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

Anonymous authors
Paper under double-blind review

Abstract

Recent advances in artificial intelligence (AI), high-performance computing, and systems biology have accelerated the development of AI-powered virtual biological systems, from virtual cells to multiscale organ models and programmable virtual humans. These systems promise transformative applications in drug discovery, precision medicine, and in silico clinical trials. This review provides a critical synthesis of current progress, key technologies, and future directions across this spectrum. We explore hybrid modeling strategies that combine mechanistic models—such as ordinary and partial differential equations—with deep learning methods including convolutional, recurrent, and graph neural networks. We emphasize the importance of robust uncertainty quantification, simulation validation, and multiscale integration across molecular, cellular, organ-level, and systemic processes. A core contribution is the introduction of the SIM-CARD framework, a standardized simulation accountability protocol to document data provenance, modeling assumptions, performance metrics, and regulatory alignment. We propose a three-phase translational roadmap: (1) validated AI-augmented virtual cells and organs (by 2030), (2) interoperable multi-organ physiological systems (by 2040), and (3) programmable full-body virtual humans supporting personalized simulations and regulatory use cases (by 2055). We identify key enablers—including high-fidelity multiscale data, computational scalability, and simulation governance—as well as bottlenecks such as algorithmic bias, explainability, and regulatory uncertainty. Finally, we call for collaborative efforts to establish minimal benchmarking suites, FAIR-compliant simulation metadata, and cross-institutional federated learning infrastructure. This review aims to guide the scientific, regulatory, and clinical communities in navigating the complex yet promising trajectory toward clinically actionable programmable human simulations.

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

1 Introduction

The intersection of artificial intelligence (AI) and computational biology has ushered in a transformative era in biomedical research, enabling the simulation and understanding of biological systems with unprecedented fidelity. As the field of systems biology has matured, researchers have increasingly sought to represent the complexity of cellular, tissue, and organ-level dynamics using advanced computational models. Traditional mechanistic approaches, while grounded in physical laws, often lack scalability and flexibility when faced with high-dimensional, nonlinear, and noisy biological data. In contrast, AI-driven models—ranging from deep neural networks to hybrid mechanistic-learning systems—offer scalable tools to bridge the gap between raw data and predictive biological insights Rajkomar et al. (2019); Brunton & Kutz (2019).

This review provides a comprehensive roadmap for understanding the landscape of AI-powered virtual biology, focusing on modeling frameworks that span from virtual cells to programmable virtual humans. We examine mathematical foundations and algorithmic strategies such as ordinary differential equations (ODEs), partial differential equations (PDEs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), and the increasingly prominent universal differential equations (UDEs) Chen et al. (2018); Rackauckas (2020). In particular, we elaborate on how these techniques enable multiscale simulations by integrating experimental data with mechanistic understanding—crucial for applications in personalized medicine, drug development, and clinical decision support Camacho et al. (2018); Eriksson & Wallerstedt (2020).

Moreover, computational infrastructure considerations are central to the practical deployment of virtual simulations. High-performance computing (HPC), including GPU clusters and tensor processing units (TPUs), remains essential for training large-scale models and conducting inference across population-level datasets. For example, training a physics-informed neural network (PINN) on cardiac electrophysiology data can require weeks of GPU time, depending on model complexity and spatial resolution Raissi et al. (2019); Blanchard & Willcox (2021). The memory footprint and runtime for simulating whole-organ systems further emphasize the need for efficient numerical solvers, parallelization strategies, and cloud-based architectures Jones et al. (2022).

In addition to technical considerations, ethical and regulatory frameworks are critical for the safe and equitable deployment of virtual human simulations. The use of patient-specific data for model personalization raises concerns about algorithmic bias, data sovereignty, and clinical accountability Morley et al. (2020). In response, we propose a structured governance framework termed SIM-CARD (Simulation Card for Accountability, Reproducibility, and Disclosure), which is designed to serve as a metadata-driven documentation and validation toolkit. This framework complements emerging guidelines such as TRIPOD-AI and CONSORT-AI by providing model-specific simulation assumptions, drift monitoring criteria, and safety update policies Liu et al. (2020); Faes et al. (2020).

This manuscript is structured to provide both breadth and depth. In Section 2, we discuss foundational mathematical and computational methods for biological simulation. Section 3 examines AI-powered virtual cells, while Section 4 explores virtual organs and their integration. Section 5 offers a revised translational roadmap with explicit readiness gates. Sections 6 through 9 discuss governance, benchmarking, and ethical dimensions, while Section 10 concludes with a forward-looking agenda. Together, this review synthesizes the technological, methodological, and regulatory advances necessary to transition from theoretical models to real-world clinical impact through virtual biology.

2 Mathematical and Computational Foundations

Mathematical modeling forms the backbone of virtual biological systems, enabling simulation of cellular and physiological behavior across spatial and temporal scales. This section reviews core modeling paradigms—ranging from mechanistic formulations like ordinary and partial differential equations (ODEs, PDEs), to data-driven and hybrid learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), and universal differential equations (UDEs). We emphasize not only their mathematical underpinnings, but also computational demands, deployment considerations, and relevance to biological phenomena.

2.1 Mechanistic Modeling Approaches

Ordinary Differential Equations (ODEs) have long been used to describe the temporal evolution of biological systems. A general ODE model of a cell signaling pathway can be expressed as:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \theta) \tag{1}$$

where $\mathbf{x} \in R^n$ is the state vector representing molecular species (e.g., mRNA, protein concentrations), and θ denotes kinetic parameters such as reaction rates. Systems of ODEs can model transcriptional regulation, metabolic flux, and pharmacokinetics (PK) Klipp et al. (2016).

Partial Differential Equations (PDEs) capture both spatial and temporal dynamics, essential for simulating tissue-level transport and diffusion processes. For example, modeling oxygen transport in tissues can be represented as:

$$\frac{\partial c}{\partial t} = D\nabla^2 c - \gamma c \tag{2}$$

where $c(x,t)$ is oxygen concentration, D is the diffusion coefficient, and γ models cellular uptake Keener & Sneyd (1998). PDEs are critical in virtual organ modeling (e.g., liver zonation, cardiac electrophysiology), where spatial heterogeneity cannot be ignored.

Finite Element and Spectral Methods are commonly used to discretize PDEs, with trade-offs between computational cost and accuracy. Finite element methods (FEM) are particularly useful for irregular geometries such as organ meshes derived from CT/MRI Hughes (2012).

2.2 Machine Learning Approaches

Convolutional Neural Networks (CNNs) are suited for spatially structured data like histopathology slides or spatial transcriptomics. A CNN layer transforms input features via:

$$\mathbf{y}_{i,j,k} = \sigma\left(\sum_{m,n,p} W_{m,n,p,k} \cdot \mathbf{x}_{i+m,j+n,p} + b_k\right) \tag{3}$$

where W denotes learned filters, σ an activation function (e.g., ReLU), and b_k a bias term. CNNs have been applied to 2D organoids, immunofluorescence images, and 3D voxel representations of tissues Esteva et al. (2021).

Recurrent Neural Networks (RNNs) and **Long Short-Term Memory (LSTM)** architectures are ideal for time-series data, such as dynamic cell responses to drug perturbations. These models capture dependencies across time steps:

$$\mathbf{h}_t = \phi(W_h \cdot \mathbf{h}_{t-1} + W_x \cdot \mathbf{x}_t + \mathbf{b}) \tag{4}$$

RNNs are increasingly used to model gene expression trajectories, electrophysiology waveforms, and cytokine response curves Chandak et al. (2022).

Graph Neural Networks (GNNs) model biological systems as networks of interactions. Given a graph $G = (V, E)$, where nodes represent proteins or cells and edges capture interactions, a GNN updates node features via message passing:

$$\mathbf{h}_v^{(k+1)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} M(\mathbf{h}_v^{(k)}, \mathbf{h}_u^{(k)}, e_{uv})\right) \tag{5}$$

GNNs excel in protein-protein interaction networks, cell–cell communication, and tissue graph representations Zitnik et al. (2018).

