Environmental Science & Technology

pubs.acs.org/est

Article

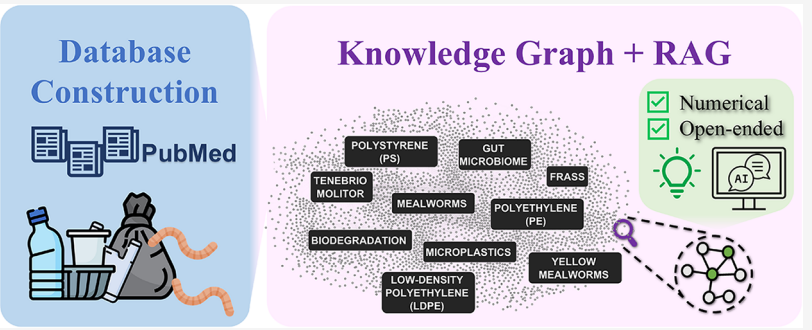
Leveraging Retrieval-Augmented Generation to Accelerate Discoveries on Mealworm Larvae and Plastic Degradation

Bingdi Liu,[‖] Wenyu Li,*[,‖] Runyu Zhao, Ross R. Klauer, Alex Hansen, Nathan Miller, Yixin Chen, Mark A. Blenner,* and Yinjie J. Tang*

Cite This: *Environ. Sci. Technol.* 2025, 59, 27437−27448

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Large language models (LLMs) are transforming broad research areas, yet concerns about their trustworthiness remain. This study explored the use of Retrieval-Augmented Generation (RAG) to improve LLMs' knowledge extraction in the field of mealworm-mediated plastic degradation. We integrated publications up to June 2024 (75 papers) to evaluate the model performance using a curated data set of 100 queries. GraphRAG, LightRAG, and a traditional RAG were examined with five LLM models (GPT-4o, GPT-5, Deepseek-V3.1, Qwen-plus, and Llama-3.3). Our results reveal that LightRAG improved LLMs the most in information extraction. Specifically, for quantitative information extraction, the best performing RAG + LLM pipeline achieves over 92% accuracy. Meanwhile, for open-ended queries, LightRAG + Llama answers the questions with the best balance of precision and information coverage. Moreover, empirical results validated the answers about the mealworm gut microbiome composition and plastic deconstruction patterns through the LightRAG + Llama pipeline. In designing plastic biodegradation experiments, the original LLMs outperformed RAG-trained LLMs. The expandable nature of RAG enables timely updates to the knowledge base. This study demonstrates a reliable application of advanced LLMs in the emerging environmental science field. Our findings identify challenges, such as conflict handling, and guide future research in scientific artificial intelligence.

KEYWORDS: *plastic degradation, Tenebrio molitor, gut microbiome, retrieval-augmented generation, large language model*

1. INTRODUCTION

The global production of plastics represents an escalating environmental challenge. In 2018, the United States generated 35.7 million tons of plastic waste, with a recycling rate of only 8.7%.[1] Annually, plastic pollution kills approximately 100,000 marine mammals.[2] Plastic pollution persists mainly because the common polymers, mostly nonhydrolyzable plastics, have chemically inert carbon−carbon backbones that resist microbial degradation.[3] Traditional waste management strategies, such as landfilling and incineration, cannot mitigate the situation and may even produce more pollutants.[4,5] Therefore, researchers keep looking for more environmentally benign approaches. Enzymes and microbes capable of degrading plastics are of increasing interest.[6] However, very few of them have been clearly characterized to deconstruct nonhydrolyzable plastics without pretreatment like UV exposure.[7] Insects, such as the mealworms (*Tenebrio molitor*),[8,9] the waxworms (*Galleria mellonella*),[10] and superworms (*Zophobas morio*),[11] have been reported to feed on a variety of plastics. Among them, mealworms are favored by researchers due to their resilience and ease of mass rearing,[12,13] offering potential for developing engineered systems to upcycle plastic waste.[14] Currently, the exact mechanisms of mealworm-based degradation are still under investigation. Their gut microbiome could play a crucial role in degrading plastics[15] with certain key bacteria like *Klebsiella*.[16] Various approaches have assessed the extent of plastics degradation,[17,18] such as weight loss

Received: October 9, 2025
Revised: November 18, 2025
Accepted: November 19, 2025
Published: December 9, 2025

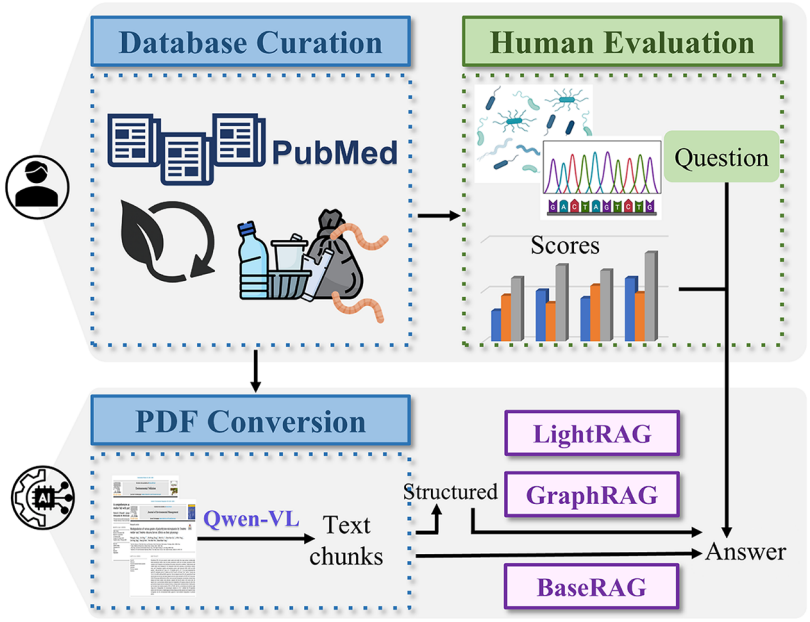


Figure 1. Workflow of our study. After constructing the paper database, the multimodal converter was used to transform PDF files into structured texts for GraphRAG and LightRAG. Next, knowledge graph-based RAGs were tested against our Q&A pool (including Transcriptome, FTIR, Microbiome, GC-MS, etc.) with a traditional RAG as baseline. Then, LLMs' performances for each RAG were evaluated by comparing them with the original models.

measurements, Gel Permeation Chromatography (GPC),[19] and spectroscopic techniques like Fourier Transform Infrared Spectroscopy (FTIR).[20] The use of more robust methods like isotopically labeled plastics enabled direct tracing of plastic-derived carbon into microbial biomass or CO_2.[21] Additionally, omics technologies enhanced the identification of key enzymes and metabolic pathways involved in plastic breakdown.[22,23]

Collectively, the rapid growth of the literature on plastic biodegradation contains an increasing wealth of information that requires systematic interpretation. Large Language Models (LLMs) are artificial intelligence (AI) systems trained on vast collections of text to learn statistical patterns of words and have shown impressive applications.[24,25] Rather than being explicitly programmed for each task, they generate text by predicting the most likely sequence of words based on context. This enables them to perform a variety of language-related tasks. A natural follow-up question, however, is how well these general-purpose models translate to specialized domains. Due to the complexity of plastics biodegradation, the organization of multilevel information is essential for guiding future research directions, making LLMs a promising tool to perform multilevel knowledge extraction. The impact of LLMs on environmental science involves streamlining scientists' workflows, freeing time for experimental design, analysis, and idea generation, and thereby accelerating progress on key environmental problems.[26]

Trained on broad and noisy data sets, LLMs often produce outdated or factually incorrect responses due to their inefficiency in long context reading, with a tendency to hallucinate.[27,28] Retrieval-Augmented Generation (RAG)[29] enhances the trustworthiness of LLMs by integrating up-to-date, verifiable information into their responses, ensuring that outputs are grounded in current, factual data.[30,31] It works by first searching a trusted textual database, such as a collection of papers, for relevant information and then using those retrieved facts to guide what LLMs generate. A smaller and curated corpus allows knowledge retrieval to be more domain-specific and precise while also being faster and less resource-intensive. Another advantage of RAG is improved transparency, since it is easier to trace back the source of an answer. Traditional RAG divides the input database into smaller chunks, embeds them, and stores them as vectors. At query time, the model finds the top similar text chunks and feeds them into the LLM as a context to generate the answer. Another way to process the database is by creating a knowledge network with nodes and edges representing the structured information extracted from it. Recently, RAGs like LightRAG[32] and GraphRAG[33] have been proposed to enhance and fact-check LLMs' reasoning ability through knowledge graphs.

