ACCSCIENCE PUBLISHING

International Journal of Bioprinting

RESEARCH ARTICLE

BioPrint-LKM: An evidence-grounded large knowledge model for bioprinting knowledge retrieval and parameter initialization

Xi Huang[1,3]*, Hanqi Su[2], Zhengjie Cui[1], Jia Min Lee[1,3], Xinchao Gao[1,3], Renzhi Hu[1,3], Jay Lee[2], and Wai Yee Yeong[1,3]*

[1]School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

[2]Center for Industrial Artificial Intelligence, Department of Mechanical Engineering, University of Maryland, College Park, MD, United States of America

[3]Singapore Centre for 3D Printing (SC3DP), Nanyang Technological University, Singapore

Abstract

Bioprinting workflows often require repeated trial-and-error to achieve acceptable print quality, while relevant process knowledge and parameter ranges are dispersed across a rapidly growing literature. General-purpose language models can assist with scientific questions, but their output may be difficult to verify and can include unsupported claims. In this work, we present BioPrint-LKM, a bioprinting large knowledge model (LKM) implemented using a retrieval-augmented generation pipeline with citation grounding to provide traceable, evidence-based responses for bioprinting tasks. A curated knowledge base was constructed from 621 bioprinting papers, which were converted to text, segmented into passages, embedded into a vector index, and retrieved using exact nearest-neighbor similarity search. Retrieved passages were assembled into an augmented context and used to constrain generation under a domain prompt guideline that enforces source and page-level citation. The LKM was evaluated using 30 bioprinting-related questions and three large language model (LLM) backbones (GPT-4o, Claude-Sonnet-4.6, and Gemini-2.5-Flash). Across the tested 30-question benchmark, retrieval and citation grounding improved question-answer accuracy compared with LLM-only baselines by an average of 24.7%, particularly for queries requiring paper specific details such as component concentrations, mixing ratios, and instrument settings. A retrieval hyperparameter sweep further showed that top-*k* and passage length affect both evidence coverage and noise, with an intermediate passage length providing the best overall performance. Beyond knowledge retrieval, the LKM was applied to bioprinting process setup by generating initial parameter sets that were used to warm-start multi-objective Bayesian optimization toward target filament width and height. Compared with manual initialization, BioPrint-LKM-assisted initialization reduced the number of calibration trials by an average of 36.0% and supported downstream printing demonstrations, including conductive cell-laden hydrogel lines and an anatomical ear model. These results suggest that BioPrint-LKM, a citation-grounded LKM, can serve as a practical lab-assistive tool to accelerate bioprinting setup and improve reproducibility.

Keywords: 3D bioprinting; Biofabrication; Large knowledge model; Retrieval-augmented generation; Citation grounding; Machine learning; Hydrogel; Process calibration

***Corresponding authors:**
Xi Huang
(huang.xi@ntu.edu.sg)
Wai Yee Yeong
(wyyeong@ntu.edu.sg)

Citation: Huang X, Su H, Cui Z, *et al*. BioPrint-LKM: An evidence-grounded large knowledge model for bioprinting knowledge retrieval and parameter initialization. *Int J Bioprint*. 2026:12(2):026110094. doi: 10.36922/IJB026110094

Received: March 13, 2026

Revised: March 23, 2026

Accepted: March 24, 2026

Published online: April 10, 2026

1. Introduction

Over the years, bioprinting has enabled the controlled spatial deposition of cells, biomaterials, and bioactive factors to fabricate living and biomimetic constructs.[1–6] In contrast to conventional scaffold fabrication or bulk cell seeding, bioprinting provides a digitally defined, layer-by-layer (or voxel-by-voxel) approach in which the geometry, local composition, and placement of multiple cell types can be programmed to better approximate native tissue organization.[7] Depending on the printing method, including extrusion,[8,9] droplet,[10–12] and light-based approaches,[13–15] bioprinting can produce features ranging from microscale patterns[16] to centimeter-scale constructs.[17] It also allows the use of structural matrices, sacrificial support materials, and spatial gradients of biochemical cues. As a result, bioprinting has become an important platform for tissue engineering, regenerative medicine, and *in vitro* models that aim to better represent human physiology, including more reproducible disease models and drug-screening systems that capture cell–matrix interactions and multicellular interfaces.[18,19] Despite rapid progress in printers, bioinks, and crosslinking strategies, routine and reproducible bioprinting remains difficult. Many laboratories still face long setup cycles, inconsistent print outcomes, and limited transferability of protocols across users, materials, and hardware.[20]

A central reason is that bioprinting performance is highly sensitive to both material state and process conditions. Small variations in bioink formulation and state, temperature, mixing history, crosslinking conditions, nozzle geometry, dispensing pressure, and motion parameters can produce large changes in deposited feature dimensions and construct fidelity.[21,22] These sensitivities are compounded by batch-to-batch variability, differences between printer platforms, and the competing requirements of printability and biological constraints. As a result, bioprinting workflows frequently rely on iterative trial-and-error to reach acceptable printing outcomes, which increases time, material use, and experimental burden.[23]

In parallel, the information required to make informed decisions is widely available but difficult to reuse efficiently. Bioprinting papers and protocols report bioink recipes, parameter ranges, printability observations, and failure modes, yet these insights are scattered across a rapidly expanding literature and are presented under heterogeneous reporting standards.[21,23,24] Even within a single laboratory, knowledge is often distributed across notebooks, spreadsheets, and informal experience. Consequently, selecting initial printing settings for a new bioink or a new target feature often involves manual literature searching and subjective judgment, and the resulting choices can be difficult to justify or reproduce.

These characteristics make bioprinting well suited for artificial intelligence (AI)-assisted decision support.[25,26] The process is high-dimensional and non-linear, where many coupled variables jointly determine outcomes such as extrusion stability, filament geometry, shape fidelity, and post-print viability. Recent reviews in AI-augmented additive manufacturing and closed-loop 3D printing have highlighted the role of AI across process modeling, monitoring, control, and optimization, supporting its relevance to bioprinting workflows.[27,28] AI methods can help with learning empirical relationships from indirect observations;[29,30] prioritize parameters that are most likely to matter for a given bioink, printer, and target feature;[31] and propose parameters to reduce the number of trial prints needed.[32,33] AI can also serve as an interface layer that converts fragmented text-based evidence (recipes, observations, and failure modes) into structured, reusable guidance.[34] Recent advances in large language models (LLMs) have created opportunities for natural-language scientific assistance, but direct use of general-purpose models is not well aligned with the needs of bioprinting. Bioprinting decisions are evidence-sensitive, and ungrounded generation can introduce unsupported claims, unclear provenance, and recommendations that are difficult to audit.[35] Conventional machine-learning approaches, while valuable, typically require curated datasets for a narrow task and do not directly provide traceable links to prior evidence. These limitations motivate a knowledge-centered approach that can both retrieve relevant technical evidence and present it in an interpretable, verifiable form.[36]

The challenges in bioprinting share similarities with other complex manufacturing domains in Industry 4.0, where outcomes depend on many coupled variables and knowledge is distributed across many sources.[37] Recent industrial AI frameworks emphasize that improving reliability and productivity is not only a matter of building better prediction models, but also of building systems that can organize domain knowledge, connect unstructured information to decision workflows, and support human operators with traceable reasoning.[38,39] These ideas suggest that a large knowledge model (LKM) can be helpful for bioprinting because it can turn scattered text-based knowledge into something that is easier to search, reuse, and apply during process setup and troubleshooting, instead of relying on manual searching and personal experience.

Here, we propose BioPrint-LKM, a bioprinting LKM designed to bridge fragmented bioprinting knowledge

and practical decision-making. In the present work, BioPrint-LKM is implemented using a citation-grounded retrieval-augmented generation pipeline over a curated bioprinting literature database. In this context, the term "large knowledge model" refers to a knowledge-centered system that combines domain-specific literature, semantic retrieval, and grounded response generation for bioprinting decision support, rather than a new standalone base language model. This design aims to provide users with answers that are not only useful but also traceable to source documents at the paper and page level, enabling verification and reducing the risk of hallucinated content. The main contribution of this study is the development and validation of a bioprinting-specific LKM workflow that integrates literature retrieval, citation grounding, and evidence-guided parameter initialization. We evaluated the proposed LKM for two practical functions. First, we assessed its ability to retrieve and summarize paper-specific bioprinting information in response to user queries, including bioink formulations, processing conditions, and reported parameter ranges, while providing source and page-level citations for traceability. Second, we assessed its ability to provide evidence-guided initial parameter suggestions that can serve as starting points for downstream tuning, for example when targeting specific deposited feature dimensions such as strand width and height. These functions are intended to reduce the overhead of manual knowledge search and accelerate early-stage bioprinting setup while maintaining transparency through source attribution.