2.3 Hybrid Modeling: Universal Differential Equations

Hybrid models embed neural networks into ODEs/PDEs, forming *Universal Differential Equations (UDEs)* Rackauckas (2020). A canonical form is:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \theta) + \mathrm{NN}_\phi(\mathbf{x}, t) \tag{6}$$

where NN_ϕ learns unknown dynamics not captured by mechanistic components. UDEs are particularly useful when partial knowledge exists (e.g., metabolic constraints), but data can refine unknown processes (e.g., transport coefficients, feedback loops).

2.4 Computational Infrastructure Requirements

Training these models at scale requires substantial computational resources. For instance, training a GNN on single-cell atlases with millions of cells and thousands of features may require multiple GPUs with at least 48GB memory each, distributed via data-parallel processing. For PDE solvers integrated into PINNs, solving cardiac electrophysiology equations over a 3D domain can require 1–2 weeks of runtime on 8 A100 GPUs Sun et al. (2022). Memory demands for 3D liver zonation simulations exceed 256 GB RAM due to fine spatial meshing.

Furthermore, multi-organ co-simulations using SBML/FMI standards typically rely on HPC clusters or cloud-based architectures such as Amazon EC2 or Google Cloud TPU pods Rios et al. (2022). Parallelization strategies such as MPI, CUDA kernels, or OpenMP are essential for solving coupled ODE-PDE-GNN hybrid systems in reasonable time frames.

In summary, a thorough understanding of mathematical models, numerical solvers, and deep learning paradigms is essential for building scalable, interpretable, and biologically grounded virtual systems.

3 Single-Cell Modeling Approaches and Virtual Cell Architectures

The modeling of biological phenomena at the single-cell level lies at the foundation of virtual biology. Single-cell modeling facilitates the simulation of intracellular dynamics, cell signaling pathways, gene regulation, and interactions with microenvironments. Virtual cells—computational replicas of biological cells—can be deployed to investigate cellular heterogeneity, perturbation responses, and molecular interventions, with growing applications in drug discovery and systems biology.

3.1 Mathematical Formulations for Virtual Cell Dynamics

Mathematical modeling at the single-cell level traditionally relies on systems of ordinary differential equations (ODEs) to describe temporal changes in molecular species. For example, transcriptional regulatory networks are often modeled using ODEs of the form:

$$\frac{dX_i}{dt} = f_i(X_1, X_2, \ldots, X_n; \theta), \tag{7}$$

where X_i denotes the concentration of molecular species i, and f_i captures production, degradation, and interaction terms parameterized by θ Karlebach & Shamir (2008). For spatial dynamics, partial differential equations (PDEs) model diffusion and active transport, typically expressed as:

$$\frac{\partial C}{\partial t} = D\nabla^2 C + R(C), \tag{8}$$

where C is the concentration of a molecule, D is the diffusion coefficient, and $R(C)$ represents reaction terms Holcman & Yuste (2015).

Stochastic effects, significant in low-copy-number regimes such as gene expression, are captured via stochastic simulation algorithms (SSA) like Gillespie's algorithm Gillespie (1977).

These models are often integrated with single-cell omics data for calibration and validation, enabling patient-specific modeling.

3.2 Deep Learning for Cellular Phenotyping and Simulation

Recent advances in machine learning have transformed virtual cell modeling. Convolutional neural networks (CNNs) are applied to single-cell imaging data for subcellular feature extraction and phenotyping, while recurrent neural networks (RNNs) and transformers model cellular trajectories from time-series transcriptomics data Zhang et al. (2019); Lopez et al. (2018). These methods have enabled predictive modeling of cell fate, lineage differentiation, and response to genetic perturbations.

Moreover, graph neural networks (GNNs) represent molecular interaction networks and cell–cell communication, offering interpretable models of cellular behavior in tissues Wang et al. (2021). For instance, Cell-GNN has been applied to spatial transcriptomics to infer cellular states in complex microenvironments.

3.3 Hybrid and Universal Modeling Approaches

Hybrid models, such as universal differential equations (UDEs), embed neural networks within mechanistic ODE systems. This allows the model to capture both known biological pathways and learn unknown dynamics from data. A UDE model may be expressed as:

$$\frac{dX}{dt} = f(X) + \mathrm{NN}(X, \phi), \tag{9}$$

where $f(X)$ is a mechanistic component and $\mathrm{NN}(X, \phi)$ is a neural network with parameters ϕ Rackauckas et al. (2020). These models are especially powerful in capturing cell signaling cascades with incomplete kinetic data.

Agent-based models (ABMs) offer another complementary approach, simulating individual cell behaviors based on rule sets or learned policies. Integration of ABMs with deep reinforcement learning enables the emergence of complex multicellular dynamics, such as tumor invasion or immune cell coordination Hamilton & Reed (2018).

3.4 Simulation Infrastructure and Data Challenges

Simulating large populations of virtual cells necessitates scalable computational frameworks. GPU-based acceleration (e.g., using PyTorch or JAX), cloud-native deployment, and support for simulation-as-a-service (SaaS) models are gaining traction. Frameworks like BioNetGen Faeder et al. (2009) and Smoldyn Andrews & Bray (2010) provide rule-based and spatial stochastic simulation engines.

Data integration remains a key challenge. Single-cell RNA-seq (scRNA-seq), ATAC-seq, and proteomics data must be harmonized and aligned. Techniques such as manifold alignment, variational autoencoders (e.g., scVI Lopez et al. (2018)), and optimal transport are being leveraged to address batch effects and modality differences. Despite these advances, issues like dropout events, sequencing noise, and annotation inconsistency continue to affect simulation fidelity.

3.5 Application Scenarios and Validation

Virtual cell models are increasingly applied in personalized medicine, such as predicting drug response in cancer treatment Zhang et al. (2022b). For example, in silico knockouts of signaling nodes can forecast resistance mechanisms. In developmental biology, virtual stem cells simulate lineage commitment under various cytokine environments.

Model validation remains challenging. Strategies include in vitro perturbation experiments, lineage tracing, and benchmarking against known gold-standard networks. Metrics such

as root mean square error (RMSE), area under the ROC curve (AUROC), and calibration curves are commonly used for performance assessment Lauter et al. (2022).

In summary, single-cell modeling has rapidly evolved from simple kinetic systems to data-driven, hybrid, and graph-based simulations. Continued progress requires innovations in data integration, interpretability, and scalable inference, ultimately supporting predictive virtual cell models deployable in research and clinical workflows.

4 Virtual Organs and Systems-Level Modeling

Virtual organs are computational representations of anatomical and physiological structures that capture organ-level behavior in health and disease. They serve as intermediate scales in the continuum from cellular processes to whole-body physiology, enabling the simulation of functional outcomes such as electrophysiological signals, metabolic fluxes, and biomechanical stress responses Hunter et al. (2003); Viceconti et al. (2016). Virtual organ modeling typically combines mechanistic principles with AI-based approximations to bridge complexity and address data sparsity.

4.1 Mechanistic and Data-Driven Synergies

Mechanistic modeling at the organ level commonly employs partial differential equations (PDEs) to capture spatially distributed processes, such as blood flow, drug perfusion, and electrical propagation in cardiac tissue Nash & Panfilov (2004). The Finite Element Method (FEM) and Finite Volume Method (FVM) are standard numerical schemes used to discretize these systems. However, purely mechanistic models face scalability issues and are sensitive to parameter estimation errors.

To address these limitations, hybrid frameworks have emerged. Physics-informed neural networks (PINNs) embed PDE constraints into deep learning architectures to enable data-efficient learning of complex physical systems Raissi et al. (2019). More recent approaches, such as operator learning (e.g., DeepONet and Fourier Neural Operators), bypass classical discretization altogether and learn mappings between function spaces, enabling rapid inference Li et al. (2021).

