

一种融合双粒度知识库的 仿生设计方案智能生成方法

张晓明^{1,2}, 刘成响¹, 钟 凯¹, 谢 锐^{3*}, 钱护磊¹

(1. 安徽大学物质科学与信息技术研究院, 安徽合肥 230601; 2. 青海省科学技术信息研究所有限公司, 青海西宁 810008;
3. 青海理工学院, 青海西宁 810016)

摘 要: 仿生学通过借鉴自然界生物的优异特性, 为工程创新提供了重要的方法支撑。然而, 如何从丰富的生物信息中提取准确的知识并与设计需求有效对接, 仍具有挑战。尽管大语言模型(Large Language Models, LLM)在自动化仿生设计过程中展现了潜力, 但其依然存在局限性, 包括: 认知缺乏导致对仿生原理解偏差; 参数化知识约束引发时效性与准确性不足; 外部知识库利用不足使得生成方案缺乏领域支撑。为此, 本文提出一种融合双粒度知识库与LLM的智能仿生设计方案生成方法(Intelligent Bio-Inspired Design, IBID), 其核心是整合双粒度仿生知识库与LLM协同生成与评审机制, 以实现从问题出发生成完整且有效的仿生设计方案。首先, 针对知识表示与检索, 本文收集生物优异特性及仿生实例, 通过将每种生物策略拆分为连贯的片段, 并依据主题相似性对其进行聚类, 形成粗细两种粒度的知识表示, 构建了双粒度仿生知识库, 并设计了双粒度检索机制以提升知识检索的效率与准确性。其次, 针对方案生成与优化, 本文设计了分层协同的LLM工作流程: 初始由基于LLM的仿生设计方案写作专家根据需求生成结构化大纲, 进而设计了专家会商评审机制, 由多个LLM形成普通专家组, 基于知识证据独立扩展草案, 再通过多轮德尔菲法进行意见交换、评估与迭代修订, 以此集成多元视角、提升一致性与可靠性; 最后, 由基于LLM的高级专家进行终审, 执行逻辑校验、细节补充与整体提升, 避免观点漂移, 提升方案的完整性与有效性。为验证IBID的有效性, 本文在自主构建的仿生数据集上开展了实验分析, 结果表明: IBID在Precision@1 (P@1)、Precision@3 (P@3)及Mean Average Precision (MAP)等检索指标上均优于最佳基线模型; 在生成质量评估中, IBID在专业质量及内容相关性等大多数指标上优于GraphRAG、RAPTOR及其他基线模型。本方法有效缓解了LLM在专业领域的幻觉与知识局限问题, 为高质量、高可靠的智能仿生设计提供了实用框架。

关键词: 仿生设计; 大语言模型; 检索增强生成; 聚类; 文本嵌入; 专家会商

基金项目: 中国科学院委托技术开发项目(No. 动合(科)2025-069)

中图分类号: TP182; TP391

文献标识码: A

文章编号: 0372-2112(2026)03-1161-17

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20250456

An Intelligent Bio-Inspired Design Solution Generation Method Integrating Dual-Granularity Knowledge Bases

ZHANG Xiaoming^{1,2}, LIU Chengxiang¹, ZHONG Kai¹, XIE Kun^{3*}, QIAN Hulei¹

(1. Institutes of Physical Science and Information Technology, Anhui University, Hefei, Anhui 230601, China; 2. Qinghai Institute of Science and Technology Information, Xining, Qinghai 810008, China; 3. Qinghai Institute of Technology, Xining, Qinghai 810016, China)

Abstract: Bio-inspired design leverages the superior characteristics of biological organisms in nature, providing essential methodological support for engineering innovation. However, extracting accurate knowledge from abundant biological information and effectively aligning it with design requirements remains challenging. Although large language models (LLMs) have demonstrated potential in automating the bio-inspired design process, they still exhibit notable limitations, including: cognitive deficiencies leading to biased understanding of bio-inspired principles; parametric knowledge constraints resulting in insufficient timeliness and accuracy; and underutilization of external knowledge bases, which leaves generated solutions lacking domain-specific support. To address these issues, this paper proposes an intelligent bio-inspired design (IBID) method that integrates a dual-granularity knowledge base with LLMs for automated design solution generation. The core of IBID lies in the synergistic integration of a dual-granularity bio-inspired knowledge base with an LLM-based collaborative generation and review mechanism, enabling the generation of complete and effective bio-inspired design solutions from problem statements. First, regarding knowledge representation and retrieval, we collect biological superior characteris-

tics and bio-inspired instances, decompose each biological strategy into coherent segments, and cluster them based on thematic similarity to form coarse- and fine-grained knowledge representations, thereby constructing a dual-granularity bio-inspired knowledge base. A corresponding dual-granularity retrieval mechanism is designed to improve the efficiency and accuracy of knowledge retrieval. Second, regarding solution generation and optimization, we design a hierarchically collaborative LLM workflow: an LLM-based bio-inspired design writing expert initially generates a structured outline based on the requirements; subsequently, an expert consultation and review mechanism is introduced, wherein multiple LLMs form a panel of general experts that independently expand drafts based on knowledge evidence, followed by multi-round Delphi-method-based opinion exchange, evaluation, and iterative revision to integrate diverse perspectives and enhance consistency and reliability. Finally, an LLM-based senior expert conducts a final review, performing logical verification, detail supplementation, and overall refinement to prevent viewpoint drift and improve the completeness and effectiveness of the solution. To validate the effectiveness of IBID, experimental analyses are conducted on a self-constructed bio-inspired dataset. The results demonstrate that IBID outperforms the best baseline models across retrieval metrics including Precision@1 (P@1), Precision@3 (P@3), and Mean Average Precision (MAP). In the generation quality evaluation, IBID surpasses GraphRAG, RAPTOR, and other baseline models on most metrics, including professional quality and content relevance. The proposed method effectively mitigates hallucination and knowledge limitation issues of LLMs in specialized domains, providing a practical framework for high-quality and reliable intelligent bio-inspired design.

Keywords: bio-inspired design; large language models; retrieval-augmented generation; clustering; text embeddings; expert consultation

Foundation Item(s): Commissioned Technology Development Project of the Chinese Academy of Sciences (No. 动合(科)2025-069)

0 引言

仿生学通过分析和模仿自然界中的生物优异特性,为材料、机械、感知、群体等领域的创新设计提供重要支撑。通过研究自然现象、生物策略等,科学家、工程师和设计师们开始逐步解锁生物启发创新的宝库,并取得了反映自然本身效率、韧性和适应性的科技进步^[1]。以生物减阻机制为例,研究者通过借鉴鲨鱼皮肤表面的齿状凸起^[2],在翼型设计中引入类似的仿生微织构,从而有效改善了气动表面的流场特性,实现了升阻比的跨越式提升。荷叶卓越的疏水性启发了超疏水表面的设计,这种表面展现了著名的“荷叶效应”^[3],在自清洁和防腐蚀材料等领域得到应用。这些实例表明,自然生物于进化过程中形成的优异结构和功能等特性,为仿生创新提供了宝贵的生物策略。通过这些生物策略转化为仿生设计方案,可以推动材料、建筑、机器人等多个领域的进步,将带来可持续且高性能的成果^[4-6]。

尽管仿生学在工程和设计领域展现出巨大潜力,但其实际应用仍面临严峻挑战^[7-8]。主要问题在于生物特性种类繁多且结构复杂,使得设计者必须在海量的生物知识中筛选出与特定需求相关的信息,以生成完整且有效的仿生设计方案^[9-10]。这一过程不仅耗时费力,而且要求设计者具备广泛的跨学科知识,从而大大增加了实施难度。因此,如何高效地将用户需求与合适的生物特性相匹配并生成高质量设计方案,成为仿生设计的一大难点^[11-13]。

随着大语言模型(Large Language Models, LLM)的发展,自动生成仿生设计方案成为可能。但是构建一个全面的仿生设计通常还需要综合大量专业知识,这导致输入数据量往往超出当前最先进 LLM 的处理能力。此外,一个精心构思的仿生设计本身可能涵盖数万个 token^[6-8],使得直接由 LLM 输出变得极不现实^[14-17]。LLM 在参数知识方面的限制也构成了阻碍^[18-19],在处理长尾话题^[20-21]时,因为模型训练数据十分稀缺,这种限制尤为明显。同样,在仿生工程这一特定领域,仅依靠 LLM 内部已有的知识难以生成全面且准确的仿生设计。LLM 可能会生成缺乏足够细节或包含不准确信息的内容^[18],从而影响设计的可靠性。这些限制削弱了利用 LLM 生成仿生设计方案的效果和可靠性,迫切需要有效整合外部知识以提高生成设计方案的准确性和全面性^[21]。

为应对上述挑战,本文提出了一种融合双粒度知识库与 LLM 的仿生设计方案生成方法,实现从问题出发生成完整且有效的仿生设计方案。本文的主要创新如下。

(1) 基于聚类的双粒度知识表示与知识库构建:该方法通过将每种生物策略拆分为连贯的片段,并依据主题相似性对其进行聚类,形成从粗细两种粒度进行仿生知识表示,应对跨多种生物策略管理大量且异质信息的问题。同时,基于此构建了双粒度仿生知识库,以支持后续的仿生设计任务:每个原始摘要内容和总结后的聚类内容都会被存储在 Macro-level 知识库中,而原始的、细粒度的文本块则保存在 Micro-level

知识库中。通过设计双粒度检索机制既可以从聚类摘要中定位粗粒度的信息,也可以从片段存储库中有效获取细粒度的知识。此方法既突出了关键信息,又保留了完整的上下文信息,使 IBID 能够生成更完整、准确的仿生设计。

(2) 基于 LLM 的专家会商评审机制:该方法利用基于 LLM 的多位普通专家协同会商与一位高级专家的决策审定,对候选设计方案进行评估、修订和完善。其核心在于先由多位普通专家分别从专业质量与结构质量对方案打分并排序,然后通过 Delphi 多轮会商机制,提升方案生成与评价的一致性与有效性。输出结果由高级专家进行最终决策并对方案进一步完善,从而避免观点漂移,提升方案生成的质量。

1 相关工作

1.1 仿生设计方案自动生成

近年来,越来越多的研究人员开始尝试自动化的生成仿生设计方案,如 Kang 等人^[22]引入了 BioSpark,该方法通过反复的分类学扩展和语义聚类来组织经过专家筛选的生物机制,从而促进了设计叙事的自动生成;Luu 等人^[23]利用 LLM 整合大量生物文献,使得设计理念能够得到更为精细的表达;Toukmaji 等人^[24]则将检索增强生成与 ReAct 智能体相结合,以减少幻觉现象并提升生成设计文本的可行性与文本的新颖性。而 BIDTrainer^[25]则在设计教育中应用 LLM,通过推动互动式“learning-by-asking”,以提高设计写作能力。以上这些研究共同为自动化仿生设计写作奠定了坚实基础,初步实现了从生物特性的观察发现到仿生设计方案生成的衔接。

