Retrieval-augmented Generation: Empowering Landscape Architects with Data-driven Design

Lacy Shelby[1], Renato Villela Mafra Alves da Silva[1]

[1]The Graduate Center, The City University of New York, NY/USA · lshelby@gradcenter.cuny.edu

Abstract: Landscape Architects and allied professions are steadily integrating artificial intelligence (AI) and machine learning (ML) in practice and deployment to enhance design processes, optimize project management, and augment analytical practices. The application of Retrieval-Augmented Generation (RAG) models in the fields of landscape architecture, planning, ecology, and architecture is still an emerging area and is not yet fully understood or widely explored. RAG models integrate a pre-trained language model with a retrieval system, effectively merging the processes of information retrieval and language generation into a cohesive framework. However, validation of information from large language models (LLMs) in Question-Answering Systems (QAS) (driven by AI/ML algorithms), such as ChatGPT and Google Gemini poses a challenge for landscape architects. The objective of this study was to assess the performance of the RAG model applied to landscape architecture literature. To address this objective, we developed a closed-domain neural network using open-source models trained on one issue of The Journal of Digital Landscape Architecture. To evaluate its performance, we queried the neural network on a series of landscape architectural tasks including design, theory, and analytical tasks. We then used quantitative measures to evaluate the performance. The results of ROUGE scores for the RAG demonstrate its effectiveness in capturing key concepts within the landscape architecture domain, particularly noting high precision values in Rouge-1 and Rouge-L metrics. While the model shows a lower performance in capturing two-word combinations as indicated by Rouge-2 scores, it successfully retrieved relevant information efficiently, as demonstrated by higher precision across other metrics. The study highlights the potential of Closed Domain Question Answering (CDQA) systems integrated with a RAG model, trained on specialized datasets, to enhance landscape architects' workflows. It also underscores the necessity of addressing challenges such as data curation, bias, and creativity limitations to maximize the utility and success of these tools in professional landscape architecture practice.

Keywords: Retrieval-Augmented Generation, landscape architecture, data-driven design, large language models (LLM), Closed-Domain Question Answering

1 Introduction

Landscape architects and allied professions are increasingly using and integrating the use of Artificial Intelligence (AI) and its many subfields including Machine Learning (ML) and Natural Language Processing (NLP) in practice. FERNBERG & CHAMBERLAIN (2023) underscore the necessity of adopting a cohesive strategy for the dissemination of knowledge regarding the interplay of AI and landscape design. This approach is suggested as pivotal for harnessing the full potential of AI in this domain. ZHANG & BOWES (2023) investigated the role of landscape designers within a cybernetic context, examining how AI's advancements are reshaping traditional design methodologies and the overall creative process. Both works collectively shed light on the transformative impacts of AI on landscape architecture, signalling a paradigm shift in both the theoretical and practical aspects of the field.

CANTRELL & ZHANG (2018) introduced the concept of a "third intelligence," advocating for the equal treatment of various intelligences and recognizing AI as an active participant in landscape design and management. Fusing material, biophysical, and machine intelligence

Journal of Digital Landscape Architecture, 9-2024, pp. 267-276. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-752-6, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537752025.

(CANTRELL 2018) with advanced computing is transforming the practice of landscape architecture and urban design. This rapid proliferation of AI/ML/NLP in landscape architecture workflows is increasingly supporting data-driven design, including optimizing aesthetics, enhancing ecological modelling, and generating planning and policy frameworks. (FERNBERG & CHAMBERLAIN 2023).

A challenge faced by landscape architects is the validation of information obtained from LLMs, such as ChatGPT and Google Gemini. The Retrieval-Augmented Generation (RAG) model, as introduced by LEWIS (LEWIS et al. 2020), is a powerful tool that combines the strengths of pre-trained LLMs with dynamic knowledge retrieval. Pairing these systems can support landscape architects in their knowledge-intensive tasks. This integration is particularly crucial in a field where the application of accurate and rigorously vetted abstract knowledge is essential (QIU 2023). Retrieval-Augmented Generation (RAG) models integrate a pre-trained large language model with a retrieval system, effectively merging the processes of information retrieval and language generation into a cohesive framework (LEWIS et al. 2020). The advantage of RAGs lies in their ability to update their knowledge base without retraining and their ability to access and incorporate knowledge sources dynamically. However, further exploration of its use for domain-specific retrieval augmentation in Question Answering Systems (QAS) is growing (SIRIWARDHANA et al. 2023). Landscape architects often work at the intersection of vast amounts of data, including site analysis data, regulatory guidelines, project reports, and design precedents, RAG models can serve to effectively organize and retrieve this diverse information, enabling access to this knowledgebase efficiently.

Most design professionals are now familiar with LLMs, a specific type of ML that utilizes NLP, as exemplified by tools like ChatGPT and Google Gemini. However, optimizing their functionality within landscape design practice presents several challenges. These range from ineffective prompt generation to uncertainty regarding the validity of information provided in an Open-Domain Question Answering (ODQA) format. The adoption of Closed-Domain Question Answering (CDQA) models could offer a solution to these issues. CDQA models can be tailored to include specialized knowledge, enabling practitioners to access trusted and referenceable information quickly and reliably (VILA 2011). For instance, FU et al. (2009) demonstrated that their Music Knowledge Question Answering system structured on concepts in the music domain achieved a precision rate of 77.25% in delivering domain-specific responses. This achievement highlights the potential of CDQA systems to effectively manage domain-specific queries with higher accuracy, suggesting a viable approach for integrating similar models into landscape design practice. Additionally, as concerns over proprietary information continue to rise, landscape architects and their firms will increasingly look to models that are closed systems, purpose-built on training data from a firm's archive or past projects.

