ACCSCIENCE PUBLISHING

International Journal of AI for Materials and Design

PERSPECTIVE ARTICLE

Model-as-evidence framework established using AI-integrated organoids and organs-on-chips systems

Peixi Wang[1,2†], Berenica Santoso[1†], Songlin He[1,2†], Liangbin Zhou[1,2], Yuwei Zhang[1,2], Yichi Zhang[1], Shing Yui Ho[1,2], Rocky S. Tuan[2,3,4*], and Zhong Alan Li[1,2,3,4*]

[1]Department of Biomedical Engineering, Faculty of Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China

[2]InnoHK Center for Neuromusculoskeletal Restorative Medicine, Hong Kong Science and Technology Park, Hong Kong SAR, China

[3]Institute for Tissue Engineering and Regenerative Medicine, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

[4]Peter Hung Pain Research Institute, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

Abstract

The predictive capability crisis in drug discovery and development remains a significant factor hindering medical progress. Among emerging solutions, organoids offer the ability to reflect the biological fidelity of tissues, whereas organs-on-chips (OoCs) demonstrate strengths in simulating dynamic tissue microenvironments and enabling multi-organ interactions. Herein, we present a unifying artificial intelligence (AI)-orchestrated framework that incorporates both platforms into verifiable and regulatable "model-as-evidence" (MAE) pathways capable of self-optimization with researcher oversight. AI is employed as an indispensable computational integration layer addressing three bottlenecks that render conventional approaches infeasible: (i) combinatorial explosion in high-dimensional parameter spaces, (ii) multi-scale coupling between molecular interactions and tissue-scale level morphogenesis, and (iii) requirement for real-time adaptive control of dynamic microenvironments exceeding human cognitive bandwidth. At the organoid level, this integration enables multi-objective optimization to improve reproducibility, reduce resource use, and achieve predictive modeling to capture extracellular matrix–cells interactions. For OoCs, active learning algorithms condense protracted design cycles into efficient, goal-directed experimentation, and deep reinforcement learning sustains physiological steady states through adaptive, real-time control. The convergence of these systems within an integrated platform enables standardized, cross-comparable phenotyping pipelines that transform heterogeneous experimental data into a universal quantitative language, and harmonized evaluation benchmarks align experimental outputs across laboratories. Together, this framework generates curated evidence packages that directly address regulatory concerns related to efficacy and safety. By systematizing experimental design, parameter control, and interpretive analytics, this AI-orchestrated approach elevates organoid and OoC technologies from specialized crafts to a robust engineering discipline. It establishes a virtuous cycle wherein AI accelerates iterative validation and optimization, freeing researchers to focus on higher-order hypothesis generation and translational strategy.

†*These authors contributed equally to this work.*

***Corresponding authors:**
Zhong Alan Li
(alanli@cuhk.edu.hk)
Rocky S. Tuan
(tuanr@cuhk.edu.hk)

Citation: Wang P, Santoso B, He S, *et al*. Model-as-evidence framework established using AI-integrated organoids and organs-on-chips systems. *Int J AI Mater Design*. doi: 10.36922/IJAMD026150008

Received: April 11, 2026

Revised: May 17, 2026

Accepted: May 29, 2026

Published online: June 12, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Thus, through AI-orchestrated quantification of biological fidelity and engineering rigor, these platforms may mature into credible, scalable, and transformative toolboxes for next-generation biomedical translation.

Keywords: Organoid; Organ-on-a-chip; Artificial intelligence; Organoid-on-a-chip

1. Introduction

The translational gap in preclinical research represents one of the most significant challenges in modern biomedical science, often referred to as the "*valley of death*" in drug discovery and development.[1] Traditional two-dimensional (2D) cell cultures, while convenient, lack the tissue-level architecture, multicellular interactions, and three-dimensional (3D) organization that are essential for accurately representing human physiology. Consequently, these monolayer systems fail to recapitulate drug responses, metabolic functions, and disease phenotypes observed *in vivo*.[2-4] Animal models, despite their widespread use, suffer from species-specific differences that frequently compromise efficacy and toxicity predictions. In 2012, Begley and Ellis demonstrated that only 11% of 53 landmark preclinical cancer studies could be independently reproduced by industry scientists, highlighting a reproducibility crisis in oncology research.[2] In parallel, industry reports indicate that nearly 90% of drug candidates entering clinical trials ultimately fail to reach the market, despite encouraging preclinical data.[5-7] This crisis imposes substantial economic costs, estimated in the billions annually, and delays the translation of effective therapeutics to patients.[8-11]

Given the limitations of conventional preclinical models, organoids and OoCs have emerged as transformative alternatives that promise to bridge the gap between oversimplified cell culture models and complex animal models. Organoids are self-organizing 3D structures derived from stem cells or tissues, which replicate biological authenticity through spontaneous morphogenesis and cell type diversity.[3,12] Clevers and coworkers demonstrated that single $Lgr5^+$ intestinal stem cells could form complete crypt-villus structures *in vitro*, establishing the foundation for organoid technology.[13,14] OoCs, pioneered by Ingber and other researchers from Europe and US, leverage microfluidic systems that allow for precise spatiotemporal regulation of chemical and mechanical stimuli, providing enhanced controllability. The landmark lung-on-a-chip study demonstrated how alveolar-capillary interfaces could be reconstituted on a microfluidic platform, enabling real-time monitoring of (patho)physiological responses.[15,16] These methods shift the focus from whether a model resembles human (patho)physiology to whether it can be engineered for reproducibility, quantitative comparison across laboratories, and stable operation over extended periods.[16-18]

Organoids excel at mimicking cellular and functional fidelity of native tissues, while OoCs excel at simulating the tissue microenvironment and multi-organ interactions. However, further development of these two platforms as well as their integration present several challenges. Organoid-associated difficulties include optimizing culture medium components, extracellular matrix (ECM), and cell type, while OoC challenges revolve around chip design, dynamic parameter control, and tissue characterization. AI offers an opportunity to solve these above-mentioned challenges, and to provide integrated solutions for organoid and OoC systems, including the development of organoids-on-chips. However, realizing these solutions through conventional experimental design faces fundamental computational and operational barriers. Manual approaches face three infeasibilities: combinatorial explosion (high-dimensional parameter spaces where, for example, numerous interacting culture medium components create nonlinear, temporally dependent control problems that rapidly exceed manual optimization capacity),[19,20] multi-scale coupling (molecular-to-tissue temporal hierarchies defy intuitive tuning),[21,22] and real-time bandwidth (microfluidic homeostasis demands millisecond precision beyond human capacity).[23,24] We propose an AI-orchestrated framework for the construction and characterization of organoids, OoCs, and organoid-OoC hybrids, aimed at advancing reliable and transformative toolboxes for innovative biomedical platforms and drug development (Figure 1).

To map biological challenges to computational solutions, Table 1 summarizes empirical barriers, method limitations, and corresponding AI approaches. Each entry highlights the algorithm's specific role and validation status, distinguishing validated outcomes, emerging evidence, and forwardlooking capabilities.

The validation status distinctions are critical for regulatory positioning: validated methods (*e.g.*, segment

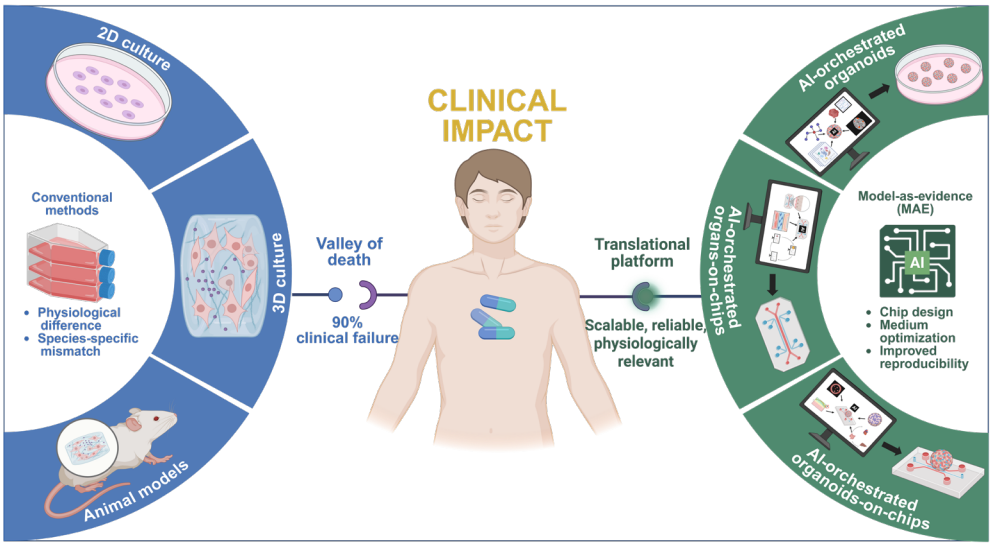


Figure 1. Evolution of preclinical models: bridging the translational gap in biomedical research through AI-orchestrated organoids, organs-on-chips (OoCs), and organoids-on-a-chip modeling platforms. The AI-orchestrated model-as-evidence (MAE) framework could potentially address the translational "*valley-of-death*" by optimizing organoids, OoCs, and organoid-on-a-chip systems. By overcoming traditional challenges related to culture medium, ECM, chip design, and tissue characterization, the AI-orchestrated MAE framework may transform these platforms into scalable, reliable toolkits for future drug innovation.
Abbreviations: 2D: Two-dimensional; 3D: Three-dimensional; AI: Artificial intelligence; ECM: Extracellular matrix.

anything model [SAM] segmentation) can be incorporated into evidence packages immediately; emerging methods require context-of-use qualification; and forward-looking capabilities demand prospective validation studies before regulatory submission. This tiered evidentiary structure prevents the common pitfall of conflating algorithmic promise with regulatory-ready performance. The following sections detail how these AI solutions operate at the organoid level (Sections 2.1–2.4), the OoC level (Sections 3.1–3.3), and their integrated convergence (Section 4).

Throughout this article, "AI-orchestrated" denotes a human-in-the-loop operational philosophy wherein AI optimizes parameters within researcher-defined boundaries and flags low-confidence decisions for review, distinct from autonomous "AI-scientist" concepts. This operational philosophy underpins our central concept, "model-as-evidence" (MAE), which we define as a regulatory-scientific framework that elevates qualified AI-integrated organoid and OoC systems from mere research instruments to primary evidence generators. Unlike AI-assisted optimization or Model-Informed Drug Development, which rely on traditional animal or clinical data as primary inputs, MAE achieves evidentiary primacy through standardized validation, closed-loop credibility with real-time uncertainty quantification, and regulatory embedment via context-of-use design, thereby extending the model-informed paradigm to complex tissue-level systems.

2. AI-orchestrated organoids

2.1. Multi-objective optimization of culture medium components

The first challenge presented in Table 1, culture medium optimization, exemplifies how combinatorial explosion renders manual approaches infeasible. Culture medium constitutes the chemical boundary conditions of organoid construction and remains the most common source of hidden variation across experiments. Standard formulations typically contain numerous growth factors and small molecules with significant temporal dependencies regarding when to add specific components, how long to maintain their concentrations, and what replacement rhythm to follow. Early studies demonstrated that defined combinations of epidermal growth factor (EGF), Noggin, and R-spondin could generate intestinal crypt-villus structures, later extended with fibroblast growth factor (FGF), hepatocyte growth factor (HGF), and gastrin for colon and gastric organoids. These formulations are not static recipes but high-dimensional control problems where dozens of parameters interact to determine phenotypes.[13,14] Gastric organoid culture similarly requires FGF, EGF, Noggin, R-spondin, and gastrin for long term

Table 1. AI method selection: biological challenges, empirical barriers, and computational solutions

Bio-challenge	Empirical barrier	Method limitation	AI solution	Specific role	Validation status	Ref.
Culture medium optimization	Greggio *et al.* (2013): 4-factor manual screening required weeks; missed long-term stability; full factorial of numerous parameters infeasible	Full factorial exploration impossible; fractional factorial requires >10^6 runs	Multi-fidelity Bayesian optimization	Low-cost metabolic assays screen candidates; RNAseq validation is reserved for the top 5 % conditions	Emerging: validated in materials science; direct organoid application pending	19,20,25
Extracellular matrix batch variation	Driehuis *et al.* (2019): passage-dependent morphometric variance >40% without quantitative tracking; months of rheological characterization per Matrigel batch	Months of rheological characterization per batch; >40% morphometric variance	Graph neural networks + SHapley additive exPlanations	Graph neural networks predict integrin binding from molecular graphs; SHapley Additive exPlanations disentangles stiffness *vs.* ligand effects	Forward-looking: graph neural networks validated on PDBbind (R=0.82); organoid extracellular matrix prediction awaiting prospective validation	21,26-28
Cell composition drift	Camp *et al.* (2015): scRNA-seq atlases exist but organoid-to-reference alignment requires domain adaptation; Pollen *et al.* (2019): transfer is non-trivial	scRNA-seq atlases exist but organoid-to-reference alignment requires retraining from scratch	Transfer learning	Aligns organoid scRNA-seq with fetal reference atlases without retraining from scratch	Emerging: validated in cerebral organoid-fetal brain alignment	29-31
3D segmentation inconsistency	Marks *et al.* (2025): inter-lab annotation bias; manual annotation infeasible for 3D volumetric data; Archit *et al.* (2025): SAM variants need microscopy-specific adaptation	Manual annotation infeasible for 3D volumetric data; inter-lab annotation bias	CellSAM/SAM-microscopy variants	Zero-shot segmentation across 9 microscopy modalities (brightfield, confocal, optical coherence tomography, holotomography, *etc.*); eliminates annotation burden	Validated: cell segment anything model zero-shot dice >0.85 on multiple organoid modalities; cross-lab domain shift under active mitigation	32-36
Chip design and culture requirements *vs.* manufacturability	Fu *et al.* (2014): manual trial-and-error faces combinatorial explosion when optimizing flow rate, shear stress, nutrient penetration simultaneously; Bein *et al.* (2018): geometric innovation relied on intuition without transferable rules	Computational fluid dynamics iterations face combinatorial explosion; manual trial-and-error	Generative models (GAN/VAE) + Bayesian optimization	Generative adversarial network learns manufacturability priors; Bayesian optimization navigates pareto front with < 12 iterations *vs.* dozens of computational fluid dynamics loops	Emerging: validated in vascular network bioprinting; organoid-specific chip design pending	37-40