For example, in cardiac electrophysiology, PINNs have been used to estimate tissue conductivity fields by fitting simulated action potentials to sparse ECG data Fritz et al. (2022). Similarly, liver lobule perfusion has been modeled using hybrid UDEs (Universal Differential Equations), combining compartmental ODEs with neural networks that learn unknown terms from imaging-derived flow fields Rackauckas (2020).

4.2 Multiscale Integration and Systems-Level Modeling

Virtual organs do not function in isolation. Systems-level models aim to couple multiple organ models to simulate whole-body physiology, drug ADME (Absorption, Distribution, Metabolism, Excretion), or disease progression. These efforts face significant challenges in scale-bridging, time resolution mismatches, and data heterogeneity Eissing et al. (2011); Holzinger et al. (2022).

Frameworks such as Physiome, OpenCOR, and SimVascular provide modular environments for multi-organ co-simulation. Interoperability is supported by model exchange formats such as CellML, SBML, and FMI Garny et al. (2008); Hucka et al. (2003a). However, alignment across time scales, boundary conditions, and parameter uncertainty remains a major hurdle.

Advanced techniques such as co-simulation scheduling, surrogate modeling, and emulator-based approximations are used to manage computational load. For example, a multi-organ pharmacokinetic model may use high-fidelity PDE models for liver metabolism while approximating renal clearance via neural network surrogates trained on simulation outputs Zhang et al. (2022a).

4.3 Clinical Use Cases and Validation

Virtual organ models are beginning to impact clinical workflows. Cardiac digital twins, for example, are used to optimize pacemaker placement, simulate arrhythmia inducibility, and evaluate risk of sudden cardiac death Arevalo et al. (2016). Liver models simulate hepatic clearance and toxicity risk for new drugs, reducing the need for animal studies Ma et al. (2023).

Validation of organ models requires task-specific ground truth. For cardiac models, ECG recordings and catheterization data serve as references. For pulmonary ventilation models, imaging (e.g., CT, MRI) and spirometry are used. Metrics such as root mean squared error, Dice similarity for segmentation-based comparisons, and calibration curves are used for performance evaluation Noe & Silva (2021).

However, model generalizability across patients remains an open challenge. Federated learning and transfer learning are promising approaches to enable personalization while respecting data privacy constraints Sheller et al. (2020); Xu et al. (2022). These methods will be essential for enabling clinical-grade simulation platforms that scale across health systems.

In sum, virtual organ modeling represents a key frontier in computational biology. It is the nexus where deep learning, biomechanics, control theory, and clinical validation converge. Continued advances will require methodological innovation, robust software ecosystems, and close partnerships between modelers and clinicians.

5 Programmable Virtual Humans

Programmable virtual humans represent the culmination of decades of progress in systems biology, AI-driven modeling, and multiscale simulation. These are not static replicas of anatomy, but dynamic, parameterizable computational frameworks capable of simulating human physiology, pathology, and therapeutic responses at individual resolution Viceconti et al. (2021); Erlich et al. (2022). By enabling in silico experimentation, these systems can accelerate translational research, reduce reliance on animal models, and support regulatory decision-making for drugs and devices Bjornsson & et al. (2022).

The concept of "programmability" refers to the capacity of virtual humans to accept input parameters—such as genetic profiles, environmental exposures, and clinical covariates—and dynamically adjust internal states to simulate individualized outcomes. For example, in the context of oncology, virtual humans may simulate tumor growth kinetics under different chemotherapy regimens based on a patient's molecular tumor profile and pharmacogenomic markers Dash et al. (2021). This programmability is made possible by integrating multiscale models spanning molecular networks, tissue biomechanics, and systemic physiology.

5.1 Architectural Framework and Interoperability

A programmable virtual human architecture must be modular, scalable, and interoperable. Each organ system (e.g., cardiovascular, hepatic, renal, CNS) is modeled using specialized subcomponents linked through standardized APIs and data exchange protocols. Standards such as CellML, SBML, and FMI (Functional Mock-up Interface) support model portability and cross-platform execution Garny et al. (2008); Hucka et al. (2003a). Simulation coordination across modules requires temporal synchronization strategies and data buffering schemes to resolve scale mismatches and ensure numerical stability.

Semantic interoperability is equally critical. Ontology-based metadata (e.g., SNOMED CT, OBI, UBERON) must annotate all simulation inputs and outputs to ensure consistency in interpretation, especially for federated applications across institutions. This also facilitates mapping between virtual simulations and real-world clinical phenotypes or regulatory endpoints Herzog et al. (2022).

5.2 Use Cases and Milestones

Initial applications of programmable virtual humans are emerging in drug development and digital clinical trials. Notable examples include FDA-acknowledged in silico trials for cardiac safety assessment (e.g., CiPA—Comprehensive in vitro Proarrhythmia Assay) and pharmacokinetic modeling for pediatric populations Polak & et al. (2020); Sagartz & et al. (2023). Companies like Dassault Systèmes, Insilico Medicine, and Unlearn.AI are actively deploying virtual patient models to generate synthetic control arms, optimize trial enrollment, and simulate therapeutic response Walsh (2023).

The roadmap toward full-body programmable virtual humans can be divided into three major phases:

- **Phase I (2025–2030):** Maturity of organ-specific models (e.g., liver, heart, lung) validated against experimental and clinical datasets; integration of uncertainty quantification, data lineage, and regulatory readiness checklists (e.g., SIM-CARDs).
- **Phase II (2030–2040):** Integration of multiple organ systems with physiological coupling (e.g., gut-liver-brain axis); deployment of federated simulation infrastructure; partial personalization via patient-derived omics and EHR data.
- **Phase III (2040–2055):** Full-body programmable virtual humans capable of simulating multi-disease progression, treatment outcomes, and population variability; regulatory acceptance for select prescriptive uses (e.g., virtual trials, device design).

These milestones are contingent upon overcoming key barriers including high-fidelity model validation, multiscale data integration, regulatory harmonization, and massive computational demand (e.g., GPU-years, HPC clusters, memory efficiency) Viceconti & et al. (2021a).

5.3 Programmability Dimensions and Control Logic

Programmability entails more than parameter tuning. It requires establishing an *operational design domain* (ODD) within which safe and valid simulations can occur. Control logic, often built upon structural causal models (SCMs), supports counterfactual inference and prescriptive interventions, such as simulating "what-if" scenarios under alternative treatment plans Pearl (2009); Ma & et al. (2023).

Examples of programmability dimensions include:

- **Biological Variability:** Age, sex, genotype, microbiome.
- **Clinical Conditions:** Comorbidities, organ dysfunction, medication history.
- **Intervention Modeling:** Drug dosages, implantable devices, surgical plans.

Closed-loop control systems, such as virtual glucose-insulin feedback models, exemplify how programmable logic can simulate real-time physiological regulation and adaptive therapy strategies Bequette (2005).

5.4 Challenges and Future Outlook

Despite progress, several challenges persist. Modeling rare phenotypes or adverse events requires massive data augmentation or generative techniques such as diffusion models and GANs Ramesh et al. (2022). Capturing emergent behaviors and feedback loops at system-wide levels remains computationally intensive and sensitive to stochastic noise.

Moreover, there is growing recognition that model transparency, explainability, and calibration must be audited continuously. This has led to calls for simulation audit trails, SIM-CARD templates, and lifecycle governance frameworks that align with ISO, FDA, and EMA regulatory standards Berners & et al. (2023); Viceconti & et al. (2021b).

Programmable virtual humans offer a transformational opportunity to move from empirical to mechanistic medicine, from trial-and-error to simulation-guided decision-making. Achiev-

ing this vision will require collaboration across computational biology, regulatory science, AI, ethics, and clinical practice.