2. Materials and methods

2.1. BioPrint-LKM Framework

BioPrint-LKM is a bioprinting LKM implemented using a retrieval-augmented generation design tailored for bioprinting, as shown in Figure 1. The framework has two parts, a retrieval layer and a generation layer. The retrieval layer finds relevant evidence from a curated bioprinting literature database. The generation layer produces a final answer using an LLM guided by a domain prompt guideline and returns an output with citations. The retrieval augmented workflow is implemented using vector indexing and similarity search backend.

2.1.1. Bioprinting knowledge base construction

A bioprinting knowledge base was built from a local folder of PDF papers. Each PDF is loaded using PyPDFLoader, which extracts the text page by page. During loading, the code adds source information to every page. It stores the PDF file name and the page number in the metadata. The page number is converted from 0-based to 1-based so it matches normal page numbering when used for citation.

After loading, the page text is split into shorter passages, so that retrieval can be done at a finer level than a full page. Each chunk has varying characters with 150 characters overlap. The overlap helps as the important sentences near the end of one chunk are likely to also appear in the next chunk, which reduces the chance of missing relevant information due to a chunk boundary. Each chunk keeps the same metadata from its original page, such that the system can still report the correct file name and page number even though it is searching and retrieving at the chunk level.

Each passage is converted into an embedding, so that the system can do semantic search. For this work, OpenAI Embeddings are used. The encoder converts raw text into embedding, a vector such that passages with similar meaning or technical content are placed close together in the embedding space, even when they do not share identical keywords. This property is important for queries where the same concept may be expressed using different terminology across papers, for example extrusion pressure versus pneumatic pressure, print speed versus feed rate, or filament width versus strand diameter. The embedding is built through FAISS (Facebook AI Similarity Search) library, and a flat index is used for the vector index as the bioprinting database is small comparatively.

To keep the output traceable, each retrieved chunk is formatted with its source tag, including the PDF file name and page number. The language model receives the retrieved chunks together with these tags, and it is instructed to cite sources using the provided tags. This allows the final answer to be checked against the original paper and page.

2.1.2. Query encoding and similarity-based retrieval

When a user submits a question, the system first encodes the question into a dense vector representation, referred to as a query embedding. The same embedding model used for the literature passages is applied to the user query such that the query and passage vectors lie in the same representation space and can be compared directly.

The query embedding is then used to retrieve relevant evidence passages through similarity search over the indexed passage embeddings. Retrieval is performed by exact k-nearest neighbor search in the embedding space, returning the most similar passages to the query. The number of retrieved passages is controlled by a top-k setting, which is treated as a tunable parameter and will be evaluated in later experiments. Increasing k can improve coverage for broad queries, while smaller k reduces the amount of irrelevant context.

The retrieved results are returned as text passages with

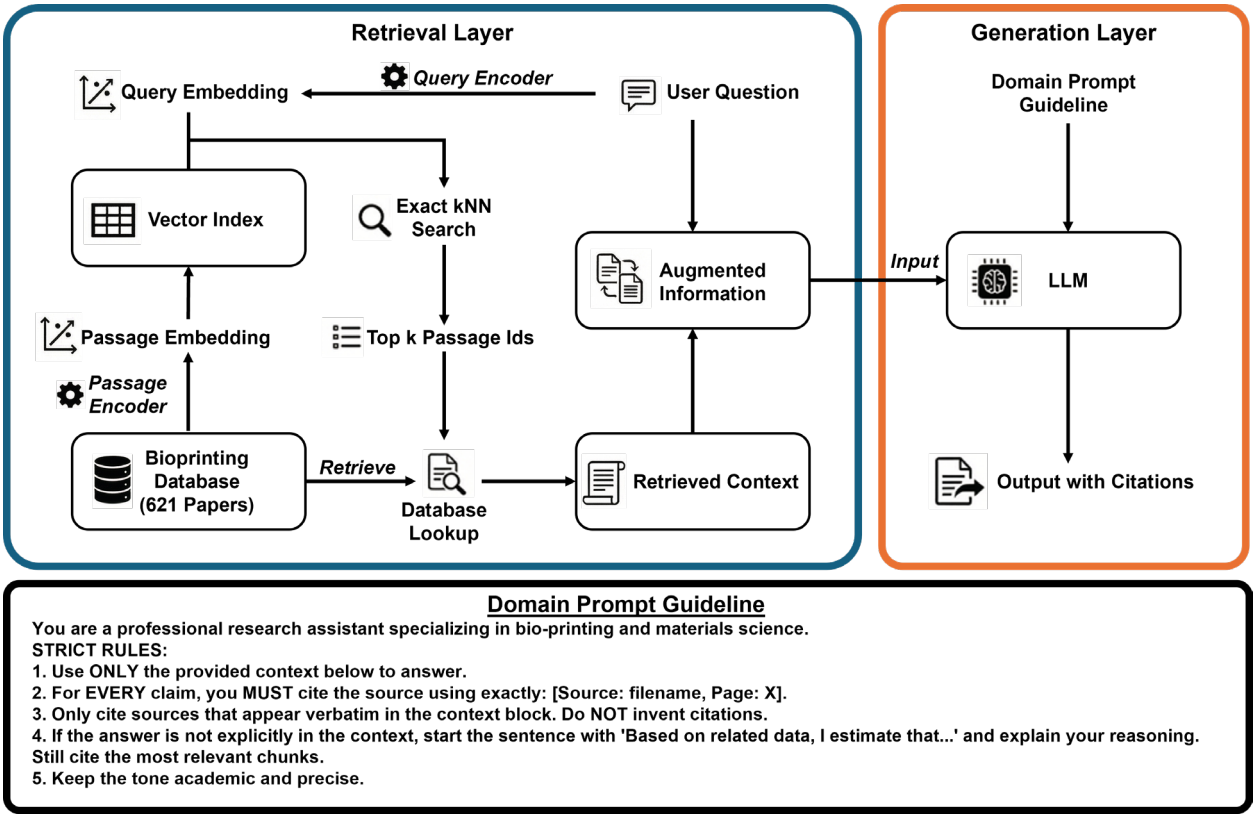

Figure 1. The workflow of the BioPrint-LKM framework for bioprinting
Abbreviations: kNN: *k*-nearest neighbors; LKM: Large knowledge model; LLM: Large language model.

their associated metadata, including the source filename and page number. This information is retained to support traceability and is later used to produce citations in the generated response.

2.1.3. Context augmentation and citation grounded generation

The retrieved passages are assembled into a retrieved context block and combined with the user question to form the augmented information passed to the generation layer. A domain prompt guideline is applied to encourage bioprinting-relevant responses and to promote grounding in the retrieved evidence. The LLM generates the final response conditioned on the augmented information. Citations are produced by mapping statements in response to the retrieved passages and their associated source papers. This design supports transparent evidence-backed answers while allowing key retrieval settings, such as passage length and number of retrieved passages, to be optimized based on downstream performance.

2.1.4. Model performance evaluation

Model performance was evaluated using 30 bioprinting-related questions. The 30-question set was designed as a task-specific proof-of-concept benchmark and does not represent a comprehensive evaluation of all bioprinting information needs. Each question was answered under two conditions: BioPrint-LKM and LLM-only (no retrieval). For the BioPrint-LKM setting, retrieval was configured with top-k = 6 and chunk size = 1000. The same evaluation protocol was applied to three LLM backbones, namely GPT-4o, Claude-Sonnet-4.6, and Gemini-2.5-Flash. Correctness was scored as a binary label ('1' correct, '0' incorrect) by manually checking each response against the ground-truth information in the relevant paper(s). Responses were marked incorrect if they missed required key details or included unsupported factual claims. The overall accuracy was reported as the percentage of correct answers across the 30 questions.

To capture answer stability across models, per-question correctness was also aggregated across the three backbones by counting how many of the three models produced a correct answer for a given question (range of 0–3). This provides a simple measure of cross-model consistency and highlights questions that are systematically difficult.