(*Cont'd...*)

Table 1. (*Continued*)

Bio-challenge	Empirical barrier	Method limitation	AI solution	Specific role	Validation status	Ref.
Real-time homeostasis	Mosadegh *et al.* (2015): proportional-integral-derivative controllers destabilize under pump drift, bubble nucleation (>10-20% shear error); Jui *et al.* (2024): organoid response has nonlinear delays	Proportional-integral-derivative assumes linear dynamics; organoid response has nonlinear delays	Deep reinforcement learning (proximal policy optimization)	Partially observable Markov decision process; adaptive policy handles non-stationary disturbances without retuning	Forward-looking: proximal policy optimization validated in microfluidic droplet control; organoid homeostasis loop pending experimental validation	23,24,41,42
Multi-modal data fusion	Wong *et al.* (2025): "multi-modal" becomes uninterpretable without unified representation; imaging + omics + sensors generate heterogeneous, unaligned data	"Multi-modal" becomes uninterpretable without unified representation	Multi-modal transformers + graph neural networks	Self-attention learns cross-modal alignment; graph neural networks model organ-organ metabolic coupling	Forward-looking: transformers validated in clinical multi-modal fusion; organoid-on-chip integration pending	43-46
Regulatory uncertainty quantification	Thompson *et al.* (2025): single-endpoint predictions unreliable for safety-critical decisions; confidence intervals required for investigational new drug submissions	Single-endpoint predictions unreliable for safety-critical decisions; confidence intervals required	Deep ensembles + Monte Carlo dropout	Quantifies epistemic uncertainty; confidence intervals required for investigational new drug submissions	Emerging: validated in surgical soft-tissue manipulation; organoid drug response uncertainty quantification in progress	47-51

Abbreviations: GAN: Generative adversarial network; RNA-seq: RNA sequencing; SAM: Segment anything model; scRNA-seq: single-cell RNA sequencing; VAE: Variational autoencoder.

maintenance.[52] These formulations represent not static recipes but high-dimensional, dynamic, constrained control problems where dozens of parameters interact to determine final phenotypes.

The AI approach frames culture medium design as multi-objective and constrained optimization problems (Figure 2a). Objectives may include morphological similarity to reference tissues, marker co-expression patterns, and functional readouts, such as barrier integrity measured by transepithelial/transendothelial electrical resistance (TEER), metabolic activity assessed through glucose consumption and lactate production rates, and secretion of tissue-specific proteins (*e.g.*, albumin from hepatocytes or mucins from intestinal cells). Constraints can be defined by amino acid consumption rates, osmolarity ranges (typically 280–320 mOsm/kg),[20] pH stability (7.2–7.4),[20] and solubility/stability conditions for growth factors that degrade rapidly at 37 °C.[25,53] Traditional medium development relies on trial-and-error approaches that cannot resolve multifactorial interactions. Greggio *et al.*'s manual screening of four pancreatic factors (FGF, EGF, HGF, gastrin) revealed nonlinear synergies,[25] a precise mathematical structure that multi-objective optimization formalizes. However, their manual screening approach could not explore the full parameter space for long-term expansion stability, a limitation that multi-fidelity Bayesian optimization addresses by efficiently searching high-dimensional formulation spaces while minimizing experimental cost.

The combinatorial explosion in high-dimensional formulation spaces necessitates AI-guided optimization. Greggio and collaborators' manual screening of four pancreatic factors (FGF, EGF, HGF, gastrin) revealed nonlinear synergies,[25] a precise mathematical structure that multi-objective optimization formalizes, yet required weeks of experimentation and missed long-term expansion stability. Extending such approaches to comprehensive formulation spaces, incorporating temporal dosing, solubility constraints, and cross-interactions, renders exhaustive screening economically

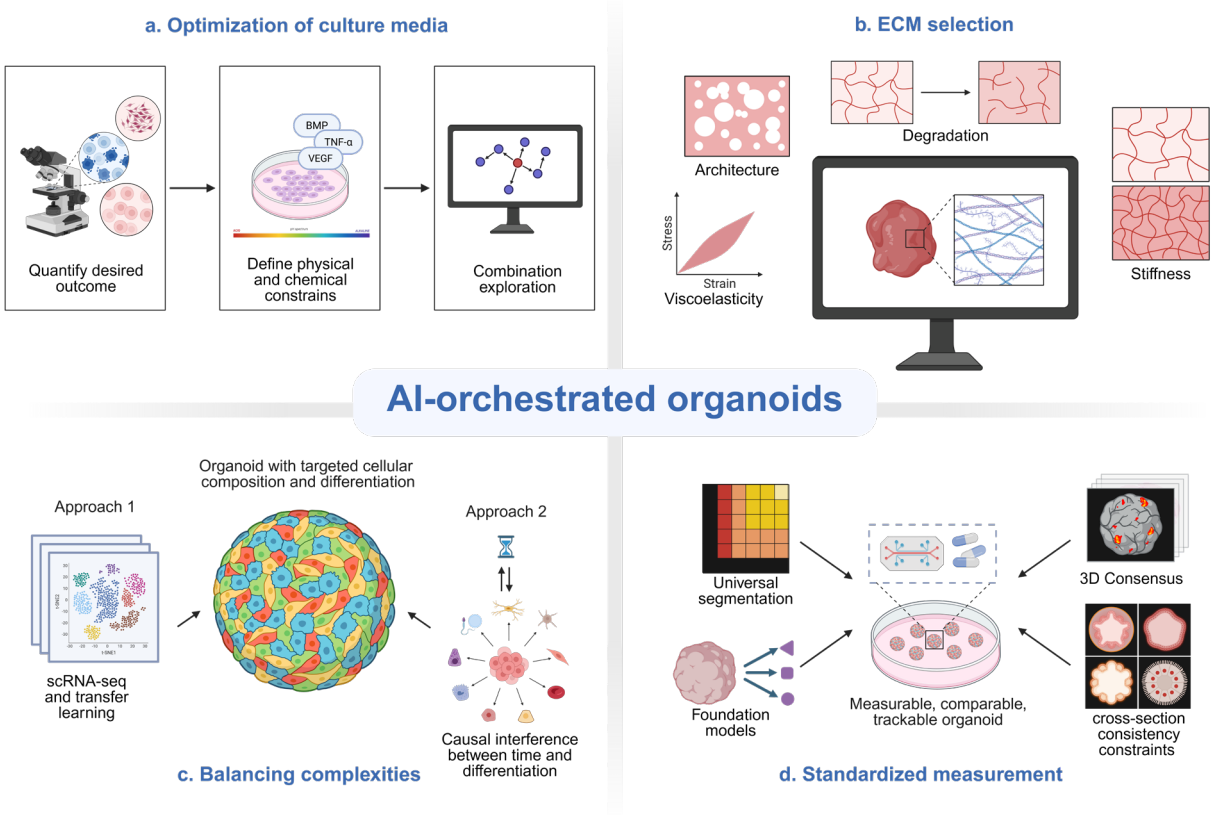


Figure 2. Four domains of AI integration in organoid systems. (a) Culture medium optimization via multi-objective problem framing. (b) ECM selection via predictive modeling to reduce batch variability. (c) Cellular composition balancing utilizing scRNA-seq and transfer learning. (d) Standardized measurements through non-invasive quantification and universal segmentation.
Abbreviations: 3D: Three-dimensional; AI: Artificial intelligence; ECM: Extracellular matrix; scRNA-seq: single-cell RNA sequencing.

prohibitive.[19,20] Each wet-lab cycle is costly: a single RNA-seq run is approximately 100-fold more expensive than a colorimetric metabolic assay,[54] and whole-transcriptome profiling or high-resolution imaging further escalates costs. Multi-fidelity Bayesian optimization[19,20] addresses this by using inexpensive metabolic assays or low-resolution microscopy as proxies to explore the design space, reserving costly validation for only the most promising candidates. This strategy reduces iteration costs by two orders of magnitude compared with conventional design-of-experiments, establishing AI not as an analytical luxury but as an economic and scientific necessity for scalable organoid engineering. Embedding thermodynamic and solubility constraints further improves search efficiency and reproducibility,[19,20] though direct application to organoid culture remains to be validated.

2.2. Intelligent selection of ECM

Beyond medium formulation, the extracellular matrix (ECM) represents the second dominant source of organoid variability, where batch-to-batch heterogeneity similarly demands AI-guided standardization. Matrigel, the most commonly used ECM substrate, is derived from Engelbreth–Holm–Swarm mouse sarcoma and exhibits significant batch-to-batch variation in composition, stiffness, and growth factor content. Previous studies often substituted material characterization with observational claims, that a particular matrix supports growth, thereby hiding critical variables within uncontrollable batch variations that undermine reproducibility. Engler and coworkers[21] seminally demonstrated that matrix elasticity alone can direct stem cell lineage specification: soft matrices (~0.1–1 kPa) promote neurogenic differentiation, intermediate stiffness (~8–17 kPa) promotes myogenic differentiation, and stiff matrices (>34 kPa) promote osteogenic differentiation. This work established stiffness as a critical design parameter, but subsequent studies by Trappmann *et al.* revealed that ECM tethering density can regulate stem cell fate independently of stiffness, adding another dimension to matrix design.[22] Sunyer and coworkers[55] further showed that collective cell durotaxis emerges from long-range intercellular force transmission through the ECM, demonstrating that matrix mechanical properties must be optimized not only for individual cell differentiation but for emergent tissue-scale morphogenesis, a multi-scale optimization problem ideally suited for graph neural network approaches that model cell-ECM interactions.[22]

The AI approach proceeds through systematic characterization, feature extraction, and predictive modeling (Figure 2b). Measurable metrics describe ECM properties, including elastic modulus, viscoelasticity, degradation rate, pore size, and ligand density. Graph neural networks (GNNs) can predict ECM-integrin binding affinities by representing ECM components as nodes and molecular interactions as edges, transforming material-cell interactions from empirical guessing to computable prediction.[21,26] For ECM-integrin binding affinity prediction, where input features comprise variable molecular graphs rather than fixed-size grids, GNNs offer architectural advantages over standard CNNs/RNNs: graph structures explicitly encode local ligand topology, avoiding the Euclidean grid assumptions that distort irregular molecular geometries;[21,26] variable node sets natively handle batch-to-batch compositional variation (*e.g.*, fluctuating laminin-111/511 ratios) without padding artifacts;[21] and attention mechanisms highlight residue-level interactions (*e.g.*, RGD-α5β1 binding), yielding experimentally verifiable hypotheses that kernel methods or multi-layer perceptrons obscure.[27] GNNs achieve Pearson R = 0.82 versus 0.74 for random forest on molecular fingerprints, with the performance gap widening for multi-domain ECM proteins where grid representations fail.[26] These advantages are task-specific to graph-structured molecular data: for Euclidean imaging data (*e.g.*, histology slides), CNNs remain optimal, and for sequential time-series (*e.g.*, growth trajectories), Recurrent Neural Networks/Long Short-Term Memory Networks (RNNs/LSTMs) are preferred. Interpretable models, such as SHAP (SHapley Additive exPlanations), then quantify the contributions of modulus, ligand density, and degradation rate to specific outcomes,[27] such as regionalized differentiation or lumen formation, generating transferable design rules that can be applied across different organoid systems. Caliari and Burdick[26] provided a practical guide to hydrogel design for cell culture, establishing design principles that can be codified into AI models. Tissue-derived ECM hydrogels, when processed through standardized decellularization and solubilization protocols, demonstrate that tissue-specific, quantitatively characterizable ECM can reduce uncontrollable variation inherent to undefined tumor-derived matrices and provide more stable, multidimensional input spaces for AI modeling.[26]

Matrigel batch validation exemplifies the operational bottleneck that motivates predictive ECM selection. Conventional rheological characterization requires months per batch, yet Driehuis *et al.*[28] demonstrated that passage-dependent morphometric variance exceeds 40% without quantitative growth tracking. This uncontrollable variation, combined with months of validation failing to ensure consistency, renders empirical guessing unsustainable for scaled manufacturing. The AI approach described above addresses this by transforming material-cell interactions into computable predictions (GNNs)[21,26] and quantifying

design-parameter contributions into transferable rules (SHAP).[27] Without such predictive tools, each new batch introduces variation that undermines reproducibility, establishing algorithmic standardization as an operational prerequisite rather than an analytical enhancement. At the cellular level, similar optimization challenges arise for composition and differentiation trajectories, requiring analogous AI-orchestrated control to balance biological authenticity with manufacturing reproducibility.