This study (Figure 1) applied LightRAG, GraphRAG, and a traditional RAG (as baseline) on LLMs, including Deepseek-V3.1, GPT-4o, GPT-5, Llama-3.3, and Qwen-plus. Comparison of their performance against their default LLM counterparts revealed different levels of improvement. We also included a reasoning LLM, Deepseek-R1, and a PubMed-finetuned model, Biomistral.[34] To ensure a rigorous and unbiased assessment, we curated a human-evaluated set of queries to better analyze complex data, identify key factors, and propose strategies for improving the efficiency of plastic degradation processes for mealworm plastic biodegradation. The data set includes 50 quantitative questions, 50 open-ended questions, and one experiment design. Empirical validation of open-ended question answers was also performed. By publishing the model codes and guidelines, we bridge the knowledge gap between environmental science knowledge and advanced LLM applications. To the best of our knowledge, this is the first work to apply RAG + LLM pipeline in environmental science and engineering. Our findings indicate that RAG can improve the credibility of information retrieval in the environmental science field and accelerate the database construction for data-driven research.

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

Table 1. Summary of the Models Used in Comparing Different RAGs[a]

model	GPT-4o	GPT-5	DS-V3	DS-R1	Qwen-plus	Llama-3.3
initial release date	May, 2024	Aug, 2025	Aug, 2025	Jan, 2025	Jan, 2025	Dec, 2024
max context	128,000	400,000	128,000	128,000	131,072	128,000
open-sourced	no	no	yes	yes	no	yes
reasoning model	no	included	no	yes	no	no
developer	OpenAI	OpenAI	DeepSeek	DeepSeek	Alibaba	Meta

[a]The reasoning model DS-R1 is not applicable in RAG settings.

2. MATERIALS AND METHODS

2.1. Literature Search and Processing. We searched papers with both "mealworm" and the names of fossil-based plastics in the article titles and abstracts. After excluding irrelevant studies and only focusing on PS (polystyrene), PE (polyethylene), PVC (polyvinyl chloride), PP (polypropylene), PU (polyurethane), and PET (polyethylene terephthalate), we downloaded 75 open-access papers from their publisher Web sites in PDF format by June 2024 (Table S1). After querying papers, all PDFs selected were processed to extract structured content using multimodal LLMs that can read images and text in literature. Each PDF was first converted to a series of images per page. To detect key structural components such as figures and tables, we employed RapidLayout.[35] After extraction, these visual elements were then cropped and stored separately for downstream processing. To extract and format the textual content from the processed images, we used Qwen2.5-VL-7B,[36] a multimodal LLM capable of interpreting text embedded in figures and tables. A structured prompting strategy was applied to guide text extraction, ensuring the accurate conversion of image content into a Markdown format. The prompts were designed to (1) Recognize and transcribe text from scientific PDFs; (2) Ensure the input is of scientific research paper context; (3) Describe detected figures and tables; and (4) Format extracted equations using LaTeX syntax. Each image was processed through tokenization and embedding transformation followed by text generation using the Qwen2.5-VL-7B model. The extracted content was postprocessed to remove redundant text artifacts and ensure Markdown compatibility. Key postprocessing steps include removing unnecessary Markdown tags, standardizing LaTeX formatting, and structuring the extracted text. If figures or tables were detected, they were referenced in the Markdown document using the appropriate syntax. This pipeline is scalable and reproducible for knowledge extraction from PDFs by integrating layout analysis, multimodal LLM-based text extraction, and structured output generation.

2.2. Configuration and Implementation of RAG and LLMs. To ensure robustness across different architectures, we tested the RAG pipelines on multiple LLM models (Table 1), including GPT-4o, GPT-5, DeepSeek-V3.1 (DS-V3), DeepSeek-R1 (DS-R1), Qwen-plus, and Llama-3.3−70b (Llama-3.3). We used a chunk size of 1200 with a 200-token overlap and fixed the training temperature at 0.3 across all models. We used the same embedding model, text-embedding-3-large, for all of the tests. We cut off the thinking output for all reasoner models for better representation. We tested three RAG pipelines for our curated dataset.

GraphRAG builds a knowledge graph with community/hierarchical summaries, while LightRAG improves it with graph-based text indexing and a lightweight dual-level retriever. Both RAGs grow the knowledge graph paper-by-paper, adding to the existing knowledge base as a union. The hyperparameters for GraphRAG and LightRAG are provided in the supplemental file (Table S2). After knowledge graph generation, we compared different search methods for answer generation to assess the Q&A performance. GraphRAG supports local and global search, while LightRAG supports local, global, hybrid, and naïve question answering strategies. Specifically, global retrieval focuses on retrieving information based on overarching themes and connections within the knowledge graph. Local retrieval targets specific entities and their immediate contexts, retrieving detailed information closely related to the query. Hybrid retrieval combines both global and local approaches to provide a comprehensive response. Naive retrieval employs a straightforward search based on basic keyword matching.

The traditional RAG (not based on the knowledge graph) for this study was implemented by using the method of Ollama and Langchain. This process included data ingestion and loading raw .txt or .md documents, splitting them into chunks, and embedding them for future vector storing. The processed data was indexed and stored in a FAISS[37] vector database for efficient retrieval. The prompt for answering questions is "You must answer strictly from the provided context." The querying process integrated the FAISS retriever with an LLM to generate context-aware answers.

2.3. Model Performance Evaluation. We manually curated 100 questions into two sets, quantitative and open-ended, to test the model's ability to extract the quantitative information and existing knowledge from the literature corpus constructed at Section 2.1. Each query set covers diverse research aspects, including microbiome, metabolomics, transcriptomics and plastic deconstruction extent. All of the reference answers required by the following model evaluation were from relevant papers or the domain experts and it took nearly 200 h/person to curate them. Since the common conclusions across the mealworm research on plastic biodegradation were limited by variations in experimental setups, we included few global questions to test the model's ability to summarize the findings. Overall, our queries focused more on specific studies, as we want to see if models can target detailed information for a certain study across a bulk of literature. We also want to avoid the oversimplification of scientific conclusions, which can be misleading to the general audience. Although it is achievable for a human to read one paper and then find one type of information in a short time, it can be labor-extensive when coming to the same type of information across numerous papers.

After the answers from different query methods were compared (e.g., local or global search), the best answer for each query was used to evaluate the LLMs trained by GraphRAG and LightRAG pipelines. To evaluate the model performance on the quantitative query set (Table S3), we compared the numeric values given by models with our reference and computed the absolute percentage error (APE).

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

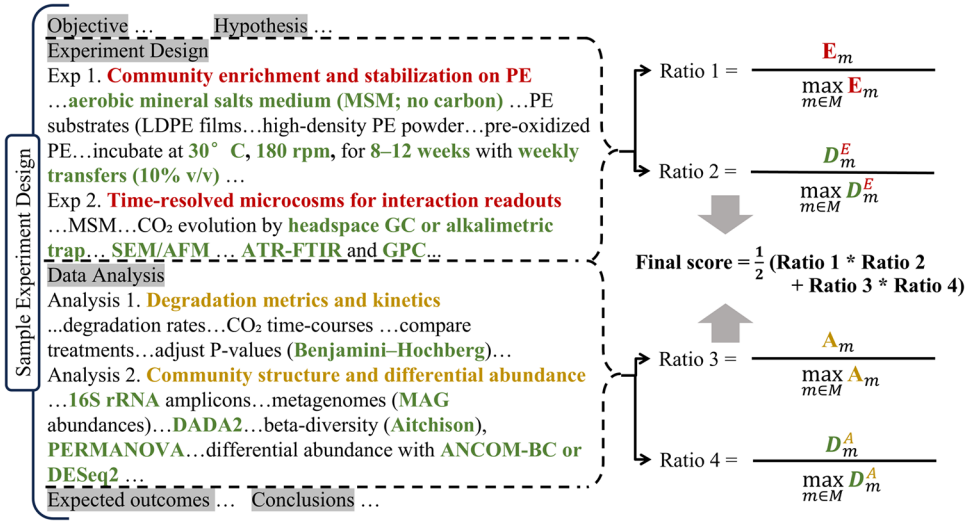


Figure 2. Structure of the model output and the scoring system. Our query instructed LLMs to generate an experiment design comprising six sections. To evaluate the model output, the numbers of experiments (red text chunks) and details (green text chunks) listed under the whole Experiment Design section were counted for each model as E_m and D_m^E. Similarly, the counts A_m and D_m^A for the Data Analysis section. Next, the maximum value for each type of counting across the set of all models (denoted by M) is found. Then, the ratio of the count for one model (denoted by m) to the maximum was calculated. Finally, four resulting ratios were used for calculating the final score ExpQ by the formula.