然而,上述方法普遍存在一些不足,如未能有效建立结构化、可持续更新的仿生知识库,缺乏对知识的精准管理与高效检索,且在知识集成与迭代优化方面存在不足,从而限制了其生成仿生设计方案的准确性和相关性。本文提出的 IBID 方法通过融合双粒度知识库与 LLM,能够更加系统地组织、检索与应用仿生知识,以期解决这些问题。

1.2 检索增强生成

检索增强生成 (Retrieval-Augmented Generation, RAG) 的一种经典方法是通过引入外部知识库,在推理阶段通过检索外部知识库来增强 LLM 生成能力,将检索到的信息与原始查询一起添加到 LLM 的上下文窗口中^[26-28]。Naïve RAG^[21]就是一种经典的 RAG 方法,它首先将文本拆分为若干块,并将这些文本块嵌入到一个向量空间中;随后,再将查询嵌入到同一向量空间中,并以与查询嵌入最近的 k 个文本块嵌入所对应的文本块作为上下文。然而,Naïve RAG 存在

精度和召回率问题,容易选取不相关或错误的信息片段,并且可能遗漏关键内容^[29]。为此,研究人员引入了 Advanced RAG^[30],以克服 Naïve RAG 的局限性,如 Ma 等人^[31]提出了改写查询的方法,即采用一个小型语言模型作为可训练的改写器,填补给定输入和检索需求之间的差距。与此同时,Gao 等人^[32]探索了完全零样本的稠密检索方案,通过在检索前生成“虚拟文档”来表示查询,从而显著提升了多种任务场景下的检索准确率和召回率。随着 RAG 的进一步发展,为了针对不同场景进行灵活调整,Modular RAG^[21]随之诞生;如 Shao 等人^[33]提出的迭代检索生成框架通过多轮次交互,利用前一轮生成的内容指导后续检索,有效弥补了一次性检索在语义覆盖上的不足。此外,Yu 等人^[34]引入了“生成-而非检索”范式,展示了 LLM 强大的上下文生成能力,从而在一定程度上减少了对外部检索模块的依赖。Sarthi 等人^[35]的 RAPTOR 模型通过递归抽象摘要构建树状检索结构,实现了对长文档层次化语义信息的捕捉和整合;同时,Edge 等人^[36]提出了 GraphRAG 方法,通过构建图形索引和社区摘要,有效支持全局查询的摘要生成。

本文的工作融合了上述 RAG 思路,并基于仿生领域数据的特点进行适配,通过基于聚类的双粒度知识表示和专家会商机制,进一步提升了检索增强系统的整体性能。

2 问题驱动的仿生设计方案生成任务

本节将介绍问题驱动的仿生设计方案生成任务,并详细阐述如何利用 LLM 与会商机制来构建智能仿生设计方案生成方法。

2.1 任务定义

问题驱动的仿生设计指:明确特定的研发或设计问题,然后系统地从生物策略中寻找对应的类比或启发,以帮助解决问题或优化相应功能。具体而言,这种问题驱动型的仿生设计方法一般遵循以下步骤^[7,12]。

(1) 问题定义

首先需要明确问题所涉及的功能、应用场景及相关约束条件,并将问题抽象成“能够准确描述需求,而不提前限定具体解决方案”的形式。这种抽象的问题定义为后续精准地寻找并评估潜在的生物学启发提供了指导和框架。

(2) 生物类比的检索

基于定义问题,系统地从生物学相关文献、数据库或专家知识库中,检索出能够与该问题对应的潜在类比对象,以作为后续仿生设计的启发来源。

(3) 生物类比的评估

在检索到大量可能相关的生物类比对象之后,需

要对这些类比对象进行评估与筛选,以确定其是否真正对当前的设计问题具有启发价值与应用潜力,保证后续设计工作的针对性和有效性。

(4) 生物类比的应用

在确认了合适的生物类比后,需要进一步提炼该生物策略的核心机理或关键特征,将其有效地融合到预先定义好的工程需求及实际约束条件之中,形成切实可行的设计方案或技术实现路径。

基于上述问题驱动型仿生设计的流程,本文设计了一种融合双粒度知识库与 LLM 的仿生设计方案生成方法,以实现仿生设计流程的智能化,并收集真实仿生数据开展方法的有效性验证。具体而言,本文首先参照上述问题定义方法,基于专业仿生知识网站 AskNature 以及仿生相关文献设计并提出了 402 个具体的仿生设计问题(Bionic Problems, BioP),随后利用这些问题,从本文构建的双粒度知识库中检索相应的生物类比对象(Bionic Analogies, BioAna),由此产生初步的仿生设计方案(Bio-Inspired Design, BID)并送入多专家会商机制(Multi-experts Seminar)进行深入评估与修订,最终生成最佳仿生设计方案(Best Bio-Inspired Design, B-BID)。

2.2 数据集构建

为了有效支持并验证基于 LLM 的问题驱动型仿生设计生成任务,本文选择公开的仿生数据源 AskNature^[37]以及仿生相关文献作为数据基础,构建本文的仿生数据集。具体而言,从 AskNature 获取的数据分为两个部分:第一部分为生物策略(Strategy),第二部分为仿生创新(Innovation)。为确保数据集涵盖完整的仿生设计链条,本文只选取 Innovation 作为初始样本,主要原因是 Innovation 均具有明确关联的 Strategy,但并不是每个 Strategy 都会关联 Innovation。针对每个获取到的 Innovation,本文使用 GPT-4o 模型作为仿生领域的专家,从每一个 Innovation 中提取并生成相应的仿生设计问题(BioP),同时将其对应的关联 Strategy 作为相应的 BioAna。为了实现对实验数据集的扩展,我们使用本实验室构建的仿生知识抽取模块,基于中国科学院动物研究所提供的部分仿生文献资料,开展了仿生数据抽取和校验,为数据集补充了来自仿生文献的 Strategy 以及对应的 Innovation。

为了进一步保证每个 BioP 定义的质量与合理性,本文制定了一套评估标准,详见 4.2 节。这套标准的核心在于确保所定义的问题能准确体现功能需求、具体使用场景以及相应的约束条件,并将问题抽象化为“需求明确但不限定具体解法”的规范表述形式。评估过程中,本文使用 GPT-4o 对 BioP 的内容进行形式审核,通过专门设计的提示词,初步评估每个生成

的 BioP 样本的质量。通过审核,本文生成了 402 个符合质量标准 and 要求的 BioP。为了分析上述 402 个 BioP 的领域覆盖情况及其跨领域交叉特征,表 1 统计了 BioP 在各类领域中的分布情况。表 2 给出了单领域样本与跨领域样本的各自数量。在构建 BioP 的答案数据时,本文使用了 AskNature 和仿生相关文献对应的仿生创新案例作为标准答案,以进一步提高数据集质量与设计方案的完整性和可靠性。仿生方案数据集的构建为后续研究提供了数据基础。

表 1 仿生设计方案数据集领域覆盖统计表

Table 1 Domain coverage statistics of the bio-inspired design dataset

大类	BioP 数量
材料与化学	143
能源与环境	48
医疗与生物技术	57
电子与信息	44
机器人与自动化	48
工业制造与工程系统	46
建筑与航空航天	37
消费与日用	37

表 2 跨领域样本统计表

Table 2 Cross-domain sample statistics

类型	样本数
单领域	357
双领域交叉	32
多领域交叉	13

3 方法

本节将详细介绍 IBID 方法。为了更好地结合 LLM 的特性,并与标准的问题驱动仿生设计写作范式保持一致,IBID 构建了三个核心的组成部分,具体包括:(1)基于聚类的双粒度知识表示与知识库构建;(2)基于语义嵌入与分库评分融合的双粒度知识检索;(3)基于 LLM 专家会商的仿生设计评估与优化。图 1 展示了 IBID 的整体流程。

3.1 基于聚类的双粒度知识表示与知识库构建

仿生知识通常具有跨领域、数量庞大、结构多样等特点,涵盖了生物特性描述、功能机制、仿生实例等多个维度,并且不同维度间具有不同的关系。这类具有多源性、维度复杂且语义分散特征的知识被定义为“异质信息”。这种知识的高度复杂性和分散性给知识检索与利用带来了挑战,迫切需要构建一种全面、一致的知识库,以系统化地组织仿生知识,实现高效的知识管理和利用。因此,本文针对仿生数据特点,设计了包含 Macro-level 与 Micro-level 两种粒度的知识表示方法,构建了双粒度知识库,以提高仿生知

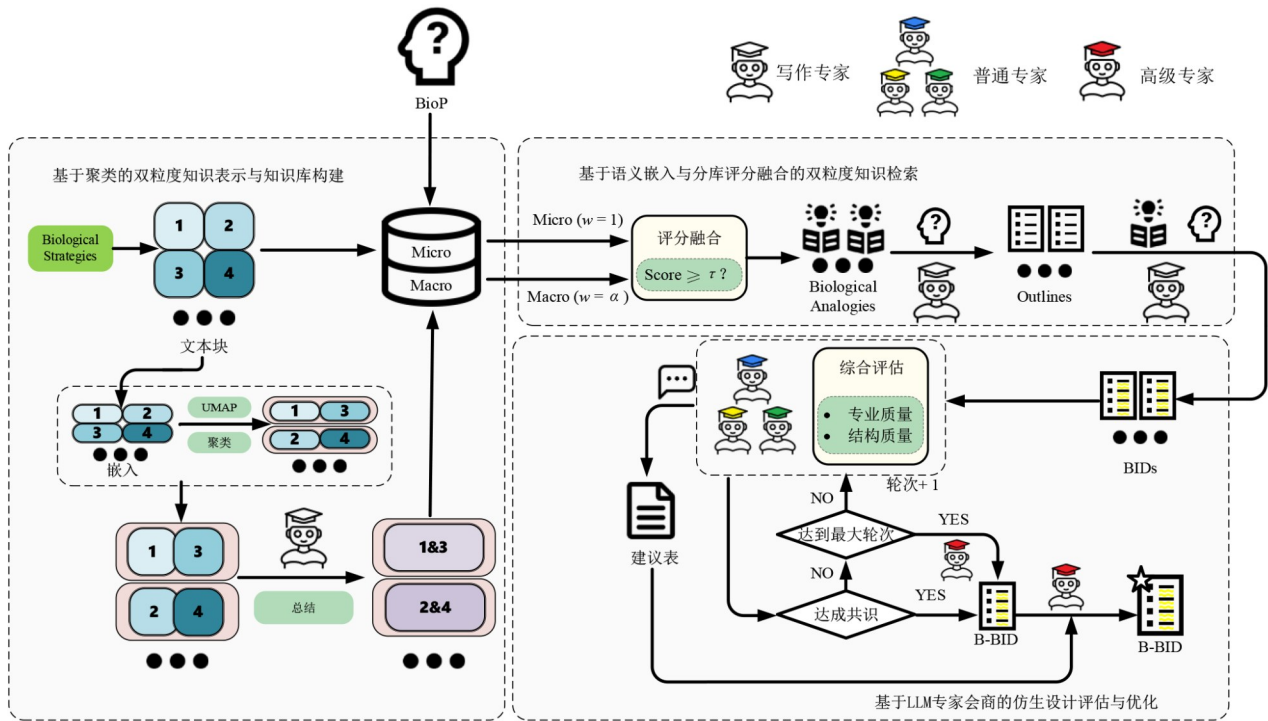


图1 IBID流程图

Figure 1 IBID flowchart

识检索的效率和质量。

针对 Micro-level 知识库, 首先将每个生物策略文本划分为长度为 200 tokens 的连续文本块, 然后使用 BGE-M3 语义向量模型为每个文本块生成对应的语义嵌入向量。这些文本块及其语义嵌入、唯一标识 ID 共同构成 Micro-level 知识库。