2.3 Optimization of cellular composition and differentiation trajectory

Organoid authenticity derives from cellular diversity and spatiotemporal organization, but increased complexity makes reproduction more difficult and manufacturing more expensive. Finding a portable balance between biological authenticity, manufacturability, and reproducibility represents a key challenge for scaled applications.

Workman *et al.*[56] engineered human induced pluripotent stem cell (iPSC)-derived intestinal tissues with functional enteric nervous systems, demonstrating that multiple cell types can be incorporated into organoid systems. Finkbeiner and coworkers performed transcriptome-wide analysis revealing hallmarks of human intestine development and maturation in organoids compared to *in vivo* tissues, establishing molecular benchmarks against which scRNA-seq-based organoid atlases can be aligned to identify missing cell types or aberrant differentiation states,[57,58] and Watson and collaborators developed an *in vivo* model of the human small intestine using PSC, providing developmental staging references that can guide *ex vivo* organoid differentiation trajectory optimization.[29]

Two AI pathways are emerging for optimizing cell composition (Figure 2c): first, single-cell RNA sequencing (scRNA-seq) combined with transfer learning aligns organoids with reference tissue atlases to infer the minimal cell set required for target functions, reducing unnecessary complexity while preserving essential functionality; second, causal inference on differentiation trajectories and perturbation time windows distinguishes correlation from intervenable causality, building culture strategies more robust to donor variation.[29-31]

These optimization approaches address design-phase prediction of cellular composition, whereas functional fidelity validation requires the measurement architectures detailed in Section 2.4.

2.4. Organoid measurements: From noninvasive monitoring to standardized validation

Patient-derived organoid (PDO) biobanks exemplify the measurement imperative: rigorous frameworks must preserve inter-patient heterogeneity for drug response assessment while eliminating technical artifacts. Standardized quantification is essential because biological variability becomes computable uncertainty only with unified measurement grammars. Sachs *et al.*'s breast cancer organoid biobank demonstrated that single-organoid resolution morphometry was required to distinguish true drug sensitivity from culture-condition artifacts, without which inter-patient heterogeneity would be confounded by technical variance.[59] Vlachogiannis *et al.*'s multi-site gastrointestinal PDO study showed that predictive accuracy for metastatic outcomes depended on standardized viability assay protocols across collection centers, whereas non-uniform endpoints rendered cross-patient comparisons statistically invalid.[60] Driehuis *et al.*'s oral mucosal organoids revealed that passage-dependent morphometric variance exceeding 40% necessitated quantitative growth tracking for reliable personalized therapy prediction.[28] Broutier *et al.*'s liver cancer organoids established that heterogeneous budding patterns required hierarchical segmentation (organoid, crypt and lumen) to link morphology with transcriptomic drug response states.[61] Standardized phenotyping transforms biological variability into computable uncertainty; lacking it, identical models yield contradictory conclusions. This section addresses three hierarchical validation challenges: noninvasive monitoring, structural segmentation, and cross-modal integration.

Noninvasive functional readouts provide the first layer of validation. Cho *et al.*, also Xu' group, demonstrated optical coherence tomography (OCT)-based quantitative morphometry of human intestinal organoids, enabling longitudinal tracking without fluorescent labeling.[62-64] Cho and colleagues developed label-free, high-resolution 3D imaging via low-coherence holotomography combined with machine learning (ML) analysis.[62] These technologies establish baseline morphodynamic signatures that flag deviations from expected growth trajectories before destructive endpoint assays are required.

For structural validation, foundation models enable zero-shot segmentation with reliability protocols ensuring deployment safety,[32-36] while cross-modal integration fuses imaging, molecular, and functional readouts to prevent single-endpoint misjudgments.[47-49] Detailed layer-specific architectures, error propagation quantification, and mitigation strategies are presented in Section 4.1.

When organoids become measurable through this standardized pipeline, from raw images to segmented objects, to quantifiable features, to uncertainty-quantified phenotypes, their integration with OoCs becomes operationally feasible. The phenotype compiler introduced

in Section 4.1 extends this architecture to hybrid systems, but its foundation rests upon the measurement primitives established here.

3. AI-orchestrated OoCs

3.1. Material-structure integrated design

Chip design faces a fundamental tension: biocompatibility objectives (cell viability, functional maintenance) and engineering feasibility constraints (manufacturing yield, structural stability) often conflict. Traditional computational fluid dynamics (CFD) iterations cannot efficiently search this multi-objective space. Tseng and colleagues' chemoresistive testing platform revealed how channel geometry governs drug concentration gradients, exposing the limitations of empirical design: when simultaneously optimizing flow rate, shear stress, and nutrient penetration, manual trial-and-error faces combinatorial explosion.[39] Ingber and coworkers' concave microchamber design demonstrated that geometric innovation can break standard channel structure-function trade-offs, yet this breakthrough relied on investigator intuition without transferable design rules.[40] AI generative design (GANs/variational autoencoders [VAEs]) addresses precisely this gap: encoding Tseng's discovered physical laws (geometry-distribution mapping) and Ingber's geometric innovations as learnable prior distributions, enabling algorithms to automatically search the viability-manufacturability Pareto frontier and compress dozens of CFD cycles to under a dozen iterations (see Bayesian optimization applications below). We advocate hybrid loss functions where Navier–Stokes residuals and reaction-diffusion terms constitute explicit penalty terms weighted against data-driven and manufacturability losses, ensuring that physical constraints dominate in data-scarce regimes (100 examples, five public datasets currently available[65,66]) where CFD-based synthetic pre-training,[37,38] multi-fidelity VAE latent spaces,[19,20] and active human-in-the-loop learning[22] ensure reliability. Latent-space interpolation tests flag non-physical patterns (*e.g.*, negative pressure), restricting deployment to validated regions until additional data are acquired.[37]

An algorithmic search that simultaneously rewards cell viability, channel manufacturability and yield, finds Pareto-front channel layouts within a dozen iterations (an order of magnitude reduction compared with traditional CFD loops). Here, material and geometric parameters serve as input variables, with biological outcomes (*e.g.*, viability and function) and engineering constraints (*e.g.*, manufacturability and reliability) formulated as dual objectives. This multi-objective formulation embeds practical considerations, such as ease of fabrication and operational stability, directly into the optimization loop.[37] Li and collaborators reviewed oxygen gradients in 3D tissue-engineered constructs, identifying critical design parameters for maintaining physiological oxygen levels.[67]

Manufacturing capabilities establish the feasible design space, while, AI-orchestrated generative methods navigate this space to optimize the viability-manufacturability trade-off. Grigoryan and coworkers[68] employed SWIFT method to achieve complex 3D vascular architectures, yet its empirical parameter tuning (speed, pressure, viscosity) exemplifies the high-dimensional search problem that generative models solve.[37] Miller and colleagues' rapid casting of patterned vascular networks[38] and Kolesky *et al.*'s multi-material bioprinting of heterogeneous constructs[42] similarly expanded the manufacturing feasible region without systematic design optimization. Recent work has integrated these capabilities with algorithmic search: Bayesian optimization studies demonstrate that coupling Grigoryan-class vascular manufacturability constraints with cell viability objectives reduces design-validation cycles from dozens of CFD iterations to under a dozen experiments, rapidly converging to Pareto-optimal channel layouts satisfying both biological performance and engineering constraints.[37] Topological generative models (GANs/VAEs) further enable inverse design, encoding discovered physical laws (*e.g.*, geometry-flow distribution mappings) as learnable priors to automatically output composite configurations that reduce dead volumes and improve gradient formation, extending structural innovation from human intuition to data-driven spaces (Figure 3a).[69-71]

3.2. Real-time regulation of microfluidic dynamic parameters and membrane interface control

Long-term perfusion and dynamic stimulation are unique assets of OoCs, yet they are also the dominant sources of instability (*e.g.*, pump drift, bubble formation, shear fluctuations, or shifts in cell state), that can drive barrier function outside its target window. Mosadegh and colleagues quantified cancer cell chemotaxis using a paper-based microfluidic platform generating oxygen gradients, proving that microenvironmental cues can be dialed in spatiotemporally.[41] Jui and coworkers, also Gao'group, embedded ML in microfluidics to dissect immune cell behavior at the single cell level under temporally varying shear and cytokine fields.[72,73] Vela and coworkers established patient-derived 3D epithelial cultures from advanced prostate tumors as a personalized drug susceptibility test bed.[74] Treating chip operation as a partially observable Markov decision process, recent studies have employed deep reinforcement learning (DRL), which integrates deep neural networks for high-dimensional state

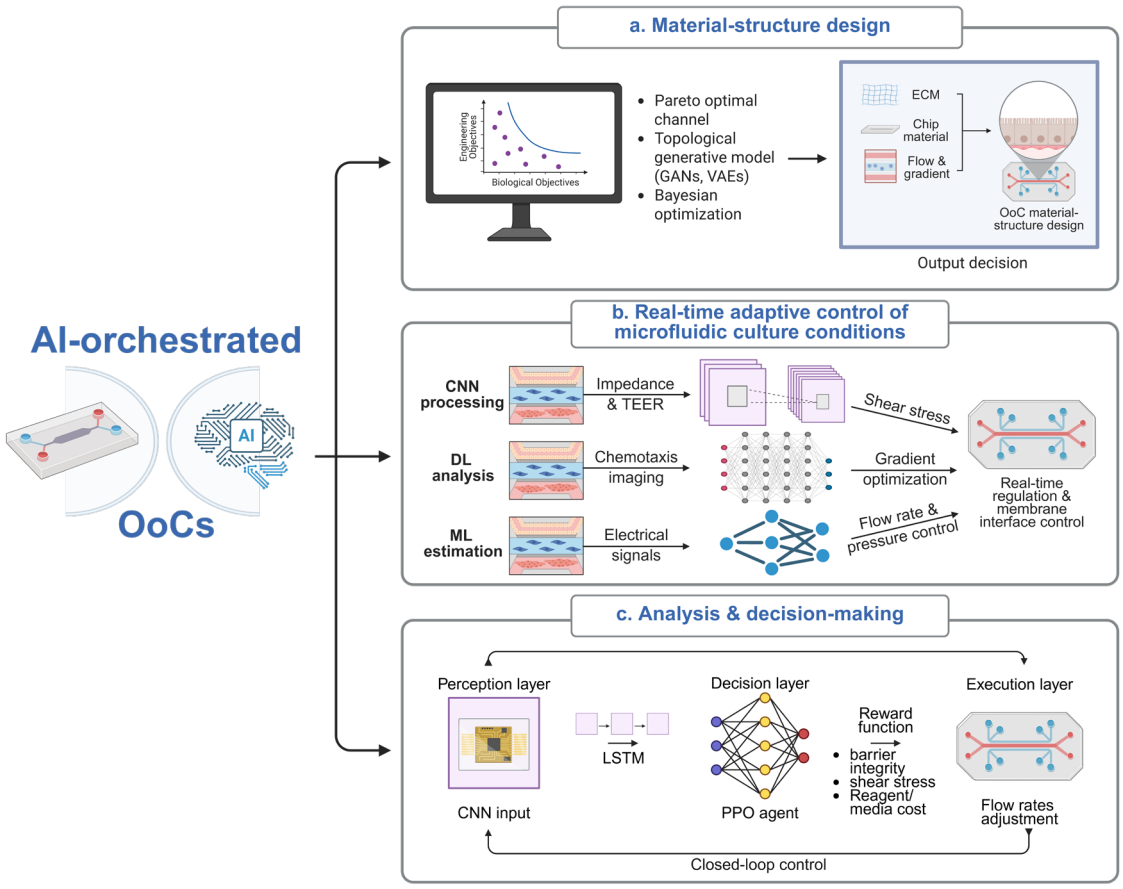


Figure 3. Three domains of AI integration in OoC systems: (a) Material-structure design accelerated by generative models and Bayesian search. (b) Real-time adaptive control of microfluidic culture conditions. (c) Closed-loop adaptive control architecture with explicit feedback pathways.
Abbreviations: CNN: Convolutional neural network; DL: Deep learning; GANs: Generative adversarial networks; LSTM: Long short-term memory networks; ML: Machine learning; OoCs: Organs-on-chips; PPO: Proximal policy optimization; TEER: Transepithelial/transendothelial electrical resistance; VAEs: Variational autoencoders.

representation with reinforcement learning for sequential action optimization, via proximal policy optimization to adaptively adjust shear stress and flow rate, achieving TEER-defined barrier integrity within 5% of the set point in proof-of-concept demonstrations.[23,24,41] In the MAE framework, this approach would be standardized with explicit uncertainty quantification and regulatory-grade validation protocols (Section 5).