6 Projected Timeline and Translational Roadmap

Building an AI-powered continuum from virtual cells to programmable human simulations demands a robust translational roadmap that balances scientific ambition with implementation realism. We outline a phased trajectory composed of three major stages—Virtual Cells (2025–2030), Virtual Organs (2030–2040), and Programmable Virtual Humans (2040–2055)—and introduce a critical path framework to systematically assess feasibility, risks, and dependencies.

Phase I (2025–2030): Virtual Cells

The initial phase focuses on establishing predictive, mechanistically grounded models of cellular behavior, including subcellular dynamics, signaling pathways, and drug responses. This requires integrating high-resolution single-cell data (e.g., scRNA-seq, proteomics, metabolomics) with hybrid modeling techniques that blend mechanistic ordinary differential equations (ODEs) with deep learning methods such as neural ODEs (NODEs) and universal differential equations (UDEs) Rackauckas (2020); Chen et al. (2018). The emphasis will be on achieving robust calibration against experimental perturbation datasets, including CRISPR screens and high-throughput drug assays Replogle et al. (2022).

Critical milestones for this phase include:

- Creation of 10–20 organ-specific virtual cell types with validated behaviors across 3–5 perturbation contexts.
- Development of standardized data schemas and uncertainty reporting mechanisms for single-cell simulations.
- Initial release of SIM-CARDs for virtual cells, including assumptions, failure modes, and sensitivity analyses.

Risks include overfitting to limited perturbation data, instability in long-range predictions, and hardware bottlenecks. Mitigation will require robust cross-validation, domain adaptation, and collaboration with cloud computing platforms (e.g., AWS HealthLake, DNAnexus).

Phase II (2030–2040): Virtual Organs

This phase aims to scale cellular simulations to mesoscopic and macroscopic tissue levels by combining agent-based models, partial differential equations (PDEs), and graph neural networks (GNNs) to simulate spatially extended organ behavior Karniadakis et al. (2021). Integration of electrophysiology (e.g., cardiac, neural), hemodynamics, and multi-scale biomechanical dynamics is required for clinically meaningful organ simulation.

Milestones include:

- Simulations of cardiac arrhythmia propagation, liver fibrosis progression, or pulmonary ventilation under disease and therapeutic perturbations.
- External validation using data from biobanks (e.g., UK Biobank), physiome models (e.g., OpenSim Delp et al. (2007)), and large-scale hospital EHR data (e.g., MIMIC-IV).
- Formal release of SIM-CARD templates for organs, including post-market surveillance KPIs (e.g., drift ¿ 5% in AUROC or calibration slope deviation ¿ 0.05).

Key risks include geometric meshing errors, numerical instability from complex boundary conditions, and non-identifiability of model parameters. Countermeasures include adaptive mesh refinement, ensemble modeling, and embedding mechanistic priors in machine learning components.

Phase III (2040–2055): Programmable Virtual Humans

The final phase envisions whole-body simulation environments, integrating multi-organ systems for longitudinal prediction of disease progression, drug interactions, and personalized interventions. Programmability implies both closed-loop control (e.g., insulin-glucose dynamics) and counterfactual inference through structural causal models (SCMs) Schölkopf (2021).

Specific milestones include:

- Deployment of virtual human platforms capable of simulating 6–12 months of clinical course for at least 3 major disease areas (e.g., cardiovascular, metabolic, neurodegenerative).
- Inclusion of multi-omics, imaging, lifestyle, and wearable sensor data in personalizing simulations.
- Integration of SIM-CARDs with regulatory frameworks like FDA's GMLP, ISO 14971, and the EU AI Act.

Gating risks encompass regulatory acceptance, ethical concerns (e.g., identity reidentification), and simulation drift under real-world complexity. Feasibility hinges on federated learning infrastructure, population-scale training data, and robust validation frameworks.

Critical Path and Contingency Planning

Table 1 outlines the critical path dependencies and projected risk mitigations across phases. Gating milestones—such as validated organ-level simulations and deployment of SIM-CARD governance templates—must be passed for downstream translation. Personnel needs scale from academic lab-sized efforts (Phase I, ∼5–10 FTEs) to consortia-level (Phase III, ∼100+ FTEs), with compute costs estimated at 0.5–1 GPU-year per organ simulation and up to 10 PF-days for whole-body inference.

Table 1: Phased Roadmap: Milestones, Risks, and Dependencies

Phase	Timeline	Key Milestones	Risks
Virtual Cells	2025–2030	Validated UDE-based cell models	Perturbation sparsity, hardware costs
Virtual Organs	2030–2040	PDE + GNN multi-scale models	Geometric meshing, numerical drift
Virtual Humans	2040–2055	Programmable control + SCMs	Regulatory acceptance, drift

Benchmarking Suite and SIM-CARD Releases

Each phase will release a benchmark suite with tasks, datasets, and reference implementations. Cell-level: gene regulatory inference; organ-level: hemodynamic response to drugs; human-level: disease progression modeling under intervention scenarios. All will follow a SIM-CARD documentation template with UQ and subgroup performance fields.

Conclusion of Roadmap

The proposed timeline provides a grounded, auditable framework with contingency planning for technical, computational, and ethical obstacles. Coordinated global efforts are required to realize these ambitions responsibly, with alignment to FAIR simulation governance principles.

7 Governance, Regulation, and Ethical Implications

The rapid evolution of AI-driven biological simulations necessitates a rigorous framework for governance, regulation, and ethical oversight. As these systems begin to influence preclinical decision-making, clinical workflows, and policy considerations, the need for formal

accountability structures becomes paramount. This section presents a practical governance architecture rooted in existing medical device regulatory pathways, expanded to accommodate simulation-specific uncertainties and lifecycle complexities.

We propose the Simulation Card for AI-Driven Research and Development (SIM-CARD) as a standardized documentation schema for simulation-based models. SIM-CARDs capture essential metadata including simulation scope, assumptions, data lineage, validation benchmarks, failure modes, uncertainty quantification (UQ), update policies, and regulatory status. Inspired by Model Cards, TRIPOD-AI, and FDA's Good Machine Learning Practice (GMLP) guidance, SIM-CARDs serve both regulatory and ethical purposes. They differ from conventional documentation by emphasizing multiscale assumptions, hierarchical UQ propagation, external generalizability, and drift monitoring across physiological domains.

Regulatory Integration. SIM-CARDs can be aligned with ISO 14971 (risk management), ISO 13485 (QMS), IEC 62304 (software lifecycle), and the FDA's SaMD framework. For simulation software used as clinical decision support (CDS), we recommend regulatory alignment with the proposed EU AI Act and U.S. FDA action plans on AI/ML-enabled software, especially concerning dynamic learning systems. Post-market surveillance of simulation models should track concept drift, calibration decay, and adverse decision correlations. We suggest adapting the U.S. FDA's Pre-Cert program for continuous simulation lifecycle auditing.

Simulation Drift and Lifecycle Management. Managing the lifecycle of simulation models demands structured drift detection and re-validation protocols. We recommend that calibration slope deviations exceeding 0.05 or Expected Calibration Error (ECE) shifts above 2% trigger lifecycle retraining, with corresponding updates to the SIM-CARD. Failure modes—including data drift, mechanistic parameter instabilities, and population subgroup shifts—must be logged and transparently addressed. We also advocate for formal "simulation bill-of-materials" listing external data, assumptions, surrogate components, and update frequencies.

Bias, Fairness, and Explainability. Algorithmic bias and fairness remain persistent challenges in AI-driven simulations, particularly when applied to underrepresented patient populations. We propose subgroup validation metrics such as calibration parity, equalized odds, and subgroup-specific AUROC differences to be included in SIM-CARDs. Explainability mechanisms—including Shapley value approximations and causal counterfactuals—should be deployed where interpretability is essential for regulatory transparency or clinical trust.