2.1.5. Hyperparameter tuning

To study the effect of retrieval settings on BioPrint-LKM performance, we conducted a grid sweep over top-k ∈{1,3,6} and chunk size ∈{600,1200,2400}. For each (k, chunk size) combination, the BioPrint-LKM answered the same 30-question set, and accuracy was computed using the same binary scoring. The results are summarized as overall accuracy across settings and as per-question correctness to identify question types that are sensitive to retrieval hyperparameters.

2.2. Materials

2.2.1. Synthesis of Pluronic F127 solution

A 25% (w/v) Pluronic F127 solution was prepared using a cold dissolution method to minimize clumping and to ensure complete solubilization. Pluronic F127 powder (Sigma-Aldrich, USA) was weighed to achieve a final concentration of 25 g per 100 mL of distilled water. The polymer was added gradually into a beaker of water in ice bath under continuous stirring to avoid formation of lumps. The mixture was stirred in ice bath until a clear and homogeneous solution was obtained, typically within 3 h. The final solution was stored at 4 °C.

2.2.2. Synthesis of sodium alginate solution

A 5% (w/v) sodium alginate solution was prepared by dissolving sodium alginate powder in distilled water under prolonged heated stirring. Sodium alginate (Sigma-Aldrich, China) was weighed to obtain a final concentration of 5 g per 100 mL of distilled water. The powder was added gradually into the water at 50 °C while stirring to minimize clumping and ensure uniform dispersion. The mixture was then stirred for 12 h to promote complete dissolution. After stirring, the solution was left standing in fridge at 4 °C for an additional 12 h to allow air bubbles to dissipate and any remaining particulates to settle. The final solution was stored at 4 °C.

2.3. Vision setup

A machine vision-based measurement system was used to quantify the width and height of printed hydrogel filaments, as shown in Figure 2a. The printed hydrogel line was placed on a transparent stage and illuminated using a green excitation light to improve contrast. Illumination was provided by a Nightsea SFA Light Head (green, 510–540 nm), and images were captured using a Logitech C310 HD Webcam mounted at a fixed position. The green excitation light was used only during short image acquisition periods for geometry measurement rather than prolonged exposure during printing or culture. A custom holder was used to maintain a repeatable camera and illumination geometry across measurements, as seen in Figure 2c.

A ResNet18 model[40] is trained to rapidly measure the width and height of printed hydrogel line. While YOLO-based approaches have been widely used in additive manufacturing for defect detection and real-time monitoring, the present study requires direct regression of continuous width and height values from images, for which a ResNet-based regressor was selected instead.[41,42] Two imaging views were used to capture filament geometry. For width measurement, a top-view image of the printed line was acquired and the lateral filament dimension was estimated from the image (Figure 2b, left). For height measurement, a side-view image was acquired to capture the filament profile and estimate the vertical dimension (Figure 2b, right). The captured images were used to train a convolutional neural network regressor based on ResNet18 to predict filament width and height. The hyperparameter of the model was tuned on 159 samples with 5-fold cross validation. Model performance was evaluated on a testing set of 39 samples by comparing predicted values with ground-truth measurements in pixel units, depicted in Figure 2d, yielding a mean absolute percentage error of 7.83% and an R^2 of 0.9756 for width, and a mean absolute percentage error of 6.39% and an R^2 of 0.9866 for height.

2.4. Printing setup

All printing experiments were performed using an extrusion-based bioprinter (Bio-Architect SR, China). Unless otherwise stated, materials were dispensed through a conical 27-gauge nozzle tip with an internal diameter of 0.21 mm. For Pluronic F127 prints, the printing temperature was maintained at 25 °C. For sodium alginate prints, the printing temperature was maintained at 20 °C.

2.5. Multiple objective Bayesian optimization

Multi-objective Bayesian optimization (MOBO)[43] was used to tune printing parameters toward a target filament width and height with a limited number of trials. Each trial corresponded to a printed line generated using a parameter setting. First, the target width and height are inputted into the MOBO model. An initial set of printing parameters, tested after being generated by BioPrint-LKM or through manual calibration, is provided to the model. In every trial, the model will suggest a set of printing parameters consisting of pressure, speed, printing height, and temperature. After printing, the achieved filament width and height were quantified using the vision system described in the previous section. These measurements were used as the objective values for the optimization loop.

The optimization was formulated as a multi-objective problem because width and height must be matched

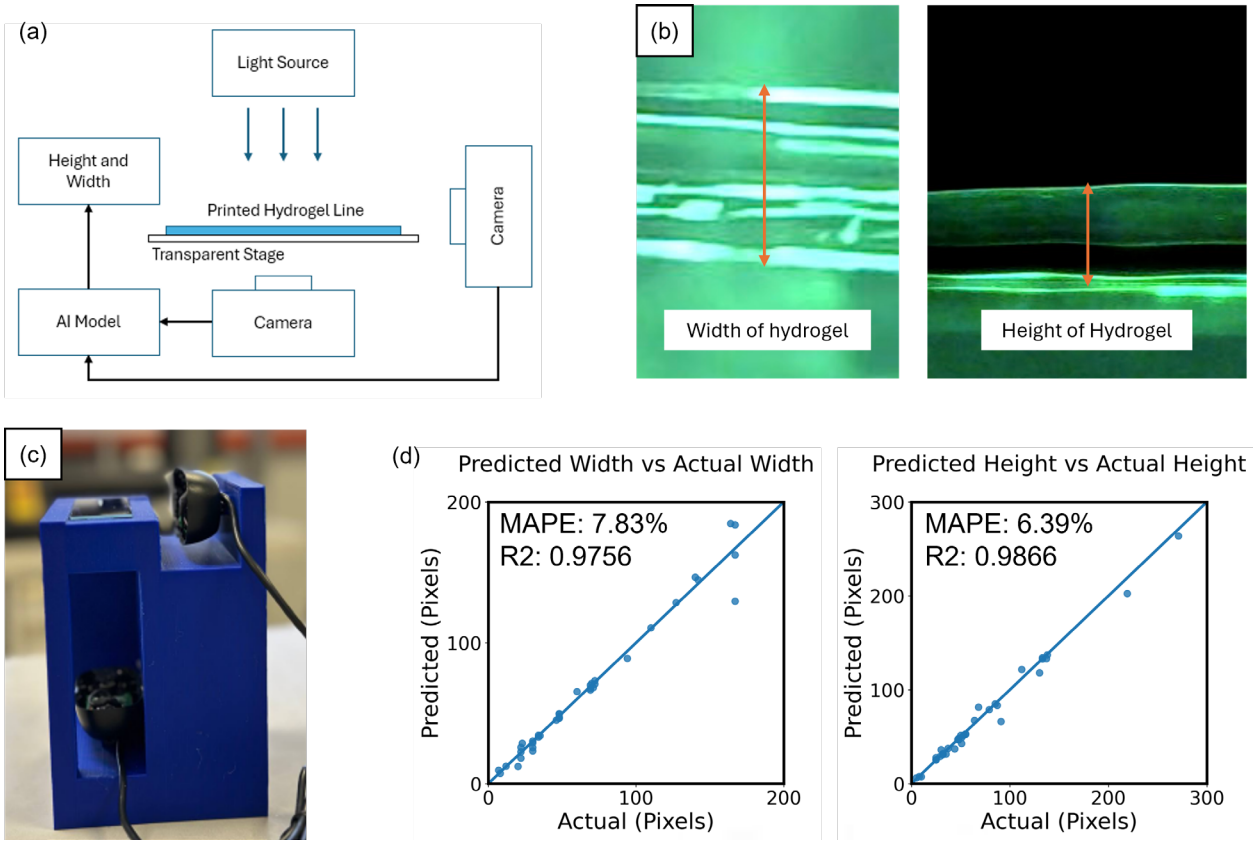

Figure 2. Vision-based setup for filament width and height measurement. (a) Schematic of the imaging configuration used to measure printed hydrogel line width and height. (b) Representative images illustrating the width view and height view of a printed hydrogel filament, with measurement directions indicated. (c) Photograph of the custom holder that fixes the camera and illumination geometry to ensure repeatable imaging conditions. (d) Comparison between model-predicted and measured width and height of the ground-truth filament using a ResNet18-based regressor.

simultaneously and may not improve together. For each trial, the achieved width and achieved height were compared with their respective targets and converted into objective errors. Two separate objectives were used, one for width, $f_w(x_i)$ and one for height, $f_h(x_i)$, using negative squared errors to be maximized by the optimizer, where and is the achieved width and height in the trial *i*, while and is the targeted width and height respectively.

$$f_w(x_i) = -(w_i - w)^2$$

$$f_h(x_i) = -(h_i - h)^2$$

A trial is considered to have met the target when both achieved width and height are within 10% of the targeted width and height.

