological data, where latent confounders and feedback loops often obscure directionality. Advances in causal discovery from time-series and mechanistic priors show promise for integration with digital twin architectures Ghassemi et al. (2021); Runge et al. (2019).

Explainability remains critical for regulatory approval and clinical adoption. While post hoc explainability tools such as SHAP Lundberg & Lee (2017), LIME Ribeiro et al. (2016), and saliency maps are commonly applied in imaging or omics tasks, their clinical relevance is limited unless tied to actionable biological insights. Recent advances in intrinsically interpretable models—e.g., symbolic regression, causal forests, or monotonic networks—offer alternatives for applications requiring accountability Molnar (2022); Rudin (2019). For example, interpretable models applied to pharmacogenomics data can suggest biomarkers governing drug response pathways, which are more likely to be accepted by clinicians.

The gap between technical interpretability and clinical trust underscores the need for human-in-the-loop interfaces and formal validation frameworks. Trust in virtual human models is shaped not just by accuracy but also by the transparency of assumptions, traceability of decisions, and alignment with domain expertise Holzinger et al. (2020); Amann et al. (2020). Trust metrics can be formalized as part of a clinical readiness score that combines: (1)

uncertainty quantification (e.g., predictive intervals), (2) explanation fidelity (e.g., alignment with expert reasoning), and (3) user satisfaction from clinician feedback trials.

Finally, we advocate for the formal adoption of Causality Cards and Explainability Datasheets—structured reporting templates documenting causal assumptions, explanation mechanisms, fidelity tests, and clinical interface elements. These complement accountability tools such as Model Cards and could be required in simulation lifecycle audits. As AI-enabled simulations evolve into prescriptive systems, causality and explainability will be indispensable for aligning predictive power with human values and clinical utility.

9 Governance, Regulation, and Societal Implications

As AI-powered simulations of biological systems advance from virtual cells to organ systems and programmable humans, their governance requires careful scrutiny. The increasing deployment of these models in preclinical and clinical contexts mandates that they be aligned with ethical norms, regulatory standards, and broader societal expectations.

A key challenge in governance is establishing regulatory clarity for models that dynamically evolve with new data. Regulatory bodies such as the U.S. FDA, European Medicines Agency, and China's NMPA have begun evaluating adaptive models, but current frameworks—such as Good Machine Learning Practice (GMLP) and Software as a Medical Device (SaMD)—do not fully address simulation-specific nuances Topol (2019); Patel & Shortliffe (2020). For instance, validation strategies for AI-driven organ simulations must account for multiscale complexity and model updates. To address this, a simulation-specific lifecycle model is proposed, combining ISO 13485 for medical devices, IEC 62304 for software lifecycle, and ISO 14971 for risk management. Moreover, the EU AI Act and MDR regulations need refinement to accommodate simulations used in synthetic control arms, virtual clinical trials, or treatment planning Schwalbe & Wahl (2020); Breton (2021).

We propose a Simulation Accountability Sheet (SIM-CARD) template—a structured form documenting model purpose, assumptions, causal scope, training datasets, validation status, known failure modes, uncertainty treatment, update policy, and clinical usage constraints. For example, a cardiac arrhythmia simulation tool using hybrid finite element and deep learning methods must disclose mesh resolution thresholds, calibration targets, sensitivity to atrial geometry variation, and performance across population subgroups. These SIM-CARDs can support regulatory audits, inform clinical use, and ensure transparency.

Governance also entails privacy preservation. Large-scale simulations of virtual humans inevitably require integrating patient-derived multi-omics, imaging, and real-world data. Risks of re-identification and misuse are non-trivial. Techniques such as federated learning, differential privacy, and secure multiparty computation can reduce exposure, but must be mandated by regulation Dwork & Roth (2014); Kaissis et al. (2021); Xu et al. (2021). Data lineage tracking and audit logs must accompany all training processes. Moreover, informed consent for virtual models must extend to inferred representations, not just raw data. The use of synthetic data to augment training is promising but should be documented transparently.

Societal implications extend to equity and misuse. Virtual simulations may inadvertently encode population biases, amplifying health disparities. Regulatory frameworks must mandate bias audits and fairness metrics at every stage of the simulation pipeline Rajkomar et al. (2018). Additionally, the possibility of misuse—such as predictive simulations for genetic selection or bioweapon research—requires proactive governance, including usage monitoring, policy restrictions, and whistleblower protections.

