ACCSCIENCE PUBLISHING

International Journal of Bioprinting

REVIEW ARTICLE

Large language models for bioink knowledge extraction, design, and biofabrication workflow automation: A critical conceptual review

Yangyang Wang[1], Jiaxun Zhang[1], Haotian Bai[1], Yufei Ren[1], Yanning Yang[1], Jihan Wang[2]*, and Jing Lv[3]*

[1]Department of Automation, School of Physics and Electronic Information, Yan'an University, Yan'an, Shaanxi, China

[2]Department of Basic Medicine, Yan'an Medical College, Yan'an University, Yan'an, Shaanxi, China

[3]Department of Clinical Laboratory, Honghui Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi, China

Abstract

Three-dimensional (3D) bioprinting offers precise spatial control for fabricating cell-laden constructs, yet bioink development remains constrained by labor-intensive trial-and-error experiments and fragmented reporting. Conventional machine-learning pipelines can support quantitative property prediction but generally require structured inputs and therefore do not natively use the formulation knowledge embedded in narrative literature. Large language models (LLMs), particularly when combined with domain-adapted information extraction, retrieval-augmented generation, and knowledge graphs, offer a complementary route for converting literature into traceable design evidence. This critical conceptual review examines how these technologies could support bioink knowledge extraction, formulation recommendation, and design–build–test–learn workflow coordination. Because direct experimental applications of LLMs to bioink design remain scarce, we explicitly distinguish demonstrated evidence in bioprinting from transferable evidence in biomedical natural-language processing, materials informatics, chemistry, and autonomous experimentation, as well as prospective biofabrication concepts. We outline a provenance-aware bioink knowledge graph (BioKG), clarify the complementary roles of LLMs, numerical generative models, Bayesian optimization, computer vision, and robotic control, and provide tissue-specific workflow scenarios and a minimum proof-of-concept evaluation framework. Key limitations include conflicting and heterogeneous evidence, cell-type-specific constraints, hallucination, limited benchmark datasets, and the absence of end-to-end validation. The resulting framework should therefore be interpreted as an evidence-bounded research roadmap rather than as a clinically or experimentally validated autonomous bioprinting system.

Keywords: Large language models; 3D bioprinting; Bioink design; Inverse design; Design–build–test–learn

***Corresponding authors:**
Jihan Wang
(jihanwang@yau.edu.cn);
Jing Lv
(lvjing-1219@163.com)

Citation: Wang Y, Zhang J, Bai H, *et al*. Large language models for bioink knowledge extraction, design, and biofabrication workflow automation: A critical conceptual review. *Int J Bioprint*. 2026;12(5):026350375.
doi: 10.36922/IJB026350375

Received: August 27, 2026

Revised: September 16, 2026

Accepted: September 21, 2026

Published online: September 21, 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.

1. Introduction

Three-dimensional (3D) bioprinting enables layer-by-layer fabrication of cell-

laden constructs with spatial control over architecture, composition, and microenvironment, supporting tissue models and regenerative strategies that are difficult to achieve with conventional scaffold fabrication alone.[1-4] Central to this process is the bioink: a cell–biomaterial formulation that must balance rheological behavior suitable for deposition, mechanical stability sufficient to preserve post-print geometry, cytocompatibility during fabrication and cross-linking, and tissue-specific biological function.[5,6] Hydrogel systems based on gelatin methacryloyl (GelMA), alginate, hyaluronic acid, and decellularized extracellular matrix (dECM) provide broad compositional flexibility, but improvements in printability or shape fidelity can compromise cell viability, matrix remodeling, or functional maturation.[7-9] Bioink development therefore remains largely empirical, and cross-study learning is further limited by inconsistent definitions of printability, heterogeneous experimental protocols, and non-standardized biological endpoints.[6,10-12]

Over the past two decades, biofabrication research has produced a substantial evidence base, but much of the information needed for formulation design—including material provenance, concentration, cross-linking conditions, printer settings, cell identity, assay method, and measurement time point—remains dispersed across narrative text, tables, figures, and supplementary files. Conventional machine-learning (ML) methods can identify useful structure–property relationships when trained on curated numerical datasets. For example, in a dataset of 210 ink formulations, a random forest classifier achieved a maximum printability classification accuracy of 88.1% under the study-specific labeling and validation scheme;[10] related studies likewise support the value of ML for formulation and process optimization.[5,13-15] However, these performance estimates arise from different tasks, datasets, labels, and validation protocols and should not be interpreted as directly comparable benchmarks. Moreover, most conventional ML pipelines depend on manually structured inputs and do not themselves recover the experimental context encoded in scientific prose. The unresolved bottleneck is therefore not simply predictive model capacity, but the conversion of fragmented literature into contextualized, machine-readable, and traceable evidence.

Large language models (LLMs) provide a potentially useful interface to this text-centered bottleneck. In biomedical and scientific domains, language models have been applied to information extraction, literature synthesis, retrieval-supported question answering, and hypothesis generation.[16-21] For bioink research, the most plausible near-term roles are to extract formulation–process–performance relations, normalize protocol descriptions, retrieve condition-matched evidence, identify contradictions and missing metadata, and generate testable candidates for evaluation using physics-based models, conventional ML, Bayesian optimization (BO), or experiments. These functions do not imply that an LLM can directly replace rheological models, numerical generative models, or experimental validation. Importantly, direct demonstrations of LLMs in bioink design and bioprinting remain scarce; much of the current support comes from adjacent fields such as biomedical natural-language processing, materials informatics, biomanufacturing, and autonomous scientific experimentation. Transfer to biofabrication must therefore be treated as a hypothesis to be validated rather than as an established capability, particularly when recommendations depend on cell type and source, biomaterial composition, printing modality, assay conditions, and conflicting evidence.

Against this background, this critical conceptual review examines three questions: which bottlenecks in bioink development are specifically amenable to language-based methods; which relevant capabilities have been demonstrated in biofabrication or adjacent scientific domains; and how those capabilities could be assembled into a testable design–build–test–learn (DBTL) workflow. Building on our previous ML-focused analysis,[13] we used the term "knowledge engine" in a functional, rather than standardized taxonomic, sense to describe an LLM-centered system that extracts, organizes, retrieves, and synthesizes evidence while preserving provenance and experimental context. Figure 1 presents the resulting conceptual architecture, in which literature-derived information is extracted and organized in a provenance-aware bioink knowledge graph (BioKG), retrieved to ground condition-specific LLM reasoning, and synthesized into candidate formulation recommendations for subsequent quantitative prediction and experimental evaluation.

This article is positioned as a critical conceptual review rather than a systematic review or meta-analysis. The literature is organized by function—text extraction, knowledge-graph construction, retrieval-grounded recommendation, numerical inverse design, optimization, monitoring, and DBTL coordination—and interpreted using three evidence categories that reflect proximity to the target domain rather than methodological quality. Category I comprises direct bioink or bioprinting studies; within this category, evidence derived from LLMs is explicitly distinguished from that generated using conventional ML, computer vision, or control systems. Category II comprises transferable evidence from biomaterials, polymer informatics, biomedical natural-language processing, and

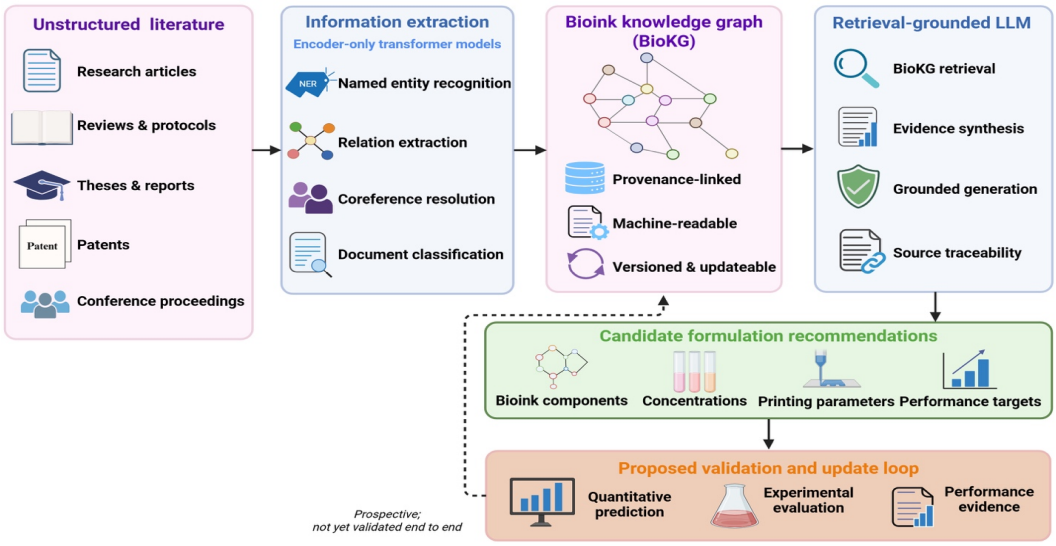


Figure 1. Conceptual architecture of the proposed literature-to-biofabrication knowledge workflow. Reading from left to right, (i) heterogeneous scientific publications provide unstructured source evidence; (ii) encoder-only transformer models extract entities, relations, coreferences, and document-level categories, with the resulting assertions linked to their source provenance; (iii) these provenance-linked assertions populate a versioned and updatable bioink knowledge graph (BioKG), from which condition-relevant evidence is retrieved to ground a generative large language model (LLM); and (iv) the retrieval-grounded LLM synthesizes candidate bioink components, concentrations, printing parameters, and performance targets for subsequent quantitative prediction and experimental evaluation. The dashed return pathway represents the proposed incorporation of quality-controlled validation results and performance evidence into the versioned BioKG, from which updated evidence can ground subsequent recommendation cycles. This feedback process is prospective and has not yet been validated in an end-to-end autonomous biofabrication workflow.

adjacent biological experimentation, whereas Category III comprises enabling demonstrations from general materials science, chemistry, and autonomous laboratories. Evidence from Categories II and III is therefore used to support bounded transfer hypotheses or prospective workflow concepts rather than established biofabrication capabilities. Because direct LLM-based bioink research remains limited, quantitative findings are reported together with their original task, dataset, comparator, and validation setting and are not treated as directly comparable across domains. Recent publications similarly characterize fully autonomous bioprinting and artificial intelligence (AI)-assisted bioink discovery as prospective rather than experimentally established end-to-end capabilities.[22,23]

2. Foundations: Large language model architectures relevant to bioprinting

2.1. Encoder-only transformer models for biomedical text understanding

Encoder-only transformer models process input sequences through bidirectional self-attention, generating dense, context-sensitive token representations that can be fine-tuned on downstream tasks—named entity recognition (NER), relation extraction (RE), sentence classification—with limited labeled data.[24] Transformer-based pre-training, introduced with bidirectional encoder representations from transformers (BERT), demonstrated that massive unsupervised text training could yield general-purpose representations transferable to diverse natural language processing (NLP) tasks with minimal task-specific supervision.[25,26] SciBERT extended this approach to scientific corpora—training on 1.14 million papers from Semantic Scholar—and achieved new state-of-the-art results on sequence tagging, sentence classification, and dependency parsing in scientific domains.[26] This pre-train–fine-tune paradigm proved transformative for biomedical NLP: BioBERT, initialized from general-domain BERT weights and continued on PubMed abstracts and PubMed Central full-text articles, improved over the general-domain baseline by 0.62% F1 on NER, 2.80% F1 on RE, and 12.24% mean reciprocal rank (MRR) on question answering.[24] PubMedBERT pushed further by pre-training exclusively on PubMed text from scratch, eliminating distributional noise from general corpora and reaching state-of-the-art on the biomedical language understanding and reasoning benchmark (BLURB).[27] The materials science community has followed the same logic: MaterialsBERT, initialized from PubMedBERT and trained on 2.4 million materials abstracts, extracted ~300,000 polymer property records from 130,000 abstracts in under 60 hours;[28] MatSciBERT outperformed SciBERT across

NER, relation classification, and abstract classification on materials-domain benchmark.[29] More recently, an NLP framework combining NER with a paragraph-level graph convolutional network (GCN) successfully resolved cross-sentence co-references in bioactive glass literature and populated a structured biomedical materials knowledge graph—a proof of concept whose transfer to bioink formulation databases requires domain-specific annotation and validation.[30-32]

These precedents define both the opportunity and the boundary for bioprinting. A domain-adapted encoder trained across biomaterials chemistry, biomedical engineering, and cell biology literature could systematically extract composition–property relationships, processing conditions, and viability outcomes from the scattered bioprinting corpus at a scale that would be difficult to achieve through manual curation alone. The critical constraint, however, is architectural: encoder-only models are generally confined to classification and extraction over fixed input windows and cannot generate novel formulation suggestions or synthesize knowledge across documents. Bridging that gap requires the generative architectures discussed in Section 2.2.

2.2. Generative large language models and scientific reasoning

Where encoder-only transformer models classify and extract within fixed input windows, generative LLMs—built on autoregressive decoder architectures—produce open-ended text by predicting each successive token conditioned on all prior context.[25,33] This mechanism endows them with qualitatively different capabilities: in-context learning, whereby the model adapts to a new task from a handful of prompt-embedded examples without any weight update; chain-of-thought reasoning, whereby intermediate steps are made explicit to guide multi-step inference; and cross-document synthesis, whereby evidence dispersed across sources is integrated into a coherent, novel output.[16,20] These emergent properties are precisely what encoder-only transformer models generally lack and what the bioprinting knowledge bottleneck demands: the ability to reason from partial or heterogeneous evidence toward a formulation hypothesis that has not been explicitly tabulated in any single source.

Proof of concept for this capability now exists in adjacent scientific domains. ChemCrow, an agent built around GPT-4 and augmented with 18 expert-designed tools, autonomously planned and executed the synthesis of an insect repellent and three organocatalysts, and guided the discovery of a novel chromophore—without any task-specific fine-tuning.[34] Coscientist, likewise GPT-4-powered, designed, planned, and physically executed a palladium-catalyzed cross-coupling reaction in a real laboratory, demonstrating that generative LLMs can close the loop between textual reasoning and wet-lab action.[35] A preliminary evaluation of GPT-4 across materials design, drug discovery, and computational chemistry demonstrated that LLMs can synthesize scientific knowledge across disciplines and propose testable hypotheses in domains far from their explicit training tasks.[36] In the biomedical domain, systematic evaluation on literature-held-out datasets confirmed that LLMs can generate novel, validated hypotheses even for concepts absent from their training data, with multi-agent interactions further improving the diversity and quality of candidates.[37] Laboratory validation in breast cancer combination therapy confirmed that LLM-generated drug combination hypotheses yielded biologically active candidates at significantly higher rates than random screening.[38] For structured classification and NER tasks, fine-tuned encoder-only transformer models retain an edge; but for open-ended synthesis, reasoning, and protocol recommendation—precisely the tasks most relevant to bioink design—generative LLMs are the appropriate architecture.[39,40]

The critical liability of generative LLMs, however, is hallucination: the autoregressive next-token prediction mechanism can produce plausible-sounding but factually incorrect outputs, a risk that is especially consequential when recommending cell-laden formulations for tissue engineering applications.[19,33] Retrieval-augmented generation (RAG), discussed in Section 2.3, addresses this vulnerability directly by grounding generation in verified, domain-specific document collections.