To model the relationship between printing parameters x_i and the two objectives, two independent Gaussian process surrogate models were trained with the entire dataset from initial points and trial run, one for f_w and one for f_h. Inputs were normalized and each objective output was standardized before fitting. The two models were combined into a multi-objective model using a hypervolume-based acquisition function. At each iteration, the current non-dominated set was identified to form a Pareto front, a reference point was calculated as 0.05 below the observed objective values, and the parameters for next trial was chosen by maximizing the log expected hypervolume improvement. Candidate parameters were selected from a discrete set of feasible parameter combinations, and previously tested settings were excluded from selection.

For the BioPrint-LKM versus manual comparison, the Bayesian optimization loop and all settings were kept identical, and only the initial parameter differed. In the LKM condition, three initial parameter settings were generated by the LKM and executed first, and their measured width and height were used to initialize the surrogate models. In the manual condition, one initial parameter setting was chosen by the operator and executed first, and Bayesian optimization proceeded from that single starting point.

3. Results and discussion

3.1. Knowledge retrieval

3.1.1. Performance of BioPrint-LKM vs LLM only

Across the tested 30-question benchmark, the BioPrint-LKM produced a clear accuracy gain over the LLM-only baseline as shown in Figure 3a, improving GPT-4o from 73% to 93%, Claude-Sonnet-4.6 from 77% to 97%, and Gemini-2.5-Flash from 80% to 97%. The improvement was consistent across the three tested backbones (one-tailed paired t-test, $p < 0.01$, $n = 3$), with an average increase in accuracy of 24.7%. As the benchmark contains only 30 manually designed questions, the reported accuracy gains should be interpreted as preliminary and may be influenced by question selection. The current evaluation compares BioPrint-LKM only with an LLM-only baseline and does not include alternative retrieval methods, such as lexical, hybrid, or scholarly search workflows. In addition, because correctness was manually scored by the authors, the current evaluation may also be subject to scoring bias and should be interpreted as a proof-of-concept assessment. This improvement is mainly driven by question types that require precise, paper-specific details rather than general domain knowledge. In such cases, LLM-only answers frequently become either generic approximations or plausible sounding guesses, while the BioPrint-LKM can ground the response in retrieved passages and provide traceable citations. The entire answer set is provided in Supplementary Texts S1 to S4. As depicted in Figure 3b, there are six questions where the LLM-only model is having trouble to solve the question as compared to BioPrint-LKM. The questions can be grouped into three categories.

The first category where the BioPrint-LKM consistently did better is numeric and configuration retrieval, where correctness depends on exact values rather than qualitative explanations. For Question 1 ("Which three types of cells were used to generate sphere skin organoids, and what was their mixing ratio?"), the LLM-only responses often listed reasonable cell types but produced conflicting or unverifiable mixing ratios, making it difficult to judge whether the ratio was correct or hallucinated. In contrast, the BioPrint-LKM retrieved the relevant source and returned a specific 2:1:1 mixing ratio with citation support. For Question 3 ("What are the specific components and their concentrations in the optimized bioink formulation used for DLP-based printing of hepatic tissue constructs?"), the LLM-only response drifted into generic "typical" components and ranges, whereas the BioPrint-LKM extracted the exact formulation details from the target paper, including the solvent and additives (*e.g.*, LAP and UV absorber concentrations) and the stated cell loading, again with citations. These examples show that retrieval is particularly valuable when the question asks for "components and concentrations," "ratios," or other hard constraints that are rarely safe to infer.

The second category is paper retrieval and paper-specific comprehension, where the question explicitly references a particular publication. For Question 7 ("In the paper '3D bioprinting of engineered exosomes secreted from M2-polarized macrophages through immunomodulatory biomaterial promotes *in vivo* wound healing and angiogenesis', what were the two types of bioinks formulated, and what were their specific applications?"), the LLM-only baseline struggled to reliably locate the correct paper context and tended to respond with plausible but incomplete or mismatched bioink descriptions. The BioPrint-LKM, by retrieving the named paper, was able to identify the two formulated bioinks and their roles (one supporting macrophage polarization/exosome secretion and the other for exosome delivery and wound-healing studies), and it provided citations tied to the retrieved document. This is a key advantage of the BioPrint-LKM in bioprinting use cases because many practical questions are framed around specific articles, protocols, or reported parameter sets. A related limitation is shown by Question 10 ("In the paper 'Bioprinting of high-performance hydrogel with *in situ* birth of stem cell spheroids', what light wavelength and intensity parameters were utilized for the DLP bioprinting process?"), where both LLM-only and BioPrint-LKM failed despite the paper is in the dataset, indicating a need for a proper hyperparameter fine-tuning to retrieve every relevant paper.

The third category is incomplete or under-specified questions, where the paper is not given and the query is under-determined. For Question 19 ("What specific growth factors are delivered by the PODS [Polyhedrin Delivery System] in this study to accelerate vascular graft development?"), LLM-only outputs often defaulted to generic growth factors commonly associated with vascularization, reflecting uncertainty about the exact study context. The BioPrint-LKM more consistently recovered the relevant paper from the literature database and returned the specific growth factors used, as supported by citations. A similar pattern was observed for Question 20 ("Which bioprinting technique did the researchers employ to fabricate the vascular grafts?"). While both provided the correct answer, LLM-only responses varied depending on the backbone, while the BioPrint-LKM more often located the intended study and identified the technique with supporting evidence. These results indicate that the BioPrint-LKM is more robust to real-world query habits, where users frequently omit full bibliographic details and expect the system to infer the most relevant source from the literature database.

3.1.2. Sensitivity analysis and error modes under top-k and chunk-size variation

Figure 3c shows that the BioPrint-LKM generally performs better at higher top-*k*. This behavior is expected because retrieving more passages increases the probability that at least one retrieved chunk contains the exact sentence or table entry needed to answer the question. Many bioprinting questions depend on a small number of paper-specific details (*e.g.*, a wavelength value, a concentration, or a named technique). With a small *k*, the correct paper may still be retrieved, but the specific passage that contains the answer can be missed. Increasing *k* adds redundancy and improves robustness, especially for questions where the relevant evidence is short, buried in a methods paragraph, or expressed with uncommon wording.

Chunk size shows a non-monotonic effect, where a larger chunk does not always increase performance. In this evaluation, a chunk size of 1200 achieves the best overall accuracy. This can be explained by a trade-off between retrieval precision and contextual completeness. Smaller chunks tend to be more focused on a single point, producing embeddings that are more discriminative for narrow questions. Larger chunks merge multiple subtopics into one embedding, which can dilute the relevance signal and reduce the chance that a chunk ranks highly for a targeted query. In addition, larger chunks can introduce more surrounding text that is not directly related to the asked detail, making generation less stable even when the correct source is retrieved. The intermediate chunk size provides enough local context to interpret technical statements while keeping the embedding concentrated enough for reliable similarity matching.

To better interpret the trends in Figure 3c,d, the representative questions that either failed across settings or showed strong sensitivity to top-*k* and chunk size are examined. These examples help explain why increasing top-*k* generally improves accuracy, while chunk size exhibits a non-monotonic effect with an optimum at an intermediate value in this dataset.

In most questions, increasing both the top-*k* and chunk size helped to retrieve the relevant text. This is notably seen in Question 3, where the parameter settings of digital light processing (DLP) is only included when both hyperparameter is increased. Performance is also affected when the query is under-specified and does not identify a unique study. These questions do not provide enough bibliographic context to uniquely determine a single source paper; therefore, retrieval must infer the intended study based on semantic similarity. This is seen in Question 12 ("Which 3D bioprinting technique did the authors employ to fabricate vascularized scaffolds?"). Since the referenced study is not named, the system may retrieve thematically related passages about vascularization and scaffold fabrication that do not explicitly state a single technique. As a result, performance depends strongly on top-*k* and chunk size. Smaller top-*k* values increase the risk that retrieved passages are relevant but incomplete, while higher top-*k* improves the chance that at least one retrieved chunk explicitly names a fabrication method, enabling a more defensible answer. A related but more challenging case is Question 20 ("Which bioprinting technique did the researchers employ to fabricate the vascular grafts?"). Here, the phrase "the researchers" implicitly assumes a specific paper, but no identifier is provided. In several settings, the system retrieves relevant sources and can cite them, yet the question itself remains under-determined because multiple studies could plausibly fit the description. This effect becomes more pronounced under lower top-*k* or smaller chunks, where the retrieved context may be too limited to disambiguate the intended study even if the retrieved source is broadly relevant.