Ethical Guardrails. Virtual biological systems could potentially be misused in dual-use scenarios (e.g., modeling virulence enhancements). We recommend adopting biosecurity screening protocols and audit trails as part of model governance. Ethical implications of virtual patient simulations also involve questions of data sovereignty, particularly when federated simulations aggregate sensitive medical data. We advocate for secure multi-party computation (SMPC) and differential privacy (DP) thresholds (e.g., $\varepsilon < 5$ for clinical simulations) to protect individual-level inferences in cross-institutional workflows.

Crosswalk with Existing Standards. Table 2 provides a summary of how SIM-CARD fields extend or complement existing standards such as Model Cards, TRIPOD-AI, and GMLP. Novel elements include multi-scale model transparency, propagation of numerical uncertainty, and integration of validation gates.

Table 2: SIM-CARD elements and their relation to existing governance standards.

SIM-CARD Field	Extension of Existing Tools	Novel Contribution
Simulation Scope	TRIPOD-AI, Model Cards	Multiscale context declaration (cell → organ → human)
Assumptions	Model Cards, ISO 14971	Causal graph declaration, surrogate bounds
Validation Metrics	TRIPOD-AI, GMLP	Physiological gates, calibration/effectiveness thresholds
Uncertainty Reporting	STUMP, GMLP	Hierarchical UQ across hybrid components
Failure Modes	SaMD, ISO 62304	Drift thresholds, retraining policies
Data Lineage	Datasheets, GMLP	External/internal dataset mapping, simulation derivation
Update Policy	Model Cards	Audit triggers for real-world retraining

In sum, governing AI-driven biological simulations requires a robust ecosystem of documentation, surveillance, and ethical oversight. SIM-CARDs, when operationalized in coordina-

tion with international standards and national regulations, offer a transparent and scalable pathway for accountability. Further collaboration with regulatory bodies, professional societies, and clinical partners will be essential to evolve these proposals into widely adopted practice.

8 Validation, Benchmarks, and Performance Assessment

Robust validation is essential to ensure the reliability, reproducibility, and clinical relevance of AI-driven biological simulations. Given the diversity of use cases—from cellular simulations for drug response to virtual organs and whole-body simulations for clinical trial emulation—distinct benchmarks and validation protocols are needed for each modeling scale. This section outlines a structured framework for validating hybrid models, introduces benchmark tasks, and proposes quantitative gates for clinical translation readiness.

8.1 Validation Gates and Translational Thresholds

To standardize performance assessment across simulation stages, we propose multilevel validation gates that correspond to translational maturity. These gates are inspired by the TRIPOD-AI framework and FDA's SaMD GMLP principles and include:

- **Gate 1: Internal Consistency.** Evaluate internal simulation reproducibility via multiple runs with fixed seeds. Require convergence diagnostics for mechanistic and learned components.
- **Gate 2: External Validation.** Compare simulation outputs to independent datasets or withheld experimental measurements. Require a minimum AUROC of 0.80 and a calibration slope in the range [0.9–1.1].
- **Gate 3: Biological Plausibility.** Ensure that model trajectories respect known physiological ranges and conservation laws. Validate reaction kinetics, diffusion coefficients, and bioelectric consistency where applicable.
- **Gate 4: Clinical Utility.** Demonstrate that simulations lead to improved decision-making or actionable insights. Metrics may include net benefit curves, decision-curve analysis, and reduction in adverse events.

Quantitative thresholds may be adjusted by domain. For example, in cardiac modeling, acceptable beat-to-beat prediction error might be ¡5 ms for arrhythmia detection, whereas in liver DILI simulation, metabolite accumulation error may need to be ¡10% for regulatory acceptance.

8.2 Benchmarking Suites by Simulation Scale

We propose a minimal benchmark suite for each simulation level to standardize comparative evaluations and encourage model generalizability:

- **Cell Level.** Tasks: transcriptional response to perturbations, metabolic flux modeling. Datasets: Human Cell Atlas, LINCS L1000, Single Cell Expression Atlas. Baselines: kinetic ODEs, GNNs on PPI networks.
- **Organ Level.** Tasks: heart rhythm simulation, liver drug metabolism, lung gas exchange. Datasets: Physiome Model Repository, OpenSim kinematic datasets, PK/PD studies. Baselines: finite element solvers, hybrid CNN-RNN models.
- **Human Level.** Tasks: synthetic trial simulation, multi-organ interaction prediction. Datasets: MIMIC-IV, UK Biobank, Simulacrum. Baselines: population-based PK models, hybrid causal simulators.

These tasks should report standard metrics: AUROC, AUPRC, mean squared error, calibration intercept/slope, and ECE. Benchmarking protocols must include train-validation-test splits with temporal segregation for generalizability testing.

8.3 Uncertainty Quantification (UQ)

Quantifying and propagating uncertainty is critical for interpreting model outputs, especially in clinical contexts. We recommend the following UQ methods for simulation pipelines:

- **Mechanistic Models:** Use global sensitivity analysis (e.g., Sobol indices), parametric bootstrapping, and profile likelihoods.
- **Black-box Models:** Apply Bayesian deep learning (e.g., variational inference, MC dropout), ensemble methods, and conformal prediction.
- **Hybrid Integration:** Propagate uncertainties from mechanistic priors to learned residuals using hierarchical Bayesian models. Report confidence bands and prediction intervals at each scale.

For deployment, we propose that simulations meet the following UQ criteria: calibrated prediction intervals (e.g., 95% coverage), ECE ¡ 0.05, and sensitivity ranking stability under perturbations.

8.4 Fairness and Subgroup Auditing

To mitigate bias and ensure equitable performance across patient subgroups, simulation pipelines must include:

- **Subgroup Metrics:** Stratify AUROC, calibration slope, and false discovery rate by gender, race, age, and comorbidity.
- **Fairness Tests:** Apply equalized odds and calibration parity tests to ensure consistent performance across groups.
- **Auditing Reports:** Include subgroup performance metrics in SIM-CARDs and flag discrepancies exceeding 5% as requiring remediation.

These practices align with ethical AI guidelines and regulatory recommendations, especially for high-stakes clinical deployment.

8.5 Towards Open Benchmarks

We call for an open simulation benchmarking initiative, modeled after ImageNet or PhysioNet, to support comparative research. Each benchmark should include:

- Standardized datasets with licenses and synthetic augmentation options.
- Baseline model implementations in multiple frameworks (e.g., PyTorch, JAX).
- Public leaderboards evaluating accuracy, calibration, fairness, and uncertainty.

Establishing these benchmarks will foster community convergence and accelerate regulatory confidence in simulation-based technologies.

9 Governance, Regulation, and Societal Implications

As AI-driven simulations transition from research prototypes to clinical decision-support systems, governance and regulation must evolve to ensure safety, accountability, and public trust. This section outlines the societal implications, regulatory pathways, and governance tools required for responsible deployment of virtual biological systems across cell, organ, and human scales.

One foundational requirement is simulation accountability. Inspired by model cards Mitchell et al. (2019) and datasheets for datasets Gebru et al. (2021), we propose Simulation-Centered Accountability and Reporting Documents (SIM-CARDs) to standardize documentation across simulation lifecycle phases. These templates include fields such as simulation

purpose, input data lineage, mechanistic and machine learning model assumptions, validation protocols, subgroup performance audits, drift monitoring strategies, update schedules, and regulatory annotations. For instance, a SIM-CARD for a cardiac electrophysiology model may document its initial validation on MIMIC-IV arrhythmia data, assumptions of homogeneous tissue conductivity, expected accuracy thresholds (e.g., AUROC 0.85), and a quarterly re-validation requirement using real-world evidence.

To operationalize SIM-CARDs, we recommend machine-readable formats such as JSON or YAML. These formats should support auditability and interoperability with regulatory systems, clinical dashboards, and electronic health records (EHRs). Worked examples for a virtual liver model for drug-induced liver injury (DILI) and a cardiovascular risk simulation system are being prepared for public release, including populated assumptions, subgroup fairness metrics (e.g., equalized odds across race), calibration slope diagnostics, and update triggers based on drift indicators.