2.3. Retrieval-augmented generation: bridging parametric knowledge and experimental evidence

Retrieval-augmented generation addresses the core liability of generative LLMs—hallucination arising from reliance on static parametric memory—by coupling a neural retriever with the generative model at inference time.[41,42] The retriever encodes a curated external corpus into a dense vector index; when a query arrives, semantically similar passages are retrieved and injected into the LLM's context window, constraining generation to evidence that can be traced back to a source document. This non-parametric augmentation allows the model to access up-to-date, domain-specific knowledge without retraining and provides the source traceability biomedical applications demand.[19,43] In public health question-answering, MEGA-RAG—which combines Facebook AI Similarity Search (FAISS)-based dense retrieval, Best Match 25 (BM25) sparse retrieval, and biomedical knowledge graph traversal—reported a >40% reduction in hallucination relative to its standalone-

LLM comparator in a task-specific public-health question-answering evaluation.[44] This result supports the value of grounded retrieval in that setting but does not establish an expected reduction for bioink queries.

A structurally important extension of vanilla RAG is graph-based retrieval, in which the retrieval step uses an explicit knowledge graph in addition to, or instead of, a flat vector index.[45,46] Entity–relationship paths can support multi-hop queries in which composition, concentration, processing, cell context, and outcome evidence are distributed across documents. A dual-pathway approach combines structured knowledge graph retrieval with unstructured corpus retrieval and has reported improved factual consistency in its original evaluation.[47] These findings motivate a BioKG architecture, but neither the task definition nor the effect size can be assumed to transfer to bioprinting without a domain-specific benchmark.

For bioprinting, where evidence is organized around material–process–cell–performance relationships, hybrid retrieval provides a plausible knowledge layer. Figure 2 illustrates how this layer could support information exchange across a proposed LLM-supported DBTL workflow. Within this architecture, the LLM serves as an evidence-synthesis and workflow-coordination interface. Medical-image segmentation and tissue reconstruction, printer-executable code validation, physical printer control, computer-vision- and sensor-based monitoring, mechanical and biological testing, real-time quality assessment, and next-cycle optimization remain distinct components that require task-specific validation.

2.4. Fine-tuning strategies for domain adaptation in bioprinting

Both encoder and generative LLMs benefit from domain adaptation, but the scale of modern generative models—often exceeding tens of billions of parameters—makes full fine-tuning prohibitively expensive in most research settings. Parameter-efficient fine-tuning (PEFT) methods address this by freezing the pre-trained weights and

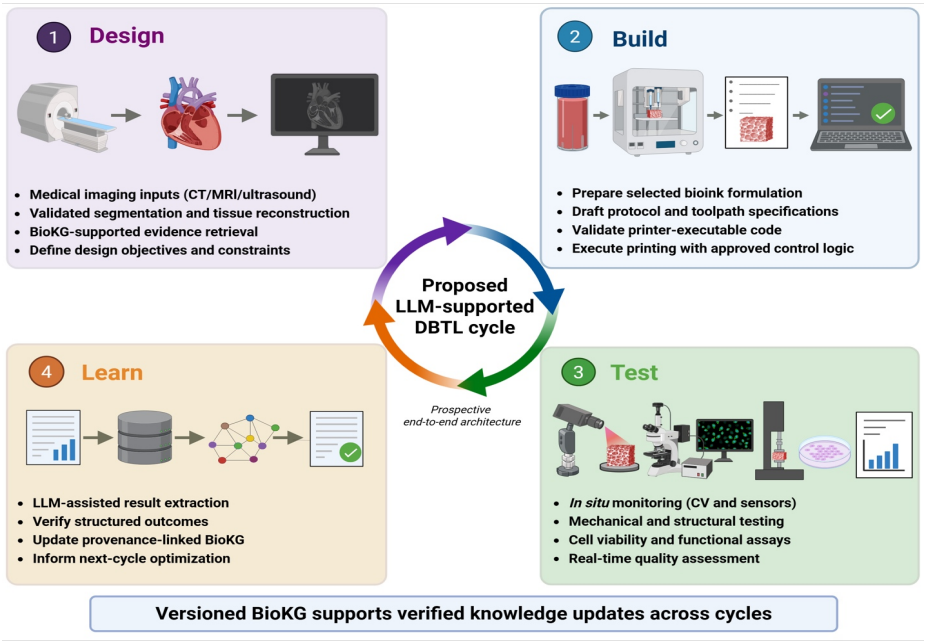


Figure 2. Proposed allocation of language-model functions within a design–build–test–learn (DBTL) biofabrication cycle. The multiphase workflow shows how bounded large language model (LLM) support could be allocated across the biofabrication pipeline. (i) Design: Medical imaging inputs, including computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, provide anatomical information. Tissue segmentation and geometric reconstruction are performed using separately validated imaging models, while BioKG-supported retrieval provides formulation evidence for defining design objectives and constraints. (ii) Build: The selected bioink formulation is prepared, and an LLM may assist in drafting protocols and structured toolpath specifications. Printer-executable code must subsequently be validated, and printing must be performed using approved control logic. (iii) Test: Computer vision (CV) and sensors generate in-process measurements and may support real-time quality assessment, whereas mechanical and structural testing and cell viability and functional assays evaluate printed constructs and biological outcomes. The LLM does not replace these measurement systems or physical controllers. (iv) Learn: An LLM may assist in extracting experimental results, but structured outcomes must be verified before they are incorporated into a provenance-linked, versioned BioKG and used to inform subsequent optimization cycles. The circular arrows indicate the proposed flow of information across the four stages rather than a currently validated autonomous control loop. Integration of all components into an end-to-end LLM-supported DBTL architecture remains prospective.

updating only a small set of injected trainable parameters, typically <1% of the total, reducing graphics processing unit (GPU) memory requirements by up to 90% while preserving the general language capabilities encoded during pre-training.[42,48,49] Among PEFT variants, low-rank adaptation (LoRA) inserts trainable low-rank decomposition matrices into the attention layers of the frozen backbone and has become the dominant approach for domain adaptation of biomedical LLMs: a LoRA-fine-tuned Llama-3-8B model achieved strong performance on structured oncology information extraction from clinical text, outperforming zero-shot GPT-4 on TNM staging and biomarker identification tasks.[50]

For bioprinting, the case for PEFT over full fine-tuning is especially compelling. Bioprinting datasets are inherently small and heterogeneous: extrusion-based systems generate diverse rheological profiles under varying shear conditions, batch-to-batch bioink inconsistencies introduce biological noise, and donor-dependent cell responses confound cross-study comparisons.[51,52] In this low-resource, high-heterogeneity regime, full fine-tuning risks catastrophic forgetting of the general scientific reasoning capabilities that make generative LLMs valuable in the first place, while LoRA-style adaptation preserves those capabilities and requires only modest curated data to shift the model's behavior toward bioprinting-specific vocabulary and reasoning patterns.[48,49,53] An alternative strategy is to combine lightweight PEFT with RAG rather than treating them as substitutes: PEFT can adapt terminology and task conventions, whereas RAG can provide current, source-linked evidence.[42] Any benefit must be measured on a bioink-specific corpus because gains reported in domain-specific question answering are dependent on the model, data split, retriever, and task definition. To systematically compare such computational paradigms, Table 1 summarizes the key characteristics, pre-training corpora, performance benchmarks, and bioprinting readiness across representative encoder-only, generative, and hybrid RAG models.

The quantitative results discussed throughout this review, including those summarized in Tables 1–3, were obtained using different corpora, tasks, output definitions, comparators, and validation schemes. Metrics such as NER F1, question-answering accuracy, hallucination rate, retrieval performance, and experimental yield are therefore not commensurate and should not be pooled or used to rank systems across studies or model classes. The assessments of bioprinting relevance and transferability provided in Tables 1–3 are consequently qualitative judgments of potential architectural applicability rather than experimentally validated performance estimates for bioink design.

3. Large language models as knowledge engines: extracting bioink intelligence from the literature

3.1. Named entity recognition and relation extraction for bioink-property mapping

The first step in converting the bioprinting literature into a usable knowledge resource is information extraction: the identification and structured linking of domain-relevant entities within unstructured text. In NLP, information extraction comprises two complementary subtasks: NER, which assigns categorical labels to text spans, and RE, which classifies the semantic relationship between identified entity pairs.[54,55] Biomedical text mining, of which bioprinting NLP is a subset, has established that transformer-based information-extraction architectures substantially outperform rule-based predecessors on complex, domain-specific entity types—a finding documented across gene–disease, drug–target, and protein–protein interaction corpora over the past decade.[56,57] For bioprinting, NER must span a tripartite vocabulary covering biomaterials chemistry (polymer names, concentrations, crosslinkers), rheological characterization (storage modulus G', yield stress σy, shear-thinning index), and biological outcomes (cell viability, printability, differentiation markers)—a combination absent from any existing biomedical corpus oriented toward genes, proteins, and diseases.

Evidence from adjacent fields demonstrates the potential transferability of transformer-based information-extraction pipelines, while bioink-specific annotation and external validation remain necessary. In polymer materials science, a MaterialsBERT-powered pipeline extracted ~300,000 property records from ~130,000 abstracts in under 60 hours.[28] More directly relevant, a hierarchical NLP framework combining LLM-based NER with a paragraph-level GCN for cross-sentence co-reference resolution achieved 84.7% accuracy at sentence level and 84.0% at paragraph level on bioactive glass literature, extracting composition–synthesis–property triplets into a structured knowledge graph[30]—instructive because the linguistic challenges in bioactive glass text (multi-component material names, co-referential abbreviations such as "GelMA" and "gelatin methacryloyl," and values distributed across sentences) closely mirror those in bioink publications. For RE specifically, a benchmark evaluation found that GPT-3.5-Turbo fine-tuned with a targeted strategy surpassed all baseline models on materials RE tasks, while joint NER and RE using fine-tuned GPT-3 on ~500 annotated abstracts produced JavaScript Object Notation (JSON)-structured composition/phase/morphology/application records with expert annotation time decreasing measurably as training samples increased.[58,59]

Table 1. Comparison of LLM architectures and domain-adapted models relevant to bioprinting research

Model	Architecture type	Pre-training corpus	Parameters	Key bioprinting task	Representative performance	Open source	Generation/Error risk (qualitative)	Potential relevance to bioprinting (qualitative)	Refs
BERT-base	Encoder-only	English Wikipedia + BooksCorpus (3.3 B tokens)	110M	Baseline classification	– (general domain baseline)	Yes	N/A (non-generative; extraction/ classification errors remain)	Low potential; general-domain vocabulary; no direct bioink validation	25
SciBERT	Encoder-only	1.14M scientific papers (Semantic Scholar; 18% biomedical)	110M	Scientific NER, sentence classification	Outperforms BERT on SciERC NER (+3.55% F1)	Yes	N/A (non-generative; extraction/ classification errors remain)	Moderate potential; scientific but not biomedical-specific; no direct bioink validation	26
BioBERT v1.1	Encoder-only	BERT init + PubMed abstracts + PMC full-text (21.3B tokens, 1M steps)	110M	Biomedical NER, RE, QA	+0.62% F1 NER; +2.80% F1 RE; +12.24% MRR QA vs. BERT	Yes	N/A (non-generative; extraction/ classification errors remain)	Moderate-to-high potential; biomedical entities may transfer to bioink extraction; no direct bioink validation	24
PubMedBERT	Encoder-only	PubMed-only from-scratch pre-training (3.1B tokens)	110M	Biomedical NLP (BLURB)	BLURB SOTA at release; F1 0.836 clinical trial NER	Yes	N/A (non-generative; extraction/ classification errors remain)	Moderate-to-high potential; biomedical and cell-related terminology; materials vocabulary requires domain adaptation; no direct bioink validation	27
MaterialsBERT	Encoder-only	PubMedBERT init + 2.4M materials science abstracts	110M	Polymer property NER and extraction	~300,000 property records from 130,000 abstracts in <60 h	Yes	N/A (non-generative; extraction/ classification errors remain)	High potential; demonstrated on polymer corpora; no direct bioink validation	28
MatSciBERT	Encoder-only	SciBERT init + materials science papers (Elsevier, Springer)	110M	Materials NER, relation classification, abstract classification	Outperforms SciBERT on all three materials benchmarks	Yes	N/A (non-generative; extraction/ classification errors remain)	High potential; materials-domain model adaptable to a bioink schema; no direct bioink validation	29

(*Cont'd...*)

Table 1. (*Continued*)

Model	Architecture type	Pre-training corpus	Parameters	Key bioprinting task	Representative performance	Open source	Generation/Error risk (qualitative)	Potential relevance to bioprinting (qualitative)	Refs
GPT-4 (zero-shot)	Generative decoder	Undisclosed large-scale web corpus	Undisclosed	Open-ended synthesis, hypothesis generation, protocol parsing	General-purpose synthesis and hypothesis generation; performance is task- and prompt-dependent	No (API)	High when used without grounding; t ask-dependent	Moderate potential; general-purpose reasoning, but no direct bioink validation	36
GPT-4 + ChemCrow	Generative + tool-augmented agent	GPT-4 backbone + 18 chemistry tools	Undisclosed	Autonomous synthesis planning, chromophore discovery	Novel chromophore discovered; insect repellent and three organocatalysts synthesized	Partial (tools open)	Reduced by tool grounding but task-dependent	Moderate-to-high potential — transferable agent paradigm; no direct bioink validation	34
Llama-3-8B-Instruct + LoRA	Generative backbone + PEFT with multi-task classification heads	General pre-training + oncology-domain LoRA adaptation on 10,677 expert-verified pathology reports	8B	Structured oncology information extraction	Macro F1 = 0.976 for TNM staging, histologic grade, and biomarker extraction in the study-specific evaluation	Yes	Task-dependent; structured classification heads constrain output errors	Moderate-to-high potential; demonstrates parameter-efficient domain adaptation; no direct bioink validation	50
GraphRAG	Hybrid KG-traversal + LLM	Entity graph constructed from document corpus	Backbone-dependent	Global query-focused summarization and corpus-level sensemaking	Improved comprehensiveness and diversity over naive RAG for global sensemaking questions; no bioink benchmark	Yes (Microsoft)	Task-, corpus-, and retrieval-dependent	High potential; supports graph-based evidence integration; no direct bioink validation	45
Dual-pathway KG-RAG	Hybrid KG + dense retrieval	Structured KG + unstructured corpus	Backbone-dependent	Knowledge-intensive QA	Improved factual consistency in the original task; no bioink benchmark	Partial	Task-, KG-, and retrieval-dependent	Potentially high; requires a bioink-specific corpus and direct validation	47

Notes: Performance values originate from different tasks, datasets, comparators, and validation protocols and are not directly comparable across rows. Relevance assessments are qualitative and do not represent validated performance in bioink design or bioprinting.