Figure 3b illustrates the real-time adaptive control of microfluidic culture conditions. Conventional proportional-integral-derivative controller (PID) controllers assume linear dynamics with short feedback delays, yet organoid responses to flow adjustments exhibit nonlinear delays spanning multiple hours.[73] This temporal mismatch destabilizes PID tuning when disturbances are non-stationary. For example, pump drift and bubble nucleation introduce shear errors that vary across experimental runs.[41] Deep reinforcement learning (DRL) via Proximal Policy Optimization (PPO) treats these disturbances as state transitions in a partially observable Markov decision process, adapting flow rates and shear stress to maintain TEER-defined barrier integrity within operational windows.[23,24] The state space comprises multi-modal sensor readings (*e.g.*, OCT morphometry, TEER, dissolved O_2), while actions adjust peristaltic pump speeds and medium exchange rates. Hard barrier functions in the reward structure prevent osmotic shock (osmolarity changes <5 mOsm/kg/h, a conservative threshold below the ~ 10–20 mOsm/kg acute cell tolerance limit[75]), encoding safety constraints as non-negotiable boundaries beyond rule-based control.[23]

Figure 3c extends this real-time capability to a hierarchical closed-loop adaptive control architecture with explicit feedback pathways. The three-layer hierarchy

explicitly separates operational timescales to solve the bandwidth mismatch problem: (i) perception layer (update cycle ~100 ms): multi-modal sensors feed CNN-based state estimators with Monte Carlo dropout uncertainty quantification;[48,76] this layer compresses >1 kHz raw sensor streams into actionable state vectors; (ii) decision layer (computation time ~50 ms, triggered by perception updates): LSTM-based partially observable state estimators predict future states (*e.g.*, imminent lumen collapse from growth trajectory curvature 24–48 h ahead), feeding PPO with a reward function embedding hard safety barriers (osmolarity changes < 5 mOsm/kg/h, consistent with mammalian cell shock thresholds[75,77]) while balancing barrier integrity, low-level shear stress (~0.1–1.0 dyn/cm², appropriate for epithelial monolayers in minimal-flow conditions),[15,23,78] and operational cost;[23] (iii) execution layer (~200 ms, including fluid stabilization time): microfluidic pumps and valves adjust flow rates and medium exchange rates, with total loop time < 1 s for barrier maintenance. A parallel slow loop for strategic replanning operates at 0.01 Hz via multi-modal graph transformers that integrate real-time metabolic readouts,[46] with periodic molecular recalibration (*e.g.*, daily ELISA/Western blot)[79] decoupled from the fast loop to prevent computational saturation. This latency separation, with fast homeostasis (<1 s) versus slow adaptation (100 s), enables real-time stability without sacrificing strategic model refinement.

Drost and Matano further showed that clustered regularly interspaced short palindromic repeats (CRISPR)-edited human 3D epithelial cultures can be integrated into the same control architecture to model mutational signatures of cancer and colorectal tumorigenesis, supplying a genetically programmable tissue layer for multi-sensor chips.[80,81]

Porous membranes govern barrier integrity, absorption, migration, and inflammatory signaling. Pore diameter (0.4–8 μm), density, thickness (10–50 μm), and surface chemistry collectively set trans-membrane resistance and cell adhesion, yet empirical membrane selection introduces batch-to-batch scatter.[82] Convolutional Neural Networks (CNNs) learn the non-linear map between pore descriptors and impedance, enabling entire membrane libraries to be screened in silico within seconds. Regression models couple plasma treatment power to contact angle, permitting surface energy to be tuned on demand and converting membrane fabrication from craft to reproducible parameter space.[41,43] To quantify these microfluidic dynamic parameters and membrane interfaces, CNN, deep learning (DL) and ML approaches are increasingly utilized for these quantification tasks. For instance, DL quantification of chemotaxis moves beyond static endpoint measurements, standardizing trajectory-level metrics (migration path, directionality and persistence), while also incorporating uncertainty bounds. Chen and coworkers, also Zhao's group, employed human 3D epithelial cultures to demonstrate that such metrics quantify cell-cell interactions and directly guide personalized therapy in advanced breast cancer.[83,84] ML interpretation of impedance and electrical signals further supplies high-frequency state estimates, allowing membrane barrier function to be coupled back to fluidic control loops. Once structure, operation and interface are captured in a single model, data integration and knowledge discovery replace the traditional "high-throughput, high-noise" deadlock. Yet, true adaptive experimentation demands more than optimized components, and it requires systematic synthesis of multi-modal data streams into actionable, self-correcting decisions.

3.3. High-throughput data analysis and closed-loop decision making for OoCs

While Sections 3.1 and 3.2 establish how AI optimizes individual chip components, the unique value of OoCs lies in synthesizing multi-modal data streams into actionable decisions, transforming "high-throughput, high-noise" experiments into adaptive, self-correcting systems. Conventional platforms set one condition and read one endpoint, yet real physiology and drug responses are dynamic. Early events (cell contact, infiltration, or killing kinetics) often determine the outcome in immune cocultures or anti-tumor assays, making temporal resolution essential for truly adaptive experimentation.

OoCs natively generate heterogeneous data: imaging delivers spatiotemporal architecture, omics uncover molecular mechanisms, and sensors report continuous physiological states. Yet, "multi-modal" does not guarantee "interpretable." Without unified representations and trusted benchmarks, studies collapse into irreproducible storytelling. Wong and coworkers developed a lung-cancer-intravasation-on-chip with ML-based automatic quantification of rare metastatic events, fusing high-content imaging with computational flow cytometry to resolve tumor-immune dynamics.[44] This exemplifies how OoC-native, multi-modal data ensembles can link microfluidic control, biosensing, and AI-orchestrated interpretation to transform raw data streams into mechanistic insights unavailable in static organoid culture.

At the integration level, graph transformers and self-supervised learning provide pathways to fuse these heterogeneous data streams. For instance, radiomics quantifies tissue structural heterogeneity with 3D CNNs, achieving intraclass correlation coefficient (ICC) > 0.8

for repeatability benchmarking,[85] while LSTM-based surrogate models for predicting polymerase chain reaction (PCR) amplification endpoints, or CNN-based surrogate models for quantifying Western blot band intensities, shorten iteration cycles without replacing validation.[46] Computer vision pipeline studies further show that harmonized endpoints, versioned parameters, and public evaluation sets greatly improve inter-lab comparability.[43] Publicly released "good-*vs*-bad" annotated OoCs image datasets provide a common training and QC ground, freeing automated QC from single lab private data.

True adaptive experimentation demands a three-layer architecture (Figure 3c): (i) GNNs fuse chemical structures, transcriptomes, and imaging features to predict drug responses, with error and uncertainty guiding the next condition to be tested, as demonstrated by multi-modal graph transformers that integrate heterogeneous molecular and imaging data;[46] (ii) Bayesian online learning, mirroring the Bayesian network approaches used in clinical survival prediction,[86] continuously refines disease models by updating probability distributions with each new sample, enabling patient-specific adaptation while monitoring for model drift; and (iii) key biological events are converted into computable metrics (*e.g.*, invasion or intravasation events are counted automatically to remove human bias). The foundational capabilities for this architecture are emerging: Shin and coworkers established an anoxic–oxic interface chip for longitudinal coculture of obligate anaerobic gut microbes with human epithelium,[45] providing stable multi-species data streams; Kim and colleagues developed a gut-microbiome-chip with peristalsis-like flow,[78] demonstrating programmable dynamic control; and Wong and coworkers' lung-cancer-intravasation-on-chip with ML-based automatic quantification[44] shows that rare events can be computationally tracked to enable feedback-driven experimentation. At this point, OoCs are already designable, operable, and decision-capable prototypes. However, merging the biological fidelity of 3D tissues with chip controllability still requires organoid-on-chip level evidence where these three layers are fully integrated, a corpus that remains limited, making benchmark establishment imperative before wider adoption.

4. Organoids-on-chips: integration framework

The convergence of organoids and OoCs represents a paradigm shift from choosing between "biological authenticity" and "engineering controllability" toward synergistic platforms that leverage self-organization within precisely controlled microenvironments. Critically, organoids excel at recapitulating native tissue architecture, cellular heterogeneity, and organ-specific function, whereas OoCs excel at reconstituting the dynamic microenvironment (shear, gradients, mechano-stimulation) and probing multi-organ interactions. Therefore, integrating the two, preserves the "tissue realistic" biology of organoids while embedding it in a "physiology realistic" context, creating an evidentiary continuum from single organ physiology to systemic pharmacology.

Pioneering work by Takebe *et al*. first demonstrated that iPSC-derived hepatic organ buds possess intrinsic capacity to engraft, vascularize, and perform liver functions *in vivo*, establishing the biological feasibility of organoid-based liver regeneration.[87] Critically, the translation of this approach to scalable, reproducible platforms required explicit engineering of boundary conditions: subsequent studies incorporated interstitial perfusion through 3D porous scaffolds to enhance nutrient delivery and waste removal,[87,88] co-culture microcompartments for endothelial-mesenchymal interactions,[89] and mechanical conditioning matching physiological liver stiffness to promote maturation.[88] These engineered parameters enabled mass production with functional consistency (CV < 15% for albumin secretion),[88] while Camp's group elucidated how spatiotemporally patterned signaling gradients govern multi-lineage cellular crosstalk during bud maturation.[89] More recently, microfluidic adaptation of the organ bud protocol has demonstrated real-time monitoring of vascular network formation under controlled shear stress,[85,90] bridging the gap between *in vivo* transplantation and *ex vivo* chip-based maturation. These studies exemplify how coupling self-organizing tissue with quantifiable engineering boundary conditions, including perfusion rates, matrix stiffness, and gradient geometries, can simultaneously enhance batch-to-batch reproducibility and accelerate functional maturation, offering an MAE route with greater clinical translatability.

Yet, integration introduces new complexity, that organoid variability must be quantified and stabilized before chip-based control becomes effective, while chip parameters (channel height, pore size, membrane stiffness) must be reoptimized for 3D tissue requirements rather than for 2D monolayers. AI serves as the orchestration layer that bridges these complementary strengths and compensates for their respective limitations, unifying heterogeneous readouts through standardized, cross-comparable measurement pipelines. This enables the compression of design cycles via Bayesian optimization and maintenance of homeostasis through closed-loop decision systems, thereby turning biological variability into computable uncertainty and converting passive observation into

interpretable intervention. To operationalize this integration, we frame the discussion around three pillars: measurement standardization, evaluation benchmarks, and regulatory pathways. These pillars transform hybrid organoid-on-chip systems from experimental curiosities to credible evidence generators.

4.1. Phenotype compilers for standardized quantification

The standardized phenotyping pipeline, which we term a "phenotype compiler" here, is the prerequisite for merging organoids with OoCs. It is an AI-orchestrated measurement pipeline that converts raw multimodal outputs (*e.g.*, 3D images, transcriptomes, barrier readouts, secretion rates) into standardized, cross-comparable, and uncertainty-quantified phenotypic vectors. Chip-based controllers can only operate on trustworthy feedback once this compiler has eliminated batch-to-batch heterogeneity, such as morphologically divergent organoids derived from the same cell line or patient. Moreover, when the compiler encodes microenvironmental parameters (*e.g.*, shear, oxygen gradients, drug exposure kinetics) into the same phenotypic space, "tissue intrinsic" and "tissue contextual" data can be linked into a portable evidence chain across laboratories. In short, the compiler is the "metrology layer" of the hybrid platform, forcing biological authenticity and engineering controllability to speak the same numeric language.

The core challenge is that organoid phenotypes are exquisitely sensitive to measurement context, yet standardized quantification frameworks remain underdeveloped. Driehuis and coworkers demonstrated that oral mucosal organoids from the same patient exhibited divergent growth trajectories and drug sensitivities depending on culture passage number, with morphometric variance exceeding 40% between early and late passages.[28] Broutier and colleagues revealed that primary liver cancer organoids from identical tumors displayed heterogeneous budding patterns and lumen morphologies that correlated with divergent transcriptomic states, confounding drug response interpretation when simple viability averaging was applied.[61] Sachs and coworkers established a breast cancer organoid biobank where inter-organoid morphological heterogeneity within single patient-derived lines necessitated single-organoid resolution for accurate drug profiling.[59] Collectively, these studies demonstrate that without a unified measurement grammar capturing multidimensional morphological and molecular phenotypes, the same biological model yields operationally contradictory conclusions depending on measurement methodology.