Regulatory pathways must adapt to simulation-specific concerns not fully covered by traditional Software as a Medical Device (SaMD) guidance. For instance, simulation models may evolve post-deployment due to new biological knowledge or retraining, necessitating alignment with software lifecycle standards such as IEC 62304, risk management under ISO 14971, and quality systems under ISO 13485. Furthermore, simulation models that serve as synthetic control arms or prescriptive virtual trials must be externally validated, transparent, and auditable. We recommend that simulated endpoints and outputs be reported in accordance with emerging guidelines such as TRIPOD-AI and CONSORT-AI Collins et al. (2022).

Fairness, transparency, and explainability are also critical. Complex hybrid models combining mechanistic and deep learning components may lack intuitive interpretability. Therefore, we advocate for multi-layered explanation systems, including mechanistic interpretability, feature attribution, and post-hoc simulation diagnostics. Regulatory frameworks should include thresholds for explainability loss—e.g., if a high-performing model is less explainable than a marginally less accurate model, then a safety-accuracy trade-off audit must be documented.

Equally important are data privacy, sovereignty, and governance for federated simulation workflows. Simulation models trained on distributed datasets across hospitals or countries should incorporate privacy-preserving technologies such as differential privacy Dwork (2008), secure multiparty computation (MPC), and federated learning Yang et al. (2019). SIM-CARDs must document privacy budgets (e.g., target $\epsilon < 1$), participation frequencies, and institutional review board (IRB) oversight protocols.

Societal acceptance also hinges on mitigation of misuse risks. AI-based virtual humans could potentially be exploited for unauthorized experimentation or even synthetic biology misuse. Therefore, simulation development environments should integrate guardrails, secure access controls, and audit trails. Governance bodies must define red lines for dual-use technologies and invest in horizon scanning to identify emerging threats.

Ultimately, responsible governance requires multi-stakeholder collaboration. Clinicians, patients, ethicists, regulators, and technologists must co-create guidance on acceptable use, validation criteria, and liability frameworks. Public engagement, transparency portals, and participatory audits will be essential to foster trust in virtual human simulations used for healthcare decision-making.

10 Future Directions and Research Challenges

The evolution of AI-powered biological simulations from virtual cells to programmable virtual humans opens transformative opportunities for biomedical research and personalized medicine. However, this ambitious vision requires addressing numerous unresolved scientific and engineering challenges across modeling, validation, integration, and governance. In this section, we highlight key future directions and research frontiers.

First, the integration of multi-omics data—including genomics, transcriptomics, proteomics, and metabolomics—into unified simulation platforms remains a critical frontier. Despite advances in multi-modal deep learning Gill et al. (2022), challenges persist in aligning heterogeneous data modalities with differing spatial, temporal, and noise characteristics. Future research should develop robust data fusion architectures capable of learning latent biological manifolds across scales. Additionally, ontology alignment and knowledge graph embedding can support biologically consistent representations of cell states and organ functions.

Second, structural causal modeling (SCM) and counterfactual inference should be embedded within simulation architectures to enhance interpretability and enable interventional reasoning. SCMs can be integrated with differentiable simulators to model disease progression and simulate virtual treatments Pearl (2009). For example, SCM-enhanced cardiac models may simulate the electrophysiological effects of pharmacological interventions. These efforts must be validated against longitudinal cohort data or randomized clinical trials to ensure biological and clinical validity.

Third, advancing the definition and operationalization of programmable virtual humans requires standardized intervention modeling frameworks. These include parametric tuning (e.g., for drug dosages), structural interventions (e.g., gene knockouts), and closed-loop control (e.g., insulin-glucose regulation). Future research should define the operational design domain (ODD) of virtual human simulations, including intended use cases, performance bounds, and acceptable uncertainties. Regulatory guidance must evolve to address these emerging prescriptive use cases.

Fourth, simulation interoperability across organs and software ecosystems is vital for scaling toward whole-body simulations. Community-endorsed standards such as SBML, CellML, FMI, and OMOP must be harmonized through metadata ontologies and schema mappings Hucka et al. (2003b); Cuellar et al. (2003). Interoperability profiles should define variable exchange formats, time step synchronization, and shared ontologies across submodels. Platforms like OpenCOR and BioUML can serve as testbeds for co-simulation integration Garny & Hunter (2009).

Fifth, lifecycle management and drift monitoring of deployed simulation models require ongoing research and tooling. Simulation updates triggered by software patches, retrained AI modules, or new clinical data must be version-controlled and auditable. We recommend adopting lifecycle documentation standards aligned with IEC 62304 and ISO 13485, integrated into the SIM-CARD accountability templates.

Finally, equitable access and sustainability are critical. AI-based virtual human simulations must be inclusive, avoiding overfitting to data from high-resource settings. Investment is needed in federated simulation infrastructure, domain adaptation techniques, and open benchmarks inclusive of underrepresented populations. Furthermore, cross-disciplinary training programs are essential to cultivate expertise in hybrid modeling, regulatory science, and computational ethics.

By addressing these research challenges, the community can move closer to a future where AI-driven simulations enhance biomedical discovery, accelerate translational research, and enable safer, more effective clinical interventions.

11 Conclusions and Future Outlook

AI-driven simulations of biological systems—from cells to organs to virtual humans—are rapidly emerging as transformative tools in biomedical research and healthcare. This paper has presented a comprehensive framework outlining the technical, regulatory, and ethical considerations required for the responsible development and deployment of such systems.

Our review emphasizes the hybrid integration of mechanistic models with machine learning, addressing the unique demands of biological complexity and scale. Across Sections 2 to 9, we have surveyed a variety of modeling approaches, ranging from ordinary and partial differential equations to graph neural networks and hybrid physics-informed neural networks. We discussed practical considerations for multiscale simulation, uncertainty quantification,

benchmarking, clinical translation, and governance through tools like SIM-CARDs. These elements are critical not only for advancing research but also for fostering public and regulatory trust in AI-derived models.

Looking forward, several key trajectories will define the field's maturation:

- **Computational scaling and reproducibility:** As models grow in complexity, demands on compute, storage, and reproducibility will intensify. Collaboration with cloud infrastructure providers and investment in open-source simulation engines will be vital to democratize access Peng et al. (2021); Samala et al. (2021).
- **Interdisciplinary and inter-institutional cooperation:** Effective integration of clinical, biological, and computational expertise is paramount. Institutions should invest in translational simulation centers where physicians, biologists, and computer scientists co-design models and workflows Zhang et al. (2022c).
- **Enhanced benchmarks and community standards:** To enable model comparison, reproducibility, and validation, standardized benchmark suites and validation protocols must be developed. These should cover simulation accuracy, calibration, robustness, fairness, and computational efficiency Ma et al. (2023).
- **Personalized and programmable virtual humans:** Future systems will need to be programmable by clinical end-users to simulate personalized interventions. This will require the development of explainable, interactive interfaces, and modular simulation architectures supporting control, counterfactual reasoning, and regulatory transparency Holzinger et al. (2022).
- **Governance and societal integration:** As these technologies move toward clinical and commercial use, transparent frameworks for model accountability, update cycles, and post-market surveillance will be essential. Policymakers must co-evolve regulatory standards (e.g., ISO 14971, IEC 62304) and ensure interoperability with digital health infrastructures Topol (2019).

Despite progress, key challenges remain: ensuring generalizability across populations, integrating multimodal and sparse datasets, and mitigating risks of misuse or algorithmic bias. Moreover, successful deployment requires public and clinician trust, which can only be earned through transparency, validation, and stakeholder engagement.