Model accuracy was quantified by the score 1/(1+APE). For quantitative questions with no answer, we assigned 100% APE to them as a penalty. The F1 score is a standard LLMs evaluation metric representing the harmonic mean of precision (relevancy of information) and recall (coverage of information); therefore, it is suitable for evaluating semantics of the answers to open-ended queries. To assess the model output of open-ended queries (Table S4), the F1 score was calculated using the Bidirectional Encoder Representations from Transformers (BERT)[38] with deberta-xlarge-mnli.[39] Within the BERT package, the texts of the reference answer (reference tokens) and the model answer (candidate tokens) can be transformed into vectors. Next, the maximum cosine similarity of pairs between the tokens can be determined and then averaged across the reference tokens for calculating the precision (average across the candidate tokens for calculating the recall). In computing the F1 score, we treated qualitative text outputs without a clear conclusion to queries as "no answer." To summarize, the average 1/(1+APE) score across 50 quantitative queries was calculated to assess the accuracy of numerical information extraction for each model; while the average F1 score of 50 open-ended queries was computed to evaluate the meaning consistency between the model answer and the human answer.

To assess differences in model scores, the pairwise comparisons were performed with a nonparametric significance test, Wilcoxon signed-rank test.[40] The score improvement of an RAG pipeline (Score $_{LLM+RAG}$) on the original LLM (Score $_{LLM}$) was calculated as the equation below

$$\text{Improvement}(\%) = \frac{\text{Score}_{LLM+RAG} - \text{Score}_{LLM}}{\text{Score}_{LLM}} \times 100 \quad (1)$$

When the improvement calculation for the quantitative query set was based on 1/(1+APE) scores, F1 scores were used for the improvement estimation on the open-ended query set. Such improvements for three different RAG pipelines on five LLMs have been computed (Table S5).

2.4. Experiment Validations. Dr. Blenner's group conducted a preliminary study and incubated mealworms at 25 °C and 80% relative humidity under dark conditions for 30 days. Mealworms were purchased from Rainbow Mealworms (Compton, CA, USA) and subject to two-day starvation to clear their gut before plastics degradation experiments. To investigate the gut microbiome, nonstripped LDPE films (Mw:111.0 kDa, crystallinity: 52.6%) were purchased from Sigma-Aldrich (St. Louis, MO, USA) and fed to 100 mealworms. The genomic DNA of gut microbes was extracted by a New England Biolabs Monarch Genomic DNA extraction kit and sent to Joint Genome Institute (Berkeley, CA, USA) for metagenomic sequencing on PacBio. To evaluate the deconstruction of PE by mealworms, stripped LDPE films (Mw: 95 kDa, Mn: 26.6 kDa, crystallinity: 52.0%) were prepared from PE pellets purchased from the same source of nonstripped films and then fed to mealworms along with bran. The molecular weights of residual PE in the frass and the original feedstock were quantified by HT-GPC (TOSOH HLC-8312GPC/H with two TSK_{gel} GMHHR-H(20)HT columns and one TSK_{gel} G2000H_{HR}). The functional groups of the residual PE and the feedstock were detected by a FTIR spectrometer (Thermo Scientific Nicolet iS5 FTIR spectrometer, Pittsburgh, PA, USA) with the absorbance mode in the wavenumber range of 4000−400 cm^{-1} (32 scans, spectral resolution: 0.482 cm^{-1}). The experiment results were used to compare with the predictions generated by the leading model of the open-ended query set. The queries for experiment validations are listed in Table S6. Detailed experimental methods are consistent with previous work.[41]

2.5. Experiment Design Evaluation. We tested LLMs to write an experimental design for a study that has not been previously conducted in the field of mealworm-based plastic degradation. Because the structures of the long text outputs given by different models can vary greatly without detailed prompts, we added instructions in our query to generate outputs consisting of six sections (Objectives, Hypothesis, Experiment Design, Data Analysis, Expected Outcomes, and Conclusions) and having a method section under each

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

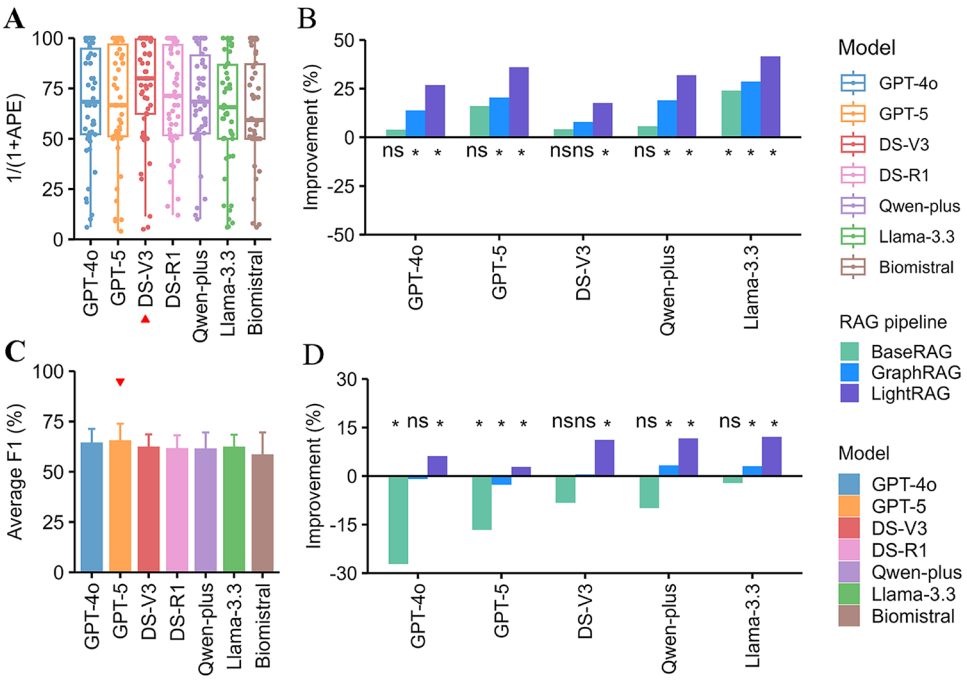


Figure 3. Model performance comparison for two different query sets. The LLM that has the highest score among original models is highlighted by a red triangle. The significance level for the score comparison against the original LLM for each RAG pipeline was denoted as below. Significance (Wilcoxon signed-rank test) $p < 0.05$ indicated by *, no statistical significance indicated by ns. (A) 1/(1+APE) scores of quantitative queries for original LLMs. (B) Improvement from three RAG pipelines on the average 1/(1+APE) scores of original LLMs for quantitative query set. (C) Average F1 scores with standard deviations across 50 open-ended queries for original LLMs. (D) Improvement from three RAG pipelines on the average F1 scores of original LLMs for open-ended query set.

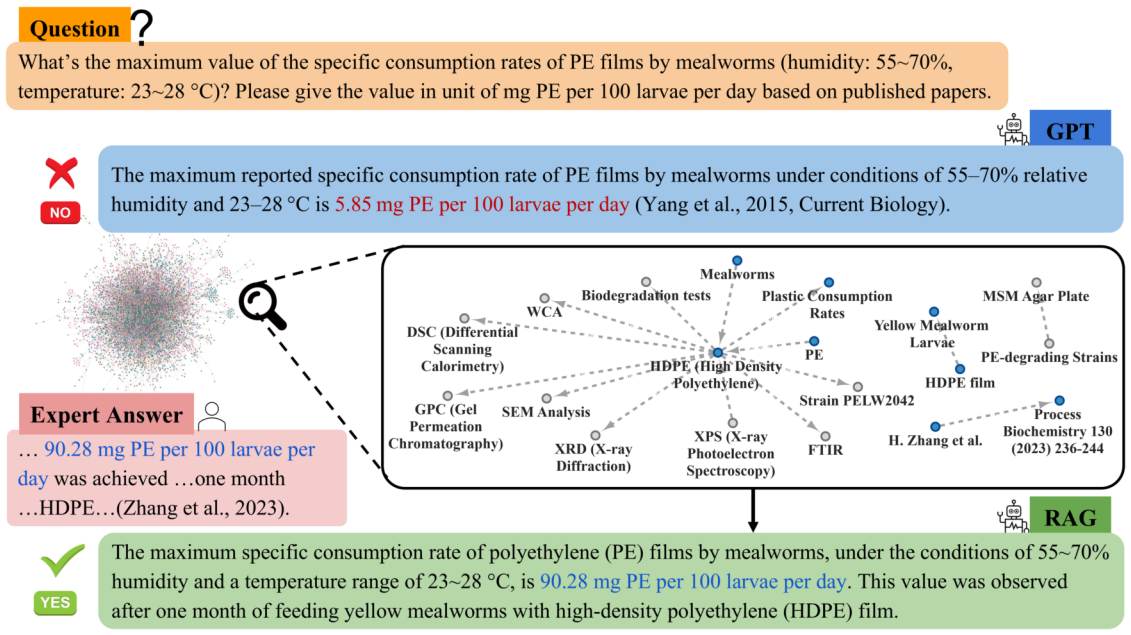


Figure 4. Quantitative query example of how LightRAG improved the performance of GPT-4o. Details of the extracted subgraph show that knowledge graph-based RAG identifies relevant information from the paper database: It "examines" the entire knowledge network to find the key nodes "Plastic Consumption Rates" and "PE" which were highly related to the query. The nearby entities and linked document sources were embedded and screened for extracting values. A simple function to find the maximum was applied and the supporting entities (text chunks) were located over the knowledge network as shown. Those entities and their supporting document sources were embedded and summarized. The resulting report was integrated into the original query as the context. Finally, the updated prompt was forwarded to LLM for a response. By this way, the answer from LLMs trained by knowledge graph-based RAG pipeline can be more reliable.