针对 Macro-level 知识库, 本文将每个 Strategy 的概述部分视作一个独立的文本单元, 使用 BGE-M3 生成嵌入向量, 并将其存入知识库。为进一步优化知识库结构并提高检索效率, 采用聚类算法识别生物策略间的相似信息并进行分组。由于原始嵌入向量维度过高, 直接聚类可能导致计算复杂度增加及度量不稳定性, 本文首先使用 Uniform Manifold Approximation and Projection (UMAP) 对数据进行降维处理, 以最大程度保留原始数据的局部邻域结构。UMAP 将优化高维空间中样本间的邻接概率 p_{ij} 与低维空间对应概率 q_{ij} 之间的交叉熵损失 L :

$$L = - \sum_{i < j} [p_{ij} \log(q_{ij}) + (1 - p_{ij}) \log(1 - q_{ij})] \quad (1)$$

降维后的数据具备更适合聚类分析的空间结构, 可有效减少高维数据带来的距离度量失真与不稳定性。在降维后的数据空间, 本文进一步采用 HDBSCAN 聚类算法进行分组。HDBSCAN 无需预先指定簇数量, 可以自动挖掘数据的多尺度密度结构,

并识别噪声数据, 适配了仿生数据存在的“数据分布不均、类别数量未知、内容复杂多噪声”特点。为优化聚类效果, 本文实施网格搜索对 HDBSCAN 算法参数进行自动调优, 以轮廓系数 S 为目标, 实现其最大化, 具体定义如下:

$$S = \frac{1}{N_c} \sum_{i \in I} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

$I = \{i | q_i = 1\}$ 为非噪声样本索引集, $N_c = |I|$, $a(i)$ 代表样本 i 到其所在簇内其他样本的平均距离, $b(i)$ 代表其到最近相邻簇样本的平均距离。通过最大化整体轮廓系数, 本文实现簇内样本的紧密性与簇间区分度的最大化。

如图 2 所示, 本文将 330 个 Strategy、1 008 个 chunk 嵌入后, 经 UMAP 降维至二维空间进行可视化, 并在该低维表示上采用 HDBSCAN 进行密度聚类。图中不同颜色对应不同聚类簇 (以 C0、C1... 表示), 灰色点为 HDBSCAN 判定的噪声/离群样本 (Noise)。从可视化结果可以观察到, 大多数样本在二维空间内形成若干相对紧凑的局部密度团簇, 簇间存在一定间隔, 同时仍存在分布较分散的点被归为噪声, 这与 HDBSCAN 能够在非均匀密度数据上识别核心簇并剔除离群点的特性一致。为进一步量化聚类效果, 本文在非噪声样本集合上计算聚类质量指标。统计结果显示: 所有样本被稳定划分到 52 个簇中。平均簇大

小为 15.48, 中位数为 13, 最小簇规模为 8、最大为 45, 表明聚类未出现被单一超大簇主导的极端不均衡现象。此外轮廓系数 $S=0.626$ 表明簇内紧密性与簇间可分性处于合理水平^[38]。

完成 HDBSCAN 聚类后, 每个簇内的多个文本块会送入写作专家来生成高度凝练的摘要, 进一步利用 BGE-M3 生成摘要的嵌入向量并存入 Macro-level 知识库中^[39]。

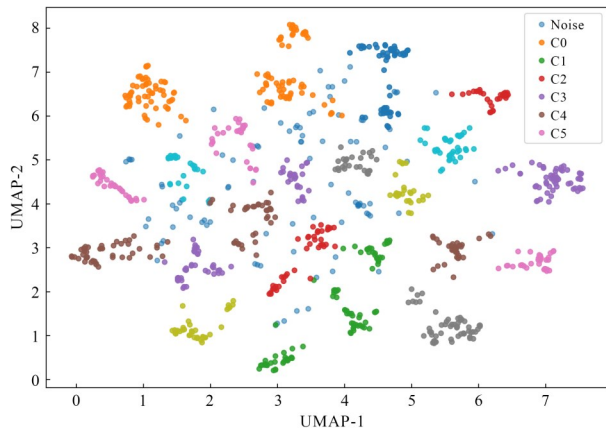


图2 UMAP-HDBSCAN 聚类可视化图

Figure 2 UMAP-HDBSCAN clustering visualization

3.2 基于语义嵌入与分库评分融合的双粒度知识检索

双粒度知识库构建后, 本文将基于具体定义的 BioP 对 Micro-level 与 Macro-level 知识库并行检索, 以高效地获得潜在的 BioAna。为了有效利用两类知识库的知识, 本文提出了一种基于向量嵌入语义检索和分库评分融合的双粒度知识检索机制, 以提升所获得的仿生类比对象的准确性和全面性。

本文定义所有候选的生物策略集合为

$$\mathbf{B} = \{b_1, b_2, \dots, b_n\} \quad (3)$$

首先每个候选的生物策略 b_i 通过知识库检索得到一个初始的相关性评分 $S(b_i)$ 。然后进一步定义了一个条目集合 E , 集合中的每个条目 e 对应于一组可能关联若干生物策略的簇。

为了调节不同粒度知识块的融合比例, 本文设计了一个权重函数 $w(e)$ 。其中, Micro-level 知识块贡献被设定为 1; Macro-level 知识块的贡献则被设定为超参数 α , 若一个 Macro-level 知识块关联多个 Micro-level 的子知识块时, 则其权重值 α 将平均分配给构成该 Macro-level 知识块的每个子知识块所关联的生物策略。权重函数 $w(e)$ 定义如下:

$$w(e) = \begin{cases} 1, & e \in \text{Micro-level} \\ \frac{\alpha}{|e|}, & e \in \text{Macro-level} \end{cases} \quad (4)$$

其中, $|e|$ 表示 Macro-level 条目 e 中 chunk 的数量。Macro-level 摘要由同一聚类簇内的多个 Micro-level 文本块(chunk)凝练生成, 而这些 chunk 所属的 Strategy 可能相同也可能不同。本文将 Macro-level 的总体贡献权重 α 按簇内 chunk 的数量进行归一化分配: 每个 chunk 贡献 $\alpha/|e|$, 再按 chunk 与 Strategy 的对应关系累加到各个 Strategy 上。当某一 Strategy 在 Macro-level 中对应的 chunk 更多时, 其累计获得的 Macro-level 增益会更高; 同时归一化分配可保证单个 Macro-level 条目的总增益不随簇规模无限增大, 从而避免在融合时忽略细粒度知识。

基于以上定义, 本文所提出的融合 Micro-level 与 Macro-level 评分的方法可以形式化地表示为

$$S(b_i) \leftarrow S(b_i) + \sum_{b_i \in e} w(e) \quad (5)$$

通过上述融合方法, 本文能更加全面而准确地评估每个候选策略 b_i 与给定 BioP 的实际语义相关性。完成评分融合后, 本文设定一个最低相关性阈值 τ , 并依据此阈值将所有评分低于该阈值的生物策略予以剔除, 最终获得满足要求的候选生物策略集合:

$$\mathbf{B}^* = \{b_i | S(b_i) \geq \tau\} \quad (6)$$

然而, 由于 LLM 在生成较长文本时易出现上下文语义丢失和结构混乱的问题, 简单直接地以 BioP 和生物策略 b_i 为输入生成仿生设计方案往往难以确保最终文本的逻辑严谨性与结构完整性。为有效应对这一问题, 本文设计了仿生设计大纲 \mathbf{O} 。这一结构化大纲生成过程可形式化为

$$\mathbf{O} = \arg \max_{\mathbf{O}} P(\mathbf{O} | \text{BioP}, b_i) \quad (7)$$

大纲 \mathbf{O} 是在给定的 BioP、具体生物策略 b_i 条件下的精炼结构文本, 并确保大纲高度贴合生物启发来源与仿生设计问题。

在初步的 BID 生成阶段, 本文综合了所有满足要求的候选生物策略集合 \mathbf{B}^* 和 BioP, 在已生成的结构化大纲基础上逐节展开细化, 最终生成完整且严谨的初步 BID, 如下式所示:

$$\text{BID} = \mathcal{F}(\{\mathbf{O}_i\}_{b_i \in \mathbf{B}^*}, \text{BioP}) \quad (8)$$

其中, $\mathcal{F}(\cdot)$ 表示融合和细化各子大纲 \mathbf{O}_i 及 BioP 而最终生成完整 BID 的函数。

3.3 基于 LLM 专家会商的仿生设计评估与优化

单一 LLM 往往存在领域知识不足、幻觉或偏见等固有问题, 我们通过设计多 LLM 决策融合以减轻该问题的影响。在本阶段, 本文提出了一种多专家会商机制, 以综合多位专家的观点对生成的候选 BID 进行评价与优化。本文将 LLM 划分为两类专家: 普通专家 E_1, E_2, E_3 和高级专家 E^* 。候选方案集合为