In conclusion, AI-driven virtual biology stands at a critical inflection point. By embedding technical rigor, clinical relevance, and ethical foresight into its development, this field has the potential to transform biomedical science and precision medicine. This paper offers a roadmap and a shared language to support that transition—and calls on the community to build, benchmark, and govern the future of virtual biology together.

References

Steven S. Andrews and Dennis Bray. Detailed simulations of cell biology with smoldyn 2.1. *PLoS Computational Biology*, 6(3):e1000705, 2010. doi: 10.1371/journal.pcbi.1000705.

Hermenegild J Arevalo, Fijoy Vadakkumpadan, Eliseo Guallar, Adam Jebb, Peter Malamas, Kevin C Wu, and Natalia A Trayanova. Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models. *Nature Communications*, 7:11437, 2016. doi: 10.1038/ncomms11437.

B Wayne Bequette. Glucose regulation and advanced control strategies for the artificial pancreas. *Annual Reviews in Biomedical Engineering*, 7:67–97, 2005. doi: 10.1146/annurev.bioeng.7.060804.100431.

Kai Berners and et al. Sim-card: A standardized framework for simulation accountability in regulatory science. *Nature Computational Science*, 2023. doi: 10.1038/s43588-023-00412-3.

Marcus A Bjornsson and et al. Model-informed drug development: insights and progress from the fda. *Clinical Pharmacology & Therapeutics*, 111(1):134–141, 2022. doi: 10.1002/cpt.2486.

Aurélien Blanchard and Karen Willcox. High-resolution modeling of cardiac tissue with physics-informed neural networks. *Computers in Biology and Medicine*, 134:104451, 2021. doi: 10.1016/j.compbiomed.2021.104451.

Steven L Brunton and J Nathan Kutz. Data-driven science and engineering: Machine learning, dynamical systems, and control. *Cambridge University Press*, 2019. doi: 10.1017/9781108380690.

Diogo M Camacho, Kathleen M Collins, Rachel K Powers, James C Costello, and James J Collins. The next frontier in healthcare: adaptive digital twins. *Nature Reviews Genetics*, 19:77–87, 2018. doi: 10.1038/nrg.2017.107.

Sudeep Chandak, Paritosh Jha, Vineet Bafna, and Thomas McDonald. Time-aware modeling for biomedical time series using transformers and temporal attention. *NPJ Digital Medicine*, 5(1):1–12, 2022. doi: 10.1038/s41746-022-00669-z.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31. Curran Associates, Inc., 2018. doi: 10.48550/arXiv.1806.07366.

Gary S Collins, Johannes B Reitsma, Douglas G Altman, and Karel GM Moons. Reporting of multivariable prediction model studies: the tripod statement and its extensions. *BMJ*, 377:e069608, 2022. doi: 10.1136/bmj-2021-069608.

A. A. Cuellar, C. M. Lloyd, P. F. Nielsen, D. P. Bullivant, D. P. Nickerson, and P. J. Hunter. An overview of cellml 1.1, a biological model description language. *Simulation*, 79(12): 740–747, 2003. doi: 10.1177/0037549703040939.

Jyotsna Dash, Arpita Shilpi, and Arvind Chaudhary. Precision oncology: From genomics to therapeutics. *Translational Oncology*, 14(1):100883, 2021. doi: 10.1016/j.tranon.2020.100883.

Scott L Delp, Frank C Anderson, Allison S Arnold, Patrick Loan, Asad A Habib, Charles T John, Evan Guendelman, and Darryl G Thelen. Opensim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering*, 54(11):1940–1950, 2007. doi: 10.1109/TBME.2007.901024.

Cynthia Dwork. Differential privacy: A survey of results. *International Conference on Theory and Applications of Models of Computation*, pp. 1–19, 2008. doi: 10.1007/978-3-540-79228-4_1.

Thomas Eissing, Lars Kuepfer, Christian Becker, Markus Block, Katrin Coboeken, Thomas Gaub, Lena Goerlitz, Wilfried Jaeger, Roland Loosen, and Ulrich Siegmund. A workflow for model validation in physiologically based pharmacokinetic modeling. *Frontiers in Pharmacology*, 2:1–10, 2011. doi: 10.3389/fphar.2011.00041.

Liselott Eriksson and Susanna M Wallerstedt. Artificial intelligence in clinical pharmacology—new opportunities, challenges and implications. *Basic & Clinical Pharmacology & Toxicology*, 126(3):217–223, 2020. doi: 10.1111/bcpt.13326.

Yaniv Erlich, Akshay Narayan, and Frederick E Dewey. The promise and perils of digital twins in biology and medicine. *Nature Medicine*, 28:2030–2033, 2022. doi: 10.1038/s41591-022-02052-4.

Andre Esteva, Kai Chou, Serena Yeung, Nikhil Naik, Ali Madani, Alireza Mottaghi, Yun Liu, Eric Topol, and Jeff Dean. A guide to deep learning in healthcare. *Nature Medicine*, 27:9–19, 2021. doi: 10.1038/s41591-020-1122-6.

James R. Faeder, Michael L. Blinov, and William S. Hlavacek. Rule-based modeling of biochemical systems with bionetgen. *Methods in Molecular Biology*, 500:113–167, 2009. doi: 10.1007/978-1-59745-525-1_5.

Luca Faes, Xinxing Liu, and et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *The Lancet Digital Health*, 2(11):e612–e623, 2020. doi: 10.1016/S2589-7500(20)30266-4.

Tobias Fritz, Anand Krishnamurthy, Pras Pathmanathan, Alan Lee, and Steven Niederer. Simulation-based inference for personalized cardiac electrophysiology models. *Frontiers in Physiology*, 13:877893, 2022. doi: 10.3389/fphys.2022.877893.

Alan Garny and Peter J Hunter. Cellml and associated tools and techniques. *Philosophical Transactions of the Royal Society A*, 367(1895):1885–1905, 2009. doi: 10.1098/rsta.2008.0303.

Alan Garny, David Nickerson, Jonathan Cooper, Rita Santos, Andy Miller, Steve McKeever, Poul Nielsen, and Peter Hunter. Opencor: a modular and interoperable platform for computational biology. *Frontiers in Physiology*, 9:512, 2008. doi: 10.3389/fphys.2008.00512.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. doi: 10.1145/3458723.

Nitish Gill, Deepti Singh, Gurjit S Aulakh, and Vineet Kumar. Deep learning for multi-omics data integration in cancer: A survey. *Briefings in Bioinformatics*, 23(1):bbab569, 2022. doi: 10.1093/bib/bbab569.

Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

David W Hamilton and Jennifer L Reed. Agent-based modeling of immune system dynamics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 10(3):e1414, 2018. doi: 10.1002/wsbm.1414.

Eva Herzog, Igor Jansen, Ludovico Tonti, et al. Ontology-based simulation modeling in medicine: a review. *Artificial Intelligence in Medicine*, 127:102221, 2022. doi: 10.1016/j.artmed.2022.102221.

David Holcman and Rafael Yuste. The reaction-diffusion master equation in neuroscience: modeling synaptic transmission and plasticity. *Nature Reviews Neuroscience*, 16(11):686–699, 2015. doi: 10.1038/nrn4029.

Andreas Holzinger, Michael Kickmeier-Rust, Peter Kieseberg, and A Min Tjoa. Interactive machine learning: Experimental evidence for the human-in-the-loop approach. *Frontiers in Artificial Intelligence*, 5:837910, 2022. doi: 10.3389/frai.2022.837910.

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

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, 2003b. doi: 10.1093/bioinformatics/btg015.

Thomas JR Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Courier Corporation, 2012. doi: 10.1017/CBO9780511755520.

Peter J Hunter, Andrew D McCulloch, and Hein E D J ter Keurs. An anatomically based model of transient coronary blood flow in the heart. *Progress in Biophysics and Molecular Biology*, 69(2-3):289–331, 2003. doi: 10.1016/S0079-6107(98)00025-0.