experiment or data analysis (Figure 2). To incorporate the necessary details into the methods proposed by models, we instructed them to include experimental conditions, technologies, and techniques. Finally, the model output can be refined by assigning the role of "an expert in the field of plastics biodegradation" to LLMs and asking about the rationales. We

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

used detailed prompts along with the local search as the query method for GraphRAG-trained LLMs and the hybrid search as the query method for LightRAG-trained LLMs. To evaluate the model output, we defined "Details" in methods as information about the technologies and software to apply, statistical techniques, and experimental conditions, such as pH, temperature, humidity, medium name, incubation duration, etc. Next, we manually highlighted the corresponding text chunks to count the number of experiments and the number of "Details" under the Experiment Design section for each model. Then, the maximum number across all models was determined. Similar counting steps were repeated for the Data Analysis section. Finally, the ExpQ score, defined specifically for this study, was calculated using the equation shown below (Figure 2, eq 2)

$$\mathrm{ExpQ}(\%) = \frac{1}{2} \times \left(\frac{E_m}{\max\limits_{m\in M} E_m} \times \frac{D_m^E}{\max\limits_{m\in M} D_m^E} + \frac{A_m}{\max\limits_{m\in M} A_m} \times \frac{D_m^A}{\max\limits_{m\in M} D_m^A} \right) \times 100 \quad (2)$$

The numerators E_m, $D_m{}^E$, A_m, and $D_m{}^A$ are the counts for experiments, details of experiments, data analyses, and details of data analysis for a model (m), respectively. The denominators are the maximums of the corresponding counts across a full set of models (M).

3. RESULTS AND DISCUSSION

3.1. Model Performances for Quantitative Queries. We first tested the original LLMs and RAG-trained LLMs (LLM+RAGs) for quantitative information extraction. Their performances were evaluated by the score 1/(1+APE), illustrated in Figure 3A. Among the seven original models, DS-V3 extracts numeric information best, with the highest average score of 75%, followed by DS-R1 and Qwen-plus. All three RAG pipelines improve the overall accuracy of the original models. LightRAG provides the largest improvement (Figure 3B) and results in overall model accuracy ranging between 86 to 93%, followed by GraphRAG (77 to 83%, Table S5) and BaseRAG (71 to 79%). Among all of the LLMs, the highest accuracy achieved is 92% by the combination of GPT-5 and LightRAG, which significantly enhanced the original GPT-5 model by 36%.

From the perspective of query scopes, most original LLMs and LLM+RAGs can accurately answer experimental condition parameters, such as temperature, humidity and pH. When targeting experiment results from specific studies, their answers are relatively accurate, such as plastics weight loss by specific strains (Table S3 Q7 and Q11−12). The global questions (Table S3 Q1−3 and Q32, etc.), however, are more challenging to both original LLMs and LLM+RAGs. For example, when asked about the highest specific consumption rate of PE films (Table S3 Q1, Figure 4) across all of the mealworm studies, most models replied with values below 10 mg per 100 larvae per day. The reference answer is 90.28 mg per 100 larvae per day.[42−45] One possible explanation could be the large variation of reported plastic consumption rates from the curated database with different experimental setups, such as diverse plastic properties.[46−48] In addition, the training data for most LLMs might be noisy. Notably, LightRAG-trained models can provide accurate answers even for global questions, including the question about the consumption rate (Figure 4).

3.2. Model Performances for Open-Ended Queries. In addition to the quantitative information extraction, the model's ability to retrieve domain-specific knowledge was tested by a set of open-ended queries. The F1 score from BERT, which captures the meaning-based similarity between a human written reference and the LLM reply, was used to evaluate the model performance of original LLMs and LLM+RAGs (see Section 2.3). The seven tested original models showed similar overall performance in answering the 50-question query set in terms of average F1 score. The recently released GPT-5 obtained the highest average F1 score of 65% (Table S5), followed by GPT-4o and DS-V3 (Figure 3C). Among different RAG pipelines, knowledge graph-based RAGs improved the overall answer quality, while the traditional RAG even deteriorated the model performance. The average F1 scores of original LLMs landed between 58 to 66%. After training by GraphRAG, the scores were slightly improved to a range of 62 to 64%. LightRAG enhanced the answer quality the best (Figure 3D), reaching scores from 66 to 70%. The combination of Llama-3.3 and LightRAG obtained the highest average F1 score of 69.5%, which was 12.1% higher than that of the original Llama-3.3.

For questions about the general knowledge of mealworm-driven plastic degradation, answers of most LLMs agreed with human-generated answers. Nonetheless, for questions requiring information collection and summarization, the performance of LLMs drops prominently. When asked about the antibiotic suppression effect on PS degradation (Table 2), original LLMs and GraphRAG + LLMs tended to give direct conclusions from papers or reply with "no published evidence." Notably, LightRAG + LLMs clearly listed the strong evidence from literature, like molecular weight change and isotopic tracing.[49−51] By this way, the LightRAG pipeline assisted LLMs to more reliably summarize the cooperation between mealworms and their gut microbes across different studies. The LightRAG-trained LLMs also gave relatively precise information on strains isolated from plastic-eating mealworms and performed better in pulling out a specific analysis result, when pertaining to a certain study. For example, Lou et al. was the only one comparing the LDPE foam + bran diet with the sole bran diet for mealworms, and it reported *Mucispirillum* and *Lactobacillus* were upregulated in their gut.[52] The LightRAG-trained GPT-4o answered with exactly those two bacterial genera (Figure 5), while original GPT-4o gave four bacterial genera (*Enterococcus*, *Lactococcus*, *Pseudomonas*, and *Serratia*) either common in mealworm gut[9,47,52] or more associated with the sole PE diet in mealworms[47] or waxworm.[10]

Although LightRAG-trained LLMs are excellent at retrieving the existing knowledge from the literature, they lack the ability to synthesize new ones. The exact biochemical mechanisms of mealworm-driven PE degradation have not been revealed yet, and direct information for corresponding pathways is not available in the paper collection. Thus, no matter what LLMs or what RAG pipelines were used in our test, each had trouble in the task of identifying enzymes that initiate PE biodegradation in the mealworm gut (Table S4 Q48). Some enzymes they proposed, e.g., esterase and lipase, can be closely related to PE degradation, considering the upregulated expression and high activity level in mealworms.[53,54] Whereas those enzymes do not directly act in the very first step. Instead, alkane monooxygenases, peroxidases, and dioxygenases are

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

Table 2. Model Output Examples of PS Degradation in Open-Ended Queries for Llama-3.3 (Showing Only the Key Texts);[a] The Pipelines Present Evidence with Different Degrees of Details When Answering the Two Complementary Questions

example	original LLM	GraphRAG + LLM	LightRAG + LLM
what published evidence indicated that mealworms cannot degrade PS under severe antibiotic suppression?	...**Yang et al.**.., found that mealworms are unable to degrade polystyrene (PS) when subjected to severe antibiotic suppression, which highlights the crucial role of gut microbiota in the degradation process.	...but the provided data **does not include published experimental results** confirming this specific outcome for mealworms and PS degradation.	... a study by **Yang et al.** that found gentamicin-treated mealworms **lost the ability to depolymerize PS** and **mineralize PS into CO_2**, as determined by characterizing worm fecula and feeding with **13C-labeled PS**.
what published evidence indicated that the PS can still be partially degraded in mealworms under severe antibiotic suppression?	... a study by **Yang et al.** (2018) that found mealworms can degrade PS even when antibiotics are present···	...the study by **Yang et al.** found that mealworms can still partially degrade polystyrene even under severe antibiotic suppression···	... the **Mw, Mn, and Mz of PS residues** extracted from frass of antibiotic-treated larvae fed with PS were **slightly decreased** compared to the raw PS feedstocks.