Even when answers remain correct across most settings, chunk size can noticeably change how directly the retrieved evidence supports the response. For certain questions, multiple settings yield correct answers, but the retrieved evidence differs in how directly it supports the response. This is illustrated by Question 14 ("What is the primary role of the PDMS substrate in the DLP-based bioprinting process?"). With smaller chunks, retrieval is more likely to give a focused passage that directly describes the polydimethylsiloxane (PDMS) substrate's functional role, resulting in a clearer answer. With larger chunks, retrieval can shift toward broader background descriptions of PDMS or general setup details, which may still allow a correct answer but provide weaker direct support in the retrieved text. A similar effect is observed in Question 17 ("What is the purpose of bioink with GelMA + PEDOT?"). Smaller chunks tend to retrieve the sentence that matches the exact wording of the question, whereas larger chunks sometimes retrieve longer discussion paragraphs where the answer is only implied. This distinction matters for citation-grounded systems, where the quality of evidence alignment is important even when all the outcome is marked as correct.

Finally, some questions require retrieving a passage that explicitly enumerates multiple items, making them sensitive to whether the exact sentence is retrieved. In Question 22 ("In additive-free hyaluronic acid-based bioink for 3D bioprinting of bone marrow microenvironments, what are the two functional groups introduced onto the hyaluronic acid (HA) backbone during the one-pot synthesis?"). Many settings retrieve context that is generally related to HA-based bioinks but fail to recover the specific statement

that clearly shows two distinct approaches. In contrast, higher top-*k* combined with an intermediate chunk size increases the likelihood of retrieving the key passage containing the explicit two-item structure, enabling a complete and well-supported answer. This behavior is consistent with the overall trend that higher top-*k* improves robustness, particularly for questions where the answer is concentrated in a short passage that may not appear in the top ranks under low *k*.

A clear failure case is Question 10 ("In the paper 'Bioprinting of high-performance hydrogel with *in situ* birth of stem cell spheroids', what light wavelength and intensity parameters were utilized for the DLP bioprinting process?"). This question remains incorrect across the sweep, suggesting a retrieval coverage limitation rather than a generation limitation. This likely reflects a limitation in retrieving highly specific numeric information, especially when such values are embedded in narrow methodological text, tables, or figure-associated content. Notably, across all tested top-*k* and chunk-size configurations, the system did not retrieve passages from the intended paper and therefore could not find the specific wavelength and intensity values required to answer the question.

As a summary, these cases explain why higher top-*k* improves robustness by reducing retrieval misses and increasing the chance of capturing short but critical passages. Chunk size shows an optimum because overly small chunks can fragment necessary context, while overly large chunks dilute semantic focus and reduce retrieval precision for targeted, detail-seeking questions.

3.1.3. Performance of BioPrint-LKM at extreme ends of retrieval hyperparameters

To examine the upper limits of retrieval context, we evaluated extreme setting where each entire paper was treated as a single retrieval unit instead of splitting into chunks. This setting increases the retrieved context length substantially and can exceed the context window of some LLMs. In this experiment, using GPT-4o with full-paper retrieval reached the 128k token limit at top-$k = 3$, therefore GPT-4.1 was used for the full-paper experiments to enable a proper comparison across larger *k* values.

The outcome matrix is shown in Figure 3e. Using full papers as context improved performance on specific questions that were previously unsolved due to missing or fragmented evidence. Question 10, which is not correctly answered in all previous attempts, was solvable only when the full paper was retrieved, suggesting that the required values were not captured reliably in the smaller retrieved chunks or were in parts of the document not surfaced under chunk-based retrieval.

However, increasing context size also introduced new failure modes. For Question 3, which is answered correctly at lower context size, the system failed to retrieve the relevant paper at lower top-*k* from 1 to 6. In contrast, increasing top-*k* to 12 enabled the system to retrieve the correct source and recover the complete formulation details, resulting in a fully correct answer across all questions in this setting.

Despite the improved correctness at top-$k = 12$ with full-paper retrieval, this configuration is highly inefficient. Because each retrieved item is a full paper, the prompt length grows rapidly and can approach hundreds of thousands of tokens per question, which is impractical for many LLMs and substantially increases inference cost and latency. This observation motivates the use of chunking as a more scalable default, where the system can retrieve only the most relevant sections rather than entire documents.

Finally, since a chunk size of 1200 performed best in the earlier sweep, we also tested whether pushing top-*k* to an extreme value would further improve performance under chunk-based retrieval. With chunk size = 1200 and top-$k = 24$, performance did not improve and in some cases degraded. For example, Question 3 remained incomplete even with the larger retrieved set, with the model indicating that specific concentrations were not available in the provided context. This suggests that simply adding more retrieved chunks can introduce enough irrelevant or weakly related content to obscure the key passage, and that excessively long augmented context can reduce the model's ability to identify and use the most relevant evidence. In practice, beyond a certain point, higher *k* can increase noise faster than it increases useful coverage.

Overall, these extreme-setting results highlight a trade-off. Full-paper retrieval can rescue certain failure cases by maximizing evidence availability, but it scales poorly and can exceed context windows. Conversely, very large top-*k* under chunk-based retrieval can overload the model with redundant or distracting context. These findings support the earlier conclusion that an intermediate chunk size with a moderate top-*k* provides a better balance between retrieval specificity, evidence coverage, and computational efficiency.

3.2. BioPrint-LKM to assist bioprinting

3.2.1. BioPrint-LKM to suggest initial parameters

Figure 4a shows the trial history for three target filament geometries using MOBO. The three targets were selected to demonstrate that Bayesian optimization can control both filament width and height. Across all targets, BioPrint-LKM-assisted initialization reached the acceptable error band in fewer trials than manual initialization, with