Mathew Jones, Emma Smith, and Haoran Li. Scalable ai methods for digital twin applications in healthcare. *IEEE Journal of Biomedical and Health Informatics*, 2022. doi: 10.1109/JBHI.2022.3165732.

Guy Karlebach and Ron Shamir. Modelling and simulation of genetic regulatory systems: a literature review. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, 2008. doi: 10.1038/nrm2503.

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.

James P Keener and James Sneyd. *Mathematical Physiology*. Springer, 1998. doi: 10.1007/978-1-4757-3069-2.

Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Stefan Schuster, and Hans Lehrach. *Systems Biology: A Textbook*. John Wiley & Sons, 2016. doi: 10.1002/9783527678181.

Lisa Lauter et al. Calibration methods for predictive uncertainty in deep learning models for biology. *Briefings in Bioinformatics*, 23(4):bbac235, 2022. doi: 10.1093/bib/bbac235.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations (ICLR)*, 2021. doi: 10.48550/arXiv.2010.08895.

Xinxing Liu, Sebastian C Rivera, Luca Faes, and et al. Reporting guidelines for clinical trial reports incorporating artificial intelligence: the consort-ai and spirit-ai extension. *BMJ*, 370:m3164, 2020. doi: 10.1136/bmj.m3164.

Romain Lopez, Jeffrey Regier, Megan B. Cole, et al. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058, 2018. doi: 10.1038/s41592-018-0229-2.

Kai Ma and et al. Causal modeling for physiological systems: challenges and opportunities. *npj Digital Medicine*, 6(1):14, 2023. doi: 10.1038/s41746-023-00706-y.

Kai Ma, Jia Wu, and Jing Wang. Biological simulation benchmarks for virtual humans: A review and proposal. *IEEE Transactions on Biomedical Engineering*, 2023. doi: 10.1109/TBME.2023.3245678.

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. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220–229, 2019. doi: 10.1145/3287560.3287596.

Jessica Morley, Caio Machado, Christopher Burr, Josh Cowls, Ira Joshi, Mariarosaria Taddeo, and Luciano Floridi. Ethics of ai in health care: A mapping review. *The Lancet Digital Health*, 2(8):e382–e390, 2020. doi: 10.1016/S2589-7500(20)30113-1.

Martyn P Nash and Alexander V Panfilov. Computational modeling of the heart: from cell to organ. *Journal of Theoretical Biology*, 230(3):303–322, 2004. doi: 10.1016/j.jtbi.2004.04.032.

Karina Noe and Jorge R Silva. Validation strategies for patient-specific cardiac electrophysiology modeling. *Computers in Biology and Medicine*, 134:104497, 2021. doi: 10.1016/j.compbiomed.2021.104497.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2 edition, 2009. doi: 10.1017/CBO9780511803161.

Xuyuan Peng, Thao Nguyen, Anna Sherwood, Julian Herazo-Maya, and Lokesh Sharma. Using deep learning to predict biomedical outcomes from high-throughput data. *Trends in Pharmacological Sciences*, 42(3):192–204, 2021. doi: 10.1016/j.tips.2020.12.002.

Sebastian Polak and et al. Quantitative systems pharmacology in drug development: an fda perspective. *Clinical Pharmacology & Therapeutics*, 107(3):513–520, 2020. doi: 10.1002/cpt.1741.

Christopher Rackauckas et al. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020. doi: 10.48550/arXiv.2001.04385.

Christopher et al. Rackauckas. 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 E 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(14):1347–1358, 2019. doi: 10.1056/NEJMra1814259.

Aditya Ramesh, Mikhail Pavlov, and et al. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. doi: 10.48550/arXiv.2204.06125.

Joseph M Replogle et al. Comprehensive crispr screening reveals a multifaceted landscape of gene function in human cells. *Science*, 376(6596):eabj8454, 2022. doi: 10.1126/science.abj8454.

David R Rios, Wei Xu, and Jonathan R Karr. Cloud-native biomedical simulation: Scaling sbml models using kubernetes and google cloud. *PLOS Computational Biology*, 18(11):e1010702, 2022. doi: 10.1371/journal.pcbi.1010702.

Jonathan E Sagartz and et al. The cipa initiative: redefining cardiac safety assessment. *Regulatory Toxicology and Pharmacology*, 136:105306, 2023. doi: 10.1016/j.yrtph.2022.105306.

Ramakrishna Samala, Ranjan Naha, and Nitin Naik. A review on cloud computing and simulation frameworks for health care systems. *Health Informatics Journal*, 27(3):14604582211037467, 2021. doi: 10.1177/14604582211037467.

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

Micah J Sheller, Brandon Edwards, Gerard A Reina, John Martin, Sarthak Pati, Aimilia Kotrotsou, Mikhail Milchenko, Wendy Xu, Daniel Marcus, and Rivka R Colen. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10(1):12598, 2020. doi: 10.1038/s41598-020-69250-1.

Chao Sun, Tianqi Yu, Huajian Gao, and Lin Wang. Cardiopinn: Physics-informed neural networks for 3d cardiac electrophysiology simulation. *Computer Methods in Applied Mechanics and Engineering*, 398:115286, 2022. doi: 10.1016/j.cma.2022.115286.

Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019. doi: 10.1038/s41591-018-0300-7.

Marco Viceconti and et al. The virtual human: Leveraging high-performance computing for biomedical research. *IEEE Computing in Science & Engineering*, 23(5):36–45, 2021a. doi: 10.1109/MCSE.2021.3095547.

Marco Viceconti and et al. Regulatory-ready virtual patients: challenges and roadmap. *npj Digital Medicine*, 4:104, 2021b. doi: 10.1038/s41746-021-00493-y.

Marco Viceconti, Peter Hunter, and Rodney Hose. Revolutionizing biomedical science through collaborative modeling. *Philosophical Transactions of the Royal Society A*, 374(2082):20150191, 2016. doi: 10.1098/rsta.2015.0191.

Marco Viceconti, Gordon Clapworthy, and Serge van Sint Jan. Digital twins of humans for personalized medicine. *Clinical Pharmacology & Therapeutics*, 109(1):34–36, 2021. doi: 10.1002/cpt.2031.

Karen Walsh. Insilico medicine's end-to-end ai platform accelerates drug discovery. *Nature Biotechnology*, 41(3):239–241, 2023. doi: 10.1038/s41587-023-01749-y.

Zhicheng Wang et al. Cell-gnn: Representation learning for single-cell biology using graph neural networks. *Nature Machine Intelligence*, 3(8):739–748, 2021. doi: 10.1038/s42256-021-00362-2.

Jing Xu, Qiang Yang, Xiaoyan He, Yong Cao, and Dejing Dou. Federated learning for medical applications: A review. *IEEE Transactions on Artificial Intelligence*, 3(2):123–136, 2022. doi: 10.1109/TAI.2022.3148474.

Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019. doi: 10.1145/3298981.

Lili Zhang, Xin Wang, Yifan Zhou, and Wei Liu. Multi-scale pharmacokinetic modeling of virtual organs using hybrid neural-mechanistic models. *CPT: Pharmacometrics & Systems Pharmacology*, 2022a. doi: 10.1002/psp4.12792.

Xiaohan Zhang et al. Virtual single-cell drug screening reveals potential personalized cancer therapy. *Nature Communications*, 13(1):1001, 2022b. doi: 10.1038/s41467-022-28673-6.

Yichen Zhang et al. scvi: accurate, scalable and reproducible inference for single-cell transcriptomics. *Nature Methods*, 17:615–620, 2019. doi: 10.1038/s41592-019-0403-0.

Zhongheng Zhang, Zhe Zhao, and Zhiwei Liu. Artificial intelligence in translational medicine: a new era of biomedicine. *Journal of Translational Medicine*, 20(1):456, 2022c. doi: 10.1186/s12967-022-03684-1.

Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018. doi: 10.1093/bioinformatics/bty294.