Abbreviations: API: Application programming interface; BERT: Bidirectional encoder representations from transformers; BLURB: Biomedical Language Understanding and Reasoning Benchmark; KG: Knowledge graph; LLM: Large language model; LoRA: Low-rank adaptation; MRR: Mean reciprocal rank; NER: Named entity recognition; NLP: Natural language processing; PEFT: Parameter-efficient fine-tuning; QA: Question answering; RAG: Retrieval-augmented generation; RE: Relation extraction; SOTA: State-of -the-art.

Three bioink-specific challenges must inform schema design. First, "printability" conflates shape fidelity, geometric accuracy, and cell viability inconsistently across research groups,[60-63] requiring ontology-anchored normalization. Second, the biofabrication window—the trade-off between polymer concentration and shear-induced cell stress—is an inherently cross-sentence relational concept that sentence-level RE alone cannot capture, precisely the problem GCN-augmented pipelines address. Third, quantitative value extraction must bind numerical measurements (G' = 1.2 kPa, viability >85%) to their experimental conditions (concentration, extrusion pressure, nozzle geometry), a structured extraction task particularly well-suited for few-shot-prompted generative LLMs. These pipelines collectively provide the technical foundation for the BioKG construction addressed in Section 3.2. A detailed evaluation of existing NLP information extraction pipelines, their quantitative benchmarks, and their potential transferability and associated evidence limits in the bioink domain is presented in Table 2.

3.2. Construction of BioKG: from unstructured text to structured formulation databases

The information-extraction outputs described in Section 3.1—including entity spans, relation triples, and quantitative value records—provide the raw material for constructing a BioKG. Following the organization shown in Figure 3, the proposed core schema comprises material/formulation, process, and biological/performance domains, supplemented by cross-cutting cell/experimental context and provenance and quality metadata. Material/Formulation nodes include polymers and precursors, crosslinkers and chemical modifiers, additives and nanomaterials, and bioactive factors. Process nodes include printing modality, nozzle geometry, temperature, extrusion pressure, print speed, crosslinking conditions, and post-processing. Biological/Performance nodes include rheological and mechanical properties, shape fidelity and stability, degradation, cell viability and phenotype, and tissue-specific function. Cell information should not be reduced to a single "cell type" label: species,

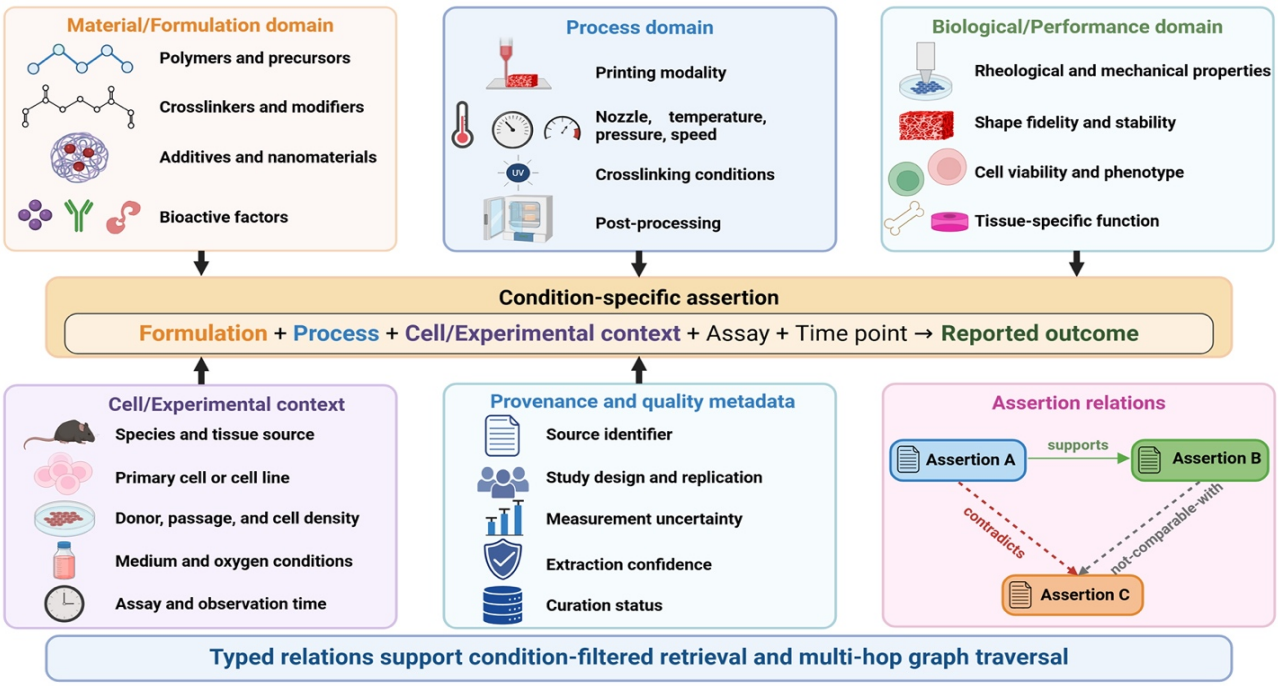


Figure 3. Proposed provenance- and context-aware schema for the BioKG. Core entities are organized into three interconnected domains: material/formulation, including polymers and precursors, crosslinkers and modifiers, additives and nanomaterials, and bioactive factors; process, including printing modality, nozzle geometry and operating parameters, crosslinking conditions, and post-processing; and biological/performance, including rheological and mechanical properties, shape fidelity and stability, cell viability and phenotype, and tissue-specific function. Reported results are represented as condition-specific assertions linking a defined formulation and process with the relevant cell context, assay method, observation time point, and reported outcome. Cell and experimental contexts include species and tissue source, primary cell versus cell line, donor, passage, cell density, culture medium, oxygen conditions, assay, and observation time. Provenance and quality metadata include the source identifier, study design and replication, measurement uncertainty, extraction confidence, and curation status. Separate assertion nodes may link via "supports," "contradicts," or "not-comparable-with" relations to preserve concordant, discordant, and contextually incompatible findings. These typed relations support condition-filtered retrieval and multi-hop graph traversal. The proposed schema requires bioink-specific ontology development and domain validation.

Table 2. NLP information extraction pipelines applicable to bioink literature mining: performance benchmarks and domain transferability

Model/System	NLP task	Source domain	Training/ Evaluation corpus	Key quantitative result	Bioink-relevant entity types covered	Transferability to bioink IE (qualitative)	Ontology/ Schema provided	Code/Data availability	Refs
BioBERT v1.1	NER, RE, QA	Biomedical (PubMed, PMC)	BC5CDR, JNLPBA, NCBI-Disease (NER); ChemProt, DDI (RE)	+0.62% F1 NER; +2.80% F1 RE; +12.24% MRR QA vs. BERT	Disease, drug, gene, chemical	High potential; chemical and biomedical entity types may transfer to polymer/cell vocabulary; no direct bioink validation	No bioink-specific schema	Weights public; GitHub	24
PubMedBERT	NER, RE, sentence similarity, document classification	PubMed biomedical literature	BLURB benchmark (8 tasks, 13 datasets)	BLURB SOTA; F1 0.836 on clinical trial NER corpus	Disease, drug, gene, protein, clinical entity	High potential; aligned with biomedical and cell-related terminology; materials vocabulary requires domain adaptation; no direct bioink validation	No	Weights public	27
MaterialsBERT	Property NER and database population	Polymer materials science	2.4M materials abstracts; 130,000 for extraction eval	~300,000 polymer property records in <60 h; high recall across concentration, Tg, bandgap entities	Polymer name, property value, processing condition	High potential; polymer concentration and property extraction is adaptable to a bioink-specific schema; no direct bioink validation	Polymer property schema (PolymerNLP)	Model and pipeline public	28
MatSciBERT	NER, relation classification, abstract classification	Materials science (Elsevier, Springer)	MatSci-NLP benchmark (7 tasks)	Outperforms SciBERT across the evaluated materials-domain benchmarks	Material, property, characterization method, application	High potential; process-condition entity types are relevant, and the schema is adaptable; no direct bioink validation	MatSci-NLP schema (open)	Model public	29
NER + paragraph-level GCN	NER + cross-sentence RE	Bioactive glass (biomaterials)	Domain-specific bioactive glass corpus	84.7% accuracy (sentence-level); 84.0% (paragraph-level)	Composition, synthesis condition, property, application	High potential; bioactive-glass text shares material–process–property structure with bioink publications; no direct bioink validation	Custom KG schema (nodes: material, process, performance)	Partial	30

(*Cont'd...*)

Table 2. (*Continued*)

Model/System	NLP task	Source domain	Training/ Evaluation corpus	Key quantitative result	Bioink-relevant entity types covered	Transferability to bioink IE (qualitative)	Ontology/ Schema provided	Code/Data availability	Refs
GPT-3.5-Turbo (fine-tuned)	RE benchmark evaluation	Materials science	MaterialsBERT RE benchmark	Surpasses all baseline models on materials RE tasks	Material–property pairs, material application, composition processing	High potential; fine-tuning strategy and relation types may transfer; no direct bioink validation	No	API access	58
GPT-3 (NERRE, fine-tuned)	Joint NER + RE → structured records	Materials chemistry	~500 annotated materials science abstracts	JSON-structured records (composition, phase, morphology, application); expert annotation time reduced significantly	Composition, phase, morphology, application	High potential; the NERRE JSON schema is adaptable to bioink formulation records; no direct bioink validation	NERRE schema (open; JSON)	Code public; GitHub	59
Fine-tuned LLM (ORD schema)	Protocol extraction from synthesis text	Organic chemistry	ORD	91.25% accuracy on reaction information extraction; compound IDs, mass quantities, reaction roles	Compound name, concentration, reaction condition, role	High potential; protocol structure is analogous to bioink methods sections; no direct bioink validation	ORD schema (open; machine-readable)	Code and data public	64
Claude 3 Opus, GPT-4, LLaMA-3 (benchmarked)	Synthesis condition extraction from full-text	MOF chemistry	Full-text MOF synthesis papers	Claude 3 Opus: best overall extraction coverage; GPT-4: strong on structured fields	Synthesis condition, temperature, solvent, time, yield	Moderate potential; benchmark methodology may transfer, but MOF synthesis text differs from bioink methods; no direct bioink validation	No (evaluation schema only)	Benchmark public	65

Notes: Quantitative results originate from different IE tasks and evaluation schemes and are not directly comparable. Transferability ratings are qualitative; none constitutes direct validation on a bioink-specific corpus.

Abbreviations: API: Application programming interface; BERT: Bidirectional encoder representations from transformers; BLURB: Biomedical Language Understanding and Reasoning Benchmark; GCN: Graph convolutional network; IE: Information extraction; JSON: JavaScript Object Notation; KG: Knowledge graph; LLM: Large language model; MOF: Metal–organic framework; MRR: Mean reciprocal rank; NER: Named entity recognition; NERRE: Named entity recognition and relation extraction; NLP: Natural language processing; ORD: Open Reaction Database; QA: Question answering; RE: Relation extraction; SOTA: State-of-the-art.

tissue source, primary cell versus cell line, donor, passage, cell density, culture medium, oxygen conditions, assay method, and post-print observation time should be encoded as contextual qualifiers. A reported viability or performance value should therefore be represented as a condition-specific assertion linking a defined formulation and process with the relevant cell context, assay, time point, reported outcome, and source, rather than as an unconditional property of the bioink.

Conflicting findings should be preserved rather than collapsed into a single averaged fact. Each assertion should retain its source identifier, experimental conditions, extraction method, study design and replication information, reported measurement uncertainty, extraction confidence, and curation status. Assertion-level relations can distinguish "supports," "contradicts," and "not-comparable-with" evidence. "Supports" can connect concordant findings obtained under sufficiently comparable conditions, whereas "contradicts" can identify discordant findings obtained under comparable conditions. "Not-comparable-with" should be used when differences in formulation, cell source, printing conditions, assay method, or observation time prevent valid direct comparison. A conflict-aware RAG system could then retrieve both concordant and discordant evidence, stratify results according to cell and process context, and identify potential sources of disagreement. Transparent, predefined evidence-quality criteria may assign greater confidence to directly measured, adequately replicated, and fully reported experiments, but minority, null, and negative findings should remain retrievable and should not be suppressed. This provenance-aware representation is particularly important when the same nominal GelMA concentration produces different viability or printability outcomes because cell source, photoinitiator, light dose, nozzle geometry, extrusion pressure, assay, or observation time differ across studies. Figure 3 summarizes the proposed BioKG schema, in which the three core domains are connected through condition-specific assertions, while cross-cutting cell and experimental context, provenance and quality metadata, and assertion-level relations preserve contextual comparability and disagreement across studies.

With these qualifications in place, structured queries could retrieve, for example, gelatin-based formulations printed at an extrusion pressure of ≤30 kPa that achieved >85% cell viability for a specified cell source, assay method, and observation time, while separately returning studies that used incompatible endpoints or experimental contexts. Standardized materials ontologies and the FAIR (findable, accessible, interoperable, and reusable) principles provide partial foundations for interoperability.[66,67] Materials-science knowledge graphs demonstrate the technical feasibility of literature-scale graph construction,[68,69] but their transfer to bioink research requires a purpose-built ontology, bioink-specific annotation guidelines, and domain validation rather than an assumption of equivalent performance.

One of the central obstacles to realizing a BioKG is the absence of a community-endorsed ontology and an openly available annotated corpus for the bioink literature—a gap closely related to the lack of standardized printability definitions discussed in Section 3.1. Existing materials-data-science ontologies provide partial scaffolding but do not yet represent bioprinting-specific formulation–process–cell–performance relationships at the required level of granularity.[66] The open research knowledge graph (ORKG) and analogous cross-domain infrastructure demonstrate the technical feasibility of constructing shared, machine-actionable knowledge resources, but comparable progress in bioprinting will require sustained coordination across the research community.[70] Development of a BioKG therefore depends on several coordinated efforts: community-curated annotation schemas that explicitly represent formulation–process–cell–performance relationships and biofabrication-window constraints; FAIR-compliant data-deposition practices that make experimental records machine-readable at the point of publication; and iterative LLM-assisted graph population with human verification. The structured record extraction demonstrated by the named entity recognition and relation extraction (NERRE) framework in materials chemistry provides a relevant technical precedent for the latter strategy, although transferring it to bioink literature requires domain-specific annotation and validation.[59] The protocol-extraction approaches discussed in Section 3.3 could further support this pipeline by converting methods-section text into candidate structured graph entries, thereby reducing, rather than eliminating, the need for manual transcription and curation.