[a]The major differences in model outputs were shown in bold.

more likely to be the key enzymes in the beginning of biodegradation by incorporating oxygen into PE and making carbon chains more vulnerable to other enzymes.[14] Solving a scientific question requires a strong multihop reasoning ability which is ambitious for the sole LLM to achieve. Promisingly, the information retrieval ability of LightRAG-trained LLMs can be utilized for preparing data sets that could be essential to discovering a new finding through computational approaches.

3.3. Model Performances of Experiment Validations.

The leading model from the open-ended query test, the LightRAG-trained Llama-3.3, was used to predict experimental outcomes for an unpublished preliminary study. Prediction was generated by answering the three queries in Table S6. Given the specific incubation conditions, properties of PE, and the cutoff of relative abundance, our model predicted *Pseudomonas*, *Acinetobacter*, and *Bacillus* as the abundant genera (>20%) of gut microbiome from mealworms consuming LDPE films as their only food source. Instead, the metagenomics results of the preliminary study reported *Staphylococcus* and *Enterococcus* as abundant genera. In terms of the presence in the sole PE diet group, the bacterial genus given by our model was logical: A few strains under the genus *Acinetobacter* and *Bacillus* were isolated from mealworms and claimed to be capable of deconstructing PE;[42,55] *Pseudomonas*, known to produce esterases, can also be related to PE degradation in mealworms.[47] However, predicted bacteria were reported at low abundance from previous studies[43,44,47] and did not overlap with abundant genera in our preliminary study. The prediction ability of our model may be limited by the dynamic nature of microbiome structures, which can vary due to factors like diet and incubation conditions. Differences in experimental settings (humidity, PE properties, etc.) between our experiment and the collected papers introduced variability in both taxonomic information and abundance of taxa. The inference of abundant bacteria is a multistep reasoning task; our model tended to capture key entities like "PE" initially but often overlooked descriptions like "abundant" and "> 20%", which were not directly presented as nodes in the knowledge network. Consequently, the biological variability and the complexity of inference tasks have likely increased the uncertainty in our model predictions.

The predicted FTIR analysis results for residual PE from frass of mealworms with a PE + Bran diet corresponded well with the actual experiment outcomes. When asked about new peaks, our model listed the carbonyl group and hydroxyl group, as the experimental result did, and emphasized the incorporation of oxygen. The predicted peak position of the C═O stretch (1700 cm^{-1}) was within the empirical range of its wavenumber (1700−1750 cm^{-1}). Additionally, our model mentioned a decrease in original peak intensities for C−H stretch bonds, which is a common observation across plastic degradation studies. The model prediction was precise for FTIR analysis on PE, likely because of its qualitative nature and the relatively consistent findings in our paper collection. In other words, the model prediction of the functional groups of residual plastics can be more uncertain if the initial reactions of plastic biodegradation are diverse and lack relevant studies.

In the case of GPC analysis, both our model and the preliminary study reported a decrease in molecular weight, comparing the residual PE with the original feedstock. Mw (number-average molecular weight) and Mn (weight-average molecular weight) of residual PE predicted by our model were 47.5−57.0 and 15.9−16.6 kDa, which approximate the values

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

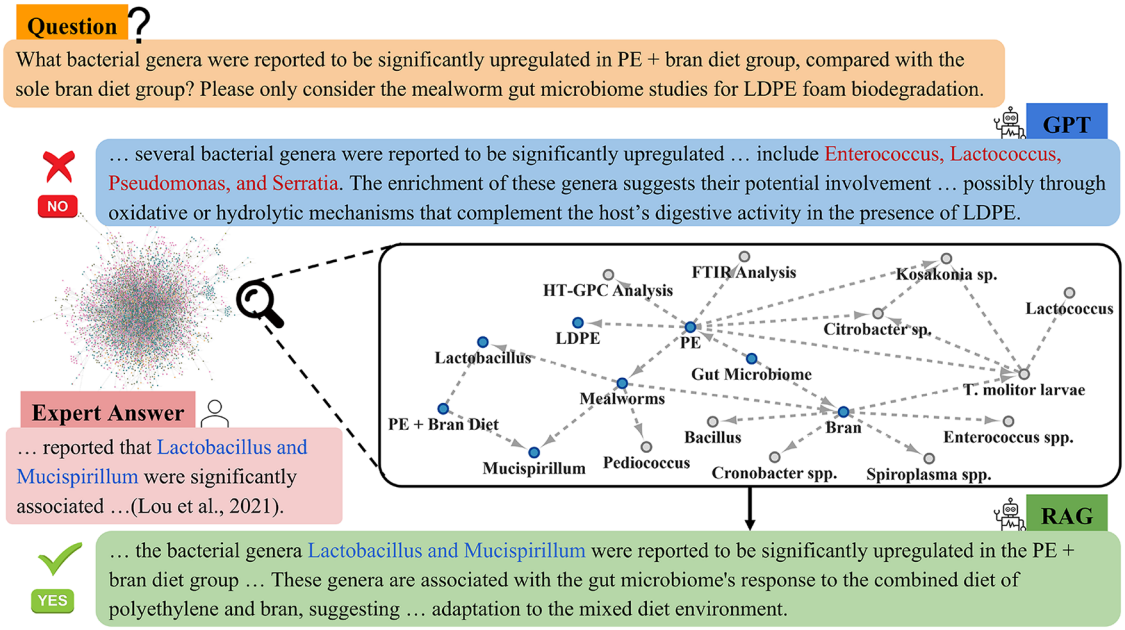


Figure 5. Open-ended query example of how LightRAG improved the performance of GPT-4o. Details of the extracted subgraph show that knowledge graph-based RAG identifies relevant information from the paper database: It "zoom in" the entire knowledge network to find entities near the key nodes "PE + Bran Diet", "LDPE" and "Gut Microbiome" which were highly related to the query. The descriptions attached to the nearby entities and relevant document sources were embedded and screened. The summarized report was incorporated into the original query as the context. Finally, the updated query was sent to LLM. By this way, responses from LLMs trained by knowledge graph-based RAG pipeline can be enhanced.

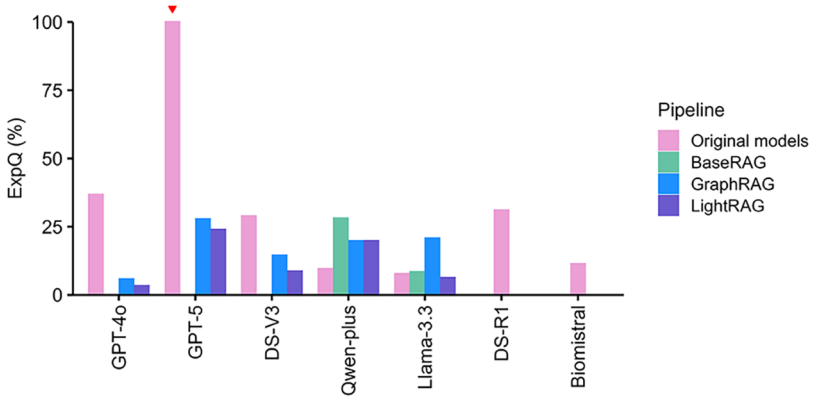


Figure 6. Model performance comparison for application in experimental design. The reasoning model DS-R1 and the finetuned model Biomistral were not integrated with any RAG pipeline, when GPT-4o, GPT-5, and DS-V3 trained from the BaseRAG pipeline did not give any experiment plan (ExpQ = 0). Vector-based RAG (e.g., BaseRAG) sometimes fails to retrieve text chunks meaningfully supporting the answer. The model giving the most specific plan relative to others was noted with red triangle. The final score ExpQ was calculated as the formula in eq 2.

given by the actual analysis, 79.8 and 18.4 kDa. Such an extensive breakdown of LDPE by mealworms was commonly found in our paper collection. In summary, the model predictions for different types of analysis, including GPC, not only align with scientific principles but are also quantitatively accurate. Meanwhile, caution should be taken when encountering complex inference tasks and when there is a lack of supporting studies.