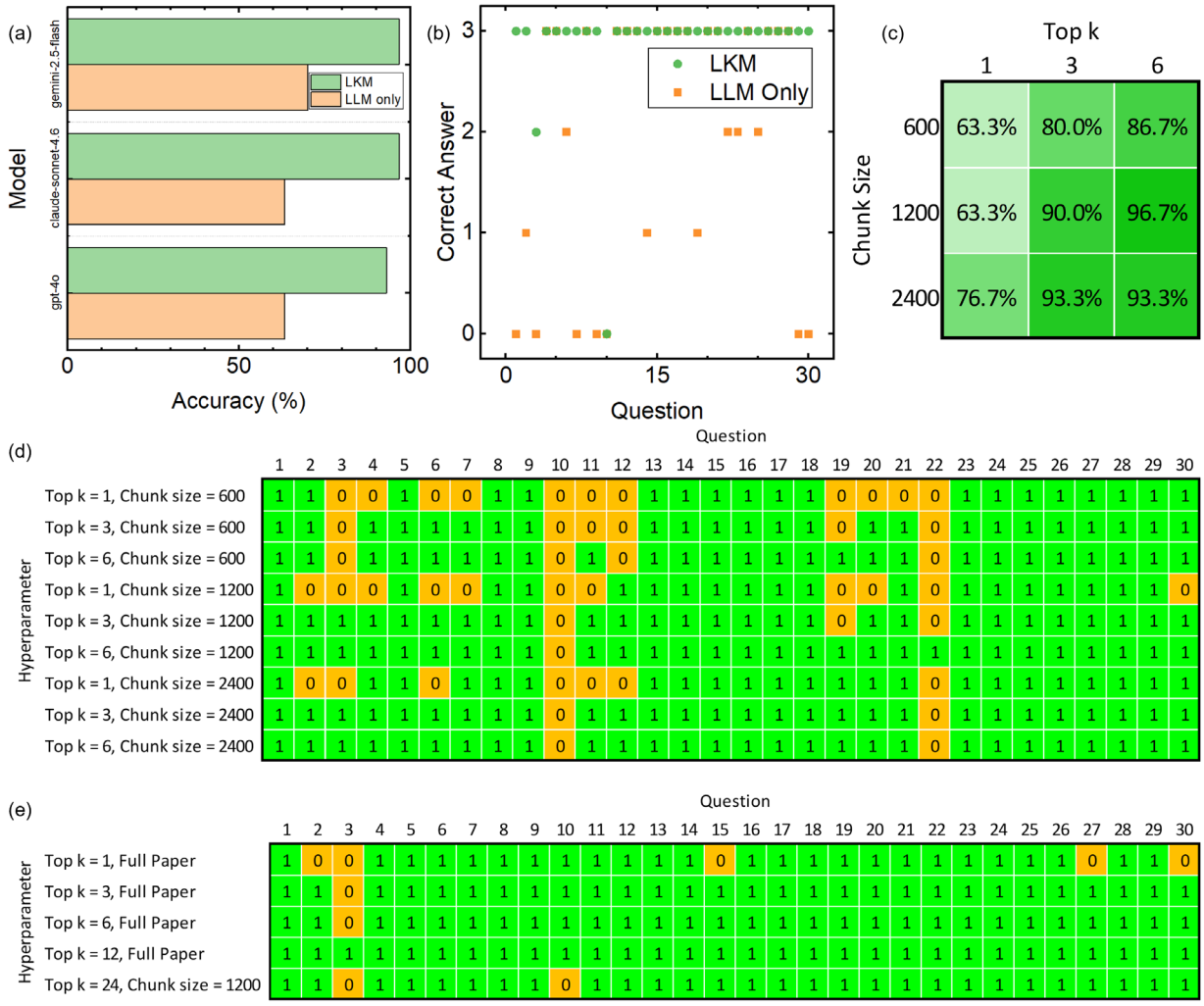

Figure 3. Performance of BioPrint-LKM for bioprinting knowledge retrieval. (a) Overall accuracy of the BioPrint-LKM and the LLM-only baseline across three LLM backbones. (b) Per-question correctness aggregated across the three backbones. For each question, the y-axis shows the number of models (out of 3) that produced a correct answer, reported separately for the BioPrint-LKM and LLM-only conditions (range 0–3). Higher values indicate more consistent correctness across models. (c) BioPrint-LKM's accuracy as a function of retrieval hyperparameters, showing the effect of top-*k* and chunk size on overall accuracy. (d) Per-question outcome matrix for different retrieval hyperparameter settings. Each cell indicates whether the response was correct (1) or incorrect (0) for a given question under each hyperparameter condition. (e) Per-question outcome matrix for extreme retrieval hyperparameter settings. Each cell indicates whether the response was correct (1) or incorrect (0) for a given question under each hyperparameter condition.
Abbreviations: LKM: Large knowledge model; LLM: Large language model.

an average 36.0% reduction in trial count (one-tailed paired *t*-test, , $p < 0.05$, $n = 3$). The single manual initial point does not capture the full practical effort of manual calibration, which is preceded by additional unrecorded trial and error attempts, including failed prints that do not yield measurable lines. By contrast, BioPrint-LKM provides multiple plausible starting points that improve early search-space coverage in Bayesian optimization and support faster convergence. With manual initialization, the first few trials often showed large deviations, including clear over- or under-deposition, which delayed convergence while MOBO searched for a feasible region. In contrast, the BioPrint-LKM suggested starting points placed the search closer to the target from the beginning, leading to fewer unproductive trials and a steadier reduction in error.

This improvement is useful for bioprinting calibration for two reasons. First, fewer trials reduce time and material use, which becomes especially important when inks are expensive or when printing involves cells. Second, reliable control of filament width and height improves

layer stacking, since small errors in line dimensions can accumulate and lead to larger shape deviations in multi-layer prints.

3.2.2. Printing of cell-laden conductive hydrogel

After calibration, the same workflow was applied to a cell-laden Pluronic F127–PEDOT:PSS conductive hydrogel formulated with 25% (*w*/*v*) Pluronic F127, 0.2% PEDOT:PSS, and a cell density of 1×10^7 cells/mL, printed using a straight 22-gauge tip (internal diameter 0.41 mm). Using the BioPrint-LKM together with Bayesian optimization, printing parameters were adjusted until the printed line width and height matched the target, with the calibration graph shown in Figure 4b. The calibrated conductive hydrogel line is shown in Figure 4c. This calibration step is useful for cell-laden printing because repeated trial-and-error increases bioink and cell consumption while prolonging the time cells spend outside controlled conditions. Reaching the target geometry more quickly and consistently improves repeatability and makes it easier to print constructs with predictable strand spacing and overall shape.

3.2.3. Printing of anatomical model

To demonstrate downstream use after calibration, Pluronic F127 was used to print a 3D ear model, with the calibration run shown in Figure 5a and two ears were printed using the calibrated parameters shown in Figure 5ab,c. Accurate filament width and height are important for anatomical printing because toolpaths are closely spaced, and small dimensional errors can lead to strand overlap, overcrowding, or fusion between neighboring lines. With calibrated geometry, line deposition was more consistent, reducing unintended overlap and improving stacking consistency. The ear geometry could be reproduced across two prints, indicating stable printing performance with

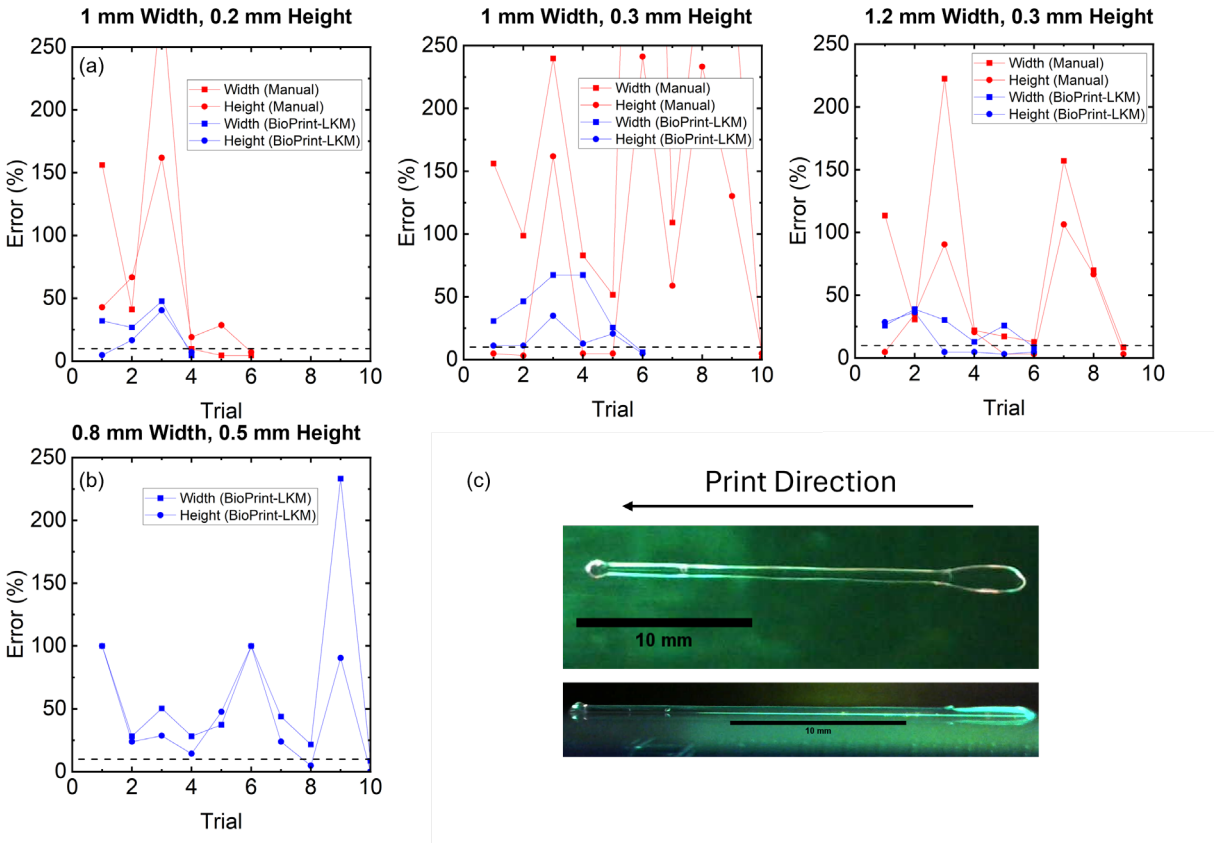

Figure 4. BioPrint-LKM-assisted calibration for target filament geometry. (a) Width and height error (%) versus trial number for three target geometries (1.0 mm width/0.2 mm height, 1.0 mm width/0.3 mm height, and 1.2 mm width/0.3 mm height), comparing manual and BioPrint-LKM initial points in the Bayesian optimization loop. The dashed line indicates the error threshold of 10%. (b) Width and height error (%) versus trial number of calibration run for calibration of a cell-laden conductive hydrogel line using BioPrint-LKM initial points. The dashed line indicates the error threshold of 10%. (c) Representative image of the printed cell-laden conductive hydrogel line.
Abbreviation: LKM: Large knowledge model.