3.3. Automated protocol extraction and standardization

A persistent structural barrier to cross-laboratory knowledge transfer in bioprinting is the heterogeneity of experimental reporting. A round-robin study conducted across 12 academic laboratories using standardized extrusion-based printing conditions found that operator-dependent variation remained a significant source of inter-laboratory discrepancy, directly limiting the comparability of bioink characterization data.[71-73] The ASTM F3659-24 Standard Guide for Bioinks, published in April 2024,

formalizes definitions and terminology for extrusion-based bioprinting but does not yet specify quantitative criteria for printability assessment or experimental reporting requirements.[74] In this context, the methods sections of published bioprinting studies constitute a largely untapped structured knowledge source: they routinely contain the complete parameter set needed for replication—polymer concentration, crosslinker type and concentration, printing temperature, nozzle diameter, extrusion pressure, print speed, and post-print crosslinking conditions—yet this information is buried in free-text prose and varies in completeness and vocabulary across research groups. The Open Reaction Database (ORD), which provides a standardized schema for structuring and sharing organic reaction data to enable downstream ML applications, demonstrates the value of such infrastructure when systematically implemented—and provides a direct template for a future "Open Bioprinting Protocol Database".[75] LLMs offer a potentially scalable extraction approach, although bioink-specific validation remains necessary. In organic chemistry, a fine-tuned LLM achieved 91.25% accuracy on extracting reaction information from synthesis procedure text into structured ORD records, including compound identifiers, mass quantities, and reaction roles.[64] In materials science, LLM-based pipelines have been benchmarked on extracting synthesis conditions for metal–organic frameworks from full-text papers, with Claude 3 Opus providing the most complete extraction coverage across diverse synthesis paragraphs.[65] Applied to bioprinting, analogous pipelines could systematically parse methods sections to populate structured protocol records—converting "GelMA (10% w/v; photoinitiator: 0.5% w/v; UV crosslinking: 30 s at 365 nm; nozzle: 25G; pressure: 40 kPa)" into machine-readable entries directly ingested by the BioKG described in Section 3.2.

The broader value of automated protocol extraction lies not only in database population but in enabling reproducibility at scale. When extracted protocol records are cross-referenced against reported outcomes—printability scores, shape fidelity indices, and viability percentages—the resulting dataset could support data-driven identification of parameter combinations that are under-reported or systematically confounded across the literature, an analysis that is difficult to perform manually at bioprinting-literature scale. Furthermore, verified LLM-extracted protocols could provide structured inputs for separately validated robotic workflows. Translation into printer-executable instructions would still require syntax validation, hardware-specific constraints, simulation or dry-run checks, and human approval.

4. Large language model–driven bioink design: from knowledge retrieval to formulation recommendation

4.1. Retrieval-augmented generation–enhanced formulation recommendation systems

The BioKG constructed in **Section 3** provides a structured retrieval substrate for a proposed RAG-enhanced bioink recommendation system. When a researcher specifies a tissue application and performance envelope, the system can retrieve formulation assertions together with their material, process, cell, assay, and provenance metadata, and pass the evidence to a generative LLM for synthesis.[42,43] Outputs should include claim-level citations, compatibility warnings, and a separation between directly reported values and model-generated extrapolations. When studies disagree, the system should present the competing results and their contextual differences rather than generate a single apparently authoritative value. Such traceability is an architectural objective that must be evaluated through citation precision, evidence coverage, contradiction recall, and expert review; it is not guaranteed merely by adding RAG. Evidence from nanostructured biomedical materials supports the feasibility of domain-specific retrieval and iterative query refinement,[76] but performance in two-photon polymerization literature cannot be assumed to transfer unchanged to cell-laden bioinks.

Two limitations of RAG in this context deserve explicit acknowledgment. First, retrieval quality degrades when a corpus is sparse, heterogeneous, or poorly indexed—a systemic challenge in biofabrication, where experimental reporting remains unstandardized, and parameter definitions vary across laboratories.[43] Second, the retrieval component cannot provide evidence that is absent from the documented design space. A generator may still extrapolate beyond retrieved records, but such output is no longer literature-grounded and must be labeled and experimentally tested. Table 3 therefore compares representative retrieval architectures only within their original evaluation domains; the table is not a cross-study ranking of expected bioink performance.

4.2. Generative artificial intelligence for inverse bioink design

Retrieval-augmented generation systems retrieve and synthesize existing records, whereas numerical generative models can sample candidate formulations that were not directly observed. These model classes should not be conflated. Variational autoencoders, conditional variational autoencoders (cVAEs), generative adversarial networks, and diffusion models learn numerical or

Table 3. Representative RAG systems in biomedical and materials domains: architecture, retrieval strategy, and relevance to BioKG-powered bioink recommendation

System	Retrieval mechanism	Knowledge source	LLM backbone	Domain of evaluation	Hallucination metric	Performance gain/reduction (study-specific)	Source traceability	Multi-hop capability	Relative latency (qualitative)	Open implementation	Potential BioKG relevance (qualitative)	Refs
Naive RAG	DPR; FAISS index	Unstructured text corpus	BART/T5	Knowledge-intensive NLP (Natural Questions, TriviaQA)	Factual accuracy on open-domain QA	Improved performance over parametric sequence-to-sequence baselines on several knowledge-intensive QA tasks; gains varied by dataset	Retrieved passages available; explicit citation generation is not intrinsic	No	Fast	Yes (Hugging Face)	Partial; flat retrieval may be insufficient for multi-constraint bioink queries; no direct bioink validation	41
MEGA-RAG	Hybrid: FAISS dense + BM25 sparse + biomedical KG traversal	PubMed + biomedical KG	GPT-3.5	Public health QA	Hallucination rate (human annotation)	>40% hallucination reduction versus a standalone LLM in the reported public-health QA evaluation; improved precision, recall, and F1	Yes (PubMed PMIDs)	Partial (KG traversal)	Moderate	Partial	Moderate-to-high potential; combines hybrid retrieval and KG traversal; no direct bioink validation	44
GraphRAG	Community detection + hierarchical graph summarization	Entity graph constructed from corpus	GPT-4	Global query-focused summarization and corpus-level sensemaking	Comprehensiveness and diversity assessed by task-specific human/LLM judgments	Improved comprehensiveness and diversity over naive RAG for global sensemaking questions; no bioink evaluation	Entity-level graph evidence	Graph-based traversal	Slower	Yes (Microsoft, open-source)	Moderate-to-high potential; graph-based evidence integration may support material–process–performance queries; direct validation required	45

(*Cont'd...*)

Table 3. (*Continued*)

System	Retrieval mechanism	Knowledge source	LLM backbone	Domain of evaluation	Hallucination metric	Performance gain/reduction (study-specific)	Source traceability	Multi-hop capability	Relative latency (qualitative)	Open implementation	Potential BioKG relevance (qualitative)	Refs
Dual-pathway KG-RAG	Parallel: structured KG entity-path + unstructured dense retrieval	Pre-built KG + text corpus	GPT-4	Knowledge-intensive QA	Task-specific factual-consistency and hallucination evaluation	Reported improvement over standard RAG in the original task; the effect size is not transferable to bioprinting	KG path/triple plus retrieved text	Yes (KG path reasoning)	Moderate	Partial	High potential; dual retrieval may suit BioKG-supported literature synthesis; direct validation required	47
LLaMA-2 + LoRA + RAG	BM25 sparse retrieval	PubMedQA corpus	LLaMA-2-7B	Biomedical QA (PubMedQA, MedMCQA)	Accuracy on multiple-choice QA	71.8% PubMedQA accuracy versus a 55.4% zero-shot baseline; approximately 60% fewer unsupported statements in the reported evaluation	Passage citation	No	Fast	Partial	Moderate potential; demonstrates task-specific PEFT and RAG benefits; no direct bioink validation	77
Nano structured materials RAG	Agent-based iterative RAG with dynamic query refinement	Two-photon poly merization literature	GPT-4	Nanostructured biomedical material–cell interaction QA	Task-specific answer relevance, precision, and misinformation evaluation	Improved answer relevance and reduced misinformation in the study-specific evaluation	Yes (publication-level)	Partial	Moderate	No	High potential; a materials-focused analog, but no direct bioink validation	76

(*Cont'd...*)

Table 3. (*Continued*)

System	Retrieval mechanism	Knowledge source	LLM backbone	Domain of evaluation	Hallucination metric	Performance gain/reduction (study-specific)	Source traceability	Multi-hop capability	Relative latency (qualitative)	Open implementation	Potential BioKG relevance (qualitative)	Refs
Self-reflective RAG	Adaptive retrieval (model decides when to retrieve)	General corpus	LLaMA-2	Open-domain QA and fact verification	Factuality and citation support on open-domain tasks	Improved factuality and citation support in the original benchmarks; no clinical or bioink validation	Yes (cited passages)	Partial	Variable	Yes	Moderate potential; adaptive retrieval may help with variable-complexity bioink queries; direct validation required	78
KG-RAG	KG retrieval combined with text retrieval	Custom knowledge graph	GPT-based	Knowledge-intensive QA	Task-specific factual consistency and precision	Reported gains over a conventional RAG baseline in the original task; no bioink validation	Source traceability depends on the implementation	Potential multi-hop retrieval	Moderate-high	Partial	Moderate-to-high potential; knowledge-graph retrieval may support BioKG queries; direct validation required	79

Notes: Performance metrics, latency labels, and comparisons are study-specific and should not be used to rank systems across rows. BioKG relevance is a qualitative transfer hypothesis, not validated bioink performance.

Abbreviations: BioKG: Bioink knowledge graph; BM25: Best Match 25 (sparse retrieval); DPR: Dense passage retrieval; FAISS: Facebook AI Similarity Search; GNN: Graph neural network; KG: Knowledge graph; LLM: Large language model; LoRA: Low-rank adaptation; PEFT: Parameter-efficient fine-tuning; QA: Question answering; RAG: Retrieval-augmented generation.

molecular distributions and are the appropriate engines for composition generation under quantitative constraints. By contrast, a generative LLM operates primarily on language and symbolic representations; it can extract constraints, integrate literature, generate hypotheses, explain trade-offs, or prepare structured inputs, but it should not substitute for a validated numerical generator or property predictor. In inverse bioink design, the target space spans rheology, viability, phenotype, shape fidelity, and tissue-specific function, and candidate generation must remain coupled to quantitative validation.

The central limitation of numerical generative approaches in the bioink context is data volume and consistency. LLMs may assist as data curators, constraint extractors, and hypothesis generators that seed an active-learning loop, but LLM-generated pseudo-records must be labeled as synthetic and must not be treated as experimental observations. Their value should be tested against baselines that use the same measured dataset without LLM augmentation. This division of labor—language models for evidence integration and proposal, numerical models for constrained generation and prediction, and experiments for adjudication—is the most defensible near-term architecture.[80-85]

4.3. Multi-objective optimization of bioink formulations via large language model–guided active learning

The central challenge of bioink formulation is not finding a composition that maximizes any single property but navigating a Pareto frontier between competing objectives: higher polymer concentration may improve filament stability and shape fidelity but can increase viscosity, the required extrusion pressure, and cellular shear exposure; similarly, stiffer gels can improve structural stability while restricting cell spreading and matrix remodeling.[86,87] Classical BO addresses multi-objective formulation problems by maintaining a probabilistic surrogate model over the design space and selecting the next experiment using an acquisition function that balances exploration and exploitation.[88,89] BO has been benchmarked across several experimental materials domains, including carbon nanotubes, silver nanoparticles, perovskites, and additively manufactured polymer structures, illustrating its potential to improve experimental efficiency under domain-specific conditions.[90] Evidence for direct LLM-driven experimental design remains mixed. Gupta *et al.*[91] found that LLM agents were insensitive to experimental feedback and were generally outperformed by classical optimization methods, whereas a hybrid strategy that separated LLM-derived prior guidance from data-driven candidate acquisition achieved competitive performance. These findings support using LLMs to provide prior knowledge and search-space guidance rather than replacing the surrogate model or acquisition function. In a wet-laboratory pharmaceutical reaction study, ChemBOMAS identified a 96% yield condition, whereas a chemist using a traditional control-variable strategy achieved 15%.[92] This comparison was not against domain-agnostic BO and was not performed in bioprinting; it therefore demonstrates task-specific feasibility in chemistry rather than an expected effect size for bioink optimization. LLM-accelerated BO (LABO) likewise provides evidence that LLM predictions can act as low-cost auxiliary observations within BO, but its results remain domain- and benchmark-dependent.[85] This proposed active-learning architecture—coupling an LLM agent with BO to navigate multi-objective bioink design trade-offs—is schematically illustrated in Figure 4. Within this proposed division of responsibilities, the LLM retrieves and synthesizes prior evidence and suggests candidate regions or search-space constraints, whereas the BO surrogate model and acquisition function determine which experiments should be performed next. Whether this hybrid strategy can reduce the experimental burden relative to random, grid-based, or conventional BO strategies requires direct benchmarking using bioink-specific datasets and experimental workflows.