The phenotype compiler addresses this challenge through a three-layer hierarchical pipeline that progressively transforms raw multimodal data into standardized, uncertainty-quantified phenotypic vectors (Figure 4a). This layered architecture systematically eliminates batch-to-batch heterogeneity, yet each layer introduces distinct uncertainty sources requiring explicit mitigation. Layer 1 uncertainty propagation: minute boundary displacements cause negligible volume error but substantial fractal dimension deviation,[46,91] which can be mitigated by uncertainty-aware feature weighting—down-weighting surface complexity variance and prioritizing robust descriptors (principal axis lengths).[47,48] Layer 2 temporal misalignment: imaging-transcriptomic timestamp mismatches (typically hours apart) are reduced to minutes via growth trajectory prediction.[62,63] Layer 3 model-form uncertainty: architecturally diverse deep ensembles (CNN, transformer, graph-based) capture epistemic uncertainty beyond parameter variance.[47,48] Empirical validation demonstrates that the three-layer architecture substantially reduces phenotypic variance compared to uncontrolled measurements, though layer-specific contributions vary by organoid type and experimental context.[53,61,87]

Layer 1 addresses image-based quantification. Traditional 2D microscopy captures only partial information, whereas light sheet fluorescence microscopy and OCT provide 3D volumetric data, but manual annotation is prohibitively time-consuming. As detailed in Section 2.4, foundation models (SAM/biomedical variants) provide zero-shot segmentation capabilities.[32-36] In OoCs contexts, these models require adaptation for 3D volumetric data from light-sheet fluorescence microscopy and OCT, where hierarchical segmentation algorithms first identify organoid boundaries, then resolve crypt-villus architectures, enabling quantification of lumen formation, epithelial thickness, and budding frequency as inputs for chip-based control systems.[92] For intestinal organoids containing crypt villus structures, hierarchical segmentation algorithms first identify organoid boundaries, then segment individual crypts and lumens, enabling quantification of crypt budding frequency, lumen area, and epithelial thickness.[46,93]

Layer 2 extracts quantitative features from segmented structures. Beyond basic morphometrics (diameter, volume, aspect ratio), advanced features include texture descriptors (Haralick features from gray-level co-occurrence matrices), shape-complexity metrics (fractal dimension, skeleton-based features), and dynamic features from time-lapse sequences (growth rate, deformation tracking). DL-based feature extractors capture morphological patterns invisible to traditional feature engineering, including crypt

budding polarity (asymmetric budding indicating Wnt/Notch dysregulation), lumen eccentricity (deviation from circularity correlating with epithelial tension and barrier dysfunction), and pericryptal fibroblast density gradients (texture features quantifying mesenchymal-epithelial crosstalk preceding TEER changes by 24–48 h).[53,56,57]

Layer 3 performs cross-modal integration and uncertainty quantification. Phenotypes from imaging must be correlated with molecular readouts (*e.g.*, transcriptomics, proteomics, metabolomics) and functional assays (*e.g.*, barrier integrity, secretion rates, drug response). Building on Section 2.4's technical framework, regulatory-grade evidence generation requires uncertainty quantification to meet evidentiary standards for clinical and industrial applications. Here, deep ensembles extend Monte Carlo dropout by quantifying model-form uncertainty across multiple architectural hypotheses, enabling confidence intervals on phenotypic measurements.[47-49]

4.2. Evaluation of benchmarks and reference standards

Without standardized benchmarks, progress in organoid and organ-on-chip research becomes difficult to assess, and cross-laboratory comparisons remain unreliable. The community has recognized this gap, leading to initiatives that establish reference datasets, evaluation protocols, and performance metrics for algorithm validation and inter-laboratory harmonization.[43]

Benchmark construction requires three components (Figure 4b). First, reference datasets with expert annotations and quality-controlled metadata. For organoid segmentation, this includes diverse organoid types (*e.g.*, cerebral, intestinal, hepatic, pancreatic, renal), imaged under multiple modalities (brightfield, fluorescence, confocal, OCT, holotomography) with pixel-wise ground truth annotations. Montes-Olivas and coworkers established a benchmark dataset of intestinal organoids with synchronized brightfield, fluorescence, and

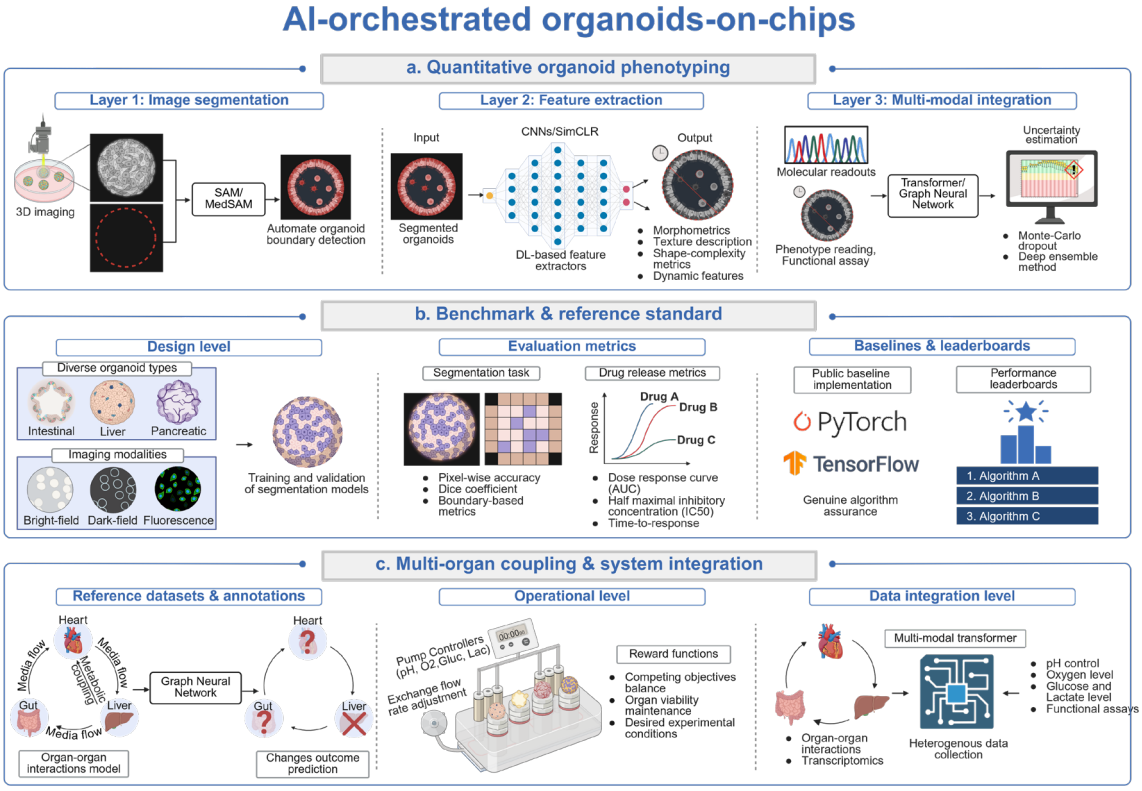


Figure 4. Three domains of AI integration in organoid-on-a-chip systems: (a) Quantitative phenotyping using advanced 3D segmentation and multimodal feature extraction. (b) Benchmarks and reference standards to ensure algorithmic performance and reproducibility. (c) Multi-organ coupling driven by GNNs and transformers for predictive modeling of drug responses and disease progression.
Abbreviations: 3D: Three-dimensional; AI: Artificial intelligence; CNN: Convolutional neural networks; DL: Deep learning; SAM: Segment anything model; SimCLR: Simple framework for contrastive learning of visual representations.

confocal imaging, providing hierarchical annotations at organoid, crypt, and single-cell levels.[46] Cho and colleagues developed a low-coherence holotomography dataset with paired 3D refractive index maps and fluorescence ground truth, enabling label-free segmentation benchmarking.[62] Wang and collaborators released an OCT dataset of human organoids with deep learning-based segmentation benchmarks for non-invasive longitudinal tracking.[63] Marks and coworkers compiled CellSAM training data spanning nine microscopy modalities with standardized annotation protocols for cross-modality generalization testing.[32]

Second, standardized evaluation metrics that capture both accuracy and clinical relevance. For segmentation tasks, beyond pixel-wise accuracy and Dice coefficient, boundary-based metrics (Hausdorff distance, average surface distance) assess the precision of organoid delineation.[57,58] For drug response, metrics should include area under the dose-response curve (AUC), half maximal inhibitory concentration (IC_{50}), and time-to-response, with confidence intervals reported.[94-96] Conflicting rankings across AUC, IC_{50}, and time-to-response are resolved via Pareto-dominant assessment with context-dependent weights—oncology prioritizes rapid kill, chronic modeling sustained exposure.[91,97] Persistent conflicts undergo value-of-information analysis: mechanistic investigation when VOI exceeds experimental cost, otherwise pre-defined clinical priorities apply.[95,96] This ensures transparent, auditable regulatory criteria. Chen and colleagues demonstrated the payoff of strict standardization: once all centers adopted identical culture medium, imaging settings, and dose-response analysis pipelines, the inter-site coefficient of variation for IC_{50} collapsed from 45% to 12%, turning cross-laboratory comparisons into statistically meaningful readouts.[97]

Third, regarding baseline implementations and leaderboards, public baseline implementations (*e.g.*, in PyTorch or TensorFlow) ensure that reported improvements reflect genuine algorithmic advances rather than implementation details.[32] Dynamic leaderboards track state-of-the-art performance across tasks, motivating continued innovation while providing clear performance targets.[33] Rolling benchmarks quarterly refresh evaluation data from unseen domains, with excessive domain transfer degradation triggering model revision.[32,33] Uncertainty-aware composite scoring penalizes high-variance predictions, calibrated to assay criticality.[47,48] Provenance verification mandates full reproducibility documentation for inclusion,[98] while human-in-the-loop adjudication reserves expert review for low-confidence predictions.[99] Borten and colleagues' OrganoSeg platform exemplifies this: open-source distribution of standardized segmentation parameters and batch-processing workflows enabled reproducible brightfield morphometry across independent laboratories.[100]

4.3. Multi-organ coupling and system integration

Single organ models, whether organoids or OoCs, cannot capture systemic effects, such as drug metabolism, immune cell trafficking, or endocrine signaling. Multiorgan OoC systems address this limitation by physically or functionally connecting multiple organ models. Scaling to multi-organ systems introduces substantial complexity in design, operation, and data interpretation. The necessity of such integration is underscored by longstanding challenges in modeling organ crosstalk: for instance, the gut-liver axis in drug metabolism and the bone-vascular axis in skeletal homeostasis, where physiological interdependence is well recognized yet experimentally inaccessible through reductionist approaches. Traditional approaches to validate these multi-organ relationships rely heavily on animal models or clinical observations, entailing protracted research cycles and limited mechanistic resolution. Recent advances demonstrate how organoid-on-chip platforms can compress these timelines. A gut-liver-on-a-chip system revealed that microbial-derived metabolites directly modulate hepatic drug-metabolizing enzyme expression, establishing a causal link previously inferred only from population pharmacokinetic studies.[85] These examples illustrate how integrated microphysiological systems can translate relevant clinical hypotheses into testable mechanistic models, accelerating the validation of systemic physiological relationships that are difficult to achieve using single-organ approaches.

AI provides tools for managing this complexity at multiple levels (Figure 4c). At the design level, GNNs can model organ-organ interactions, predicting how changes in one organ (*e.g.*, liver metabolism) affect others (*e.g.*, heart electrophysiology). The nodes represent organ models with their internal states, while edges represent physical connections (medium flow) or functional relationships (metabolic coupling). At the operational level, reinforcement learning can optimize flow rates and medium composition to maintain homeostasis across all organs. The state space includes sensor readings from each organ (*e.g.*, pH, oxygen, glucose, lactate), while actions adjust pump speeds and medium exchange rates. The reward function balances competing objectives, maintaining viability in all organs while achieving desired experimental conditions (*e.g.*, drug concentration at therapeutic levels in the target organ). At the data integration level, multi-modal transformers can fuse heterogeneous data streams into unified representations. For a liver-kidney chip, this

includes hepatocyte albumin secretion and urea synthesis rates from the liver compartment; glomerular filtration rate and tubular secretion from the kidney compartment; plus circulating drug/metabolite concentrations from liquid chromatography mass spectrometry (LCMS) analysis. The integrated representation enables prediction of systemic pharmacokinetic parameters (clearance, volume of distribution, bioavailability) that can be validated against clinical data.[101] A notable example demonstrating AI's indispensable role is the development of gut-liver-on-chip systems for oral drug bioavailability prediction. By coupling intestinal absorption (Caco-2 or intestinal organoids) with liver metabolism (hepatocyte spheroids or iPSC-derived hepatocytes), these systems can predict first-pass metabolism effects. Physiologically-based pharmacokinetic (PBPK) modeling achieves R^2 = 0.78 – 0.85 when augmented by ML-mediated multi-modal fusion; without ML integration, physiological coupling alone yields R^2 = 0.65–0.70, insufficient for FDA evidentiary standards.[102,103] The cost implication is substantial: phase I failures due to pharmacokinetic issues (including poor bioavailability) represent a significant contributor to the ~$2.6 billion average cost per approved drug.[104,105] AI-enhanced chip prediction could has the potential to reduce such failures based on retrospective validation.[106] ML extensions of these models, integrating drug physicochemical properties with chip-derived absorption and metabolism kinetics, promise further improvement in first-pass effect prediction, though large-scale clinical validation remains ongoing.

5. Challenges and perspectives

Although the "organoid × organ-on-chip × AI" integration framework is emerging, its entry into mainstream drug development pipelines within the next few years hinges on three bottlenecks: technical reproducibility, data credibility, and regulatory acceptability. Technically, organoids from the same human iPSC line can exhibit substantial variation in morphogenetic trajectories and functional maturation across production batches and laboratories. On the chip side, pump drift, bubble nucleation, and material batch-to-batch variation can introduce substantial error in shear stress (often exceeding 10–20%),[23,24,82] sufficient to alter barrier integrity readouts. Data-wise, fewer than five publicly available multi-modal "organoid-on-chip" datasets exist, cross-site imaging protocols, metadata formats, and uncertainty annotations remain nonuniform, creating a "same look, different numbers" dilemma.[65] Regarding the regulatory aspect, while Food and Drug Administration (FDA)/European Medicines Agency (EMA) already accept some organoid data as part of Investigational New Drug (IND) packages, guidance specific to "AI-Orchestrated + multi-organ + closed-loop" composite platforms is still missing, and the four operational pillars of MAE—validation benchmarks, uncertainty budgets, end-to-end traceability, and mechanistic interpretability—have yet to be codified.[50] These bottlenecks are not independent but share a common root: the fundamental mismatch between canonical machine learning paradigms, optimized for abundant, standardized data (*e.g.*, ImageNet, clinical imaging), and biological systems characterized by irreducible variance, small sample sizes, and context-dependent generalization. Before articulating the MAE four-pillar implementation roadmap, we must therefore examine how AI's epistemic constraints, knowledge boundaries that limit what regulatory-scientific claims can legitimately be made, shape the feasibility of transforming organoid and OoC platforms from experimental tools into credible evidence generators. These constraints manifest across four interconnected dimensions that directly inform pillar design.