Lastly, public trust hinges on transparency and participation. Stakeholders—including patients, clinicians, ethicists, and regulators—must be involved in co-designing simulation standards, testing frameworks, and deployment policies. Engagement platforms such as open model repositories, version tracking, public dashboards, and simulation impact assessments (SIAs) can improve accountability. A forward-looking governance strategy recognizes that as programmable humans become more accurate and consequential, societal engagement must scale in parallel.

10 Conclusions and Future Perspectives

AI-driven virtual biological systems, spanning from single-cell simulations to programmable virtual humans, are poised to redefine the frontiers of biomedical research and clinical practice. This review has highlighted the scientific foundations, numerical and computational techniques, regulatory landscapes, and ethical concerns surrounding these evolving systems. While rapid advances have been made, critical gaps in validation, standardization, and explainability remain that must be addressed before widespread clinical deployment.

A central insight from this analysis is the need for robust hybrid modeling frameworks that can accommodate both mechanistic and data-driven paradigms. Mechanistic models offer physical interpretability and control, while deep learning methods excel at extracting patterns from high-dimensional data. Bridging these paradigms requires novel numerical solvers, multi-scale architectures, and uncertainty-aware training pipelines. The future of virtual biology will likely depend on the systematic fusion of structural causal modeling, physics-informed learning, and ensemble methods to represent the inherent complexity and stochasticity of biological systems.

Moreover, there is an urgent need to institutionalize simulation governance mechanisms such as SIM-CARDs, benchmarking standards, and simulation validation gates. These tools are not merely procedural: they are foundational to ensuring trust, reproducibility, and clinical safety. Future work should prioritize the creation of open simulation repositories with metadata-standardized reporting (analogous to model cards) and interoperable formats such as CellML, SBML, and SED-ML. Collaborative platforms that support federated modeling and secure simulation validation will also be essential.

Looking ahead, we envision a translational roadmap where virtual cells achieve validated utility in drug screening within 3–5 years, multi-organ simulations become clinically embedded by 2035, and programmable virtual humans inform personalized medicine by 2050. However, realizing this vision depends on concurrent advances in data curation, regulatory clarity, interdisciplinary training, and public engagement. We must also maintain a critical stance toward technological hype, acknowledging that many modeling challenges—such as emergent behavior, rare-event prediction, and real-world generalizability—are still open research problems.

Ultimately, AI-powered virtual biology offers a transformative lens on life sciences. If carefully governed and rigorously validated, these technologies hold the potential to democratize discovery, accelerate therapeutics, and deepen our understanding of the human body in silico.

References

Naomi E Allen, Cathie Sudlow, Tim Peakman, and Rory Collins. Uk biobank: Current status and what it means for epidemiology. *Health Policy and Technology*, 4(3):169–182, 2015. doi: 10.1016/j.hlpt.2015.07.003.

Julia Amann et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1):310, 2020. doi: 10.1186/s12911-020-01332-6.

Christof Angermueller, Tanel Pärnamaa, Leopold Parts, and Oliver Stegle. Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 2016. doi: 10.15252/msb.20156651.

Elisabeth S Berner et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of ai algorithms in healthcare. *BMJ Health & Care Informatics*, 28(1):e100204, 2021. doi: 10.1136/bmjhci-2020-100204.

Thierry Breton. Ai regulation and the risk of missed opportunity. *Nature Electronics*, 4(7): 430–431, 2021. doi: 10.1038/s41928-021-00611-x.

Kyle R Campbell et al. Multimodal single-cell data integration with totalvi. *Nature Biotechnology*, 39:1229–1236, 2021. doi: 10.1038/s41587-021-01050-3.

Junyue et al. Cao. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566:496–502, 2020. doi: 10.1038/s41586-019-0969-x.

Sophia et al. Chen. scgen predicts single-cell perturbation responses. *Nature Methods*, 17: 715–721, 2020. doi: 10.1038/s41592-020-0901-2.

Janusz Chmielecki, Anna Fornasiero, and Marija Poljak. Hybrid ai models in computational biology. *Nature Computational Science*, 3(5):310–321, 2023. doi: 10.1038/s43588-023-00418-y.

Sayantan Chowdhury, Qing Yan, and Chong Liu. Artificial intelligence in the age of synthetic biology: Applications and challenges. *Trends in Biotechnology*, 2022. doi: 10.1016/j.tibtech.2022.05.006.

European Commission. Medical device regulation (eu) 2017/745. `https://eur-lex.europa.eu/eli/reg/2017/745/oj`, 2021.