5. Integrating large language models into the biofabrication workflow: toward DBTL automation

5.1. Large language models in the design phase: hypothesis generation and target-to-blueprint translation

The design phase of the DBTL cycle encompasses two analytically distinct tasks for bioprinting: generating candidate hypotheses about which formulations may satisfy a target application, and translating a clinical or tissue-engineering specification into a concrete fabrication blueprint. In adjacent biomedical and materials domains, LLMs have shown measurable capacity for the first task: evaluation on biomedical literature found that LLMs could generate externally testable hypotheses for concepts withheld from the supplied literature, with multi-agent configurations improving candidate diversity and acceptability.[37] In materials discovery, the goal-driven ACCELMAT agent system combined a hypothesis generator, multiple critic models, and an evaluation agent to generate and refine hypotheses on a structured benchmark.[93] These studies support hypothesis generation as a transferable capability, not as direct validation in bioink design. Applied prospectively, an analogous pipeline could combine a target tissue profile with BioKG evidence to

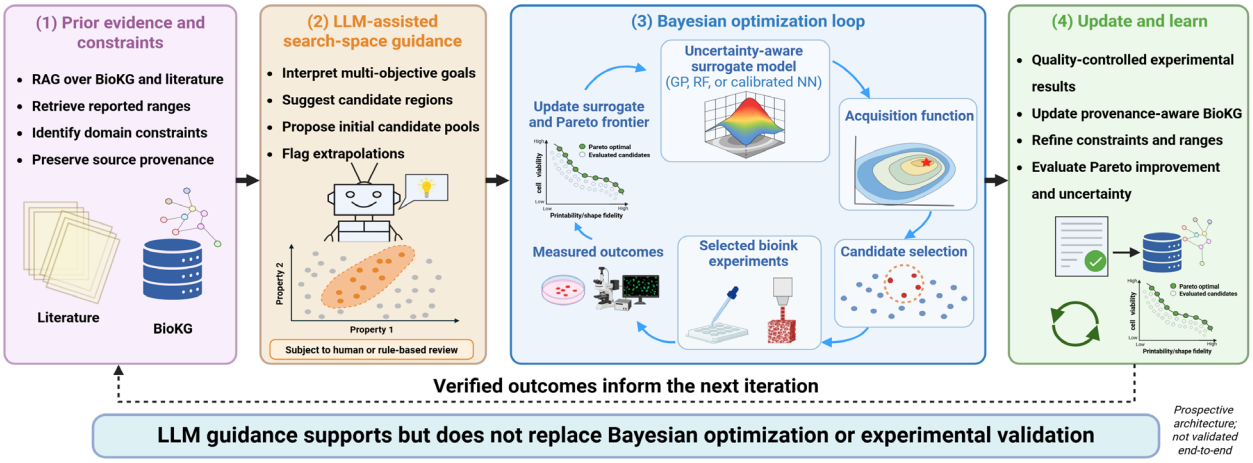


Figure 4. Proposed large language model (LLM)-assisted active-learning workflow using Bayesian optimization for multi-objective bioink formulation. (1) Prior evidence and constraints: Retrieval-augmented generation (RAG) over the bioink knowledge graph (BioKG) and relevant literature retrieves reported formulation ranges, identifies domain constraints, and preserves source provenance. (2) LLM-assisted search-space guidance: The LLM interprets multi-objective design goals, suggests candidate regions, proposes initial candidate pools, and flags extrapolations or unsupported assumptions. These outputs remain subject to human or rule-based review and provide prior guidance rather than replacing quantitative optimization. (3) Bayesian optimization loop: An uncertainty-aware surrogate model, such as a Gaussian process (GP), random forest (RF), or appropriately calibrated neural network (NN), estimates the objective functions and their uncertainty. An acquisition function uses these estimates to select candidate formulations for experimental evaluation. The measured outcomes are subsequently used to update the surrogate model and estimated Pareto frontier, supporting iterative navigation of trade-offs such as those between printability or shape fidelity and cell viability. (4) Update and learn: Quality-controlled experimental results update the provenance-aware BioKG, refine formulation constraints and parameter ranges, and enable assessment of Pareto improvement and remaining uncertainty. Verified outcomes then inform evidence retrieval and search-space guidance in the subsequent iteration. LLM-derived guidance supports but does not replace Bayesian optimization, experimental measurement, or domain validation. The complete workflow is a proposed architecture and has not been validated as an end-to-end bioink-optimization system.

generate bounded formulation hypotheses. Vascular tissue engineering is an instructive example because perfusable lumen formation requires rapid stabilization of channel geometry while maintaining endothelial-cell viability under extrusion-induced shear stress.[94,95]

For the second task, a recent review described AI-assisted integration of multimodal computed tomography and magnetic resonance imaging data with 3D-printing workflows for anatomical segmentation, reconstruction, and process planning.[96] These functions are performed through dedicated imaging and geometric-processing models rather than through language models alone. Connecting patient-specific radiological data to tissue specifications and then matching those specifications to BioKG-supported bioink candidates is therefore a prospective research direction; no cited study demonstrates an end-to-end multimodal LLM pipeline that converts clinical images into validated bioprinting instructions.

5.2. Large language models in the build and test phases: protocol automation and quality assessment

The build phase converts the design blueprint into a physical print. LLMs can assist with extracting and drafting protocol records, but converting them into printer-executable parameter files or G-code requires a separately validated translation layer, syntax and range checks, hardware-specific constraints, simulation or dry runs, and human authorization; direct end-to-end LLM validation for this task in bioprinting has not yet been established. The test phase requires real-time assessment of print quality against the intended design. Here, the dominant AI approach is computer vision rather than language modeling: convolutional neural network–based systems have been used to detect and correct extrusion anomalies in bioprinting settings.[97,98] A modular *in situ* monitoring platform using the Segment Anything Model (SAM) as a segmentation backbone enabled layer-by-layer assessment of embedded bioprints without modifying the support bath, providing proof of concept for printer-agnostic quality monitoring.[99]

The role of LLMs in the test phase is to interpret structured outputs from validated computer-vision and sensor models, document possible causes, retrieve relevant corrective evidence, and propose candidate adjustments for human or rule-based approval. This role must not be conflated with real-time control: an LLM is

not itself a sensor, deterministic controller, or actuator. Any online adjustment requires bounded control logic, latency and safety testing, and a predefined fallback state. Thus, computer vision performs spatial detection, control software executes validated parameter changes, and the LLM provides evidence-grounded coordination and explanation.[100]

5.3. Large language models in the learn phase: closing the loop with automated knowledge update

The learn phase closes the DBTL cycle by converting validated experimental outcomes into structured records and model updates. In biological optimization systems, automated testing can still require human intervention when results must be interpreted and mapped to reusable knowledge.[101] LLM–RAG methods have been explored for identifying outdated knowledge-graph triples and proposing updates from retrieved evidence.[102] In a future bioprinting implementation, a completed experiment could be parsed into candidate formulation–process–outcome assertions, but numerical values, units, assay context, and provenance would require automated validation rules and human verification before graph insertion. Dynamic knowledge graphs provide a potential memory substrate for accumulating such verified experimental records.[103]

If each update is quality-controlled, verified, and stored with appropriate provenance, iterative learn-phase records could progressively expand the coverage and traceability of the BioKG and increase the evidence available to subsequent design cycles. This constitutes between-run learning rather than autonomous closed-loop control. Autonomous in-process feedback would additionally require measurements acquired during the same fabrication run to trigger validated state estimation, safety checks, and physical adjustments through deterministic or human-approved control logic. Figure 5 therefore distinguishes a

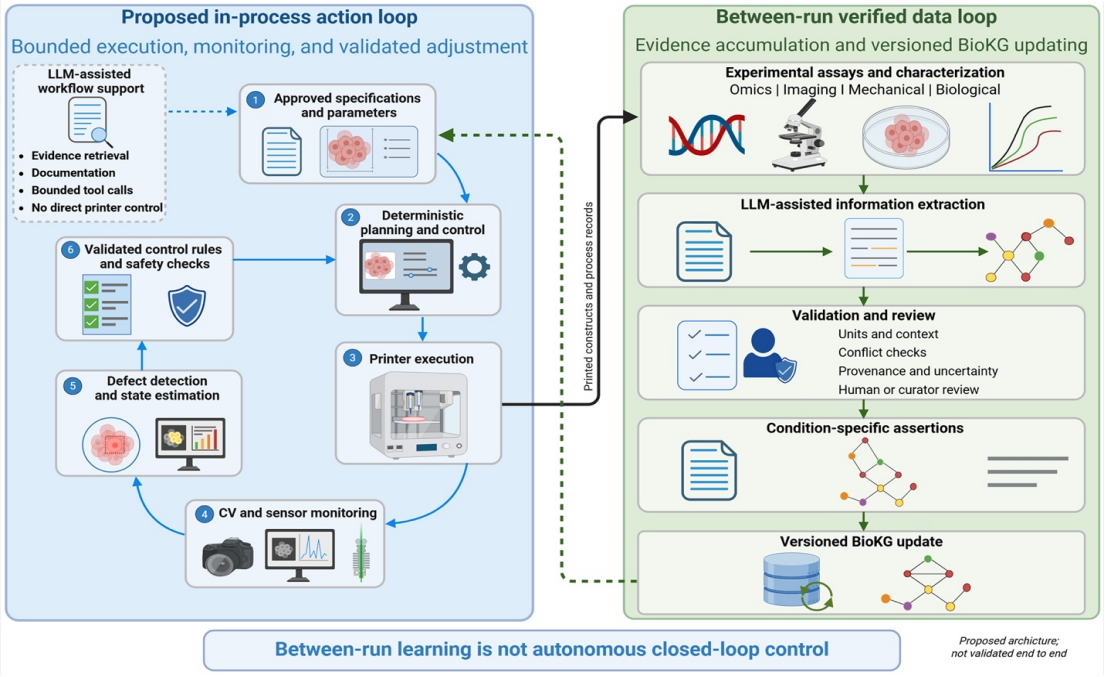


Figure 5. Proposed dual-loop architecture separating in-process execution and monitoring from between-run evidence accumulation in progressively automated biofabrication. In the proposed in-process action loop (left), approved tissue specifications and process parameters are translated using deterministic planning and control software into printer actions. Computer vision (CV) and sensors acquire process measurements, while separately validated defect-detection and state-estimation models identify deviations. Physical parameter adjustments are authorized only through validated control rules operating within predefined safety limits; an unconstrained large language model (LLM) does not directly control the printer. Instead, an LLM may support evidence retrieval, documentation, and bounded tool calls within predefined permissions. Printed constructs and associated process records are transferred from the action loop to the between-run verified data loop (right), where experimental assays and characterization—including omics, imaging, mechanical testing, and biological measurements—generate candidate evidence records. LLM-assisted information extraction may convert these records into condition-specific assertions. Before incorporation into the BioKG, however, the extracted assertions must undergo validation of units, experimental context, provenance, uncertainty, and conflicting evidence, together with human or curator review. Quality-controlled assertions may then be incorporated into a versioned BioKG and used to inform approved specifications and parameters in subsequent experimental runs. This return pathway represents between-run learning rather than autonomous real-time closed-loop control, and the complete architecture remains prospective rather than an experimentally validated end-to-end system.

proposed in-process action loop for bounded execution, monitoring, and validated adjustment from a between-run verified data loop for evidence extraction, review, and versioned BioKG updating.

5.4. Case scenarios: large language models–assisted bioink development for representative tissue targets

The following scenarios are illustrative workflow specifications, not reports of completed LLM-guided bioink experiments. Each applies the same evidence-bounded pipeline—context definition, conflict-aware retrieval, candidate generation, numerical optimization, experimental testing, and provenance-preserving update—to a different biological target.

5.4.1. Vascular tissue engineering: maintaining lumen patency while protecting endothelial cells

(a) Input and retrieval

The user specifies vessel diameter, target compliance, printing modality, endothelial cell source, assay time point, and an acceptable extrusion-pressure range. The BioKG retrieves GelMA–alginate, fibrin–gelatin, and related coaxial records, but stratifies them by nozzle geometry, crosslinking sequence, cell density, light or ionic dose, and viability assay. Conflicting lumen-patency or viability results are displayed with their conditions rather than averaged.[87,94,95]

(b) Design, build, test, and learn

The LLM converts the evidence into explicit constraints and proposes a bounded candidate set; a multi-objective BO model selects experiments that maximize endothelial viability and lumen patency while minimizing pressure and gelation time. Coaxial printing and post-print assays quantify lumen diameter, wall thickness, short-term viability, vascular endothelial-cadherin expression, and perfusability. Only measured results are added to the BioKG, with failures and negative outcomes retained to reduce publication and survivorship bias.

5.4.2. Skin and wound repair: cell-resolved, multi-material formulation planning

(a) Input and retrieval

The specification distinguishes epidermal keratinocytes from dermal fibroblasts and records their source, passage, density, medium, oxygen exposure, and assay schedule. Retrieval covers dECM, fibrin, collagen, GelMA, and hyaluronic-acid systems while preserving dECM tissue source and batch characterization. The system flags studies in which "viability" was measured with different assays or at incompatible time points.[104-109]

(b) Design, build, test, and learn

The LLM proposes a zone-specific sequence and hands numerical concentration and process variables to a cVAE or BO module. A multi-nozzle experiment evaluates print fidelity, keratinocyte differentiation, fibroblast migration, collagen deposition, and barrier function. Updates are recorded as cell- and zone-specific assertions; a single overall viability score is not used to represent both compartments.

5.4.3. Tumor microenvironment models: prioritizing biological fidelity over printability alone

(a) Input and retrieval

The query specifies tumor type, stromal and immune-cell composition, matrix-stiffness range, oxygen gradient, culture duration, and functional readouts. The BioKG links collagen, Matrigel, alginate–gelatin, and stiffness-gradient records to hypoxia, immune infiltration, and spatial-omics evidence, while distinguishing native-tissue observations from bioprinted-model measurements.[110-114]

(b) Design, build, test, and learn

Candidate formulations are ranked for both manufacturability and preservation of the intended tumor phenotype. Experiments quantify geometric fidelity, hypoxic-zone extent, hypoxia-inducible factor 1-alpha (HIF-1α), immune-cell distribution, cytokines, and spatial gene-expression patterns. The learn phase records where a formulation fails to reproduce the target phenotype; printability without biological fidelity is treated as an unsuccessful outcome.

Table 4 compares these three scenarios using common DBTL dimensions while preserving tissue- and cell-specific constraints.

5.5. Minimum proof-of-concept implementation and evaluation framework

A minimum viable demonstration could begin with a bounded corpus of open-access extrusion-bioprinting papers focused on one material family and one cell class. A manually annotated subset would define entities, relations, numerical-condition bindings, cell metadata, and contradictory assertions. An encoder-only extraction pipeline, followed by terminology normalization, entity linking, and duplicate resolution, would populate a small property graph; hybrid text and graph retrieval would return condition-matched evidence; and a generative LLM would produce a structured recommendation containing candidate ranges, supporting and conflicting citations, uncertainty flags, and an explicit statement of extrapolation.