The MAE framework's feasibility hinges on acknowledging that AI in biology faces fundamental knowledge boundaries, epistemic constraints, that differ qualitatively from canonical ML domains such as computer vision. Organoid and OoC data suffer from irreducible biological variance compounded by technical variance, with fewer than five publicly available multi-modal datasets globally and severe demographic skew in existing iPSC lines.[23,24,28,107] This creates a small-data-high-dimensionality paradox: >20 interacting medium factors with <500 experimental runs render Bayesian surrogates vulnerable to >40% predictive variance at unexplored formulations.[19,20] Deep learning architectures with millions of parameters trained on small samples exhibit computational irreproducibility (cosine similarity 0.72–0.89 across U-Net initializations),[108,109] while foundation models suffer 15–30% Dice degradation across laboratories due to domain shift.[32-36] Beyond technical barriers, pharmaceutical-scale implementation requires substantial qualification investment per organ system, organizational restructuring into hybrid computational-experimental teams,[110] resolution of workforce training gaps,[111] and clear IP governance frameworks for multi-institutional AI model deployment.

Therefore, an MAE four-pillar implementation roadmap for 2026–2030 is illustrated by early exemplars (Table 2). Pillar 1 pertains to model validation.[112,113] Platforms must match gold standards within a predefined "context of use" versus clinical endpoints and animal toxicokinetics/toxicology data, followed by cross-laboratory replication at three or more sites.[112] Under locked protocols, multicenter hepatocyte screening exercises have reported substantial inter-laboratory coefficient of variation (CV) reduction,[112]

though such improvements typically reflect controlled conditions using identical iPSC lines with synchronized passages.[107] Patient-derived organoids introduce additional biological variance that may elevate the technical floor above controlled-condition baselines,[28,59] suggesting that initial qualification targets should account for this irreducible heterogeneity. Dynamic microenvironments further contribute shear stress variability,[23,24] necessitating active compensation algorithms that remain to be prospectively validated. If multi-site consortia formation is delayed, Pillar 1 can operate independently through existing drug development tool (DDT) infrastructure for single-endpoint assays (*e.g.*, hepatotoxicity)[113] without requiring full multi-organ integration.

Pillar 2 addresses uncertainty quantification.[86,87,114]All predictions must carry confidence intervals for safety-critical assays (acute toxicity, pro-arrhythmia), with false negative rates maintained below 5%. Global sensitivity analysis ranks input parameters and establishes quality control limits. Preliminary modeling suggests that embedding Bayesian posterior networks into vascularized liver-bud models may compress survival prediction intervals considerably compared to uncontrolled baselines. The regulatory timeline for novel biomarkers with added AI scrutiny typically spans multiple years;[113,115] for context, EMA Innovation Task Force (ITF) scientific advice procedures often require extended durations.[51] The integrated timeline proceeds as follows: In Phase 1 (Years 1–2), sponsors submit DDT Stage 1 applications with analytical validation across three or more sites (CV target below 20%).[112,115] In Phase 2 (Years 2–3), retrospective clinical correlation with substantial compound libraries is conducted alongside prospective blinded prediction, establishing physiologically-based pharmacokinetic coupling per International Council for Harmonisation (ICH) guideline M12.[115,116] In Phase 3 (Years 3–4), investigational new drug (IND) or new drug application (NDA) integration occurs with predefined performance

Table 2. Mapping of MAE pillars to existing regulatory pathways

MAE pillar	Pillar 1: validation	Pillar 2: uncertainty	Pillar 3: provenance	Pillar 4: interpretability
Core requirement	Inter-lab reproducibility; clinical correlation	Confidence intervals; false-negative control	Audit trails; version control	Mechanistic grounding; SHapley Additive exPlanations (SHAP) attributions
Food and Drug Administration (FDA) pathway	Drug Development Tool (DDT) Qualification (21st Century Cures Act)	Emerging Technology Program (ETP) –Chemistry, Manufacturing, and Controls (CMC) review; FDA Artificial Intelligence (AI)/Machine Learning (ML) guidance	ETP – data integrity (ALCOA+: Attributable, Legible, Contemporaneous, Original, Accurate + Complete, Consistent, Enduring, Available)	ETP – process understanding
Model-as-Evidence (EMA) pathway	Qualification of Novel Methodologies	Innovation Task Force (ITF) scientific advice; EMA AI reflection paper	Good Manufacturing Practice (GMP) Annex 11; EU GMP Chapter 4	EMA AI reflection paper (human oversight)
International Conference on Harmonisation (ICH) guideline	ICH S5(R3): Safety testing	ICH Q9(R1): Quality risk management	ICH E6(R2): Good Clinical Practice (GCP) data integrity	ICH M12: Drug interactions
Example context of use	Hepatotoxicity screening for Investigational New Drug (IND)-enabling studies	Risk-based control strategy for cardiotoxicity assays	Complete audit trail for IND submission	Mechanistic Drug-Drug Interactions (DDI) prediction for labeling

Notes: The four-pillar structure explicitly positions MAE within established regulatory pathways rather than creating parallel requirements. Pillar 1 maps onto FDA's Drug Development Tool (DDT) qualification program under the 21st Century Cures Act, where inter-laboratory CV < 20% mirrors FDA's published acceptance threshold for *in vitro* toxicology assays;[112,113] Pillar 2 aligns with ICH Q9(R1) quality risk management, requiring explicit prediction intervals for safety-critical endpoints in IND submissions;[122] Pillar 3 satisfies FDA's ALCOA+ data integrity guidance and EMA's GMP Annex 11 via tamper-proof audit trails; Pillar 4 addresses FDA Emerging Technology Program (ETP) expectations and EMA's 2024 AI reflection paper mandating human oversight.[50,51,112] We propose submitting MAE packages as ETP briefing documents with pre-defined Context of Use (CoU).
Abbreviations: AI: Artificial intelligence; ALCOA+: Attributable, legible, contemporaneous, original, accurate, complete, consistent, enduring, available; CMC: Chemistry, manufacturing, and controls; DDT: Drug development tool; EMA: European Medicines Agency; ETP: Emerging technology program; FDA: Food and Drug Administration; GMP: Good Manufacturing Practice; ICH: International Council for Harmonisation; IND: Investigational new drug; ITF: Innovation Task Force; NDA: MAE: Model-as-evidence; ML: Machine learning; New drug application; SHAP: SHapley additive exPlanations.

criteria (sensitivity and specificity targets exceeding 85% and 90%, respectively, with negative predictive value above 95%).[115] Relaxed confidence intervals may be accepted for non-critical endpoints during initial adoption phases.

Pillar 3 relates to data provenance.[98] Every step from raw images to AI model weights must be time-stamped, hashed, and versioned. Blockchain or Git-like signed commits provide tamper-proof records; versioned pipelines (RO-Crate + ML Flow) suffice for most submissions.[98,117] However, publicly available multi-modal organoid-on-chip datasets remain limited in number,[65] and cross-site protocol nonuniformity creates "same look, different numbers" dilemmas. As a fallback, process analytical technology[20] through in-line monitoring of glucose, lactate, and dissolved oxygen provides real-time quality flags, potentially increasing effective sample sizes and mitigating small-data constraints.[20,28]

Pillar 4 pertains to interpretability.[118] Black-box predictions are insufficient; models must highlight which features, pathways, or parameters drive conclusions, with causal inference separating correlation from causation. Emerging evidence from pancreatic cancer organoid studies combining SHapley Additive exPlanations (SHAP) attributions with intervention experiments demonstrates the potential for mechanistic insight validation through transcriptomic perturbation.[119,120] Nevertheless, deep learning on small samples exhibits computational irreproducibility,[121] and foundation models may suffer substantial performance degradation across laboratories due to domain shift.[32,33] To address this limitation, architecturally diverse deep ensembles (CNN, transformers, graph-based models) capture model-form uncertainty beyond parameter variance,[47,48] providing robustness against single-architecture failures.

6. Conclusions

Distinct from prior reviews,[4,17] the framework presented herein maps algorithmic choices to biological bottlenecks where conventional methods face combinatorial, temporal, or bandwidth infeasibility (**Table 1**), elevating organoid and OoC systems from supportive adjuncts to primary evidence generators contingent upon the four-pillar validation architecture (Section 5). AI serves as the orchestration layer that converts biological fidelity (organoids) and engineering controllability (OoCs) into quantitative, cross-comparable evidence, positioning human researchers to focus on higher-order hypothesis generation and translational strategy rather than parameter tuning. AI eliminates three long-standing bottlenecks: high-dimensional design spaces that transcend manual screening, batch-to-batch variability that prevents result accumulation, and data torrents that overwhelm human analysts. On the organoid side, multi-objective optimization converts culture medium and ECM selection into an engineering discipline, significantly reduce medium costs, enhance batch consistency, improve ECM matching, and minimize the number of experiments. On the OoC side, generative topology optimization enables inverse co-design of materials and geometry: DRL closed loop control maintains very low steady state error while significantly shortening design cycles and enhancing multi-organ system's accuracy in predicting human drug response. At the integration level of organoids-on-chips, AI-orchestrated phenotype compilers, cross-site benchmarks, and an MAE regulatory framework now place "biological authenticity" and "engineering controllability" in the same coordinate system. Looking to 2030, a "validation-before-scale" roadmap is articulated. Once realized, this computable evidence factory will turn the "*valley of death*" into a data highway, delivering safer and more effective therapies to patients faster than ever before.

Abbreviations

AI:	Artificial Intelligence
ECM:	Extracellular Matrix
TEER:	Transepithelial/ Transendothelial Electrical Resistance
EGF:	Epidermal Growth Factor
FGF:	Fibroblast Growth Factor
GNN:	Graph Neural Networks
HGF:	Hepatocyte Growth Factor
SAM:	Segment Anything Model
GANs:	Generative Adversarial Networks
VAEs:	Variational Autoencoders
CFD:	Computational Fluid Dynamics
CNNs:	Convolutional Neural Networks
LCMS:	Liquid Chromatography Mass Spectrometry
iPSC:	induced Pluripotent Stem Cells
CRISPR:	Clustered Regularly Interspaced Short Palindromic Repeats
OoCs:	Organs-on-Chips
MAE:	Model-as-Evidence
IND:	Investigational New Drug
FDA:	Food and Drug Administration
EMA:	European Medicines Agency
CV:	Coefficient of Variation

ML:	Machine Learning
2D:	Two-Dimensional
3D:	Three-Dimensional
PID:	Proportional-Integral-Derivative Controller
DRL:	Deep Reinforcement Learning

Acknowledgments

None.

Funding

This work was supported in part by: (1) CUHK Peter Hung Pain Research Institute (to ZAL and RST, PHPRI/2024/122); (2) InnoHK Center for Neuromusculoskeletal Restorative Medicine (to RST and ZAL), under the Health@InnoHK program, Innovation and Technology Commission (ITC), Hong Kong SAR; (3) National Natural Science Foundation of China (to ZAL, 82302753); (4) Hong Kong Research Grants Council (to ZAL, 24203523); and (5) the Mainland-Hong Kong Technology Cooperation Funding Scheme (MHKTCFS) of ITC, Hong Kong SAR (to RST and ZAL, project #GHP-260-23SZ and project #MHP/101/23; to ZAL, project #GHP/140/22GD). ZAL acknowledges support from the CUHK Vice-Chancellor Early Career Professorship Scheme. RST is supported by the CUHK Lee Quo Wei and Lee Yick Hoi Lun Professorship in Tissue Engineering and Regenerative Medicine. The funders had no role in study design, data collection and analysis, decisions to publish, or preparation of the manuscript.