European Commission. Proposal for a regulation laying down harmonised rules on artificial intelligence (artificial intelligence act). `https://artificialintelligenceact.eu`, 2023.

Megan Dahlquist, Ajay Patel, and James A Byrne. Ai-assisted mechanistic modeling of hepatic drug metabolism and toxicity. *Bioinformatics*, 38(22):4903–4912, 2022. doi: 10.1093/bioinformatics/btac590.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. In *Foundations and Trends in Theoretical Computer Science*, volume 9, pp. 211–407, 2014. doi: 10.1561/0400000042.

Maha Elbadawi, Ye Zhu, and Michael P Menden. Livernet: Deep learning for liver zonation prediction and toxicity modeling. *Toxicological Sciences*, 184(1):121–134, 2021. doi: 10.1093/toxsci/kfab111.

FDA. Good machine learning practice for medical device development: Guiding principles. `https://www.fda.gov/media/153486/download`, 2021a.

FDA. Artificial intelligence/machine learning (ai/ml)-based software as a medical device (samd) action plan, 2021b. Available at `https://www.fda.gov/media/145022/download`.

Michael Fralick et al. Evaluation of using real-world evidence and synthetic control arms in clinical trials. *BMJ Evidence-Based Medicine*, 25(4):135–139, 2020. doi: 10.1136/bmjebm-2019-111205.

Tim Fritz, Francisco Sahli Costabal, and Ellen Kuhl. Graph-based surrogates for patient-specific cardiac simulations. *Computers in Biology and Medicine*, 145:105428, 2022. doi: 10.1016/j.compbiomed.2022.105428.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *ICML*, 2016. doi: 10.48550/arXiv.1506.02142.

Timnit Gebru et al. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. doi: 10.1145/3458723.

Ayman Ghallab, Reham Hassan, and Jan G. Hengstler. Predicting drug-induced liver injury using hybrid virtual hepatocyte models. *Frontiers in Pharmacology*, 13, 2022. doi: 10.3389/fphar.2022.843181.

Marzyeh Ghassemi et al. Causal inference across latent variables in healthcare. *Nature Biomedical Engineering*, 5:1163–1172, 2021. doi: 10.1038/s41551-021-00791-y.

Ahmad Ghoniem, Matthias Gsell, Alexander Neic, and Gernot Plank. Model exchange and co-simulation of cardiac electrophysiology using the functional mock-up interface. *Frontiers in Physiology*, 6:508, 2015. doi: 10.3389/fphys.2015.00508.

Shamik Ghosh et al. Quantifying uncertainty in multiscale cardiac models. *Computer Methods in Applied Mechanics and Engineering*, 391:114535, 2022. doi: 10.1016/j.cma.2021.114535.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

Casey S. Greene, Lana X. Garmire, and Jaqueline A. Gilbert. Transforming medicine with machine learning. *Nature Reviews Genetics*, 20(7):442–456, 2019. doi: 10.1038/s41576-019-0124-6.

Andreas Holzinger et al. Measuring the quality of explanations: the system causability scale (scs). *KI - Künstliche Intelligenz*, 34:193–198, 2020. doi: 10.1007/s13218-020-00636-z.

Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, and et al. The systems biology markup language (sbml): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003. doi: 10.1093/bioinformatics/btg015.

Peter Hunter and Thomas Borg. The iups human physiome project. *Journal of Physiology*, 552:1–3, 2003. doi: 10.1113/jphysiol.2003.049841.

ISO. Iso 62304:2006: Medical device software — software life cycle processes, 2006.

ISO. Iso 14971:2019: Medical devices — application of risk management to medical devices, 2019.

Jose Jimenez-Luna, Fabio Grisoni, and Gisbert Schneider. Artificial intelligence in biology and medicine: Promises and challenges. *Trends in Pharmacological Sciences*, 43(4):276–291, 2022. doi: 10.1016/j.tips.2022.01.007.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016. doi: 10.1038/sdata.2016.35.

Georgios A Kaissis et al. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 3(6):473–484, 2021. doi: 10.1038/s42256-020-0186-1.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Anuj Karpatne et al. Theory-guided data science: A new paradigm for scientific discovery. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 2017. doi: 10.1109/TKDE.2017.2720168.

Jonathan R. Karr, Jayodita C. Sanghvi, Derek N. Macklin, Miriam V. Gutschow, Jon M. Jacobs, Brian Jr. Bolival, Nacyra Assad-Garcia, John I. Glass, and Markus W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012. doi: 10.1016/j.cell.2012.05.044.