$\{BID_1, BID_2, \dots, BID_m\}$, 每位普通专家 E_i 分别从专业质量和结构质量两个方面对 BID 进行综合评估。其中, 专业质量要求方案清晰阐述所借鉴生物现象背后的功能机理、全面考虑生物系统的多种功能需求并合理权衡, 同时明确描述实现特定功能时的因果关系; 结构质量则要求各章节之间具备严谨的逻辑衔接和连贯性。基于这两个指标, 普通专家 i 对候选方案 BID_j 的评分定义为 $S_i(BID_j)$ 。每位普通专家随后选出其认为评分最高的方案。所有专家在发言时会输出针对当前审核的 BID 建议, 并共同维护一张建议表作为会商记录保存。

如果三位普通专家的初步最优选择一致, 则该方案直接被确定为最终的仿生设计方案 (B-BID), 无需进一步讨论; 反之, 则进入多轮讨论过程。在讨论中, 设第 t 轮时普通专家 E_i 的选择为 $BID_{opt}^{(i,t)}$, 并更新选择的过程如下式所示:

$$BID_{opt}^{(i,t)} = Seminar\left(BID_{opt}^{(1,t-1)}, BID_{opt}^{(2,t-1)}, BID_{opt}^{(3,t-1)}\right) \quad (9)$$

其中, Seminar 表示专家在参考其他两位专家意见后对自身选择的调整函数。当在某一讨论轮 t 中, 所有普通专家意见一致, 即

$$BID_{opt}^{(1,t)} = BID_{opt}^{(2,t)} = BID_{opt}^{(3,t)} \quad (10)$$

则该方案被确定为会商结论 B-BID, 无需再进行后续轮次。如果在预设的最大会商轮次 t_{max} 内仍无法达成一致, 则由高级专家 E^* 读取所有普通专家在最后一轮的选择与建议, 重点分析普通专家的相似问题, 并进行最终决策, 其过程形式化为

$$B-BID = E^*\left(BID_{opt}^{(1,t_{max})}, BID_{opt}^{(2,t_{max})}, BID_{opt}^{(3,t_{max})}\right) \quad (11)$$

在确定最终的 B-BID 后, 高级专家会根据建议表, 对该方案进行最后的修订完善, 以进一步提升其在专业性、逻辑严谨性和整体结构连贯性方面的水平。

在多轮会商过程中, 普通专家在阅读其他专家的评分与建议后对自身判断进行更新, 可能出现阶段性的观点变化。本文将该现象视为交叉评估下的正常“理性调整”, 其作用是促使专家在不同证据与论证角度之间综合思考, 从而提升最终方案的合理性与一致性。为保证讨论过程不发生偶发性观点漂移, 本文在会商中包含两层稳定机制: (1) 三位普通专家在每轮发言时均会输出针对当前候选方案的具体修改建议, 并共同维护一张建议表作为会商记录保存, 后续轮次的调整均需以建议表中累积的意见为参考推进。当仅个别专家发生观点变化时, 其他专家在相同评价维度(专业质量与结构质量)下给出的相对稳定的评分与建议, 会在有限轮次内形成一致性约束, 使讨论逐步向更具共同支撑的方案收敛; (2) 若在预设的最

大讨论轮次内仍未达成一致, 则由高级专家根据最后一轮各普通专家的选择与意见进行裁定, 并结合建议表对最终方案进行统一修订完善。

4 实验与分析

4.1 数据集

在 3.2 节已详细描述仿生设计方案数据集的构建方法, 仿生设计方案数据集包含 Innovation、BioP、Strategy 三种类型的数据, 其中 BioP 可能会对应多个 Strategy, 即存在多种 Strategy 可以解决某一 BioP, 数据集规模如表 3 所示。

表 3 仿生设计方案数据集规模数据统计表

Table 3 Scale statistics of the bio-inspired design dataset

	Innovation	BioP	Strategy
数量	402	402	330

4.2 实验环境和评价指标

实验环境为: 操作系统为 Ubuntu 22.04.5 LTS, CPU 为 AMD EPYC 7K62 48-Core Processor, GPU 为两张 NVIDIA RTX A6000, 单张显存为 48 GB。

实验从检索和生成两个方面对 IBID 进行评测。生成模块的仿生设计评分细则如表 4 所示。引入专业质量与结构质量两大评价维度。专业质量包含相关仿生文献^[7,12]中提到的三种关键仿生指标, 包括可行性与可持续性(F&S)、基于功能的类比匹配(Func-Match)与机制抽象与跨域转化(M&C), 分别用于评估生成设计的可行性与适用性、功能匹配程度及作用机制与应用背景的一致性; 结构质量涵盖连贯性和相关性^[19], 用以衡量文本的逻辑连贯性和主题聚焦度。实验使用 Prometheus 2(7B)^[40]基于表 4 中的评分细则对仿生设计进行评分, 该模型是一个用于评估其他语言模型的开源语言模型。所有评分均为五分制, 分值越高表明设计在相应维度上的表现越优。

实验中所有的文本嵌入维度大小均设置为 1 024 维, 系统在 Qwen2.5-7B-Instruct(以下简称 Qwen2.5-7B)上采用 LoRA 微调, 使用仿生领域科研文献作为微调语料, 提升 Qwen2.5-7B 在仿生领域的知识理解能力和领域适应性。训练阶段每卡批量为 1, 结合 8 步梯度累积实现等效批量 8, 初始学习率设为 1×10^{-4} , 配合余弦退火调度器并在线性预热前 10% 步骤中完成预热, 整个过程共计 3 个 epoch; 同时启用了 bfloat16 混合精度以减少显存占用。

为全面量化检索模块的效果, 本文采用 P@1、P@3 及 MAP 三项指标, 其中 P@K 用于评估候选结果前 K 名中目标策略的命中率, 而 MAP 通过对所有检索请求的平均精确率进行统计以衡量整体检索质量:

表 4 仿生设计评分细则

Table 4 Scoring criteria for bio-inspired design evaluation

结构质量	<p>连贯性:连贯性与组织结构:设计是否组织良好、结构合理? 得分 1:描述杂乱无章,缺乏逻辑结构和连贯性。 得分 2:描述基本有一定组织,存在基本结构但未能始终遵循。 得分 3:描述较为有条理,大体上遵循清晰结构,但在连贯性上有些许缺失。 得分 4:描述组织良好,结构清晰,仅有轻微的连贯性问题。 得分 5:描述组织极为出色,设计逻辑严密、结构合理,过渡自然,论述清晰。</p>
	<p>相关性:相关性:与聚焦程度:设计是否紧扣主题并保持清晰的重点? 得分 1:偏离主题;内容与标题或核心议题不符。 得分 2:部分切题,但存在多处跑题;核心议题虽可辨识,但未能始终坚持。 得分 3:总体切题,尽管包含少量无细节。 得分 4:大多紧扣主题且重点明确;叙述内容与核心议题保持较高相关性,仅偶有偏离。 得分 5:高度聚焦且完全切题;设计紧密围绕主题展开,每一条信息都有助于对主题的全面理解。</p>
专业质量	<p>可行性与可持续性:该设计概念是否涵盖了现实可行的制造方式、成本与环境影响评估,并利用了低能耗、可循环的生物过程? 得分 1:未考虑制造、成本或可持续性,纯属概念层面。 得分 2:对某一因素(如成本)有浅尝辄止的提及,但忽略了其他方面。 得分 3:对制造可行性或成本进行了基本分析,但忽视了环境或可持续性方面的内容。 得分 4:对制造、成本与环境影响进行了全面评估,但生物可持续循环的整合程度有限。 得分 5:具备完整的实施方案,详细规划了制造路径、成本估算、环境影响评估,并明确采用了低能耗、可循环的仿生过程。</p>
	<p>基于功能的类比匹配:在选择生物原型时,是否进行了严谨的功能-行为-结构映射,而非仅仅基于形态相似性? 得分 1:仅因外观或材料相似性选择原型,缺乏功能层面的依据。 得分 2:提及了一定的功能对应关系,但未验证其行为表现或结构的可扩展性。 得分 3:建立了功能等价关系,但对可转移的行为或可扩展的结构仅进行了部分考量。 得分 4:具有明确的功能-行为映射,并验证了大部分相关的结构特征,尽管在可扩展性或应用条件方面仍存在小部分不足。 得分 5:在功能、行为与结构层面实现了全面的同源对应:功能等价性已证实,类比条件下的行为表现得到验证,结构的可扩展性亦得到充分证明。</p>
	<p>机制抽象与跨域转化:该设计是否提炼出了可迁移的核心原理,并提供了详尽的工程实现路径? 得分 1:仅进行直接的结构仿制,缺乏原理抽象与工程实现策略。 得分 2:表面识别出某一原理,但未能提出具体的工程实现路径。 得分 3:完成了原理抽象,并初步提出工程转化设想,但在材料选择、制造工艺或性能评估方面仍显不足。 得分 4:清晰地完成了机制抽象,并提出了可行的工程路线图,包括材料选择、工艺流程和初步性能预测,尽管在部分细节上略有缺失。 得分 5:机制抽象与工程转化堪称典范:详尽的材料与工艺选择、深入的可行性分析、量化的性能预测,且明确避免了结构层面的拘泥复制。</p>

$$P@K = \frac{1}{K} \sum_{i=1}^K \text{rel}_i \quad (12)$$

$$\text{MAP} = \frac{1}{|Q|} \sum_{q \in Q} \text{AP}(q) \quad (13)$$

其中, rel_i 为第 i 条结果的二值相关性标记; Q 为检索请求集合; $\text{AP}(q)$ 表示查询 q 的平均精确率。

4.3 对比实验

如表 5 与表 6 所示,为验证 IBID 在仿生设计任务中的有效性,本文从检索和生成两个环节设计了分级对比方案。

在检索结果对比中以 Qwen2.5-7B 为统一主干构建 Naïve RAG,其中生物策略整体嵌入为“Full”,固定长度切块嵌入为“Chunked”,根据知识块的所属生物

表 5 不同方法的检索结果对比

Table 5 Comparison of retrieval results across different methods

		检索指标		
		P@1	P@3	MAP
Naïve RAG (Qwen2.5-7B)	Full	0.471	0.454	0.493
	Chunked	0.403	0.436	0.482
IBID		0.785	0.465	0.578

注:加粗数据表示最优结果。

策略来判断其是否命中目标策略,Naïve RAG 剔除了 IBID 中的相关性评分 $S(\mathbf{b}_i)$,以此考察 IBID 中的相关性评分 $S(\mathbf{b}_i)$ 的影响以及不同 Naïve RAG 方案对召回和排序的影响。在生成结果对比中则设置三条基线:一是 Only-LLM,即在完全无外部检索的条件下直接