Table 4. Illustrative LLM-assisted DBTL workflow requirements and constraints for three representative tissue-engineering targets

DBTL optimization dimension	Vascular tissue engineering	Skin/Wound repair bioprinting	TME modeling	Refs
Representative bioink systems	GelMA–alginate (coaxial); fibrin–gelatin; Alginate/HA/Gelatin	dECM–fibrin; GelMA–HA; collagen–fibrin multi-layer	Stiff ECM mimic (collagen I + Matrigel); alginate–gelatin with tunable stiffness gradient	88,94,106-109,114
Illustrative mechanical and transport constraints	Compliant construct, stable lumen geometry, and a formulation-specific gelation window; no universal G' target	Layer-specific mechanics and barrier function; targets depend on tissue source, construct design, and assay conditions	Matrix stiffness and oxygen transport matched to the tumor model; targets are tumor-, cell-, and assay-dependent	87,94,106-114
Critical formulation challenge	Biofabrication window is narrow: gelation must be rapid enough to maintain lumen geometry yet slow enough to avoid nozzle blockage; endothelial cell viability under coaxial extrusion shear	Multi-material co-culture: keratinocyte and fibroblast compartments require spatially distinct formulations; dECM batch variability confounds reproducibility	Recapitulating hypoxic gradients and ECM-immune cell interactions simultaneously; stiffness must be tunable without compromising cell encapsulation	94,95,107,111
Key parameters for LLM-BioKG query	Gelation time; lumen and coaxial-nozzle dimensions; endothelial-cell species/source, passage, density, medium, viability assay and time point; G' at physiological temperature	Keratinocyte and fibroblast source, passage, density, medium, oxygen exposure, viability assay and time point; migration; dECM tissue source and batch; crosslinking homogeneity; spatial resolution	Tumor, stromal, and immune-cell source and ratio; passage; oxygen conditions and culture duration; collagen concentration; oxygen diffusivity; modulus; immune infiltration; hypoxic-zone fraction	[94,95,106,107,110,111]
LLM hypothesis generation priority	Retrieve condition-matched GelMA–alginate and coaxial records; present concordant and conflicting evidence; propose bounded crosslinking-kinetics variants	Synthesize dECM and fibrin literature while preserving tissue source, batch, cell, and assay context; propose zone-specific formulation sequences	Integrate ECM biochemistry, spatial-omics, and rheology evidence while distinguishing native-tissue observations from bioprinted measurements; propose candidate stiffness-gradient formulations	27,59,94,106,107,110,111
Multi-objective BO Pareto objectives	Maximize endothelial viability and lumen patency; maintain gelation time within a predefined feasible window; minimize extrusion pressure subject to safety and printability constraints	Maximize cell-type-specific viability, migration, differentiation, and barrier function; minimize between-batch performance variance and crosslinking heterogeneity	Maximize tumor-phenotype and hypoxia-profile fidelity; minimize deviation from prespecified matrix-stiffness and spatial oxygen targets	86-91
DBTL build phase specifics	Coaxial or core–shell extrusion under formulation-specific temperature and crosslinking conditions	Multi-nozzle or multi-material deposition followed by formulation-appropriate thermal, enzymatic, ionic, or photo-crosslinking	Extrusion or multi-material deposition, with sacrificial or embedded support where required by geometry and formulation	94,95,106-109,114
DBTL test-phase imaging, structural, and functional metrics	Lumen diameter and wall-thickness uniformity using validated imaging and segmentation; endothelial monolayer integrity and perfusability	Epidermal/dermal layer thickness; barrier function; re-epithelialization; cell-type-specific viability and differentiation	Hypoxic-zone extent; spheroid compaction; immune-cell spatial distribution; tumor-phenotype markers	94,95,97,99,108-114

(*Cont'd...*)

Table 4. (*Continued*)

DBTL optimization dimension	Vascular tissue engineering	Skin/Wound repair bioprinting	TME modeling	Refs
DBTL learn phase BioKG update	Condition-specific viability–gelation, lumen-geometry, and crosslinking–mechanics assertions with provenance, uncertainty, and failed outcomes retained	Cell- and zone-specific formulation–viability assertions, dECM source–property records, and print-sequence outcomes with provenance and conflicts retained	Condition-specific stiffness–immune-infiltration, spatial-omics, and oxygen-gradient assertions with provenance and failed outcomes retained	101-103,110
Primary biological validation readouts	Endothelial cell tube formation (angiogenesis assay); NO production; VE-cadherin expression	Keratinocyte differentiation markers (KRT10, involucrin); fibroblast collagen secretion (Col I, Col III)	HIF-1α; PD-L1 expression; cytokine profile (IL-6, TNF-α, IFN-γ)	94,95,106,107,111
Key unresolved knowledge gap for LLM	Sparse coaxial bioprinting records in BioKG; crosslinking kinetics data rarely reported in machine-readable format	dECM characterization endpoints inconsistent across sources; no community consensus on keratinocyte–fibroblast co-culture viability definition	Spatial omics–bioink integration at very early stage; no benchmark dataset linking formulation to quantitative TME recapitulation metrics	52,71,110,111
Proposed computational architecture (not yet validated for the stated use case)	BioKG-grounded RAG and LLM-supported constraint definition coupled to numerical multi-objective BO	BioKG-grounded RAG and LLM-supported evidence synthesis coupled to cVAE or BO for numerical candidate generation	BioKG-grounded RAG and spatial-omics-informed evidence synthesis coupled to numerical inverse design or BO	42,45,76,80,88-91

Notes: These entries are illustrative workflow specifications, not reports of completed LLM-guided bioink experiments. Numerical targets, fabrication settings, and biological endpoints require condition-matched evidence and prospective validation for each use case.
Abbreviations: BioKG: Bioink knowledge graph; BO: Bayesian optimization; CV: Computer vision; cVAE: Conditional variational autoencoder; DBTL: Design–build–test–learn; dECM: Decellularized extracellular matrix; ECM: Extracellular matrix; G′: Storage modulus; GelMA: Gelatin methacryloyl; HA: Hyaluronic acid; HIF-1α: Hypoxia-inducible factor 1α; IFN-γ: Interferon gamma; IL-6: Interleukin-6; KRT10: Keratin 10; LLM: Large language model; NO: Nitric oxide; PD-L1: Programmed death-ligand 1; RAG: Retrieval-augmented generation; TME: Tumor microenvironment; TNF-α: Tumor necrosis factor alpha; VE: Vascular endothelial.

doi: 10.36922/IJB026350375

Evaluation should separate pipeline components. Information extraction should be assessed with entity- and relation-level precision, recall, and F1 scores, together with numerical value-unit-condition binding accuracy. Retrieval should be assessed with Recall@k, expert-judged evidence-context compatibility, and contradiction-retrieval precision, recall, and F1. Generation should be assessed with claim-level citation correctness and completeness, unsupported-claim rate, confidence calibration using predefined confidence outputs and an appropriate calibration metric, and blinded expert review. A prospective formulation test should then compare an LLM-assisted workflow with a matched BO-only baseline using the same initial data, search space, and experimental budget; where feasible, a manual literature-search-plus-BO baseline should also be included. Relevant endpoints include the number of candidates satisfying prespecified feasibility constraints, the number of experiments required to reach a prespecified target, cell-type-specific viability, shape fidelity, reproducibility across batches, and the frequency of unsafe or infeasible recommendations. No end-to-end results of this kind are claimed in the present review; the framework defines what would constitute a credible proof of concept.

6. Evidence boundaries, translational challenges, and research priorities

6.1. Limitations of the evidence base

The principal limitation of this review is the immaturity of the direct evidence base. Most published demonstrations concern biomedical NLP, polymer or materials informatics, chemical optimization, computer vision, or autonomous experimentation rather than LLM-guided bioink formulation followed by bioprinting and biological validation. Transfer across these domains is plausible but not guaranteed because bioinks couple time-dependent rheology, living cells, culture conditions, and tissue-specific function. This review is therefore not a meta-analysis, and the cross-domain metrics in Tables 1–3 cannot be pooled or interpreted as expected bioprinting performance.

The literature is also vulnerable to publication and availability bias: successful workflows and positive model comparisons are more likely to be reported, whereas failed formulations, unsafe recommendations, and negative automation results may remain unpublished. Closed-access methods, incomplete parameter reporting, preprints, changing model versions, and inconsistent definitions of printability and viability further limit reproducibility. The proposed evidence framework and provenance-aware BioKG can reduce the risk of overinterpretation but cannot compensate for missing or selectively reported experiments.

6.2. Technical, biological, and regulatory limitations

Hallucination remains the most immediate language-model risk. RAG and knowledge graph grounding can reduce unsupported generation, but their effect is task-, corpus-, retriever-, and evaluation-dependent. A sparse or biased corpus can produce well-cited but still misleading answers. Systems should therefore expose retrieved evidence, incompatible conditions, and uncertainty; enforce material–cell compatibility constraints; and require expert approval before physical execution. Conflict-aware retrieval is especially important when apparently identical formulations differ in cell source, printing conditions, or assay definition.

Biological heterogeneity creates a second limitation. Viability must be interpreted by cell type, donor, passage, culture condition, assay, and time point, and short-term survival cannot substitute for phenotype or tissue function. A third limitation is auditability. Attention maps and chain-of-thought text do not provide regulatory-grade explanations; defensible oversight requires source provenance, structured decision logs, model and prompt versioning, predefined acceptance criteria, and validated control boundaries.[115,116] Finally, clinical translation requires compatibility with quality-management, biocompatibility, data-governance, and manufacturing requirements, none of which has been established for LLM-generated bioink recommendations.

6.3. Research priorities

The most actionable near-term priority is a community-curated, ontology-anchored BioKG accompanied by an open benchmark corpus. The resource should include negative results, condition-qualified cell data, source provenance, and explicit conflict relations. Benchmark tasks should separately evaluate extraction, retrieval, evidence synthesis, candidate ranking, and experimental decision support. Prospective multicenter studies should then compare LLM-assisted workflows with matched baselines under identical data and experimental budgets.

Self-driving bioprinting remains a horizon goal rather than a demonstrated general capability.[22,23] Progress should proceed through bounded autonomy: first literature extraction and documentation, then human-approved experiment planning, then between-run model updating, and only finally validated online control for narrowly specified variables. This staged pathway preserves the potential of LLM-enabled coordination while keeping biological safety, reproducibility, and causal experimental evidence central. Table 5 summarizes the corresponding challenges, mitigations, residual gaps, and validation priorities.

Table 5. Key challenges in LLM integration for bioprinting, current mitigation strategies, and unresolved research gaps

Challenge	Specific manifestation in bioprinting	Current best mitigation	Demonstrated efficacy of mitigation	Residual gap	Research priority	Relevant framework/ standard	Refs
Hallucination	LLM generates plausible but incorrect bioink formulations (e.g., erroneous polymer–crosslinker pairing, fictitious viability values); especially consequential for cell-laden constructs in pre-clinical translation	RAG with BioKG grounding; claim-level source citation; retrieval-path display; contradiction retrieval; expert verification	RAG, KG grounding, and self-reflection improve factuality in task-specific benchmarks, but reported effect sizes are not directly comparable and no bioink-specific benchmark exists	Cannot be eliminated with a sparse or biased corpus; retrieved evidence may be condition-incompatible; performance is limited by corpus completeness and retrieval quality	Critical	ASTM F3659-24; FAIR data principles	42-45,47,78
Data heterogeneity and inaccessibility	Bioink records use incompatible vocabularies and endpoints; substantial relevant literature may be inaccessible, incompletely accessible, or insufficiently machine-readable	FAIR deposition; provenance- and conflict-aware BioKG; condition-qualified cell and assay metadata; community ontology; human-verified protocol extraction	FAIR implementation improves downstream ML performance in adjacent domains; automated extraction achieves 91.25% accuracy on ORD records	No community-agreed bioink ontology or open annotated corpus; incomplete parameter reporting and unavailable negative results remain major gaps	Critical	FAIR principles; ASTM F3659-24; MDS-Onto	64,66,67,71,74
Interpretability and regulatory opacity	LLM reasoning chains opaque; formulation recommendations cannot be audited for GMP compliance; no explainability standard applicable to LLM-generated bioink proposals	Claim-level citations; retrieval-path display; structured decision logs; model and prompt versioning; predefined acceptance criteria; human approval and validated control boundaries	Source provenance and audit logs improve traceability; attention maps and CoT text are not regulatory-grade explanations	No GMP-applicable LLM auditability framework; regulatory applicability depends on the system's intended use and role in the product or manufacturing process	High	GMP guidelines; ISO 13485; EU AI Act (classification depends on intended use)	115,116
Sparse annotated bioink NLP corpus	NER and RE models lack bioink-specific training data; transfer from general biomedical corpora misses printability-specific entity types (e.g., biofabrication window, extrusion pressure as condition modifier)	Transfer learning from MaterialsBERT/MatSciBERT; few-shot prompting with generative LLMs; synthetic data augmentation via LLM	MaterialsBERT extracts ~300k records from polymer literature; NERRE achieves high-quality extraction from ~500 annotated abstracts	No publicly available bioink NLP benchmark dataset; no standard annotation schema for printability, viability, and process entities	High	–	28-30,59

(*Cont'd....*)

Table 5. (*Continued*)

Challenge	Specific manifestation in bioprinting	Current best mitigation	Demonstrated efficacy of mitigation	Residual gap	Research priority	Relevant framework/ standard	Refs
Absence of bioprinting NLP benchmark	No standardized task definitions, evaluation corpora, or metrics for bioink NER, RE, protocol extraction, or formulation recommendation; impossible to compare LLM methods rigorously	Borrow evaluation methodology from materials NLP benchmarks (MatSci-NLP, BLURB); adapt existing chemistry datasets	MatSci-NLP provides methodology template; BLURB covers broader biomedical NLP; neither covers bioink-specific entities	No bioprinting-specific annotated corpus; no leaderboard; no community-agreed evaluation protocol	High	MatSci-NLP benchmark; BLURB benchmark	16,20,117
Reproducibility gap in protocol reporting	Methods-section text heterogeneous in completeness and vocabulary; operator-dependent variation documented across 12 laboratories for the same nominal printing conditions	ASTM F3659-24 terminology; standardized reporting templates; round-robin benchmarking; automated extraction with human verification	Round-robin study quantified inter-lab variability; automated extraction achieves 91.25% on chemistry protocols; no bioink-specific validation yet	No "Open Bioprinting Protocol Database"; ASTM F3659-24 is a guide and does not establish universal quantitative printability criteria	Moderate	ASTM F3659-24	64,71,74
Clinical translation and safety	LLM-generated formulations lack a safety validation pipeline; risk of recommending biomaterials with insufficient cytotoxicity or immunogenicity characterization for clinical-grade constructs	Restrict recommendations to materials with documented regulatory status, biocompatibility evidence, and manufacturing controls appropriate to the intended use; require expert review before testing or execution	RAG can restrict outputs to literature-documented formulations, but it does not establish safety, regulatory acceptability, or clinical effectiveness	No established regulatory pathway for LLM-assisted bioink design; applicable requirements depend on the intended product, use, jurisdiction, and degree of automation	Moderate	ISO 10993; FDA Quality Management System Regulation (21 CFR Part 820); EU MDR 2017/745; EU AI Act, as applicable	80,118

Notes: Mitigation efficacy is task- and implementation-dependent. No listed strategy has been validated as an end-to-end safety or regulatory solution for LLM-guided bioink development.
Abbreviations: AI: Artificial intelligence; BioKG: Bioink knowledge graph; CoT: Chain of thought; EU: European Union; FAIR: Findable, accessible, interoperable, and reusable; FDA: Food and Drug Administration; GMP: Good manufacturing practice; KG: Knowledge graph; LLM: Large language model; MDS-Onto: Materials Data Science Ontology; ML: Machine learning; NER: Named entity recognition; NLP: Natural language processing; ORD: Open Reaction Database; RAG: Retrieval-augmented generation; RE: Relation extraction.

doi: 10.36922/IJB026350375

7. Conclusion

Large language models could make an important contribution to 3D bioprinting by converting narrative literature into structured, queryable, and traceable evidence and by coordinating information across design, fabrication, testing, and learning. Their most credible near-term roles are information extraction, provenance-aware retrieval, constraint elicitation, hypothesis generation, and workflow documentation. Numerical generative models remain appropriate for quantitative candidate generation, whereas BO provides uncertainty-aware experiment selection; computer vision and sensors provide measurements, and validated control rules and deterministic control systems—not LLMs alone—authorize and execute physical adjustments.