Daniel S Kermany et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131, 2018. doi: 10.1016/j.cell.2018.02.010.

Steffen et al. Klamt. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 10:356, 2009. doi: 10.1186/1471-2105-10-356.

Philippe Lambin, Hugo JWL Aerts, and Ralph TH Leijenaar. Clinical applications of ai models in virtual organs: From theory to practice. *The Lancet Digital Health*, 3(4): e190–e198, 2021. doi: 10.1016/S2589-7500(21)00025-0.

Xuefeng Liu, Samantha Cruz Rivera, David Moher, Melanie J Calvert, and Alastair K Denniston. Guidelines for clinical trial protocol and reporting in ai medical interventions: Tripod-ai and consort-ai. *BMJ*, 370:m3164, 2020. doi: 10.1136/bmj.m3164.

Tabea Lohoff, Shila Ghazanfar, and Anna Missarova. Simultaneous measurement of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology*, 40:759–768, 2022. doi: 10.1038/s41587-021-01191-0.

Romain Lopez, Jeffrey Regier, Maxwell Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, 2018. doi: 10.1038/s41592-018-0229-2.

Yu Lu, Rui Zhang, and Ziqiang Wang. Generative simulation with cross-modal transformers for biomedical modeling. *Nature Machine Intelligence*, 5(3):227–241, 2023. doi: 10.1038/s42256-023-00609-w.

Malte D. et al. Luecken. Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, 19:41–50, 2022. doi: 10.1038/s41592-021-01336-8.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017. URL `https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html`.

Jun Ma, Jinyuan Zhou, Lu Wang, and Jing Sun. Deep learning for cellular image analysis. *Nature Cell Biology*, 20:1340–1350, 2018. doi: 10.1038/s41556-018-0221-5.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220–229, 2019. doi: 10.1145/3287560.3287596.

Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2): 2053951716679679, 2016. doi: 10.1177/2053951716679679.

Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE Symposium on Security and Privacy (SP)*, pp. 19–38. IEEE, 2017. doi: 10.1109/SP.2017.12.

Christoph Molnar. *Interpretable Machine Learning*. Leanpub, 2022. URL `https://christophm.github.io/interpretable-ml-book/`.

Urs Muehlematter et al. Evaluating synthetic control arms in clinical trials. *Nature Reviews Clinical Oncology*, 18:475–476, 2021. doi: 10.1038/s41571-021-00499-y.

Steven A. Niederer et al. Multi-physics simulation of cardiac function: From cell to organ. *Annual Review of Biomedical Engineering*, 21:243–267, 2019. doi: 10.1146/annurev-bioeng-062117-121042.

Denis Noble et al. The cardiac electrophysiology web lab. *Philosophical Transactions of the Royal Society A*, 370(1965):2965–2975, 2012. doi: 10.1098/rsta.2011.0576.

Thomas M. et al. Norman. Exploring transcriptional programs in high-throughput crispr screens with scrna-seq. *Cell*, 175:377–389, 2019. doi: 10.1016/j.cell.2018.09.056.

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447–453, 2019. doi: 10.1126/science.aax2342.

Vishal L Patel and Edward H Shortliffe. Regulatory frameworks for clinical decision support systems and predictive analytics. *Journal of the American Medical Informatics Association*, 27(5):759–763, 2020. doi: 10.1093/jamia/ocaa003.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge university press, 2009.

Christopher Rackauckas, Yingbo Ma, Jakob Martensen, Carlo Warner, Kirill Zubov, Rohit Supekar, David Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020. doi: 10.48550/arXiv.2001.04385.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

Alvin Rajkomar, Jeff Dean, and Isaac Kohane. Machine learning in medicine. *New England Journal of Medicine*, 380:1347–1358, 2019. doi: 10.1056/NEJMra1814259.

Alvin Rajkomar et al. Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12):866–872, 2018. doi: 10.7326/M18-1990.

Aviv et al. Regev. The human cell atlas. *eLife*, 6:e27041, 2017. doi: 10.7554/eLife.27041.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. doi: 10.1145/2939672.2939778.

Nicola Rieke, Jonathan Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, MJ Cardoso, Antonio Criminisi, Qi Dou, Ben Glocker, et al. The future of digital health with federated learning. *NPJ Digital Medicine*, 3(1):1–7, 2020. doi: 10.1038/s41746-020-00323-1.

Samantha Cruz Rivera, Xuefeng Liu, Ana-Wu Chan, and Melanie J Calvert. Guidelines for reporting machine learning predictive models in biomedical research: Minimar. *BMJ Open*, 11(4):e041721, 2021. doi: 10.1136/bmjopen-2020-041721.