表 6 不同方法的生成结果对比

Table 6 Comparison of generation results across different methods

		生成质量				
		专业质量			结构质量	
		F&S	FuncMatch	M&C	连贯性	相关性
AskNature		<u>4.26</u>	<u>4.77</u>	4.70	3.98	4.78
Only LLM	GPT-4o	3.84	3.64	2.99	4.82	4.31
	Qwen2.5-7B	3.33	3.75	2.62	4.36	4.46
	Qwen2.5-7B*	3.02	3.74	2.81	4.51	4.13
	Qwen2.5-14B	3.06	3.78	2.84	4.61	4.27
	DeepSeek-Chat	3.59	3.76	2.73	4.72	4.11
Naïve RAG (Qwen2.5-7B)	RAG-Full	2.72	3.64	2.85	4.44	4.68
	RAG-Chunked	2.70	3.51	2.66	4.28	4.71
GraphRAG	GPT-4o	4.17	4.35	3.46	4.69	4.30
	Qwen2.5-7B*	3.70	4.01	3.41	4.30	4.64
	Qwen2.5-7B	3.85	3.99	3.32	4.51	4.55
RAPTOR	GPT-4o	4.11	4.60	4.02	4.66	4.51
	Qwen2.5-7B*	3.77	3.98	4.21	3.69	4.50
	Qwen2.5-7B	3.80	4.01	4.06	4.01	4.39
IBID	GPT-4o	4.43	<u>4.61</u>	4.61	4.78	4.93
	Qwen2.5-7B	4.22	4.46	<u>4.72</u>	4.76	4.82
	Qwen2.5-7B*	4.24	4.52	4.77	4.62	<u>4.90</u>
	Qwen2.5-14B	4.22	4.46	4.71	<u>4.79</u>	4.83

注:加粗数据表示最优结果,下划线数据表示次优结果。

使用 GPT-4o、Qwen2.5-7B/14B 及 DeepSeek-Chat 生成 BID,其中带“*”标注的 Qwen2.5-7B 代表仿生领域上经 LoRA 微调的版本,用以评估轻量域适配带来的增益;二是 Naïve RAG (Full 与 Chunked),在获得检索结果后直接拼接上下文让 Qwen2.5-7B 生成设计文本;三是本文提出的 IBID,在相同 LLM 主干上引入“双粒度知识库+分层加权检索+结构化大纲驱动写作+多专家迭代”全流程,并分别测试三种写作专家和高级专家的 LLM 基座, Qwen2.5-7B/14B 与 GPT-4o。同时给出数据集集中的 Innovation 数据,作为人工撰写的参考 BID。

在检索指标方面, IBID 取得 $P@1 = 0.785$ 、 $P@3 = 0.465$ 与 $MAP = 0.578$,较最佳基线 Naïve RAG-Full 分别提升 31.4、1.1 与 8.5 个百分点。相比之下, Naïve RAG-Chunked 在三项指标上均劣于 Full,表明对 Strategy 盲目切块会削弱全局语义,而 IBID 借助 Macro-level 与 Micro-level 双粒度索引及权重融合,既保持整段语义完整性,又捕获细粒度细节,提升了命中率与整体排序质量。

生成质量方面, IBID 系列模型展现出显著优势。IBID-Qwen2.5-7B 在三项专业质量指标 (F&S、FuncMatch、M&C) 上均突破 4.2 分,其中在反映知识匹配度的 M&C 指标上达到 4.72,较同主干的 Only-LLM 提升了

2.1 分,较 Naïve RAG-Full 提升了 1.87 分。该提升表明, IBID 能够更稳定地将检索到的生物机理证据转化为可复用的设计约束,并在生成过程中维持“需求—生物原理—工程实现”的一致性。其原因在于,双粒度知识库检索同时提供细粒度的关键属性证据与粗粒度的摘要性语义框架,使模型在组织方案时既具备具体可引用的机制细节,又具备全局逻辑的上位约束,从而显著降低了“仅表层类比”或“证据使用不充分”导致的匹配偏差。在与先进 RAG 方法的对比中, IBID 依然保持领先。实验数据表明,尽管 GraphRAG 和 RAPTOR 较 Naïve RAG 在专业质量上有明显改善,但仍逊于 IBID。以 GPT-4o 为主干模型时, IBID 在 F&S (4.43) 和 FuncMatch (4.61) 上均优于 GraphRAG (4.17, 4.35) 与 RAPTOR (4.11, 4.60) 的最高表现;其中, FuncMatch 的优势主要体现为生成方案对目标功能的对齐更为准确。相较于仅依赖通用检索或结构化链路的基线方法, IBID 在检索阶段通过“细粒度证据—粗粒度摘要”的联合约束,能够更精确地筛选与需求功能直接相关的生物特性与作用机制,并在生成阶段以结构化大纲对功能链条进行显式组织,从而减少了“形态相似但功能不匹配”或“功能描述泛化”的情况。尤其在衡量深度仿生逻辑的 M&C 指标上, IBID-GPT-4o (4.61) 较 GraphRAG (3.46) 和 RAPTOR (4.02) 分别高出 1.15 分和 0.59 分。该结果说明, IBID 不仅能够提供与需求相关的证据,还能促进模型对“机理—结构—行为”关系的表达。相比之下,部分基线方法虽能改善检索覆盖,但在机理解释与工程映射的表达中仍容易出现证据断裂或推理跳跃,导致 M&C 得分受限。此外, IBID-GPT-4o 在内容相关性上达到 4.93,已超过 AskNature 条目 (4.78)。这一结果证实, IBID 在保持高相关性和连贯性的同时,增强了生成方案的专业匹配度。F&S 指标的提升反映了 IBID 对可行性与可持续性约束的处理更为充分。IBID 在生成过程中引入多专家会商与建议表机制,使方案能够在材料选择、结构可制造性、成本与应用环境等维度得到补充与纠偏,从而减少“可实施细节缺失”与“工程约束考虑不足”的问题。

值得注意的是,即便使用 Qwen2.5-7B 等较小规模模型, IBID 的整体生成质量在多项指标上已超越 GPT-4o 和 DeepSeek-Chat 等通用大模型基线。通过微调后的 IBID-Qwen2.5-7B* 在 FuncMatch (4.52) 和 M&C (4.77) 上均优于未经微调的版本,证明了利用特定仿生语料进行领域微调能有效提升模型在该领域的表征能力。该结果表明,领域微调增强了模型对仿生术语、机理描述模式以及功能映射表达的稳定性,使其更容易在生成中调用高质量的领域知识,从而在 FuncMatch 与 M&C 等强调专业质量的指标上获得持续增益。此

外, Only-LLM 虽然在连贯性指标上表现出色, 但在相关性上普遍偏低; 这是由于 LLM 在一次性生成文本时虽能维持上下文衔接, 但极易出现主题偏离和“幻觉”现象。综上所述, IBID 无论在检索准确性还是方案生成质量上, 优于现有的先进 RAG 框架及 Only-LLM 基线。

除生成质量外, 本文进一步统计不同方法在仿生方案生成任务中的端到端推理效率, 以衡量其工程可用性。我们在统一模型 Qwen2.5-7B 的条件下, 记录 100 个 BioP 从输入设计需求到输出完整 BID 方案文本的总耗时, 并对 100 个总耗时取平均值并向上取整, 结果如表 7 所示。可以观察到, RAPTOR 与 GraphRAG 的平均耗时分别为 57 s 与 65 s, 而 IBID 的平均耗时为 117 s。IBID 相比两种对比方法耗时更高, 主要原因在于其在同一主干模型上引入了更完整的生成链路, 包括双粒度知识检索、结构化大纲驱动写作以及多专家迭代评审等步骤, 从而带来额外的模型调用与推理开销; 但结合前文生成质量对比结果, IBID 在付出一定时间成本的同时显著提升了方案的专业质量与结构质量, 对于仿生设计方案而言, 通过一定的时间成本换取高质量方案通常是可接受的。

表 7 不同方法的端到端推理效率对比

Table 7 Comparison of end-to-end inference efficiency across different methods

方法	平均耗时/s	相对耗时/s
RAPTOR	57	1
GraphRAG	65	1.14
IBID	117	2.05

4.4 消融实验

为进一步验证 IBID 各关键模块对整体性能提升的作用, 本文进行了消融实验, 使用 IBID-Qwen2.5-7B*, 分别在检索阶段和生成阶段逐步去除或简化特定模块, 以观察其对系统性能产生的影响。

在检索阶段, 本文设计了双粒度知识库消融实验如表 8 所示, 以验证 Micro-level 与 Macro-level 知识库的作用。具体实验设置为: (1) 去除 Macro-level 知识库(记作 w/o Macro), 只保留细粒度的 Micro-level 文本块进行检索; (2) 去除 Micro-level 知识库(记作 w/o Micro), 仅利用聚类生成的 Macro-level 摘要知识库进行检索。从实验结果中可以观察到, 完整的双粒度知识库系统在三个检索指标上 ($P@1 = 0.785$, $P@3 = 0.465$, $MAP = 0.578$) 均优于仅使用单一粒度的消融版本, 其中去除 Macro-level 后系统性能明显下降, 尤其在 $P@3$ 和 MAP 上分别降低至 0.401 和 0.499, 表明聚类摘要在捕获全局语义、提高检索排序准确性方面的显著作用, 但与 Naïve RAG 相比, 去除 Macro-level 后系统性能

仍有提升, 从而体现了评分融合机制的重要性; 而去除 Micro-level 时性能同样有所降低 ($P@1 = 0.743$, $P@3 = 0.449$, $MAP = 0.508$), 但相比去除 Macro-level 影响相对较小, 说明细粒度知识块在补充局部细节、提供更精准的知识匹配上同样必不可少。这一实验明确了 Macro-level 和 Micro-level 之间的互补效应与各自的重要性, 进一步验证了本文所设计的双粒度检索策略的有效性。

表 8 知识库消融实验

Table 8 Knowledge base ablation experiments

	P@1	P@3	MAP
IBID	0.785	0.465	0.578
w/o Macro	0.713	0.401	0.499
w/o Micro	0.743	0.449	0.508