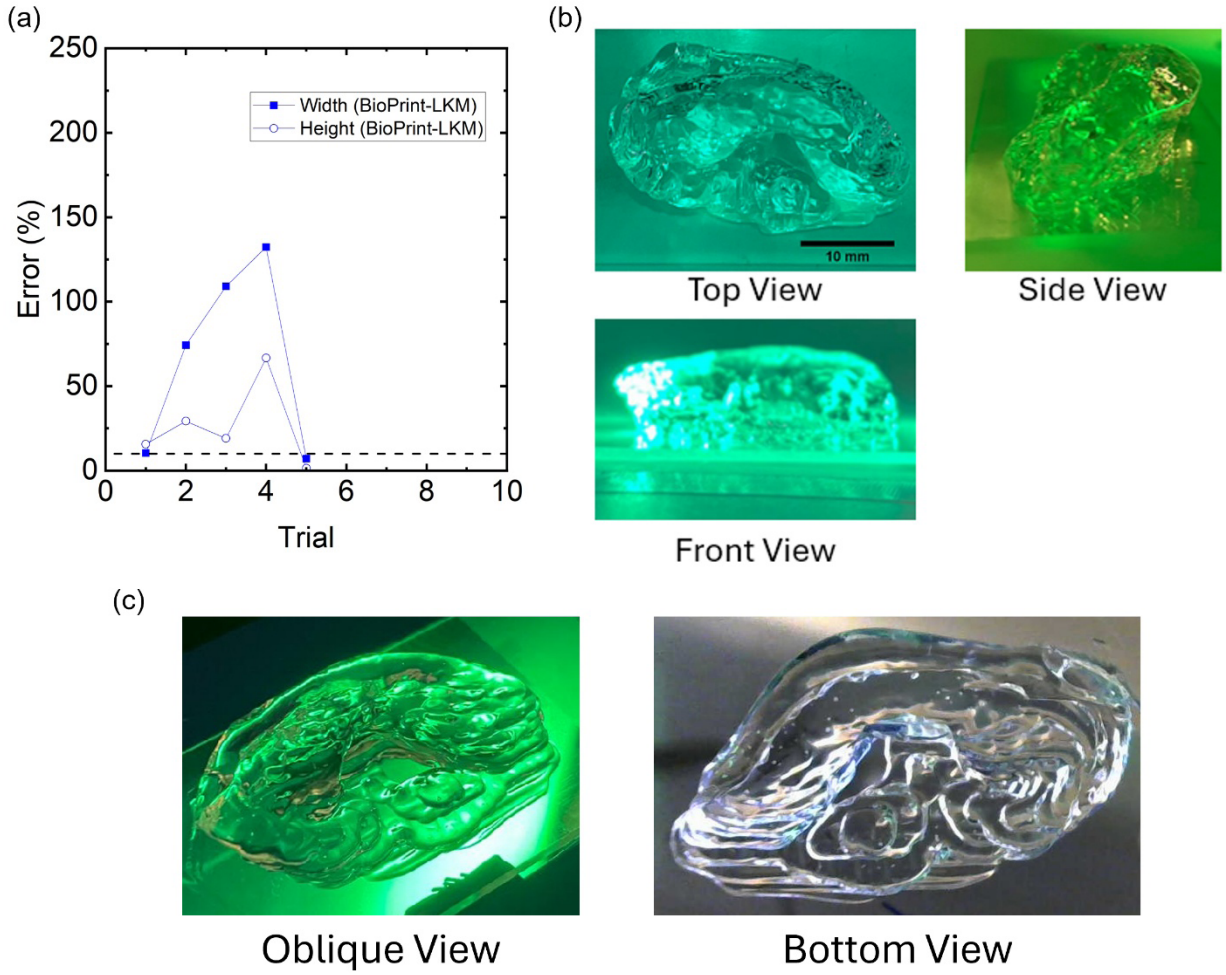

Figure 5. Calibration and anatomical printing using Pluronic F127. (a) Width and height error (%) versus trial number of calibration run using BioPrint-LKM initial points for printing a 3D ear model. The dashed line indicates the error threshold of 10%. (b) 3D-printed ear after calibration, shown in top, side, and front views. (c) Oblique view and bottom view of the other 3D-printed ear.
Abbreviation: LKM: Large knowledge model.

this calibration. This makes it easier to move from single-line tests to full 3D prints and improves the shape quality of the final model.

4. Conclusion

This study presented BioPrint-LKM, a bioprinting LKM that combines literature retrieval with citation-grounded generation to support bioprinting decision-making and reduce reliance on manual searching and trial-and-error. Using a database of 621 bioprinting-related papers and a 30-question benchmark, the BioPrint-LKM consistently improved question-answer accuracy across three LLM backbones compared with LLM-only baselines, especially for paper-specific details such as formulations, ratios, and process settings. The hyperparameter sweep further showed that retrieval settings affect both coverage and noise, with a chunk size of 1200 and top-k of 6 providing the best balance between accuracy and context length. Finally, the BioPrint-LKM was demonstrated as a practical tool for process setup by providing initial parameter suggestions that helped MOBO reach target filament width and height in fewer trials, supporting faster calibration and downstream printing tasks.

Future work will focus on improving both retrieval quality and practical usability. First, the database search workflow can be refined using a two-stage strategy, where a wider candidate set is retrieved initially (larger k) and then filtered or re-ranked to a smaller, higher-quality subset before being passed to the language model. This would reduce long-context noise while keeping the benefit

of higher recall. In addition, the current 30-question benchmark will be expanded and refined to improve topic coverage and reduce potential question-selection bias. Future work will also include independent evaluation or multiple raters to improve reproducibility and reduce potential scoring bias. Comparison with alternative retrieval strategies, including lexical, hybrid, and reranking-based approaches, will also be explored. Second, the current system only indexes text, but key bioprinting details are often contained in figures, tables, and schematic workflows. Extending the knowledge base to include image-derived information and multimodal retrieval would improve coverage for questions where critical parameters are reported visually. Third, the framework can be extended to incorporate user experimental data, such as printer settings, batch-specific material properties, and measured outcomes. Allowing the LKM to retrieve from both published literature and user-provided lab records would enable more personalized recommendations and support continuous improvement as new experiments are added. Finally, an agentic AI extension can be explored in which the system can plan and execute multi-step tasks, such as iteratively refining queries, validating evidence across sources, and updating the knowledge base as new results are produced, with human-in-the-loop oversight.[44]

Acknowledgments

None.

Funding

This research is supported by the National Research Foundation for NRF Investigatorship Award No.: NRF-NRFI07-2021-0007.

Conflict of interest

Wai Yee Yeong serves as an Associate Editor of this journal, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. The authors declare they have no competing interests.

Author contributions

Conceptualization: Xi Huang, Hanqi Su
Formal analysis: Xi Huang
Investigation: Xi Huang, Zhengjie Cui, Xinchao Gao, Renzhi Hu
Methodology: Xi Huang, Jia Min Lee
Writing–original draft: Xi Huang
Writing–review & editing: Xi Huang, Hanqi Su, Zhengjie Cui, Jia Min Lee, Xinchao Gao, Renzhi Hu, Jay Lee, Wai Yee Yeong

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data are available from the corresponding author upon reasonable request.