3.4. Application for Experiment Design. The original LLMs and RAG-trained LLMs were used to generate an experimental plan for a future study about the interactions between species enriched from the mealworm gut and with PE as the sole carbon source. Our detailed prompts instructed models to give outputs with a consistent structure (Figure 2) and then we calculated the score ExpQ for evaluating major sections, Experiment Design and Data Analysis according to Section 2.5. The experimental plan generated by the original GPT-5 was the most detailed for both experiments and data analysis among all other models, while the same LLM trained by BaseRAG pipeline failed to generate the plan and only responded with "I don't know". The GraphRAG-trained models scored similarly to the LightRAG-trained models, and the original models scored higher (Figure 6). Generally, original LLMs gave a more specific plan than RAG-trained ones, except for Qwen-plus and Llama-3.3. This limitation likely stems from the narrow scope of the document collection used in the RAG pipeline: when the system encounters a study lacking sufficient methodological details within the available domain knowledge, the LLM's reasoning capability is constrained.

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

Although the original GPT-5 provided more details in the experimental plan than others, the design generated by it still lacks a few critical elements. Microbial community enrichment is the very beginning of an investigation into the interactions between gut species *in vitro*. GPT-5 mentioned gut extraction and set up sufficient control groups and culture incubation conditions, but it did not specify experimental conditions such as the humidity and temperature for mealworms before sampling, which can impact the species diversity *in vivo*. The assessment on the deconstruction of PE was important, and a set of experiments that GPT-5 suggested is sufficient for investigating the degradation ability of the microbial community. Nevertheless, when the number of experiments increases, it tends to leave out details in the later section and even loses track of controlling the experimental variables. For example, in experiments like community enrichment, the synthetic consortia construction, and stable-isotope tracing, the substrates GPT-5 selected were inconsistent. Even among the same type of plastics, property differences like molecular weight can change the plastics biodegradability, introducing variations to the microbial community.[47] Therefore, the experiments proposed by LLMs, even the most detailed ones, still need expert supervision.

4. ENVIRONMENTAL IMPLICATIONS, LIMITATIONS, AND FUTURE DIRECTIONS

In specialized fields such as mealworm-driven plastic degradation, we show that RAG-enhanced models outperform their non-RAG counterparts in answering expert questions, except for experimental design. RAG + LLM can improve information transparency and factual accuracy without the massive computational cost required for fine-tuning. Specifically, our work indicated great potential for RAG + LLM pipelines to provide practical aid by automating the feature extraction (e.g., the strain names, plastic type, and incubation conditions) from bulky literature in a specific domain. Such a framework can help with data curation and reduce researcher workload. It can also shorten the discovery cycles through context-aware retrieval. For example, redundant search can be minimized with basic inferences from RAG + LLM applied (fewer iterations, semantic cues, real-time refinement, etc.). By integrating unstructured data like policy documents with structured environmental data sets, the pipeline may enable faster comparisons of technological options based on environmental impacts.

In this study, the papers we collected for model training can be a representative subset of the plastic biodegradation domain because the conceptual foundation and methodologies applied are similar. The core challenges and primary goals are shared among the broader fields, such as the marginal degradation rate for nonhydrolyzable plastics and the investigation on degradation pathways. However, mealworm-driven plastic degradation is still a young field and has not yet progressed to include metabolic engineering and strain engineering. Currently, the metabolic pathways of mealworm plastic degradation are unclear, and it is difficult to differentiate the degradation process in gut microbes from that in the larvae themselves due to the complexity of their interactions. Thus, even multiple gut microbes capable of deconstructing plastics have been isolated from mealworms, but their application to waste treatment was limited by uncertainty and conflicting outcomes. With more relevant papers emerging in the future, the RAG can easily evolve with updated information and allow models to access recent reports, making them particularly valuable in rapidly evolving fields of engineering technology. This dynamic retrieval capability ensures that responses remain current without the need for frequent model retraining.

On the other hand, we note that RAG remains unreliable for higher-order tasks, such as hypothesizing mechanisms, explaining biological pathways in depth, or reconciling contradictory findings. Most "errors" in the generated answers likely reflect conflicting preliminary findings inherent to a field that is still in its nascent stage. Even the best models merely aggregate information without distinguishing which study says what, undermining trust in their synthesized outputs from the perspective of humans who read these papers. For example, early research on *KlebCP*'s PVC-degrading activity attributed new FTIR peaks to polymer breakdown, whereas a later paper showed those peaks stemmed from enzyme residues on the plastic surface.[56] After adding this new conflicting finding, the knowledge graph does not reflect the change by deleting the edge between *KlebCP* and PVC (Figure S1) because RAG treats the next paper as direct additions to a union rather than as corrections to existing structures (Section 2.2). The observed weaknesses in open-ended questions also highlight risks in relying solely on current AI for generating expert-level scientific content. The tendency toward oversimplification, the omission of critical details, and a lack of nuanced understanding regarding the weight of different types of evidence necessitate rigorous expert oversight. These limitations highlight the need for advanced information extraction methods.

AI outputs in specialized scientific domains should be viewed as preliminary drafts or informational starting points, requiring thorough verification against primary, high-impact literature, and critical evaluation by domain experts. The following three future directions can help mitigate the poor conflict reconciliation problem inherent in RAG pipelines: (1) apply a temporal filter, such as a temporal knowledge graph, that prioritizes newer studies over older ones, (2) act as intelligent, self-evolving agents that can check and weigh conflicting claims, or (3) detect and flag discrepancies, prompting users to choose which interpretation to follow. RAG + LLM may empower machine learning[57,58] and mechanistic modeling[59] to enhance the prediction of enzymatic reactions, microbial specialists, residual plastic distributions, and biodegradation processes.

■ ASSOCIATED CONTENT

Data Availability Statement

The code used and data sets generated during the current study are available in the following repository: https://github.com/wenyuli23/Mealworm-plastic-degradation-RAG. The original repository for LightRAG can be found here: https://github.com/HKUDS/LightRAG. The original repository for GraphRAG can be found here: https://github.com/microsoft/graphrag. Sequencing data is available upon request from Dr. Blenner's group. The detailed model outputs are available upon request from Dr. Tang's group.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.5c14258.

Details about queries and method related sections. An explanation of each table is on the first sheet (XLSX)

Auxiliary figures mainly related to knowledge graph extraction. Knowledge graph growing (Figure S1);

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448

knowledge graph comparison by entity types (Figure S2); detailed example of ExpQ score calculation (Figure S3) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Wenyu Li − *Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0009-0009-0220-2259; Email: l.wenyu@wustl.edu

Mark A. Blenner − *Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States;* orcid.org/0000-0001-9274-3749; Email: blenner@udel.edu

Yinjie J. Tang − *Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0000-0002-5112-0649; Email: yinjie.tang@wustl.edu

Authors

Bingdi Liu − *Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

Runyu Zhao − *Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

Ross R. Klauer − *Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States*

Alex Hansen − *Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States*

Nathan Miller − *Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States*

Yixin Chen − *Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.5c14258

Author Contributions

[‖]B.L. and W.L. contributed equally to this work. Conceptualization and ideas: Y.J.T., M.B., YC. and W.L.; Methodology and programming: GraphRAG team, LightRAG team, W.L. and B.L.; Database construction and processing: B.L., R.Z., R.K., A.H. and N.M.; Experimental design analysis: B.L., R.K., A.H., N.M. and M.B.; Experiment validation: B.L., R.K., A.H., N.M. and M.B.; Writing: B.L., W.L. and R.Z.; Review and editing: all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0023085. R.K. was funded in part by the Delaware Environmental Institute, University of Delaware. A.H. was funded in part by the U.S. Department of Education GAANN (P200A210065).

REFERENCES

(1) Plastics: Material-Specific Data U.S. Environmental Protection Agency 2024 https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/plastics-material-specific-data (accessed May 8, 2025).

(2) Santos, R. G.; Machovsky-Capuska, G. E.; Andrades, R. Plastic ingestion as an evolutionary trap: Toward a holistic understanding. *Science* **2021**, *373* (6550), 56−60.

(3) Yang, X. G.; Wen, P. P.; Yang, Y. F.; Jia, P. P.; Li, W. G.; Pei, D. S. Plastic biodegradation by in vitro environmental microorganisms and in vivo gut microorganisms of insects. *Front. Microbiol.* **2022**, *13*, No. 1001750.

(4) Kurniawan, T. A.; Hassan, G. K.; Al-Hazmi, H. E.; Othman, M. H. D.; Goh, H. H.; Aziz, F.; Anouzla, A.; Ali, I.; Khan, M. I.; Khan, M. M. H.; Mąkinia, J. Landfill mining: A step forward to reducing CH4 emissions and enhancing CO2 sequestration from landfill. *J. Hazard. Mater. Adv.* **2024**, No. 100512, DOI: 10.1016/j.hazadv.2024.100512.

(5) Pathak, G.; Nichter, M.; Hardon, A.; Moyer, E.; Latkar, A.; Simbaya, J.; Pakasi, D.; Taqueban, E.; Love, J. Plastic pollution and the open burning of plastic wastes. *Global Environ. Change* **2023**, *80*, No. 102648, DOI: 10.1016/j.gloenvcha.2023.102648.