Conflict of interest

Zhong Alan Li serves as the Editorial Board Member of this journal, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Conceptualization: Peixi Wang, Liangbin Zhou, Yuwei Zhang, Yichi Zhang, Shing Yui Ho, Rocky S. Tuan, Zhong Alan Li

Visualization: Peixi Wang, Berenica Santoso

Writing–original draft: Peixi Wang, Berenica Santoso, Songlin He

Writing–review & editing: Peixi Wang, Rocky S. Tuan, Zhong Alan Li

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Parrish MC, Tan YJ, Grimes KV, Mochly-Rosen D. Surviving in the valley of death: opportunities and challenges in translating academic drug discoveries. *Annu Rev Pharmacol. Toxicol.* 2019;59(1):405-421.

   doi: 10.1146/annurev-pharmtox-010818-021625

2. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483(7391):531-3.

   doi: 10.1038/483531a

3. Clevers H. Modeling development and disease with organoids. *Cell*. 2016;165(7):1586-1597.

   doi: 10.1016/j.cell.2016.05.082

4. Zhou L, Chen S, Liu J, *et al.* When artificial intelligence (AI) meets organoids and organs-on-chips (OoCs): Game-changer for drug discovery and development? *Innov Life.* 2025;3(1):100115.

   doi: 10.59717/j.xinn-life.2024.100115

5. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sin B.* 2022;12(7):3049-3062.

   doi: 10.1016/j.apsb.2022.02.002

6. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.

   doi: 10.1371/journal.pmed.0020124

7. Mullard A. Parsing clinical success rates. *Nat Rev Drug Discov*. 2016;15(7):447-448.

   doi: 10.1038/nrd.2016.136

8. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10(9):712-712.

   doi: 10.1038/nrd3439-c1

9. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLoS Biol.* 2015;13(6):e1002165.

   doi: 10.1371/journal.pbio.1002165

10. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533:452-454.

    doi: 10.1038/533452a

11. Errington TM, Iorns E, Gunn W, Tan FE, Lomax J, Nosek BA. An open investigation of the reproducibility of cancer

biology research. *eLife*. 2014;3:e04333.

doi: 10.7554/eLife.04333

12. Drost J, Clevers H. Organoids in cancer research. *Nat Rev Cancer* 2018;18(7):407-418.

doi: 10.1038/s41568-018-0007-6

13. Sato T, Vries RG, Snippert HJ, *et al*. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*. 2009;459(7244):262-265.

doi: 10.1038/nature07935

14. Sato T, Stange DE, Ferrante M, *et al*. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology*. 2011;141(5):1762-1772.

doi: 10.1053/j.gastro.2011.07.050

15. Huh D, Matthews BD, Mammoto A, Montoya-Zavala M, Hsin HY, Ingber DE. Reconstituting organ-level lung functions on a chip. *Science*. 2010;328(5986):1662-1668.

doi: 10.1126/science.1188302

16. Bhatia SN, Ingber DE. Microfluidic organs-on-chips. *Nat Biotechnol*. 2014;32(8):760-72.

doi: 10.1038/nbt.2989

17. Ingber DE. Human organs-on-chips for disease modelling, drug development and personalized medicine. *Nat Rev Genet*. 2022;23(8):467-491.

doi: 10.1038/s41576-022-00466-9

18. Zhang B, Radisic M. Organ-on-a-chip devices advance to market. *Lab Chip*. 2017;17(14):2395-2420.

doi: 10.1039/C6LC01554A

19. Kandasamy K, Dasarathy G, Oliva J, Schneider J, Poczos B. Multi-fidelity gaussian process bandit optimisation. *J Artif Intell Res*. 2019;66:151-196.

doi: 10.1613/jair.1.11288

20. Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: A review of Bayesian optimization. *Proc IEEE* 2015;104(1):148-175.

doi: 10.1109/JPROC.2015.2494218

21. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. *Cell*. 2006;126(4):677-689.

doi: 10.1016/j.cell.2006.06.044

22. Trappmann B, Gautrot JE, Connelly JT, *et al*. Extracellular-matrix tethering regulates stem-cell fate. *Nat Mater*. 2012;11(7):642-649.

doi: 10.1038/nmat3339

23. Quang Tran D, Bae S-H. Proximal Policy Optimization Through a Deep Reinforcement Learning Framework for Multiple Autonomous Vehicles at a Non-Signalized Intersection. *Appl Sci*. 2020;10(16):5722.

doi: 10.3390/app10165722

24. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv*. Preprint posted online 2017.

doi: 10.48550/arXiv.1707.06347

25. Greggio C, De Franceschi F, Figueiredo-Larsen M, *et al*. Artificial three-dimensional niches deconstruct pancreas development in vitro. *Development*. 2013;140(21):4452-4462.

doi: 10.1242/dev.096628

26. Caliari SR, Burdick JA. A practical guide to hydrogels for cell culture. *Nat Methods*. 2016;13(5):405-14.

doi: 10.1038/nmeth.3839

27. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf. Process. Syst*. 2017;30. Accessed May 17, 2026. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

28. Driehuis E, Kolders S, Spelier S, *et al*. Oral mucosal organoids as a potential platform for personalized cancer therapy. *Cancer Discov*. 2019;9(7):852-871.

doi: 10.1158/2159-8290.CD-18-1522

29. Watson CL, Mahe MM, Múnera J, *et al*. An in vivo model of human small intestine using pluripotent stem cells. *Nat Med*. 2014;20(11):1310-1314.

doi: 10.1038/nm.3737

30. Camp JG, Badsha F, Florio M, *et al*. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc Natl Acad Sci USA*. 2015;112(51):15672-7.

doi: 10.1073/pnas.1520760112

31. Pollen AA, Bhaduri A, Andrews MG, *et al*. Establishing cerebral organoids as models of human-specific brain evolution. *Cell*. 2019;176(4):743-756. e17.

doi: 10.1016/j.cell.2019.01.017

32. Marks M, Israel U, Dilip R, *et al*. CellSAM: a foundation model for cell segmentation. *Nat Methods*. 2025.

doi: 10.1038/s41592-025-02879-w

33. Archit A, Freckmann L, Nair S, *et al*. Segment Anything for Microscopy. *Nat Methods*. 2025;22(3):579-591.

doi: 10.1038/s41592-024-02580-4

34. Ali M, Wu T, Hu H, *et al*. A review of the Segment Anything Model (SAM) for medical image analysis: Accomplishments and perspectives. *Comput Med Imaging Graph*. 2025;119:102473.

doi: 10.1016/j.compmedimag.2024.102473

35. Zhang Y, Shen Z, Jiao R. Segment anything model for medical image segmentation: Current applications and future directions. *Comput Biol Med*. 2024;171:108238.

doi: 10.1016/j.compbiomed.2024.108238

36. Zhang L, Deng X, Lu Y. Segment Anything Model (SAM) for Medical Image Segmentation: A Preliminary Review. *IEEE BIBM* 2023:4187-4194.

doi: 10.1109/BIBM58861.2023.10386032

37. Kundacina I, Kundacina O, Miskovic D, Radonic V. Advancing microfluidic design with machine learning: a Bayesian optimization approach. *Lab Chip*. 2025;25(4):657-672.

doi: 10.1039/D4LC00872C

38. Miller JS, Stevens KR, Yang MT, *et al*. Rapid casting of patterned vascular networks for perfusable engineered three-dimensional tissues. *Nat Mater*. 2012;11(9):768-774.

doi: 10.1038/nmat3357

39. Fu C-Y, Tseng S-Y, Yang S-M, Hsu L, Liu C-H, Chang H-Y. A microfluidic chip with a U-shaped microstructure array for multicellular spheroid formation, culturing and analysis. *Biofabrication*. 2014;6(1):015009.

doi: 10.1088/1758-5082/6/1/015009

40. Bein A, Shin W, Jalili-Firoozinezhad S, *et al*. Microfluidic Organ-on-a-Chip Models of Human Intestine. *Cell Mol Gastroenterol Hepatol*. 2018;5(4):659-668.

doi: 10.1016/j.jcmgh.2017.12.010

41. Mosadegh B, Lockett MR, Minn KT, *et al*. A paper-based invasion assay: Assessing chemotaxis of cancer cells in gradients of oxygen. *Biomaterials*. 2015;52:262-271.

doi: 10.1016/j.biomaterials.2015.02.012

42. Kolesky DB, Truby RL, Gladman AS, Busbee TA, Homan KA, Lewis JA. 3D Bioprinting of Vascularized, Heterogeneous Cell-Laden Tissue Constructs. *Adv Mater*. 2014;26(19):3124-3130.

doi: 10.1002/adma.201305506

43. Gracioso Martins AM, Wilkins MD, Ligler FS, Daniele MA, Freytes DO. Microphysiological System for High-Throughput Computer Vision Measurement of Microtissue Contraction. *ACS Sens*. 2021;6(3):985-994.

doi: 10.1021/acssensors.0c02172

44. Wong CWT, Lee JZX, Jaeschke A, *et al*. Lung cancer intravasation-on-a-chip: Visualization and machine learning-assisted automatic quantification. *Bioact Mater*. 2025;51:858-875.

doi: 10.1016/j.bioactmat.2025.06.028

45. Shin W, Wu A, Massidda MW, *et al*. A Robust Longitudinal Co-culture of Obligate Anaerobic Gut Microbiome with Human Intestinal Epithelium in an Anoxic-Oxic Interface-on-a-Chip. *Front Bioeng Biotechnol*. 2019;7:13.

doi: 10.3389/fbioe.2019.00013

46. Montes-Olivas S, Legge D, Lund A, *et al*. In-silico and in-vitro morphometric analysis of intestinal organoids. *PLoS Comput Biol*. 2023;19(8):e1011386.

doi: 10.1371/journal.pcbi.1011386

47. Thompson J, Koe R, Le A, Goodman G, Brown DS, Kuntz A. Early Failure Detection in Autonomous Surgical Soft-Tissue Manipulation via Uncertainty Quantification. *arXiv*. Preprint posted online 2025.

doi: 10.48550/arXiv.2501.10561

48. van Leeuwen PJ, Chiu JC, Yang C-KK. Uncertainty quantification for deep learning. *Environ Data Sci*. 2024;4.

doi: 10.1017/eds.2025.10027

49. Zeevi T, Venkataraman R, Staib LH, Onofrey JA. Monte-Carlo Frequency Dropout for Predictive Uncertainty Estimation in Deep Learning. *IEEE ISBI*. 2024:1-5.

doi: 10.1109/ISBI56570.2024.10635511

50. Han DY. Artificial Intelligence in and Beyond Healthcare Psychology. *J Clin Psychol Med Settings*. 2025;32(4):600-607.

doi: 10.1007/s10880-025-10101-4

51. Chauhan SB, Gaur R, Akram A, Singh I. Artificial Intelligence Driven insights for Regulatory Intelligence in Medical Devices: Evaluating EMA, FDA and CDSCO Frameworks. *Glob Clin Eng J*. 2025;7:11-24.

doi: 10.31354/globalce.v7i2.210

52. Barker N, Huch M, Kujala P, *et al*. Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. *Cell Stem Cell*. 2010;6(1):25-36.

doi: 10.1016/j.stem.2009.11.013

53. Huch M, Bonfanti P, Boj SF, *et al*. Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J*. 2013;32(20):2708-2721.

doi: 10.1038/emboj.2013.204

54. Conesa A, Madrigal P, Tarazona S, *et al*. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016/01/26 2016;17(1):13.

55. DOI:·10.1186/s13059-016-0881-8

56. Sunyer R, Conte V, Escribano J, *et al*. Collective cell durotaxis emerges from long-range intercellular force transmission. *Science*. 2016;353(6304):1157-1161.

doi: 10.1126/science.aaf7119

57. Workman MJ, Mahe MM, Trisno S, *et al*. Engineered human pluripotent-stem-cell-derived intestinal tissues with a functional enteric nervous system. *Nat Med*. 2017;23(1):49-59.

doi: 10.1038/nm.4233

58. Finkbeiner SR, Hill David R, Altheim Christopher H, *et al*. Transcriptome-wide Analysis Reveals Hallmarks of Human Intestine Development and Maturation In Vitro and In Vivo. *Stem Cell Rep*. 2015;4(6):1140-1155.

doi: 10.1016/j.stemcr.2015.04.010

59. Lindeboom RGH, van Voorthuijsen L, Oost KC, *et al*. Integrative multi-omics analysis of intestinal organoid differentiation. *Mol Syst Biol*. 2018;14(6):MSB188227.

doi: 10.15252/msb.20188227

60. Sachs N, De Ligt J, Kopper O, *et al*. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell*. 2018;172(1):373-386. e10.

doi: 10.1016/j.cell.2017.11.010

61. Vlachogiannis G, Hedayat S, Vatsiou A, *et al*. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science*. 2018;359(6378):920-926.

doi: 10.1126/science.aao2774

62. Broutier L, Mastrogiovanni G, Verstegen MM, *et al*. Human primary liver cancer-derived organoid cultures for disease modeling and drug screening. *Nat Med*. 2017;23(12):1424-1435.

doi: 10.1038/nm.4438

63. Cho J, Lee MJ, Park J, *et al*. Label-free, High-Resolution 3D Imaging and Machine Learning Analysis of Intestinal Organoids via Low-Coherence Holotomography. *J Vis Exp*. 2025;(222):e68529.

doi: 10.3791/68529

64. Wang B, Ganjee R, Khandaker I, *et al*. Deep learning based characterization of human organoids using optical coherence tomography. *Biomed Opt Express*. 2024;15(5):3112-3127.

doi: 10.1364/BOE.515781

65. Bao D, Wang L, Zhou X, Yang S, He K, Xu M. Automated detection and growth tracking of 3D bio-printed organoid clusters using optical coherence tomography with deep convolutional neural networks. *Front Bioeng Biotechnol*. 2023;11:1133090.

doi: 10.3389/fbioe.2023.1133090

66. Gu J, Liu F, Li L, Mao J. Advances and Challenges in Modeling Autosomal Dominant Polycystic Kidney Disease: A Focus on Kidney Organoids. *Biomedicines*. 2025;13(2):523.

doi: 10.3390/biomedicines13020523

67. Bukas C, Subramanian H, See F, *et al*. MultiOrg: a multi-rater organoid-detection dataset. In: Proceedings of the Advances in Neural Information Processing Systems 37. 2024; Vancouver, BC, Canada.

doi: 10.52202/079017-3036

68. Li Y, Zhang H, Xiang Z, Yuan Z. Predictive Modeling of Oxygen Gradient in Gut-on-a-Chip Using Machine Learning and Finite Element Simulation. *Appl Sci*. 2026;16(2):571.