References

1. Mandrycky C, Wang Z, Kim K, Kim DH. 3D bioprinting for engineering complex tissues. *Biotechnol Adv*. 2016;34(4):422-434.

   doi: 10.1016/j.biotechadv.2015.12.011

2. Murphy SV, Atala A. 3D bioprinting of tissues and organs. *Nat Biotechnol*. 2014;32(8):773-785.

   doi: 10.1038/nbt.2958

3. Groll J, Boland T, Blunk T, *et al.* Biofabrication: reappraising the definition of an evolving field. *Biofabrication*. 2016;8(1):013001.

   doi: 10.1088/1758-5090/8/1/013001

4. Gungor-Ozkerim PS, Inci I, Zhang YS, Khademhosseini A, Dokmeci MR. Bioinks for 3D bioprinting: an overview. *Biomater Sci*. 2018;6(5):915-946.

   doi: 10.1039/c7bm00765e

5. Sun W, Starly B, Daly AC, *et al.* The bioprinting roadmap. *Biofabrication*. 2020;12(2):022002.

   doi: 10.1088/1758-5090/ab5158

6. Ng WL, Chua CK, Shen YF. Print Me An Organ! Why We Are Not There Yet. *Prog Polym Sci*. 2019;97:101145.

   doi: 10.1016/j.progpolymsci.2019.101145

7. Lee JM, Ng WL, Yeong WY. Resolution and shape in bioprinting: Strategizing towards complex tissue and organ printing. *Appl Phys Rev*. 2019;6(1):011307.

   doi: 10.1063/1.5053909

8. Ozbolat IT, Hospodiuk M. Current advances and future perspectives in extrusion-based bioprinting. *Biomaterials*. 2016;76:321-343.

   doi: 10.1016/j.biomaterials.2015.10.076

9. Meng Z, He J, Li J, Su Y, Li D. Melt-based, solvent-free additive manufacturing of biodegradable polymeric scaffolds with designer microstructures for tailored mechanical/biological properties and clinical applications. *Virtual Phys Prototyp*. 2020;15(4):417-444.

   doi: 10.1080/17452759.2020.1808937

10. Choe YE, Kim GH. A PCL/cellulose coil-shaped scaffold via a modified electrohydrodynamic jetting process. *Virtual Phys Prototyp*. 2020;15(4):403-416.

doi: 10.1080/17452759.2020.1808269

11. Li X, Liu B, Pei B, *et al.* Inkjet Bioprinting of Biomaterials. *Chem Rev*. 2020;120(19):10793-10833.

doi: 10.1021/acs.chemrev.0c00008

12. Weygant J, Entezari A, Koch F, *et al.* Droplet 3D cryobioprinting for fabrication of free-standing and volumetric structures. *Aggregate*. 2024;5(5):e599.

doi: 10.1002/agt2.599

13. Ng WL, Lee JM, Zhou M, *et al.* Vat polymerization-based bioprinting—process, materials, applications and regulatory challenges. *Biofabrication*. 2020;12(2):022001.

doi: 10.1088/1758-5090/ab6034

14. Lin H, Zhang D, Alexander PG, *et al.* Application of visible light-based projection stereolithography for live cell-scaffold fabrication with designed architecture. *Biomaterials*. 2013;34(2):331-339.

doi: 10.1016/j.biomaterials.2012.09.048

15. Lemma ED, Spagnolo B, De Vittorio M, Pisanello F. Studying Cell Mechanobiology in 3D: The Two-Photon Lithography Approach. *Trends Biotechnol*. 2019;37(4):358-372.

doi: 10.1016/j.tibtech.2018.09.008

16. Miri AK, Khalilpour A, Cecen B, Maharjan S, Shin SR, Khademhosseini A. Multiscale bioprinting of vascularized models. *Biomaterials*. 2019;198:204-216.

doi: 10.1016/j.biomaterials.2018.08.006

17. Lee A, Hudson AR, Shiwarski DJ, *et al.* 3D bioprinting of collagen to rebuild components of the human heart. *Science*. 2019;365(6452):482-487.

doi: 10.1126/science.aav9051

18. Parihar A, Parihar DS, Gaur K, Arya N, Choubey VK, Khan R. 3D bioprinting for drug development and screening: Recent trends towards personalized medicine. *Hybrid Adv*. 2024;7:100320.

doi: 10.1016/j.hybadv.2024.100320

19. Parihar A, Pandita V, Khan R. 3D printed human organoids: High throughput system for drug screening and testing in current COVID-19 pandemic. *Biotechnol Bioeng*. 2022;119(10):2669-2688.

doi: 10.1002/bit.28166

20. 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

21. 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

22. He Y, Yang F, Zhao H, Gao Q, Xia B, Fu J. Research on the printability of hydrogels in 3D bioprinting. *Sci Rep*. 2016;6(1):29977.

doi: 10.1038/srep29977

23. Compaired PM, García-Gareta E, Pérez MÁ. An experimental workflow for bioprinting optimization: Application to a custom-made biomaterial ink. *Int J Bioprinting*. 2025;11(3):397-415.

doi: 10.36922/IJB025120094

24. Tian S, Zhao H, Lewinski N. Key parameters and applications of extrusion-based bioprinting. *Bioprinting*. 2021;23:e00156.

doi: 10.1016/j.bprint.2021.e00156

25. Yu C, Jiang J. A Perspective on Using Machine Learning in 3D Bioprinting. *Int J Bioprinting*. 2020;6(1):253.

doi: 10.18063/ijb.v6i1.253

26. Shangguan P, Zhou H, Huang X, Su J, Yeong WY, Sing SL. Artificial intelligence-driven material development for additive manufacturing: A critical review. *Int J AI Mater Des*. 2025;2(2):1-26.

doi: 10.36922/IJAMD025100007

27. Zolfagharian A, Jin L, Ge Q, *et al.* Roadmap on Artificial Intelligence-Augmented Additive Manufacturing. *Adv Intell Syst*. 2026:e202500484.

doi: 10.1002/aisy.202500484

28. Sani AR, Zolfagharian A, Kouzani AZ. Artificial Intelligence-Augmented Additive Manufacturing: Insights on Closed-Loop 3D Printing. *Adv Intell Syst*. 2024;6(10):2400102.

doi: 10.1002/aisy.202400102

29. Huang X, Ng WL, Yeong WY. Predicting the number of printed cells during inkjet-based bioprinting process based on droplet velocity profile using machine learning approaches. *J Intell Manuf*. 2024;35(5):2349-2364.

doi: 10.1007/s10845-023-02167-4

30. Huang X, Wong YX, Goh GL, Gao X, Lee JM, Yeong WY. Machine learning-driven prediction of gel fraction in conductive gelatin methacryloyl hydrogels. *Int J AI Mater Des*. 2024;1(2):61-75.

doi: 10.36922/ijamd.3807

31. Shi X, Sun Y, Tian H, Abhilash PM, Luo X, Liu H. Material Extrusion Filament Width and Height Prediction via Design of Experiment and Machine Learning. *Micromachines*. 2023;14(11):2091.

doi: 10.3390/mi14112091

32. Ruberu K, Senadeera M, Rana S, *et al.* Coupling machine learning with 3D bioprinting to fast track optimisation of extrusion printing. *Appl Mater Today*. 2021;22:100914.

doi: 10.1016/j.apmt.2020.100914

33. 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

34. Lewis P, Perez E, Piktus A, *et al.* Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems 33. In: Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020); December 6–12, 2020; Online. Neural Information Processing Systems Foundation, Inc. (NeurIPS); 2020:9459-9474.

35. Huang L, Yu W, Ma W, *et al.* A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Trans Inf Syst*. 2025;43(2):42.

doi: 10.1145/3703155

36. Tshitoyan V, Dagdelen J, Weston L, *et al.* Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*. 2019;571(7763):95-98.

doi: 10.1038/s41586-019-1335-8

37. Lee J, Su H. A unified industrial large knowledge model framework in Industry 4.0 and smart manufacturing. *Int J AI Mater Des*. 2024;1(2):41-47.

doi: 10.36922/ijamd.3681

38. Lee J, Su H. Rethinking industrial artificial intelligence: A unified foundation framework. *Int J AI Mater Des*. 2025;2(2):56-68.

doi: 10.36922/IJAMD025080006

39. Lee J, Su H, Ji DY, Minami T. Engineering Artificial Intelligence: Framework, Challenges, and Future Direction. *Mach Learn Eng*. 2025;1(1):013001.

doi: 10.1088/3049-4761/adce0d

40. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *arXiv*. Posted online December 10, 2015.

doi: 10.48550/arXiv.1512.03385

41. Sani AR, Zolfagharian A, Kouzani AZ. Automated defects detection in extrusion 3D printing using YOLO models. *J Intell Manuf*. 2026;37(1):351-371.

doi: 10.1007/s10845-024-02543-8

42. Sani AR, Kouzani AZ, Zolfagharian A. Real-time defect monitoring in material extrusion 3D printing using optimized YOLO models. *Prog Addit Manuf*. 2026;11(1):1115-1137.

doi: 10.1007/s40964-025-01401-0

43. Daulton S, Balandat M, Bakshy E. Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization. *arXiv*. Posted online June 9, 2020.

doi: 10.48550/arXiv.2006.05078

44. Lee J, Su H. Agentic AI for smart manufacturing. *Manuf Lett*. 2025;46:92-96.

doi: 10.1016/j.mfglet.2025.10.013