(6) Liu, Z.; Chang, S. H.; Mailhot, G. Emerging Biochemical Conversion for Plastic Waste Management: A Review. *Molecules* **2025**, *30* (6), No. 1255, DOI: 10.3390/molecules30061255.

(7) Wei, R.; Zimmermann, W. Microbial enzymes for the recycling of recalcitrant petroleum-based plastics: how far are we? *Microb. Biotechnol.* **2017**, *10* (6), 1308−1322.

(8) Yang, S. S.; Ding, M. Q.; He, L.; Zhang, C. H.; Li, Q. X.; Xing, D. F.; Cao, G. L.; Zhao, L.; Ding, J.; Ren, N. Q.; Wu, W. M. Biodegradation of polypropylene by yellow mealworms (Tenebrio molitor) and superworms (Zophobas atratus) via gut-microbe-dependent depolymerization. *Sci. Total Environ.* **2021**, *756*, No. 144087.

(9) Brandon, A. M.; Gao, S. H.; Tian, R.; Ning, D.; Yang, S. S.; Zhou, J.; Wu, W. M.; Criddle, C. S. Biodegradation of Polyethylene and Plastic Mixtures in Mealworms (Larvae of Tenebrio molitor) and Effects on the Gut Microbiome. *Environ. Sci. Technol.* **2018**, *52* (11), 6526−6533.

(10) Lou, Y.; Ekaterina, P.; Yang, S. S.; Lu, B.; Liu, B.; Ren, N.; Corvini, P. F.; Xing, D. Biodegradation of Polyethylene and Polystyrene by Greater Wax Moth Larvae (Galleria mellonella L.) and the Effect of Co-diet Supplementation on the Core Gut Microbiome. *Environ. Sci. Technol.* **2020**, *54* (5), 2821−2831.

(11) Quan, Z.; Zhao, Z.; Liu, Z.; Wang, W.; Yao, S.; Liu, H.; Lin, X.; Li, Q. X.; Yan, H.; Liu, X. Biodegradation of polystyrene microplastics by superworms (larve of Zophobas atratus): Gut microbiota transition, and putative metabolic ways. *Chemosphere* **2023**, *343*, No. 140246.

(12) Ribeiro, N.; Abelho, M.; Costa, R. A Review of the Scientific Literature for Optimal Conditions for Mass Rearing Tenebrio molitor(Coleoptera: Tenebrionidae). *J. Entomol. Sci.* **2018**, *53* (4), 434−454.

(13) Riaz, K.; Iqbal, T.; Khan, S.; Usman, A.; Al-Ghamdi, M. S.; Shami, A.; El Hadi Mohamed, R. A.; Almadiy, A. A.; Al Galil, F. M. A.; Alfuhaid, N. A.; et al. Growth Optimization and Rearing of Mealworm (Tenebrio molitor L.) as a Sustainable Food Source. *Foods* **2023**, *12* (9), No. 1891.

(14) Klauer, R. R.; Hansen, D. A.; Wu, D.; Monteiro, L. M. O.; Solomon, K. V.; Blenner, M. A. Biological Upcycling of Plastics Waste. *Annu. Rev. Chem. Biomol. Eng.* **2024**, *15*, 315−342.

(15) Vital-Vilchis, I.; Karunakaran, E. Using Insect Larvae and Their Microbiota for Plastic Degradation. *Insects* **2025**, *16* (2), No. 165, DOI: 10.3390/insects16020165.

(16) Brandon, A. M.; Garcia, A. M.; Khlystov, N. A.; Wu, W. M.; Criddle, C. S. Enhanced Bioavailability and Microbial Biodegradation of Polystyrene in an Enrichment Derived from the Gut Microbiome of Tenebrio molitor (Mealworm Larvae). *Environ. Sci. Technol.* **2021**, *55* (3), 2027−2036.