At present, evidence for end-to-end LLM-guided bioink discovery is insufficient to claim an autonomous paradigm shift. Progress depends on standardized reporting, condition- and cell-resolved data, conflict-aware knowledge graphs, transparent benchmarks, prospective experimental comparisons, and human oversight. Under these constraints, the framework presented here offers a testable, evidence-bounded research agenda for moving from literature intelligence toward progressively more reproducible and auditable biofabrication workflows.

Acknowledgments

None.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82560411), Shaanxi Natural Science Foundation of China (No. 2025JC-YBMS-916), and the Startup Foundation for Doctors of Yan'an University (No. YDBK2024-91 and No. YDZKBK2025-06).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Yangyang Wang, Jihan Wang, Jing Lv
Funding acquisition: Yangyang Wang, Jihan Wang, Jing Lv
Project administration: Yangyang Wang, Yanning Yang
Supervision: Jihan Wang, Jing Lv
Visualization: Jiaxun Zhang, Haotian Bai, Yufei Ren
Writing–original draft: Yangyang Wang, Jiaxun Zhang
Writing–review & editing: Jihan Wang, Jing Lv

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Mathur V, Agarwal P, Kasturi M, Srinivasan V, Seetharam RN, Vasanthan KS. Innovative bioinks for 3D bioprinting: Exploring technological potential and regulatory challenges. *J Tissue Eng*. 2025;16:20417314241308022.

   doi: 10.1177/20417314241308022

2. Lee SJ, Jeong W, Atala A. 3D Bioprinting for Engineered Tissue Constructs and Patient-Specific Models: Current Progress and Prospects in Clinical Applications. *Adv Mater*. 2024;36(49):e2408032.

   doi: 10.1002/adma.202408032

3. Prendergast ME, Burdick JA. Recent Advances in Enabling Technologies in 3D Printing for Precision Medicine. *Adv Mater*. 2020;32(13):e1902516.

   doi: 10.1002/adma.201902516

4. Dey M, Ozbolat IT. 3D bioprinting of cells, tissues and organs. *Sci Rep*. 2020;10(1):14023.

   doi: 10.1038/s41598-020-70086-y

5. Freeman S, Calabro S, Williams R, Jin S, Ye K. Bioink Formulation and Machine Learning-Empowered Bioprinting Optimization. *Front Bioeng Biotechnol*. 2022;10:913579.

   doi: 10.3389/fbioe.2022.913579

6. Panwar A, Tan LP. Current Status of Bioinks for Micro-Extrusion-Based 3D Bioprinting. *Molecules*. 2016;21(6):685.

   doi: 10.3390/molecules21060685

7. Yin J, Yan M, Wang Y, Fu J, Suo H. 3D Bioprinting of Low-Concentration Cell-Laden Gelatin Methacrylate (GelMA) Bioinks with a Two-Step Cross-linking Strategy. *ACS Appl Mater Interfaces*. 2018;10(8):6849-6857.

   doi: 10.1021/acsami.7b16059

8. Jongprasitkul H, Turunen S, Parihar VS, Kellomäki M. Sequential Cross-linking of Gallic Acid-Functionalized GelMA-Based Bioinks with Enhanced Printability for Extrusion-Based 3D Bioprinting. *Biomacromolecules*. 2023;24(1):502-514.

   doi: 10.1021/acs.biomac.2c01418

9. Zhang CY, Fu CP, Li XY, *et al*. Three-Dimensional Bioprinting of Decellularized Extracellular Matrix-Based Bioinks for Tissue Engineering. *Molecules*. 2022;27(11):3442.

   doi: 10.3390/molecules27113442

10. Chen H, Liu Y, Balabani S, Hirayama R, Huang J. Machine Learning in Predicting Printable Biomaterial Formulations for Direct Ink Writing. *Research (Wash D C)*. 2023;6:0197.

doi: 10.34133/research.0197

11. Theus AS, Ning L, Hwang B, *et al*. Bioprintability: Physiomechanical and Biological Requirements of Materials for 3D Bioprinting Processes. *Polymers (Basel)*. 2020;12(10):2262.

doi: 10.3390/polym12102262

12. Schwab A, Levato R, D'Este M, Piluso S, Eglin D, Malda J. Printability and Shape Fidelity of Bioinks in 3D Bioprinting. *Chem Rev*. 2020;120(19):11028-11055.

doi: 10.1021/acs.chemrev.0c00084

13. Wang Y, Bai H, Ren Y, Yang Y, Wang J. Artificial intelligence-augmented bioprinting systems: Data-driven optimization and future applications in pharmacological research. *Int J Bioprint*. 2025;11(6):1-22.

doi: 10.36922/ijb025350349

14. Lee J, Oh SJ, An SH, Kim WD, Kim SH. Machine learning-based design strategy for 3D printable bioink: elastic modulus and yield stress determine printability. *Biofabrication*. 2020;12(3):035018.

doi: 10.1088/1758-5090/ab8707

15. Ramesh S, Deep A, Tamayol A, Kamaraj A, Mahajan C, Madihally S. Advancing 3D bioprinting through machine learning and artificial intelligence. *Bioprint*. 2024;38:e00331.

doi: 10.1016/j.bprint.2024.e00331

16. Zhou J, Li H, Chen S, Chen Z, Han Z, Gao X. Large language models in biomedicine and healthcare. *npj Artif Intell*. 2025;1(1):44.

doi: 10.1038/s44387-025-00047-1

17. Lin A, Ye J, Qi C, *et al*. Bridging artificial intelligence and biological sciences: a comprehensive review of large language models in bioinformatics. *Brief Bioinform*. 2025;26(4):bbaf357.

doi: 10.1093/bib/bbaf357

18. Lehman E, Hernandez E, Mahajan D, *et al*. Do We Still Need Clinical Language Models? *arXiv*. Preprint posted online 2023.

doi: 10.48550/ARXIV.2302.08091

19. Gao Y, Xiong Y, Gao X, *et al*. Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv*. Preprint posted online 2023.

doi: 10.48550/ARXIV.2312.10997

20. Li W, Mao Z, Xiao Z, *et al*. Large language model for knowledge synthesis and AI-enhanced biomanufacturing. *Trends Biotechnol*. 2025;43(8):1864-1875.

doi: 10.1016/j.tibtech.2025.02.008

21. Zhou L, Ling H, Fu C, *et al*. Autonomous Agents for Scientific Discovery: Orchestrating Scientists, Language, Code, and Physics. *arXiv*. Preprint posted online 2025.

doi: 10.48550/ARXIV.2510.09901

22. Zhang S, Liang D, Lei Y, Luo Y, Shi R. Artificial intelligence-assisted smart hydrogel bioinks in 3D bioprinting: design, optimization, and construct validation for functional tissue engineering. *Front Bioeng Biotechnol*. 2026;14:1898243.

doi: 10.3389/fbioe.2026.1898243

23. Liu S, Kaur N, Song DH, Moses JC, Ozbolat IT. Self-driving bioprinting laboratories. *Biofabrication*. 2026;18(1):013001.

doi: 10.1088/1758-5090/ae3645

24. Lee J, Yoon W, Kim S, *et al*. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234-1240.

doi: 10.1093/bioinformatics/btz682

25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics; 2019:4171-4186.

doi: 10.18653/v1/n19-1423

26. Beltagy I, Lo K, Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics; 2019:3613-3618.

doi: 10.18653/v1/d19-1371

27. Gu Y, Tinn R, Cheng H, *et al*. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. A*CM Trans Comput Healthcare*. 2021;3(1):1-23.

doi: 10.1145/3458754

28. Shetty P, Rajan AC, Kuenneth C, *et al*. A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing. *npj Comput Mater*. 2023;9(1):52.

doi: 10.1038/s41524-023-01003-w

29. Gupta T, Zaki M, Krishnan NMA, Mausam. MatSciBERT: A materials domain language model for text mining and information extraction. *npj Comput Mater*. 2022;8(1):102.

doi: 10.1038/s41524-022-00784-w

30. Li M, Zhuang Y, Hou Y, *et al*. Integrating graph convolutional networks with large language models for structured biomedical material knowledge representation. *Regen Biomater*. 2025;12:rbaf083.

doi: 10.1093/rb/rbaf083

31. Chimene D, Lennox KK, Kaunas RR, Gaharwar AK. Advanced Bioinks for 3D Printing: A Materials Science Perspective. *Ann Biomed Eng*. 2016;44(6):2090-2102.

doi: 10.1007/s10439-016-1638-y

32. Moroni L, Boland T, Burdick JA, *et al*. Biofabrication: A Guide to Technology and Terminology. *Trends Biotechnol*. 2018;36(4):384-402.

doi: 10.1016/j.tibtech.2017.10.015

33. Zhang W, Zhang J. Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review. *Mathematics*. 2025;13(5):856.

doi: 10.3390/math13050856

34. M Bran A, Cox S, Schilter O, Baldassari C, White AD, Schwaller P. Augmenting large language models with chemistry tools. *Nat Mach Intell*. 2024;6(5):525-535.

doi: 10.1038/s42256-024-00832-8

35. Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. *Nature*. 2023;624(7992):570-578.

doi: 10.1038/s41586-023-06792-0

36. AI4Science MR, Quantum MA. The Impact of Large Language Models on Scientific Discovery: a Preliminary Study using GPT-4. *arXiv*. Preprint posted online 2023.

doi: 10.48550/ARXIV.2311.07361

37. Qi B, Zhang K, Li H, *et al*. Large Language Models are Zero Shot Hypothesis Proposers. *arXiv*. Preprint posted online 2023.

doi: 10.48550/ARXIV.2311.05965

38. Abdel-Rehim A, Zenil H, Orhobor O, *et al*. Scientific hypothesis generation by large language models: laboratory validation in breast cancer treatment. *J R Soc Interface*. 2025;22(227):20240674.

doi: 10.1098/rsif.2024.0674

39. Labrak Y, Rouvier M, Dufour R. A Zero-shot and Few-shot Study of Instruction-Finetuned Large Language Models Applied to Clinical and Biomedical Tasks. *arXiv*. Preprint posted online 2023.

doi: 10.48550/ARXIV.2307.12114

40. Wei J, Wang X, Schuurmans D, *et al*. Chain-Of-Thought Prompting Elicits Reasoning in Large Language Models. *Adv Neural Inf Process Syst*. 2022:24824-24837.

doi: 10.52202/068431-1800

41. Lewis P, Perez E, Piktus A, *et al*. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020:9459-9474.

42. Fan W, Ding Y, Ning L, *et al*. A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024:6491-6501.

doi: 10.1145/3637528.3671470

43. Neha F, Bhati D, Shukla DK. Retrieval-Augmented Generation (RAG) in Healthcare: A Comprehensive Review. *AI*. 2025;6(9):226.

doi: 10.3390/ai6090226

44. Xu S, Yan Z, Dai C, Wu F. MEGA-RAG: a retrieval-augmented generation framework with multi-evidence guided answer refinement for mitigating hallucinations of LLMs in public health. *Front Public Health*. 2025;13:1635381.

doi: 10.3389/fpubh.2025.1635381

45. Edge D, Trinh H, Cheng N, *et al*. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. *arXiv*. Preprint posted online 2024.

doi: 10.48550/ARXIV.2404.16130

46. Peng B, Zhu Y, Liu Y, *et al*. Graph Retrieval-Augmented Generation: A Survey. *ACM Trans Inf Syst*. 2025;44(2):35.

doi: 10.1145/3777378

47. Xu S, Chen M, Chen S. Enhancing Retrieval-Augmented Generation Models with Knowledge Graphs: Innovative Practices Through a Dual-Pathway Approach. In: Huang DS, Si Z, Chen W, eds. *Advanced Intelligent Computing Technology and Applications*. Singapore: Springer Nature; 2024:398-409.

doi: 10.1007/978-981-97-5678-0_34

48. Xu L, Xie H, Qin SJ, Tao X, Wang FL. Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment. *IEEE Trans Pattern Anal Mach Intell*. 2026;48(6):6107-6126.

doi: 10.1109/TPAMI.2026.3657354

49. Hu EJ, Shen Y, Wallis P, *et al*. LoRA: Low-Rank Adaptation of Large Language Models. In: *ICLR 2022*. 2022. Accessed July 29, 2026. https://www.microsoft.com/en-us/research/publication/lora-low-rank-adaptation-of-large-language-models/

50. Shao J, Khan AN, Brett C, Berg T, Li X, Yao B. Multi-Task LLM with LoRA Fine-Tuning for Automated Cancer Staging and Biomarker Extraction. *arXiv*. Preprint posted online 2026.

doi: 10.48550/ARXIV.2604.13328

51. Silva Robazzi JV, Derman ID, Gupta D, *et al*. The Synergy of Artificial Intelligence and 3D Bioprinting: Unlocking New Frontiers in Precision and Tissue Fabrication. *Adv Funct Mater*. 2026;36(1):e09530.

doi: 10.1002/adfm.202509530

52. Sarah R, Schimmelpfennig K, Rohauer R, Lewis CL, Limon SM, Habib A. Characterization and Machine Learning-Driven Property Prediction of a Novel Hybrid Hydrogel Bioink Considering Extrusion-Based 3D Bioprinting. *Gels*. 2025;11(1):45.

doi: 10.3390/gels11010045

53. Luo Y, Yang Z, Meng F, Li Y, Zhou J, Zhang Y. An Empirical Study of Catastrophic Forgetting in Large Language Models During Continual Fine-tuning. *arXiv*. Preprint posted online 2023.

doi: 10.48550/ARXIV.2308.08747

54. Perera N, Dehmer M, Emmert-Streib F. Named Entity Recognition and Relation Detection for Biomedical Information Extraction. *Front Cell Dev Biol*. 2020;8:673.

doi: 10.3389/fcell.2020.00673

55. Nasar Z, Jaffry SW, Malik MK. Named Entity Recognition and Relation Extraction. *ACM Comput Surv*. 2022;54(1):20.

doi: 10.1145/3445965

56. Grouin C, Grabar N. Year 2023 in Biomedical Natural Language Processing: a Tribute to Large Language Models and Generative AI. *Yearb Med Inform*. 2024;33(1):241-248.

doi: 10.1055/s-0044-1800751

57. Kalyan KS, Rajasekharan A, Sangeetha S. AMMU: A survey of transformer-based biomedical pretrained language models. *J Biomed Inform*. 2022;126:103982.