doi: 10.3390/app16020571

69. Grigoryan B, Paulsen SJ, Corbett DC, *et al*. Multivascular networks and functional intravascular topologies within biocompatible hydrogels. *Science*. 2019;364(6439):458-464.

doi: 10.1126/science.aav9750

70. Grünewald TA, Liebi M, Wittig NK, *et al*. Mapping the 3D orientation of nanocrystals and nanostructures in human bone: Indications of novel structural features. *Sci Adv*. 2020;6(24):eaba4171.

doi: 10.1126/sciadv.aba4171

71. Shelat R, Bhatt LK, Paunipagar B, Kurian T, Khanna A, Chandra S. Regeneration of hyaline cartilage in osteochondral lesion model using L-lysine magnetic nanoparticles labeled mesenchymal stem cells and their in vivo imaging. *J Tissue Eng Regen Med*. 2020;14(11):1604-1617.

doi: 10.1002/term.3120

72. Ates GC, Gorguluarslan RM. Two-stage convolutional encoder-decoder network to improve the performance and reliability of deep learning models for topology optimization. *Struct Multidiscip Optim*. 2021;63(4):1927-1950.

doi: 10.1007/s00158-020-02788-w

73. Jui E, Kingsley G, Phan HKT, *et al*. Shear Stress Induces a Time-Dependent Inflammatory Response in Human Monocyte-Derived Macrophages. *Ann Biomed Eng*. 2024;52(11):2932-2947.

doi: 10.1007/s10439-024-03546-5

74. Gao Z, Li Y. Enhancing single-cell biology through advanced AI-powered microfluidics. *Biomicrofluidics*. 2023;17(5):051301.

doi: 10.1063/5.0170050

75. Vela I, Chen Y. Prostate cancer organoids: a potential new tool for testing drug sensitivity. *Expert Rev Anticancer Ther*. 2015;15(3):261-3.

doi: 10.1586/14737140.2015.1003046

76. Biology ASfC. Molecular biology of the cell. vol 15. American Society for Cell Biology; 2004. Accessed May 17, 2026. https://books.google.com.hk/books/about/Molecular_Biology_of_the_Cell.html?id=Z-1FAQAAIAAJ&redir_esc=y

77. Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. 2016; Proceedings of Machine Learning Research. Available from: https://proceedings.mlr.press/v48/gal16.pdf [Last accessed on May 17, 2026].

78. Esch M, King T, Shuler M. The role of body-on-a-chip devices in drug and toxicity studies. *Annu Rev Biomed Eng*.

2011;13(1):55-72.

doi: 10.1146/annurev-bioeng-071910-124629

79. Kim HJ, Huh D, Hamilton G, Ingber DE. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. *Lab Chip*. 2012;12(12):2165-2174.

doi: 10.1039/C2LC40074J

80. Edabashi H, Elghadafi R, Rajkanth N, *et al*. A hybrid technique for measurement of intra/extracellular proteins. *PLoS ONE*. 2023;18(5):e0282948.

doi: 10.1371/journal.pone.0282948

81. Matano M, Date S, Shimokawa M, *et al*. Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat Med*. 2015;21(3):256-62.

doi: 10.1038/nm.3802

82. Drost J, van Boxtel R, Blokzijl F, *et al*. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*. 2017;358(6360):234-238.

doi: 10.1126/science.aao3130

83. Karra N, Fernandes J, James J, Swindle EJ, Morgan H. The effect of membrane properties on cell growth in an 'Airway barrier on a chip'. *Organs Chip.* 2023;5:100025.

doi: 10.1016/j.ooc.2022.100025

84. Chen P, Zhang X, Ding R, *et al*. Patient-Derived Organoids Can Guide Personalized-Therapies for Patients with Advanced Breast Cancer. *Adv Sci*. 2021;8(22):e2101176.

doi: 10.1002/advs.202101176

85. Ju M, Qi A, Bi J, *et al*. A five-mRNA signature associated with post-translational modifications can better predict recurrence and survival in cervical cancer. *J Cell Mol. Med*. 2020;24(11):6283-6297.

doi: 10.1111/jcmm.15270

86. Abbas Y, van Wyk M, Sze H, *et al*. A primary human Gut/Liver microphysiological system to estimate human oral bioavailability. *Drug Metab. Dispos*. 2025;53(9):100130.

doi: 10.1016/j.dmd.2025.100130

87. Chen R, Luo L, Zhang YZ, Liu Z, Liu AL, Zhang YW. Bayesian network-based survival prediction model for patients having undergone post-transjugular intrahepatic portosystemic shunt for portal hypertension. *World J Gastroenterol*. 2024;30(13):1859-1870.

doi: 10.3748/wjg.v30.i13.1859

88. Takebe T, Sekine K, Enomura M, *et al*. Vascularized and functional human liver from an iPSC-derived organ bud transplant. *Nature*. 2013;499(7459):481-484.

doi: 10.1038/nature12271

89. Takebe T, Sekine K, Kimura M, *et al*. Massive and Reproducible Production of Liver Buds Entirely from Human Pluripotent Stem Cells. *Cell Rep*. 2017;21(10):2661-2670.

doi: 10.1016/j.celrep.2017.11.005

90. Camp JG, Sekine K, Gerber T, *et al*. Multilineage communication regulates human liver bud development from pluripotency. *Nature*. 2017;546(7659):533-538.

doi: 10.1038/nature22796

91. Jalili-Firoozinezhad S, Gazzaniga FS, Calamari EL, *et al*. A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. *Nat Biomed Eng*. 2019;3(7):520-531.

doi: 10.1038/s41551-019-0397-0

92. Duivenvoorden AAM, Claes BSR, van der Vloet L, *et al*. Lipidomic Phenotyping Of Human Small Intestinal Organoids Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. *Anal Chem*. 2023;95(50):18443-18450.

doi: 10.1021/acs.analchem.3c03543

93. Chao CJ, Gu YR, Kumar W, *et al*. Foundation versus domain-specific models for left ventricular segmentation on cardiac ultrasound. *NPJ Digit Med*. 2025;8(1):341.

doi: 10.1038/s41746-025-01730-y

94. Tong L, Li X, Shu T, *et al*. Organfit: a multi-scale convolutional model with ellipse fitting for organoid identification. *Complex Intell Syst*. 2025;12(2):67.

doi: 10.1007/s40747-025-02177-0

95. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat Methods*. 2016;13(6):521-7.

doi: 10.1038/nmeth.3853

96. Pozdeyev N, Yoo M, Mackie R, Schweppe RE, Tan AC, Haugen BR. Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget*. 2016;7(32):51619-51625.

doi: 10.18632/oncotarget.10010

97. Herrera C. The Pre-clinical Toolbox of Pharmacokinetics and Pharmacodynamics: in vitro and ex vivo Models. *Front Pharmacol*. 2019;10:578.

doi: 10.3389/fphar.2019.00578

98. Chen Y, Zhang J, Zhang B, *et al*. Optimizing drug sensitivity assays in patient-derived tumor organoids: a comparison of IC50 estimation methods and experimental parameters. *Biol Methods Protoc*. 2025;10(1):bpaf012.

doi: 10.1093/biomethods/bpaf012

99. Peklaj K, Cerovšek M. Implementing laboratory computerized systems in pharmaceutical industry: regulatory compliance. *Ventil*. 2025;31(1):26-32. Accessed May 17, 2026. https://revija-ventil.si/wp-content/uploads/

cerovsek_02-2025.pdf

100. Amershi S, Weld D, Vorvoreanu M, *et al*. Guidelines for Human-AI Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; 2019; Glasgow, Scotland UK.

doi: 10.1145/3290605.3300233

101. Borten MA, Bajikar SS, Sasaki N, Clevers H, Janes KA. Automated brightfield morphometry of 3D organoid populations by OrganoSeg. *Sci Rep*. 2018;8(1):5319.

doi: 10.1038/s41598-017-18815-8

102. Chang SY, Weber EJ, Ness KV, Eaton DL, Kelly EJ. Liver and Kidney on Chips: Microphysiological Models to Understand Transporter Function. *Clin Pharmacol Ther*. 2016;100(5):464-478.

doi: 10.1002/cpt.436

103. U.S. Food and Drug Administration (FDA). Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. July 2005. *Pharmacol Toxicol.* 2020. Accessed May 17, 2026. https://api.semanticscholar.org/CorpusID:17535338

104. Kim R, Sung JH. Microfluidic gut-axis-on-a-chip models for pharmacokinetic-based disease models. *Biomicrofluidics*. 2024;18(3):031507.

doi: 10.1063/5.0206271

105. Waring MJ, Arrowsmith J, Leach AR, *et al.* An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov*. 2015;14(7):475-486.

doi: 10.1038/nrd4609

106. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ*. 2016;47:20-33.

doi: 10.1016/j.jhealeco.2016.01.012

107. Anand O, Pepin XJ, Kolhatkar V, Seo P. The Use of Physiologically Based Pharmacokinetic Analyses---in Biopharmaceutics Applications-Regulatory and Industry Perspectives: Anand *et al*. *Pharm Res*. 2022;39(8):1681-1700.

doi: 10.1007/s11095-022-03280-4

108. Quaid K, Xing X, Chen Y-H, *et al*. iPSCs and iPSC-derived cells as a model of human genetic and epigenetic variation. *Nat Commun*. 2025;16(1):1750.

doi: 10.1038/s41467-025-56569-4

109. Discher DE, Mooney DJ, Zandstra PW. Growth factors, matrices, and forces combine and control stem cells. *Science*. 2009;324(5935):1673-1677.

doi: 10.1126/science.1171643

110. Wei X, Feng T, Huang Q, Chen Q, Zuo C, Ma H. Deep learning-powered biomedical photoacoustic imaging. *Neurocomputing*. 2024;573:127207.

doi: 10.1016/j.neucom.2023.127207

111. Muhyi HA, Sukmadewi R, Chan A, Kahfi AA. Organizational Readiness for Artificial Intelligence with Systematic Mapping Study in Public and Private Sectors. *Sosiohumaniora*. 2024;26(3):483-494.

doi: 10.24198/sosiohumaniora.v26i3.56493

112. Topol E. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK; 2019. Accessed May 17, 2026. https://books.google.com.hk/books?hl=zh-CN&lr=&id=_EFlDwAAQBAJ&oi=fnd&pg=PT8&ots=BJ7DqZYxfP&sig=V_NOc8HR0TkycFhAm_2PpUpGJGM&redir_esc=y#v=onepage&q&f=false

113. Christodoulou I, Goulielmaki M, Kritikos A, Zoumpourlis P, Koliakos G, Zoumpourlis V. Suitability of Human Mesenchymal Stem Cells Derived from Fetal Umbilical Cord (Wharton's Jelly) as an Alternative In Vitro Model for Acute Drug Toxicity Screening. *Cells*. 2022;11(7).

doi: 10.3390/cells11071102

114. U.S. Food and Drug Administration. Drug development tool (DDT) qualification programs. FDA. *Silver Spring,* 2020. Accessed May 17, 2026. https://www.fda.gov/drugs/development-approval-process-drugs/drug-development-tool-ddt-qualification-programs

115. Cai ZQ, Si SB, Chen C, *et al*. Analysis of prognostic factors for survival after hepatectomy for hepatocellular carcinoma based on a bayesian network. *PLoS ONE*. 2015;10(3):e0120805.

doi: 10.1371/journal.pone.0120805

116. Clay I, Peerenboom N, Connors DE, *et al*. Reverse engineering of digital measures: inviting patients to the conversation. *Digit Biomark*. 2023;7(1):28-44.

doi: 10.1159/000530413

117. Bhavna, Ojha A, Bhargava S. Chapter 3 - International Council for Harmonisation (ICH) guidelines. In: Ali J, Baboota S, eds. *Regulatory Affairs in the Pharmaceutical Industry*. Academic Press. 2022:47-74.

doi: 10.1016/B978-0-12-822211-9.00008-3

118. Peroni S, Soiland-Reyes S, Sefton P, *et al*. Packaging research artefacts with RO-Crate. *Data Sci*. 2022;5(2):97-138.

doi: 10.3233/ds-210053

119. Yu M, Guo G, Huang L, *et al*. CD73 on cancer-associated fibroblasts enhanced by the A(2B)-mediated feedforward circuit enforces an immune checkpoint. *Nat Commun*. 2020;11(1):515.

doi: 10.1038/s41467-019-14060-x

120. B S, Bhargavi MS. Explainable AI for Pancreatic Cancer Prediction and Survival Prognosis: An Interpretable Deep

Learning and Machine Learning Approach. *Informatica*. 2024;48(4):623-640.

doi: 10.31449/inf.v48i4.5151

121. Alharbi W, Alfayez AA. Explainable artificial intelligence in pancreatic cancer prediction: from transparency to clinical decision-making. *Front Oncol*. 2025;15.

doi: 10.3389/fonc.2025.1720039

122. Nakkiran P, Kaplun G, Bansal Y, Yang T, Barak B, Sutskever I. Deep double descent: Where bigger models and more data hurt. *J Stat Mech*. 2021;2021(12):124003.

doi: 10.1088/1742-5468/ac3a74

123. ICH harmonised tripartite guideline: Quality risk management Q9. Current Step 4 version. 2005. Accessed May 17, 2026. https://www.pharmatech.nl/pdf/part3/quality_risk_management_en.pdf