在生成阶段, 本文进一步设计了四项消融实验如图 3 所示, 以验证多专家评审机制及结构化大纲生成对最终仿生设计质量的贡献。这四项设置分别是: (1) 去除普通专家(w/o 普通专家), 即取消多专家的评审反馈, 仅用高级专家进行仿生设计方案的筛选; (2) 去除高级专家(w/o 高级专家), 不设置高级专家最终裁定环节; (3) 去除结构化大纲(w/o 大纲), 检索到的知识直接以原始形式拼接输入模型; (4) 去除建议表(w/o 建议表), 取消最后阶段的设计细节完善和优化过程。从生成质量分析结果来看, IBID 在所有五个维度(F&S、FuncMatch、M&C、相关性和连贯性)均表现最优, 各消融版本均出现不同程度的性能下降。具体而言, 去除结构化大纲对相关性和连贯性的负面影响最为显著, 表明结构化大纲在组织生成内容的逻辑性和专业相关性方面发挥了核心作用; 去除多专家评审反馈机制(w/o 普通专家和 w/o 高级专家)则明显降低了 FuncMatch、M&C 和 F&S 指标, 揭示了多专家机制在精确评估与深化仿生设计方案的细节、确保方案的专业质量方面的重要价值; 去除建议表后的指标也有显著降低, 证实了建议参考环节对提升方案完整性与可行性的必要性。鉴于 Macro-level 知识库由聚类簇内的多文本块凝练生成簇级摘要并用于后续检索融合, 其摘要质量将直接影响宏观检索的可靠性与下游生成效果。为此, 本文在消融实验中对 Macro-level 簇级摘要引入通用自动评价指标, 从事实一致性与信息覆盖两个角度进行了补充量化分析。

在事实一致性方面, 本文采用 SummaCConv 与 QAFactEval 两类互补指标对簇级摘要的可支持性进行评估。SummaCConv 基于自然语言推断(NLI)机制衡量摘要陈述是否被证据文本蕴含支持; QAFactEval 则采用“从摘要抽取信息点并生成问题—在证据文本中回答—比较答案一致性”的 QA 评价范式, 对摘要

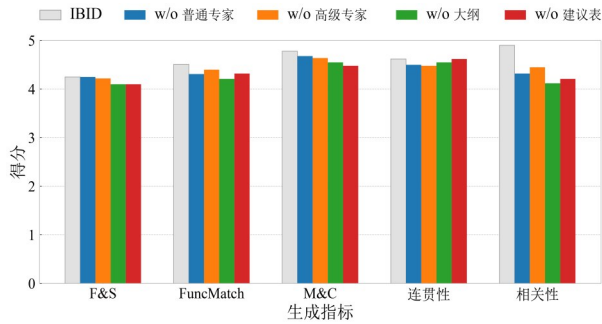


图3 生成阶段消融实验

Figure 3 Ablation experiments in the generation stage

进行更细粒度的核验。两者得分越高,表明摘要越能被源证据支持,潜在的无依据扩写与事实偏移风险越低。

在信息覆盖方面,为强调簇级摘要对簇内关键知识点的保留程度,本文采用 ROUGE-L (recall) 与 BERTScore (recall) 进行评估。ROUGE-L (recall) 基于最长公共子序列 (LCS) 刻画对参考文本的覆盖率召回;BERTScore (recall) 基于上下文语义嵌入的软匹配召回,可在同义改写场景下更稳健地反映语义覆盖程度。两者得分越高,表明簇级摘要在信息压缩的同时能够更充分地覆盖簇内核心内容。

评价设置如下:对每个聚类簇的 Macro-level 摘要 S_c ,将簇内 Micro-level 文本块按原顺序拼接得到证据文本 D_c ,用于计算 SummaCConv 与 QAFactEval;同时,将簇内各 Strategy 条目中可作为“人工摘要/概述”的字段(如 Introduction)拼接为参考摘要 R_c ,用于计算 ROUGE-L (recall) 与 BERTScore (recall)。最终对所有簇的得分取平均并统计分位数,以获得 Macro-level 摘要质量的整体量化结果,并与不同消融设置下的检索与生成性能变化进行对照分析。

如表 9 所示,Macro-level 聚类摘要在事实一致性评价中取得了 SummaCConv = 0.47、QAFactEval = 0.58 的结果,表明簇级摘要整体陈述与簇内证据文本具有较好的可支持性。具体而言,SummaCConv 的得分说明摘要句级语义与证据之间总体一致,未出现明显的大范围语义冲突;QAFactEval 的得分进一步表明,从摘要中抽取的信息点在证据文本中大多具有可回答依据,摘要中的关键表述在问答核验框架下具备较好的一致性。综合来看,Macro-level 摘要能够较可靠地承载簇内核心语义,在宏观检索与后续融合时具备较好的稳定性。

如表 10 所示,Macro-level 聚类摘要在信息覆盖评价中取得 ROUGE-L (Recall) = 0.33、BERTScore (Recall) = 0.886 的结果,说明簇级摘要在压缩簇内长文本信息的同时,能保持对关键内容的有效覆盖。其

表 9 Macro-level 聚类摘要质量验证(事实一致性)

Table 9 Macro-level cluster summary quality verification (factual consistency)

	SummaCConv	QAFactEval
Macro-level 摘要 (事实一致性)	0.47	0.58

表 10 Macro-level 聚类摘要质量验证(信息覆盖)

Table 10 Macro-level cluster summary quality verification (information coverage)

	ROUGE-L (Recall)	BERTScore (Recall)
Macro-level 摘要 (信息覆盖)	0.33	0.886

中,ROUGE-L (Recall) 反映摘要在表面对参考信息点具有覆盖能力;BERTScore (Recall) 则表明摘要在语义层面能够较充分地对齐参考摘要的核心含义,即使存在同义改写或措辞变化,仍能保持较高的语义覆盖度。总体而言,这组结果表明 Macro-level 摘要在“信息压缩—语义保留”之间达到了较为合理的平衡,为后续双粒度检索框架提供了有效的高层语义入口。

综合上述消融实验结果可知,本文提出的双粒度知识库构建、结构化大纲生成和多专家会商机制均在 IBID 框架中起到了重要的作用,各模块互补协作,提升了仿生设计方案的全面性、专业性与可靠性。

4.5 参数敏感性分析

本节对模型中 2 个关键超参数,Macro-level 知识块的贡献权重值 α 以及最低相关性阈值 τ 进行实验分析,阐述不同的超参数值对检索性能的影响。

如图 4 所示,本文考察了 Macro-level 知识块的贡献权重值 α 对检索性能的影响。实验中将 α 的值从 1 逐步提高到 5,观察到当 α 在较低范围内(1~2)逐步增加时,各检索指标(P@1、P@3 与 MAP)均表现出逐步提高的趋势,MAP 值从 0.514 提升到 0.578,P@1 指标亦达到最优点 0.785。这一趋势表明,适当增强

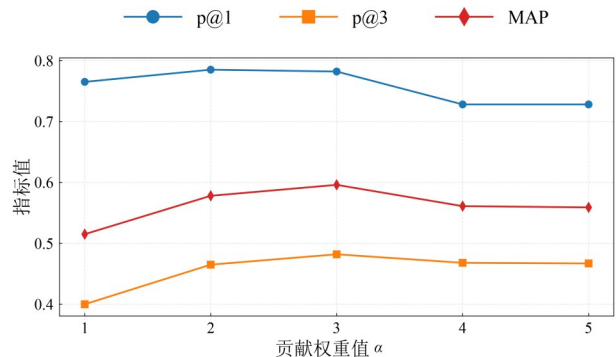


图 4 贡献权重值 α 的影响

Figure 4 Impact of contribution weight value α

Macro-level 知识块的贡献权重能够有效提高检索的准确性和全面性。然而,当 α 继续增加后,各指标逐渐出现明显的性能下降趋势。分析其原因可能在于,过高的 α 值将过度强调 Macro-level 粒度的聚类摘要知识块,使得部分与 BioP 相关性相对较低的生物策略也被赋予过高的的重要性,进而引入检索噪声。因此,Macro-level 知识库的权重应设定在一个适中的范围内,以有效平衡宏观语义信息与细节精准性,避免因权重过高而造成的信息干扰。

如图 5 所示,在实验中,当 τ 从 1 提升至 2 时,各项检索指标均有提升,特别是 P@1 达到峰值 0.785。这表明适当提高最低相关性阈值能够有效过滤掉一些弱相关或不相关的候选策略,从而提升检索结果的精准度。然而,当阈值进一步提高超过 2 时,模型的检索准确率呈明显下降趋势,特别是 P@3 和 MAP 指标迅速降低。经过进一步的分析发现,这是由于更高的相关性阈值限制了较多潜在的相关知识进入候选集合,使得检索到的生物策略变得过于稀少甚至为空,进而降低了整体的召回率与排序效果。因此,最低相关性阈值 τ 的选取需谨慎,以在过滤无关策略的同时,避免遗漏有价值的相关策略,从而实现召回与精度之间的合理平衡。

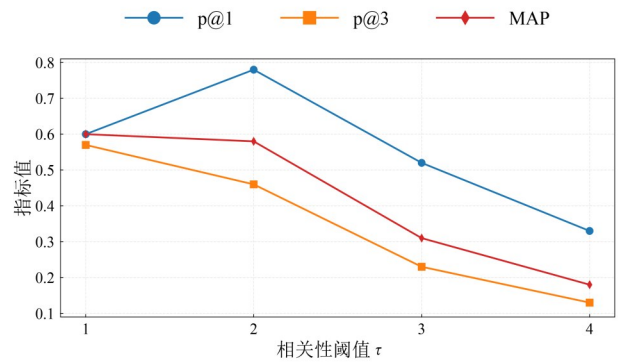


图 5 最低相关性阈值 τ 的影响

Figure 5 Impact of minimum relevance threshold

综合上述两项超参数分析可以得出结论:在实际应用中,Macro-level 知识块贡献权重 α 与最低相关性阈值 τ 均对 IBID 检索模块的性能具有较高的敏感性,且各参数存在较明显的最优范围。针对本文的数据集与任务, α 推荐取值为 2 或 3, τ 推荐取值为 2,以实现整体检索性能的最优化。

4.6 案例分析

为进一步验证本文所提出的 IBID 方法的有效性,与优势,本文选取了一项实际 BioP 作为示例,如图 6 所示。该 BioP 要求设计一种系统,以有效地选择、部