doi: 10.1016/j.jbi.2021.103982

58. Foppiano L, Lambard G, Amagasa T, Ishii M. Mining experimental data from materials science literature with large language models: an evaluation study. *Sci Technol Adv Mater Methods*. 2024;4(1):2356506.

doi: 10.1080/27660400.2024.2356506

59. Dagdelen J, Dunn A, Lee S, *et al*. Structured information extraction from scientific text with large language models. *Nat Commun*. 2024;15(1):1418.

doi: 10.1038/s41467-024-45563-x

60. Kyle S, Jessop ZM, Al-Sabah A, Whitaker IS. 'Printability' of Candidate Biomaterials for Extrusion Based 3D Printing: State-of-the-Art. *Adv Healthc Mater*. 2017;6(16):1700264.

doi: 10.1002/adhm.201700264

61. Paxton N, Smolan W, Böck T, Melchels F, Groll J, Jungst T. Proposal to assess printability of bioinks for extrusion-based bioprinting and evaluation of rheological properties governing bioprintability. *Biofabrication*. 2017;9(4):044107.

doi: 10.1088/1758-5090/aa8dd8

62. Malda J, Visser J, Melchels FP, *et al*. 25th anniversary article: Engineering hydrogels for biofabrication. *Adv Mater*. 2013;25(36):5011-5028.

doi: 10.1002/adma.201302042

63. Gillispie G, Prim P, Copus J, *et al*. Assessment methodologies for extrusion-based bioink printability. *Biofabrication*. 2020;12(2):022003.

doi: 10.1088/1758-5090/ab6f0d

64. Ai Q, Meng F, Shi J, Pelkie B, Coley CW. Extracting structured data from organic synthesis procedures using a fine-tuned large language model. *Digit Discov*. 2024;3(9):1822-1831.

doi: 10.1039/d4dd00091a

65. Shi Y, Rampal N, Zhao C, *et al*. Comparison of LLMs in extracting synthesis conditions and generating Q&A datasets for metal–organic frameworks. *Digit Discov*. 2025;4(10):2676-2683.

doi: 10.1039/d5dd00081e

66. Rajamohan BP, Bradley ACH, Tran VD, *et al*. Materials Data Science Ontology(MDS-Onto): Unifying Domain Knowledge in Materials and Applied Data Science. *Sci Data*. 2025;12(1):628.

doi: 10.1038/s41597-025-04938-5

67. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al*. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.

doi: 10.1038/sdata.2016.18

68. Venugopal V, Olivetti E. MatKG: An autonomously generated knowledge graph in Material Science. *Sci Data*. 2024;11(1):217.

doi: 10.1038/s41597-024-03039-z

69. Ye Y, Ren J, Wang S, *et al*. Construction and Application of Materials Knowledge Graph in Multidisciplinary Materials Science via Large Language Model. *Adv Neural Inf Process Syst 37*. 2024:56878-56897.

doi: 10.52202/079017-1812

70. Stocker M, Oelen A, Jaradeh MY, *et al*. FAIR scientific information with the Open Research Knowledge Graph. *FAIR Connect*. 2023;1(1):19-21.

doi: 10.3233/fc-221513

71. Grijalva Garces D, Strauß S, Gretzinger S, *et al*. On the reproducibility of extrusion-based bioprinting: round robin study on standardization in the field. *Biofabrication*. 2023;16(1):015002.

doi: 10.1088/1758-5090/acfe3b

72. Skardal A, Atala A. Biomaterials for integration with 3-D bioprinting. *Ann Biomed Eng*. 2015;43(3):730-746.

doi: 10.1007/s10439-014-1207-1

73. Zimmerling A, Zhou Y, Chen X. Synthesis of Alginate/Collagen Bioink for Bioprinting Respiratory Tissue Models. *J Funct Biomater*. 2024;15(4):90.

doi: 10.3390/jfb15040090

74. F04 Committee. ASTM F3659-24: Standard Guide for Bioinks Used in Bioprinting. ASTM International; 2024.

doi: 10.1520/F3659-24

75. Kearnes SM, Maser MR, Wleklinski M, *et al*. The Open Reaction Database. *J Am Chem Soc*. 2021;143(45):18820-18826.

doi: 10.1021/jacs.1c09820

76. Krotkov NA, Sbytov DA, Chakhoyan AA, *et al*. Nanostructured Material Design via a Retrieval-Augmented Generation (RAG) Approach: Bridging Laboratory Practice and Scientific Literature. *J Chem Inf Model*. 2025;65(20):11064-11078.

doi: 10.1021/acs.jcim.5c01897

77. Hassan T, Karim MF, Jeelani H, Behnam E, Green R, Syed FJ. Optimizing Medical Question-Answering Systems: A Comparative Study of Fine-Tuned and Zero-Shot Large Language Models with RAG Framework. *arXiv*. Preprint posted online 2025.

doi: 10.48550/ARXIV.2512.05863

78. Asai A, Wu Z, Wang Y, Sil A, Hajishirzi H. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection. In: *ICLR 2024*. 2024:9112-9141. Accessed July 29, 2026. https://proceedings.iclr.cc/paper_files/paper/2024/file/25f7be9694d7b32d5cc670927b8091e1-Paper-Conference.pdf

79. Wang S, Yang H, Liu W. Research on the construction and application of retrieval enhanced generation (RAG) model based on knowledge graph. *Sci Rep*. 2025;15(1):40425.

doi: 10.1038/s41598-025-21222-z

80. Mohammad S, Akand R, Cook KM, Nilufar S, Chowdhury F. Leveraging Deep Learning and Generative AI for Predicting Rheological Properties and Material Compositions of 3D Printed Polyacrylamide Hydrogels. *Gels*. 2024;10(10):660.

doi: 10.3390/gels10100660

81. Park H, Li Z, Walsh A. Has generative artificial intelligence solved inverse materials design? *Matter*. 2024;7(7):2355-2367.

doi: 10.1016/j.matt.2024.05.017

82. Zeni C, Pinsler R, Zügner D, *et al*. A generative model for inorganic materials design. *Nature*. 2025;639(8055):624-632.

doi: 10.1038/s41586-025-08628-5

83. Zhang Z, Tao ZZ, Du R, Huo R, Zheng X. Artificial Intelligence Informed Hydrogel Biomaterials in Additive Manufacturing. *Gels*. 2025;11(12):981.

doi: 10.3390/gels11120981

84. Lim KS, Galarraga JH, Cui X, Lindberg GCJ, Burdick JA, Woodfield TBF. Fundamentals and Applications of Photo-Cross-Linking in Bioprinting. *Chem Rev*. 2020;120(19):10662-10694.

doi: 10.1021/acs.chemrev.9b00812

85. Chen Z, Yuan X, Zhang J, *et al*. LABO: LLM-Accelerated Bayesian Optimization through Broad Exploration and Selective Experimentation. *arXiv*. Preprint posted online 2026.

doi: 10.48550/ARXIV.2605.22054

86. Malekpour A, Chen X. Printability and Cell Viability in Extrusion-Based Bioprinting from Experimental, Computational, and Machine Learning Views. *J Funct Biomater*. 2022;13(2):40.

doi: 10.3390/jfb13020040

87. Elango J, Zamora-Ledezma C. Rheological, Structural, and Biological Trade-Offs in Bioink Design for 3D Bioprinting. *Gels*. 2025;11(8):659.

doi: 10.3390/gels11080659

88. Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. In: *Proc IEEE*. 2016;104:148-175.

doi: 10.1109/JPROC.2015.2494218

89. Snoek J, Larochelle H, Adams RP. Practical Bayesian Optimization of Machine Learning Algorithms. *Adv Neural Inf Process Syst 25*. 2012. Accessed July 29, 2026. https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

90. Liang Q, Gongora AE, Ren Z, *et al*. Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. *npj Comput Mater*. 2021;7(1):188.

doi: 10.1038/s41524-021-00656-9

91. Gupta R, Hartford J, Liu B. LLMs for Bayesian Optimization in Scientific Domains: Are We There Yet? *arXiv*. Preprint posted online 2025.

doi: 10.48550/ARXIV.2509.21403

92. Han D, Ai Z, Cai P, *et al*. ChemBOMAS: Accelerated BO in Chemistry with LLM-Enhanced Multi-Agent System. *arXiv*. Preprint posted online 2025.

doi: 10.48550/ARXIV.2509.08736

93. Kumbhar S, Mishra V, Coutinho K, Handa D, Iquebal A, Baral C. Hypothesis Generation for Materials Discovery and Design Using Goal-Driven and Constraint-Guided LLM Agents. *arXiv*. Preprint posted online 2025.

doi: 10.48550/ARXIV.2501.13299

94. Vanlauwe F, Dermaux C, Shamieva S, Vermeiren S, Van Vlierberghe S, Blondeel P. Small molecular weight alginate gel porogen for the 3D bioprinting of microvasculature. *Front Bioeng Biotechnol*. 2024;12:1452477.

doi: 10.3389/fbioe.2024.1452477

95. Mazari-Arrighi E, Lépine M, Ayollo D, *et al*. Self-Organization of Long-Lasting Human Endothelial Capillary-Like Networks Guided by DLP Bioprinting. *Adv Healthc Mater*. 2024;13(14):e2302830.

doi: 10.1002/adhm.202302830

96. Zhou W, Gao L, Wu M, Liu J. AI-assisted interdisciplinary integration of multimodal CT/MRI imaging and 3D printing in medicine: Technical advancements, clinical translations, and future perspectives. *Int J Bioprint*. 2026;12(1):125-144.

doi: 10.36922/ijb025390404

97. Kelly D, Sergis V, Ventura i Blanco L, Mason K, Daly AC. Autonomous Control of Extrusion Bioprinting Using Convolutional Neural Networks. *Adv Funct Mater*. 2025;35(30):2424553.

doi: 10.1002/adfm.202424553

98. Jin Z, Zhang Z, Shao X, Gu GX. Monitoring Anomalies in 3D Bioprinting with Deep Neural Networks. *ACS Biomater Sci Eng*. 2023;9(7):3945-3952.

doi: 10.1021/acsbiomaterials.0c01761

99. Sergis V, Kelly D, Pramanick A, Britchfield G, Mason K, Daly AC. *In-situ* quality monitoring during embedded bioprinting using integrated microscopy and classical computer vision. *Biofabrication*. 2025;17(2):025004.

doi: 10.1088/1758-5090/adaa22

100. Ke D. AI-driven closed-loop extrusion-based bioprinting system: from dynamic regulation to medical translation. *Bioprinting*. 2026;54:e00459.

doi: 10.1016/j.bprint.2025.e00459

101. Spannenkrebs JB, Eiermann A, Zoll T, Hackenschmidt S, Kabisch J. Closing the loop: establishing an autonomous test-learn cycle to optimize induction of bacterial systems using a robotic platform. *Front Bioeng Biotechnol*. 2025;12:1528224.

doi: 10.3389/fbioe.2024.1528224

102. Hatem S, Khoriba G, Gad-Elrab MH, ElHelw M. Up To Date: Automatic Updating Knowledge Graphs Using LLMs. *Procedia Comput Sci*. 2024;244:327-334.

doi: 10.1016/j.procs.2024.10.206

103. Shi H, Xu Z, Wang H, *et al*. Continual Learning of Large Language Models: A Comprehensive Survey. *ACM Comput Surv*. 2025;58(5):1-42.

doi: 10.1145/3735633

104. Basara G, Ozcebe SG, Ellis BW, Zorlutuna P. Tunable Human Myocardium Derived Decellularized Extracellular Matrix for 3D Bioprinting and Cardiac Tissue Engineering. *Gels*. 2021;7(2):70.

doi: 10.3390/gels7020070

105. Lim KS, Abinzano F, Bernal PN, *et al*. One-Step Photoactivation of a Dual-Functionalized Bioink as Cell Carrier and Cartilage-Binding Glue for Chondral Regeneration. *Adv Healthc Mater*. 2020;9(15):e1901792.

doi: 10.1002/adhm.201901792

106. Zhe M, Wu X, Yu P, *et al*. Recent Advances in Decellularized Extracellular Matrix-Based Bioinks for 3D Bioprinting in Tissue Engineering. *Materials*. 2023;16(8):3197.

doi: 10.3390/ma16083197

107. Kim BS, Das S, Jang J, Cho DW. Decellularized Extracellular Matrix-based Bioinks for Engineering Tissue- and Organ-specific Microenvironments. *Chem Rev*. 2020;120(19):10608-10661.

doi: 10.1021/acs.chemrev.9b00808

108. Zhang M, Zhang C, Li Z, Fu X, Huang S. Advances in 3D skin bioprinting for wound healing and disease modeling. *Regen Biomater*. 2022;10:rbac105.

doi: 10.1093/rb/rbac105

109. Cavallo A, Al Kayal T, Mero A, *et al*. Fibrinogen-Based Bioink for Application in Skin Equivalent 3D Bioprinting. *J Funct Biomater*. 2023;14(9):459.

doi: 10.3390/jfb14090459

110. Saul D, Kosinsky RL. Spatial transcriptomics herald a new era of transcriptome research. *Clin Transl Med*. 2023;13(5):e1264.

doi: 10.1002/ctm2.1264

111. Du L, Yang H. Spatial omics in 3D culture model systems: decoding cellular positioning mechanisms and microenvironmental dynamics. *J Transl Med*. 2025;23(1):1356.

doi: 10.1186/s12967-025-07390-6

112. Rodrigues DB, Reis RL, Pirraco RP. Modelling the complex nature of the tumor microenvironment: 3D tumor spheroids as an evolving tool. *J Biomed Sci*. 2024;31(1):13.

doi: 10.1186/s12929-024-00997-9

113. Zanoni M, Piccinini F, Arienti C, *et al*. 3D tumor spheroid models for in vitro therapeutic screening: a systematic approach to enhance the biological relevance of data obtained. *Sci Rep*. 2016;6:19103.

doi: 10.1038/srep19103

114. Yi HG, Jeong YH, Kim Y, *et al*. A bioprinted human-glioblastoma-on-a-chip for the identification of patient-specific responses to chemoradiotherapy. *Nat Biomed Eng*. 2019;3(7):509-519.

doi: 10.1038/s41551-019-0363-x

115. Rizzo M, Veneri A, Marcuzzo M, *et al*. Machine learning models explanations as interpretations of evidence: a theoretical framework of explainability and its implications on high-stakes biomedical decision-making. *BMC Med Res Methodol*. 2025;25(S1):282.

doi: 10.1186/s12874-025-02703-1

116. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20(1):310.

doi: 10.1186/s12911-020-01332-6

117. Yang T, Xiao Y, Bao Z, Hao J, Peng J. The rise and potential opportunities of large language model agents in bioinformatics and biomedicine. *Brief Bioinform*. 2025;26(6):bbaf601.

doi: 10.1093/bib/bbaf601

118. Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. *npj Digit Med*. 2023;6(1):195.

doi: 10.1038/s41746-023-00939-z