(17) Obrador-Viel, T.; Zadjelovic, V.; Nogales, B.; Bosch, R.; Christie-Oleza, J. A. Assessing microbial plastic degradation requires robust methods. *Microb. Biotechnol* **2024**, *17* (4), No. e14457.
(18) Wang, Y.; Zheng, S.; Wang, F.; Peng, J.; Zhou, J.; Wang, F.; Jiang, M.; Chen, X. Advances in methods for detecting plastics biodegradation. *Chin. J. Biotechnol.* **2023**, *39* (5), 1889−1911.
(19) Vink, H. Determination of Molecular Weight Distributions by Gel-Permeation Chromatography. *Makromol. Chem.* **1968**, *116*, 241−249.
(20) Siesler, H. W. Fourier-Transform Infrared (Ftir) Spectroscopy in Polymer Research. *J. Mol. Struct.* **1980**, *59*, 15−37.
(21) Nelson, T. F.; Baumgartner, R.; Jaggi, M.; Bernasconi, S. M.; Battagliarin, G.; Sinkel, C.; Kunkel, A.; Kohler, H. E.; McNeill, K.; Sander, M. Biodegradation of poly(butylene succinate) in soil laboratory incubations assessed by stable carbon isotope labelling. *Nat. Commun.* **2022**, *13* (1), No. 5691.
(22) Zhang, C.; Mu, Y.; Li, T.; Jin, F. J.; Jin, C. Z.; Oh, H. M.; Lee, H. G.; Jin, L. Assembly strategies for polyethylene-degrading microbial consortia based on the combination of omics tools and the "Plastisphere". *Front. Microbiol.* **2023**, *14*, No. 1181967.
(23) Shafana Farveen, M.; Narayanan, R. Omic-driven strategies to unveil microbiome potential for biodegradation of plastics: a review. *Arch. Microbiol.* **2024**, *206* (11), 441.
(24) Malik, S.; Maurya, A.; Khare, S. K.; Srivastava, K. R. Computational Exploration of Bio-Degradation Patterns of Various Plastic Types. *Polymers* **2023**, *15* (6), No. 1540, DOI: 10.3390/polym15061540.
(25) Xu, B. Y.; Li, Z. H.; Yang, Y. X.; Wu, G. L.; Wang, C. Z.; Tang, X. P.; Li, Y.; Wu, Z. H.; Su, Q. X.; Shi, X. Q.; et al. Evaluating and Advancing Large Language Models for Water Knowledge Tasks in Engineering and Research. *Environ. Sci. Tech. Lett.* **2025**, *12* (3), 289−296.
(26) Rillig, M. C.; Agerstrand, M.; Bi, M.; Gould, K. A.; Sauerland, U. Risks and Benefits of Large Language Models for the Environment. *Environ. Sci. Technol.* **2023**, *57* (9), 3464−3466.
(27) Liu, N. F.; Lin, K.; Hewitt, J.; Paranjape, A.; Bevilacqua, M.; Petroni, F.; Liang, P. Lost in the Middle: How Language Models Use Long Contexts. *Trans. Assoc. Comput. Linguist.* **2024**, *12*, 157−173.
(28) Farquhar, S.; Kossen, J.; Kuhn, L.; Gal, Y. Detecting hallucinations in large language models using semantic entropy. *Nature* **2024**, *630* (8017), 625−630.
(29) Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.et al. *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 2020 DOI: 10.48550/arXiv.2005.11401.
(30) Perkins, G.; Anderson, N. W.; Spies, N. C. Retrieval-augmented generation salvages poor performance from large language models in answering microbiology-specific multiple-choice questions. *J. Clin. Microbiol.* **2025**, *63* (3), No. e0162424.
(31) Jeong, M.; Sohn, J.; Sung, M.; Kang, J. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. *Bioinformatics* **2024**, *40* (Suppl 1), i119−i129.
(32) Guo, Z.; Xia, L.; Yu, Y.; Ao, T.; Huang, C. *LightRAG: Simple and Fast Retrieval-Augmented Generation*; Association for Computational Linguistics: Suzhou, China, 2025; pp 10746−10761 DOI: 10.18653/v1/2025.findings-emnlp.568.
(33) Edge, D.; Trinh, H.; Cheng, N.; Bradley, J.; Chao, A.; Mody, A.; Truitt, S.; Metropolitansky, D.; Ness, R. O.; Larson, J. From Local to Global: A Graph RAG Approach to Query-Focused Summarization *arXiv*, Preprint, 2024 DOI: 10.48550/arXiv.2404.16130.
(34) Labrak, Y.; Bazoge, A.; Morin, E.; Gourraud, P.-A.; Rouvier, M.; Dufour, R. *BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains*; Martins, A.; Srikumar, V., Eds.; Association for Computational Linguistics: Bangkok, Thailand, 2024; pp 5848−5864 DOI: 10.18653/v1/2024.findings-acl.348.
(35) RapidLayout: Fast Hard Block Placement of FPGA-Optimized Systolic Arrays using Evolutionary Algorithms, 2020 30th International Conference on Field-Programmable Logic and Applications (FPL). 2020 DOI: 10.1109/fpl50879.2020.00034.
(36) Wang, P.; Bai, S.; Tan, S.; Wang, S.; Fan, Z.; Bai, J.; Chen, K.; Liu, X.; Wang, J.; Ge, W.et al. Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution *arXiv*, Preprint, 2024 DOI: 10.48550/arXiv.2409.12191.
(37) Johnson, J.; Douze, M.; Jegou, H. Billion-Scale Similarity Search with GPUs. *IEEE Trans. Big Data* **2021**, *7* (3), 535−547.
(38) Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K. Q.; Artzi, Y. BERTScore: Evaluating Text Generation with BERT, International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020 DOI: 10.48550/arXiv.1904.09675.
(39) He, P.; Liu, X.; Gao, J.; Chen, W. DeBERTa: Decoding-enhanced BERT with Disentangled Attention, International Conference on Learning Representations, Virtual Event, Austria, 2021 DOI: 10.48550/arXiv.2006.03654.
(40) Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1* (6), 80 DOI: 10.2307/3001968.
(41) Klauer, R. R.; Hansen, D. A.; Schyns, Z. O. G.; Monteiro, L. O.; Moore-Ott, J. A.; Williams, M.; Tarr, M.; Singh, J.; Mhadeshwar, A.; Korley, L. T. J.; Solomon, K. V.; Blenner, M. A. Biological polyethylene deconstruction initiated by oxidation from DyP peroxidases *bioRxiv*, Preprint, 2025 DOI: 10.1101/2025.02.27.640435.
(42) Zhang, H.; Liu, Q.; Wu, H.; Sun, W. X.; Yang, F.; Ma, Y. H.; Qi, Y. J. Biodegradation of polyethylene film by the Bacillus sp. PELW2042 from the guts of Tenebrio molitor (Mealworm Larvae). *Process Biochem.* **2023**, *130*, 236−244.
(43) Yang, S. S.; Ding, M. Q.; Zhang, Z. R.; Ding, J.; Bai, S. W.; Cao, G. L.; Zhao, L.; Pang, J. W.; Xing, D. F.; Ren, N. Q.; Wu, W. M. Confirmation of biodegradation of low-density polyethylene in dark- versus yellow- mealworms (larvae of Tenebrio obscurus versus Tenebrio molitor) via gut microbe-independent depolymerization. *Sci. Total Environ.* **2021**, *789*, No. 147915.
(44) Yang, L.; Gao, J.; Liu, Y.; Zhuang, G.; Peng, X.; Wu, W. M.; Zhuang, X. Biodegradation of expanded polystyrene and low-density polyethylene foams in larvae of Tenebrio molitor Linnaeus (Coleoptera: Tenebrionidae): Broad versus limited extent depolymerization and microbe-dependence versus independence. *Chemosphere* **2021**, *262*, No. 127818.
(45) Pham, T. Q.; Longing, S.; Siebecker, M. G. Consumption and degradation of different consumer plastics by mealworms (Tenebrio molitor): Effects of plastic type, time, and mealworm origin. *J. Cleaner Prod.* **2023**, *403*, No. 136842.
(46) Peng, B. Y.; Sun, Y.; Xiao, S.; Chen, J.; Zhou, X.; Wu, W. M.; Zhang, Y. Influence of Polymer Size on Polystyrene Biodegradation in Mealworms (Tenebrio molitor): Responses of Depolymerization Pattern, Gut Microbiome, and Metabolome to Polymers with Low to Ultrahigh Molecular Weight. *Environ. Sci. Technol.* **2022**, *56* (23), 17310−17320.
(47) Ding, M. Q.; Ding, J.; Zhang, Z. R.; Li, M. X.; Cui, C. H.; Pang, J. W.; Xing, D. F.; Ren, N. Q.; Wu, W. M.; Yang, S. S. Biodegradation of various grades of polyethylene microplastics by Tenebrio molitor and Tenebrio obscurus larvae: Effects on their physiology. *J. Environ. Manage.* **2024**, *358*, No. 120832.
(48) Zhong, Z.; Zhou, X.; Xie, Y.; Chu, L. M. The interplay of larval age and particle size regulates micro-polystyrene biodegradation and development of Tenebrio molitor L. *Sci. Total Environ.* **2023**, *857* (Pt 2), No. 159335.
(49) Yang, Y.; Yang, J.; Wu, W. M.; Zhao, J.; Song, Y.; Gao, L.; Yang, R.; Jiang, L. Biodegradation and Mineralization of Polystyrene by Plastic-Eating Mealworms: Part 2. Role of Gut Microorganisms. *Environ. Sci. Technol.* **2015**, *49* (20), 12087−12093.
(50) Peng, B. Y.; Su, Y.; Chen, Z.; Chen, J.; Zhou, X.; Benbow, M. E.; Criddle, C. S.; Wu, W. M.; Zhang, Y. Biodegradation of Polystyrene by Dark (Tenebrio obscurus) and Yellow (Tenebrio

molitor) Mealworms (Coleoptera: Tenebrionidae). *Environ. Sci. Technol.* **2019**, *53* (9), 5256−5265.
(51) Yang, S. S.; Wu, W. M.; Brandon, A. M.; Fan, H. Q.; Receveur, J. P.; Li, Y.; Wang, Z. Y.; Fan, R.; McClellan, R. L.; Gao, S. H.; et al. Ubiquity of polystyrene digestion and biodegradation within yellow mealworms, larvae of Tenebrio molitor Linnaeus (Coleoptera: Tenebrionidae). *Chemosphere* **2018**, *212*, 262−271.
(52) Lou, Y.; Li, Y.; Lu, B.; Liu, Q.; Yang, S. S.; Liu, B.; Ren, N.; Wu, W. M.; Xing, D. Response of the yellow mealworm (Tenebrio molitor) gut microbiome to diet shifts during polystyrene and polyethylene biodegradation. *J. Hazard Mater.* **2021**, *416*, No. 126222.
(53) Zhong, Z.; Nong, W.; Xie, Y.; Hui, J. H. L.; Chu, L. M. Long-term effect of plastic feeding on growth and transcriptomic response of mealworms (Tenebrio molitor L.). *Chemosphere* **2022**, *287* (Pt 1), No. 132063.
(54) Przemieniecki, S. W.; Kosewska, A.; Ciesielski, S.; Kosewska, O. Changes in the gut microbiome and enzymatic profile of Tenebrio molitor larvae biodegrading cellulose, polyethylene and polystyrene waste. *Environ. Pollut.* **2020**, *256*, No. 113265.
(55) Yin, C. F.; Xu, Y.; Zhou, N. Y. Biodegradation of polyethylene mulching films by a co-culture of Acinetobacter sp. strain NyZ450 and Bacillus sp. strain NyZ451 isolated from Tenebrio molitor larvae. *Int. Biodeter. Biodegr.* **2020**, *155*. DOI: 10.1016/j.ibiod.2020.105089.
(56) Stepnov, A. A.; Lopez-Tavera, E.; Klauer, R.; Lincoln, C. L.; Chowreddy, R. R.; Beckham, G. T.; Eijsink, V. G. H.; Solomon, K.; Blenner, M.; Vaaje-Kolstad, G. Revisiting the activity of two poly(vinyl chloride)- and polyethylene-degrading enzymes. *Nat. Commun.* **2024**, *15* (1), No. 8501.
(57) Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J. G.; Gu, A.; Li, B.; Ma, X.; Marrone, B. L.; Ren, Z. J.; Schrier, J.; et al. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environ. Sci. Technol.* **2021**, *55* (19), 12741−12754.
(58) Lin, C.; Zhang, H. Polymer Biodegradation in Aquatic Environments: A Machine Learning Model Informed by Meta-Analysis of Structure-Biodegradation Relationships. *Environ. Sci. Technol.* **2025**, *59* (2), 1253−1263.
(59) Peng, B. Y.; Sun, Y.; Zhang, X.; Sun, J.; Xu, Y.; Xiao, S.; Chen, J.; Zhou, X.; Zhang, Y. Unveiling the residual plastics and produced toxicity during biodegradation of polyethylene (PE), polystyrene (PS), and polyvinyl chloride (PVC) microplastics by mealworms (Larvae of Tenebrio molitor). *J. Hazard. Mater.* **2023**, *452*, No. 131326.

https://doi.org/10.1021/acs.est.5c14258
Environ. Sci. Technol. 2025, 59, 27437−27448