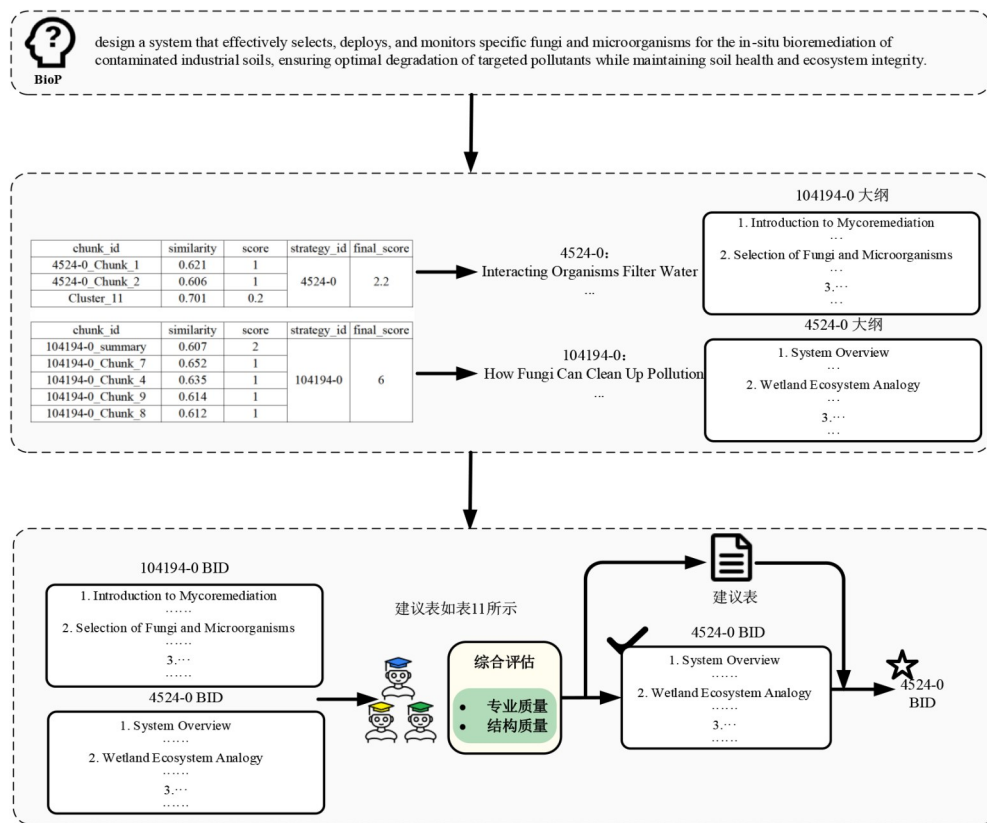


图 6 IBID 实例

Figure 6 IBID case study

署并监测特定的真菌和微生物,用于原位修复受工业污染的土壤,并确保在降解目标污染物的同时,维护土壤健康与生态系统的完整性。

在案例实施过程中,初始生成的两个候选设计方案通过专家会商环节进行了详细评审与多轮反馈迭代,如表 11 所示(表 11 中用 e 和 f 代替图 6 中的

104194-0 BID 和 4524-0 BID,完整的专家会商可见附录表 1)。专家们对每轮生成的方案分别从技术结构、功能实现、可行性及整体质量等多方面进行了评价和优化建议。经过多轮的交叉评估和会商以及高级专家的总结,最终确定了最佳设计方案。

表 11 专家会商记录表

Table 11 Expert consultation records

轮次	专家	最佳方案	方案	得分	建议
0	A	e	f	3	The design shows some potential in terms of feasibility and cost, but it lacks ...
			e	5	This design excels in all aspects, with a robust and scalable system for ...
	B	e	f	3	The design demonstrates a good understanding of the bioremediation process ...
			e	4	This design effectively addresses the core requirements. ...
	C	f	f	4	Design f meets the core engineering requirements well. ...
			e	3	Design e has fair compliance with the core engineering requirements. ...
1	A	e	f	4	The design has made significant improvements since the last round. ...
			e	5	This design continues to excel, with a robust and scalable system. ...
	B	f	f	4	The design shows a good balance between functional performance, ...
			e	3	While this design excels in many aspects, it requires more refinement to ...
	C	f	f	4	The design meets the core requirements effectively, with a strong selection ...
			e	4	The design is comprehensive and robust, but there is room for improvement ...
2	A	e	e	5	This design continues to excel, with a robust and scalable system ...
			f	4	I agree with the other experts that this design has made significant ...
	B	f	f	4	The design demonstrates a good balance between functional performance ...
			e	3	While this design excels in many aspects, it requires more refinement ...
	C	f	f	4	The design maintains its strong performance, with a well-rounded approach ...
			e	4	The design's excellence is recognized, but the feedback suggests that ...
3	A	f	e	4	I've revised my rating to reflect the feedback from other experts, ...
			f	5	After reviewing the feedback from experts B and C, I'm impressed by the ...
	B	e	e	4	This design demonstrates a robust and scalable system that effectively ...
			f	4	The design demonstrates a good balance between functional performance, ...
	C	e	e	4	The design remains strong, but improvements in cost and sustainability are ...
			f	4	This design is well-rounded and functional, with room for further ...
4	Auth	f	e	4	This design demonstrates a robust system that addresses the requirement. ...
			f	5	After reviewing the feedback from experts B and C, I'm impressed by the ...

本文对 IBID 方法生成的仿生设计方案与 GPT-4o 生成的仿生设计方案分别在“仿生”或“模仿”的阐述和应用程度对内容的侧重点和具体性(以系统组件/设计方法为例)进行了对比分析,如表 12 和表 13 所示。

在表 12 中 Design 1 从目标设定之初就非常明确地将“仿生”作为其核心方法论和灵感来源。它清晰地指出了其模仿的对象——湿地生态系统中植物、细菌与物理环境间的复杂互动。不仅仅是概念上的启发,Design 1 强调要将这种互动模式作为其土壤修复系统的“模型(Model)”,并且明确提出要“模仿(Mimic)”湿地中生物间的相互依赖和协同关系。在

其结论中,再次强调了通过“模仿(Mimicking)”实现了有效的生物修复。这种仿生思路是贯穿始终的,且是具体和主动的。

而 Design 2 虽然也深入探讨了湿地的生态功能和生物间的协同作用(例如系统协同性与互补性),但其表述更侧重于“理解(Understanding)”这些已存在的自然机制,以及如何将普适的“生态学原理(Ecological principles)”与“工程实践(Engineering practices)”相结合。它并没有像 Design 1 那样,提出要以一个特定的自然系统(如湿地中的特定互动)作为直接的“模仿对象”来构建一个全新的设计。其仿生理念是间接的,更多的是从自然中学习普适规律,而非

表 12 两篇 Design 对“仿生”或“模仿”的阐述和应用程度

Table 12 Comparison of two designs on the elaboration and application of bio-inspiration and mimicry

IBID(Design 1)	解读	GPT-4o(Design 2)	解读
Objective : ...inspired by the interactions between wetland plants, bacteria, and the physical environment.	目标直指设计灵感来源于湿地中生物与环境的相互作用	Design 2 中没有直接使用“inspired by”或“mimic”来描述其核心设计思路,其更多的是分析和阐述湿地已有的功能和过程。	未直接声明受特定自然现象“启发”或进行“模仿”,侧重分析湿地已有功能。
Wetland Ecosystem Analogy : Use the interactions between wetland plants, bacteria, and the physical environment as a model for the soil bioremediation system.	明确以湿地生态互动为修复系统的“模型”。	By aligning engineering practices with ecological principles, we can create systems that not only purify water but also restore and enhance the surrounding environment, thus demonstrating that the intersection of engineering and ecology can yield remarkable outcomes for water management.	强调工程实践与普适性生态学原理的结合,而非模仿特定模型。
Interdependence and Synergy : Mimic the interdependence and synergistic relationships between different organisms in a wetland.	设计原则明确包含“模仿(Mimic)”湿地生物间的特定关系。	System Synergies and Complementarity Understanding the interconnected roles of different organisms within wetland ecosystems reveals a rich tapestry of system synergies and complementarity that enhances remediation efficiency.	侧重“理解”湿地系统中的协同性以提升修复效率,而非主动模仿。

表 13 两篇 Design 对内容的侧重点和具体性(以系统组件/设计方法为例)

Table 13 Comparison of two designs on content focus and specificity (system components/design methods)

IBID(Design 1)	解读	GPT-4o(Design 2)	解读
Key Components : Fungi, microorganisms, and soil environment.	明确列出系统的“关键组成部分”。	## II. Exploratory Chapter 1: Interdisciplinary Perspectives on Wetland Ecosystem Functionality	章节标题体现其内容更侧重于对湿地功能进行多角度、概括性的理论探索。
For example, **Phanerochaete chrysosporium** for hydrocarbons and **Pseudomonas putida** for heavy metals.	提供具体的生物选择案例,如特定真菌和细菌及其应用。	### A. Identifying Contaminants of Concern ### B. Selection of Target Organisms ... Establishing criteria for selecting these organisms requires a nuanced understanding of their ...	生物修复设计部分侧重通用步骤,如“识别污染物”和确立生物选择的“标准(criteria)”。
This involves creating a controlled environment that mimics the conditions found in a wetland, such as maintaining a consistent moisture level and providing essential nutrients.	强调环境营造需“模仿湿地条件”,并举例说明(如湿度、营养)。	### C. Deployment Strategies In-situ bioremediation allows for the treatment of contaminated soils directly at the site of pollution ... while ex-situ methods involve the removal of contaminated materials...	讨论如原位/异位修复等生物修复领域的通用技术方法分类。

直接仿生特定模式。

在表 13 中, Design 1 的内容高度聚焦于一个具体的、可实施的生物修复系统的构建。它明确列出了系统的“关键组成部分(Key Components)”,包括真菌、微生物和土壤环境。在生物体的选择上,它不仅提出了选择原则(能降解特定污染物),还给出了具体的生物名称作为例子(如白腐菌用于碳氢化合物,假单胞菌用于重金属)。更重要的是,它强调要为这些选定的生物创造一个适宜的“土壤环境(Soil Environment)”,并且这个环境是要“模仿湿地中的条件(Mimics the conditions found in a wetland)”,例如保持一致的湿度和提供必需营养。这种具体性体现在从宏观的系统组件到微观的环境参数控制。

而 Design 2 的内容则展现出更强的广度和理论

性。其章节标题如“跨学科视角下的湿地生态系统功能性”和“湿地系统中生物体之间复杂的相互作用”就表明了其探索的范围较广。在其第五部分“为污染土壤设计生物修复系统”中,虽然也涉及了污染物识别、目标生物选择和部署策略,但这些更多的是在阐述生物修复的一般性步骤和需要考虑的普适性标准(例如,选择生物需理解其代谢能力、环境耐受性等),而非针对一个特定仿生设计方案的具体细节。它讨论的是“方法论”层面,而不是像 Design 1 那样详细描述一个“具体产品”的构成和运作。

分析可见, IBID 生成的设计方案在“仿生”与“模仿”概念的具体应用上更加明确且主动,提供了具体、可实施的设计细节和系统组件选型。相比之下, GPT-4o 的方案则更倾向于生态理论与普遍方法的阐