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019. doi: 10.1038/s42256-019-0048-x.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Inferring causation from time series in earth system sciences. *Nature Communications*, 10 (1):2553, 2019. doi: 10.1038/s41467-019-10105-3.

Francisco Sahli Costabal, Jiannan Yao, Daniel E Hurtado, and Ellen Kuhl. Digital twins in cardiovascular medicine. *npj Digital Medicine*, 4:1–3, 2021. doi: 10.1038/s41746-021-00447-y.

Bernhard Schölkopf. Causality for machine learning. *The MIT Press*, 2022. doi: 10.7551/mitpress/14230.001.0001.

Nina Schwalbe and Brian Wahl. Emerging digital health technologies and the ethics of data donation. *Nature Medicine*, 26(4):469–471, 2020. doi: 10.1038/s41591-020-0843-6.

Bernhard Schölkopf. Toward causal representation learning. *Proceedings of the IEEE*, 109 (5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954.

Qingsong Sun, Wenrui Zhao, Nicolas Smith, and Alfio Quarteroni. Physml: Scalable physics-informed machine learning for multiscale cardiac simulations. *IEEE Transactions on Biomedical Engineering*, 2023. doi: 10.1109/TBME.2023.3245678.

Eric Topol. Deep medicine: How artificial intelligence can make healthcare human again. *Basic Books*, 2019.

UK Biobank. Uk biobank: An open-access resource for identifying the causes of complex diseases, 2022. Available at `https://www.ukbiobank.ac.uk`.

Marco Viceconti, Peter Hunter, and Rod Hose. The digital twin of the human body: A roadmap for virtual physiological human projects. *IEEE Journal of Biomedical and Health Informatics*, 25(9):3285–3299, 2021. doi: 10.1109/JBHI.2021.3092822.

Dagmar Waltemath, Rebekka Adams, Frank T. Bergmann, Michael Hucka, Fedor Kolpakov, Alison K. Miller, Ion I. Moraru, David Nickerson, Sven Sahle, Jacky L. Snoep, and Nicolas Le Novère. Reproducible computational biology experiments with sed-ml - the simulation experiment description markup language. *BMC Systems Biology*, 5(1), 2016. doi: 10.1186/1752-0509-5-198.

Lin Wang et al. Physics-informed machine learning for cardiac electrophysiology and arrhythmia modeling. *Nature Biomedical Engineering*, 5:548–556, 2021. doi: 10.1038/s41551-021-00691-1.

Trevor A. Williams et al. Translating single-cell gene expression analyses to a multiscale modeling pipeline for drug discovery. *Nature Communications*, 12(1):1–12, 2021. doi: 10.1038/s41467-021-21343-z.

Jinming Wu et al. Quantifying uncertainty in deep learning for cardiac image segmentation. *Medical Image Analysis*, 68:101911, 2021. doi: 10.1016/j.media.2020.101911.

Jindong Xu et al. Federated learning for healthcare informatics. *IEEE Intelligent Systems*, 36(4):64–71, 2021. doi: 10.1109/MIS.2020.2987600.

Yunfan Xu, Soham Roy, and Hyowon Lee. Data assimilation in biomedical systems: A hybrid modeling frontier. *Annual Review of Biomedical Engineering*, 24:309–335, 2022. doi: 10.1146/annurev-bioeng-110921-061315.

Fan Yang, Jie Liu, and Wei et al. Lin. A transformer-based model for predicting cell fate from gene expression data. *Nature Machine Intelligence*, 2022. doi: 10.1038/s42256-022-00539-w.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, 2018. URL `https://openreview.net/forum?id=ByKWUeWA-`.

Chen Yu, Raj Singh, Dongjun Lee, and Bin Zhao. Towards interpretable and trustworthy ai for liver toxicity prediction. *npj Digital Medicine*, 4(1):1–12, 2021. doi: 10.1038/s41746-021-00431-6.

Yujing Yuan and Liang Cai. Imaging transcriptomics for spatially resolved transcriptome-wide profiling. *Nature Reviews Genetics*, 23:681–698, 2022. doi: 10.1038/s41576-022-00489-2.

Jiahui Zhang, Edward Black, and Xuesong Yang. Uncertainty quantification in deep learning for biomedicine. *Nature Biomedical Engineering*, 7(1):44–58, 2023. doi: 10.1038/s41551-022-00956-9.