述,未能直接体现出具体的仿生机制与设计细节。这种差异充分体现了 IBID 方法在实际仿生设计场景中针对性和操作性的优势,更有利于具体仿生工程实施和创新应用。

5 结束语

本文针对现有仿生设计方案生成方法在知识检索精度、生成针对性和专业性方面的不足,提出了融合双粒度知识库的智能仿生设计方法 IBID。该方法通过构建双粒度知识库、设计分层加权检索策略以及引入多 LLM 专家会商决策机制,有效提升了生物策略检索的召回率与准确性,并增强了生成方案的结构完整性、功能匹配能力和跨领域迁移能力。实验结果表明,IBID 在仿生设计指标和通用质量指标上均优于传统 RAG 方法和仅依赖 LLM 的生成方案。消融实验与参数敏感性分析进一步验证了各模块设置的有效性。未来可进一步结合仿生知识图谱,继续探索更精细的知识表示方法、更智能的多专家协作机制,以及在更复杂任务和多模态场景中的应用。

参考文献

- [1] Budholiya S, Bhat A, Raj S A, et al. State of the art review about bio-inspired design and applications: An aerospace perspective[J]. *Applied Sciences*, 2021, 11(11): 5054.
- [2] Domel A G, Saadat M, Weaver J C, et al. Shark skin-inspired designs that improve aerodynamic performance[J]. *Journal of the Royal Society Interface*, 2018, 15(139): 20170828.
- [3] Latthe S S, Terashima C, Nakata K, et al. Superhydrophobic surfaces developed by mimicking hierarchical surface morphology of lotus leaf[J]. *Molecules*, 2014, 19(4): 4256-4283.
- [4] Kapoor R T, Rafatullah M, Qamar M, et al. Review on recent developments in bioinspired-materials for sustainable energy and environmental applications[J]. *Sustainability*, 2022, 14(24): 16931.
- [5] Ha N S, Lu Guoxing. A review of recent research on bio-inspired structures and materials for energy absorption applications[J]. *Composites Part B: Engineering*, 2020, 181: 107496.
- [6] Dressler F, Akan O B. A survey on bio-inspired networking[J]. *Computer Networks*, 2010, 54(6): 881-900.
- [7] Helms M, Vattam S S, Goel A K. Biologically inspired design: Process and products[J]. *Design Studies*, 2009, 30(5): 606-622.
- [8] Bar-Cohen Y. *Biomimetics: Nature-based innovation*[M]. Boca Raton: CRC Press, 2016.
- [9] Zhang Taiyao. Biologically-inspired design in engineering: Current perspective on biomimicry applications[C]//*Proceedings of 2024 International Conference on Mechanics, Electronics Engineering and Automation*. [S. l.]: Atlantis Press, 2024: 637-645.
- [10] Wegst U G K, Bai Hao, Saiz E, et al. Bioinspired structural materials[J]. *Nature Materials*, 2015, 14(1): 23-36.
- [11] Ng L, Elgar M A, Stuart-Fox D. From bioinspired to bioinformed: Benefits of greater engagement from biologists[J]. *Frontiers in Ecology and Evolution*, 2021, 9: 790270.
- [12] Shu L H, Ueda K, Chiu I, et al. Biologically inspired design[J]. *CIRP Annals*, 2011, 60(2): 673-693.
- [13] Cohen Y H, Reich Y, Greenberg S. Biomimetics: Structure-function patterns approach[J]. *Journal of Mechanical Design*, 2014, 136(11): 111108.
- [14] Liu N F, Lin K, Hewitt J, et al. Lost in the middle: How language models use long contexts[J]. *Transactions of the Association for Computational Linguistics*, 2024, 12: 157-173.
- [15] Shi F, Chen X, Misra K, et al. Large language models can be easily distracted by irrelevant context[C]//*International Conference on Machine Learning*. PMLR, 2023: 31210-31227.
- [16] Li D, Shao R, Xie A, et al. How long can context length of open-source LLMs truly promise [C]//*NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*. 2023.
- [17] Li J, Wang M, Zheng Z, et al. Loogle: Can long-context language models understand long contexts?[C]//*Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2024: 16304-16333.
- [18] Wang Yidong, Guo Qi, Yao Wenjin, et al. AutoSurvey: Large language models can automatically write surveys[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. New York: Curran Associates Inc., 2024: 3655.
- [19] Shao Yijia, Jiang Yucheng, Kanell T, et al. Assisting in writing Wikipedia-like articles from scratch with large language models[C]//*Proceedings of 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. [S. l.]: Association for Computational Linguistics, 2024.
- [20] Kandpal N, Deng H, Roberts A, et al. Large language

- models struggle to learn long-tail knowledge[C]//International Conference on Machine Learning. PMLR, 2023: 15696-15707.
- [21] Gao Yunfan, Xiong Yun, Gao Xinyu, et al. Retrieval-augmented generation for large language models: A survey[PP/OL]. V5.arXiv (2024-03-27)[2025-05-26]. <https://arxiv.org/abs/2312.10997>.
- [22] Kang H B, Lin D C E, Martelaro N, et al. BioSpark: An end-to-end generative system for biological-analogical inspirations and ideation[C]//Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. New York: ACM, 2024: 61.
- [23] Luu R K, Buehler M J. BioinspiredLLM: Conversational large language model for the mechanics of biological and bio-inspired materials[J]. *Advanced Science*, 2024, 11(10): 2306724.
- [24] Toukmaji C, Tee A. Retrieval-augmented generation and LLM agents for biomimicry design solutions[J]. *Proceedings of the AAAI Symposium Series*, 2024, 3(1): 273-278.
- [25] Chen Liuqing, Jiang Zhaojun, Xia Duowei, et al. BIDTrainer: An LLMs-driven education tool for enhancing the understanding and reasoning in bio-inspired design[C]//Proceedings of 2024 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2024: 676.
- [26] Ram O, Levine Y, Dalmedigos I, et al. In-context retrieval-augmented language models[J]. *Transactions of the Association for Computational Linguistics*, 2023, 11: 1316-1331.
- [27] Izacard G, Lewis P, Lomeli M, et al. Atlas: Few-shot learning with retrieval augmented language models[J]. *The Journal of Machine Learning Research*, 2023, 24(1): 251.
- [28] Gao T, Yen H, Yu J, et al. Enabling large language models to generate text with citations[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023: 6465-6488.
- [29] Huang Lei, Yu Weijiang, Ma Weitao, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[J]. *ACM Transactions on Information Systems*, 2025, 43(2): 42.
- [30] Dong Yuxin, Wang Shuo, Zheng Hongye, et al. Advanced RAG models with graph structures: Optimizing complex knowledge reasoning and text generation[C]//Proceedings of the 2024 5th International Symposium on Computer Engineering and Intelligent Communications. Piscataway: IEEE, 2024: 626-630.
- [31] Ma X, Gong Y, He P, et al. Query rewriting in retrieval-augmented large language models[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023: 5303-5315.
- [32] Gao L, Ma X, Lin J, et al. Precise zero-shot dense retrieval without relevance labels, 2022[PP/OL]. V1.arXiv (2022-12-10)[2025-05-26]. <https://arxiv.org/abs/2212.10496>.
- [33] Shao Zhihong, Gong Yeyun, Shen Yelong, et al. Enhancing retrieval-augmented large language models with iterative retrieval-generation synergy[C]//Findings of the Association for Computational Linguistics: EMNLP 2023. [S.l.]: Association for Computational Linguistics, 2023: 9248-9274.
- [34] Yu Wenhao, Iter D, Wang Shuohang, et al. Generate rather than retrieve: Large language models are strong context generators[C/OL]//Proceedings of the 11th International Conference on Learning Representations, 2023. <https://dblp.org/db/conf/iclr/iclr2023.html#conf/iclr/0002IWXJ000023>.
- [35] Sarthi P, Abdullah S, Tuli A, et al. RAPTOR: Recursive abstractive processing for tree-organized retrieval[C/OL]//Proceedings of the 12th International Conference on Learning Representations, 2024. <https://dblp.org/db/conf/iclr/iclr2024.html#conf/iclr/SarthiATKGM24>.
- [36] Edge D, Trinh H, Cheng N, et al. From local to global: A graph rag approach to query-focused summarization[PP/OL]. V2.arXiv (2025-02-19)[2025-05-26]. <https://arxiv.org/abs/2404.16130>.
- [37] Deldin J M, Schuknecht M. The AskNature database: Enabling solutions in biomimetic design[M]//Goel A K, McAdams D A, Stone R B. *Biologically Inspired Design: Computational Methods and Tools*. London: Springer, 2014: 17-27.
- [38] Kaufman L, Rousseeuw P J. *Finding Groups in Data: An Introduction to Cluster Analysis*[M]. New York: Wiley, 1990.
- [39] 郭哲, 张智博, 周炜杰, 等. 融合引导注意力的中文长文本摘要生成[J]. *电子学报*, 2024, 52(12): 3914-3930.
Guo Zhe, Zhang Zhibo, Zhou Weijie, et al. Chinese long text summarization with guided attention[J]. *Acta Electronica Sinica*, 2024, 52(12): 3914-3930. (in Chinese)
- [40] Kim S, Suk J, Longpre S, et al. Prometheus 2: An open source language model specialized in evaluating other language models[C]//Proceedings of 2024 Conference on Empirical Methods in Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2024: 4334-4353.

附录 A

相关内容见《电子学报》官网: <https://dx.doi.org/10.12263/DZXB.20250456FJ>。

作者简介



张晓明 男,1979年12月出生于山东省济南市。现为安徽大学物质科学与信息技术研究院教授、博士生导师。主要研究方向为知识图谱与智能决策。

E-mail: zxm@ahu.edu.cn



刘成响 男,2002年8月出生于安徽省安庆市。现为安徽大学物质科学与信息技术研究院人工智能专业学生。主要研究方向为大模型与知识图谱。

E-mail: cxbetter825@gmail.com



钟凯 男,1990年4月出生于安徽省含山县。现为安徽大学物质科学与信息技术研究院副教授。主要研究方向为复杂系统建模与智能运维、人工智能方法及其工程应用。

E-mail: kaizhong0402@ahu.edu.cn

附录 B

本文开展仿生方案生成的相关提示词和详细示例请见: <https://github.com/lllcbst/Appendix>。



谢锐 男,1968年出生于青海省湟源县。现为青海理工学院教务处负责人、副教授。主要研究方向为网络动力学、概率论与数理统计、高校教学管理。

E-mail: kxie@qhut.edu.cn



钱护磊 男,2000年10月出生于安徽省无为市。现为安徽大学物质科学与信息技术研究院人工智能专业学生。主要研究方向为大模型与知识图谱。

E-mail: qianhulei1@gmail.com