In this study, we explored the potential of creating a landscape architecture tool that could serve as a collaborator to a landscape architect's design and analysis process. We developed a CDQA system paired with a RAG trained on the August 2023 issue of JoDLA, with the intention of it functioning as a landscape architectural adjunct. This adjunct is a practical solution that landscape architecture firms can readily adopt to inform future project and design development tasks. To develop this LLM, we employed a RAG framework that incorporates a retrieval component and a reader component. A retriever seeks out and identifies relevant information from a knowledge base that may contain the answer to a given question.

The reader component extracts the precise answer from the relevant retrieved information. With a narrowed set of relevant documents, the answer extraction engine, powered by pre-trained language models (e. g., BERT) parses the identified documents to pinpoint a precise answer. RAGs are a novel tool for landscape architects to leverage, in that they can be adapted to integrate with specialized knowledge bases. Our objective was to determine if a RAG model trained on a Closed-Domain dataset of landscape architecture literature would generate contextual and relevant outputs compared to human-generated responses on landscape architecture-based queries.

We trained our RAG model on the 65 articles and 3 auxiliary texts (Foreword, Introduction, and Preface) of the August 2023 Issue of The Journal of Digital Landscape Architecture. This issue covered a wide range of computational and technological landscape architecture topics. We then queried the RAG to trigger the retrieval process, which then leveraged the models' generation component to form a contextually appropriate response based on the query and retrieved information.

The research addressed two specific research questions:

1) What are the limitations of using CDQA paired with RAG models purpose-built on a landscape architectural knowledge base?
2) What impact might CDQA paired with RAGs have on the practice of landscape architecture as their performance and function improve?

2 Objective

2.1 Domain-Specific Landscape Architecture RAG

Our objective was to create a domain-specific Landscape Architecture RAG capable of assisting landscape architects in their design and analytical tasks. We hypothesized that the RAG, trained on the August 2023 issue of The Journal of Digital Landscape Architecture and employing a CDQA system integrated with a RAG model, would provide contextually relevant responses to a range of landscape architecture-specific queries. The queries included design idea generation, analysis of site conditions, and recommendations for enhancement of ecological services. We assessed the RAG's proficiency using quantitative metrics (e. g., precision-recall). The results may contribute to establishing a framework for future research and development of AI/LLM applications in the field of landscape architecture.

2.2 Unlocking AI Potential in Practice

Using a domain-specific RAG model in landscape architecture practice and ecological design can unlock new data insights. We propose the CDQA paired with a RAG approach has the potential to impact three categories of design practice:

- Propel creative thinking and amplify ideation through the analysis of past projects, identifying patterns, relationships, and design features or elements native to the design firm's approach. The RAG can function to explore variations and iterate on this archival dataset to generate variations on design concepts encouraging the exploration of alternative schemes or approaches to a design problem.
- Influence efficiency and boost productivity through the automation of repetitive tasks like generating content for responses to Requests for Proposals or Qualifications, create-

ing project cut sheets for prospective clients, developing project descriptions, and summarizing site data. Functionally the RAG can act as a firm's brain trust or central knowledge repository enhancing collaboration between team members and collaborative partners.

- Enhance communication and deepen client interactions through the analysis of client preferences and historical client data to support the development of design proposals and presentations specific to a client's mission or goals. The RAG can craft compelling narratives on a project's impact and purpose enhancing client uptake and investment.

3 Methodology

3.1 Development of the CDQA System Paired with a RAG

We built a CDQA system using a Retriever-Reader framework trained on the August 2023 issue of The Journal of Digital Landscape Architecture (JoDLA). The Retriever-Reader framework introduced by DAS et al. (2019) enables iterative interactions between the retriever and the reader. The intention of using this framework and the training data from JoDLA for the RAG was to demonstrate a proof of concept on a broad range of computational topics specific to the field of landscape architecture in a closed-domain setting. Further, the Journal provides a good source of foundational research themes in computational urban landscapes and ecology with a high content quality. To build the RAG we undertook the development of key components using open-source tools, which include the development of a database (the documents and data), implementation of a retriever (the function that scans and searches the information), and integration of a reader (the language comprehension expert that analyses the documents to extract the response). We built the system using the following steps:

1) Establish the database using the installation Elasticsearch and organize the data to make for more efficient retrieval (TAWARE et al. 2018). The study's database was trained on the August 2023 issue of The Journal of Digital Landscape Architecture.
2) Implement the Retriever using the Haystack framework and Elasticsearch's capabilities to search relevant documents and rank those documents for relevance using the search algorithm BM25 for document ranking.
3) Apply the Reader using the pre-trained language model (BERT) to understand the text, using BERT for answer extraction (DEVLIN et al. 2019).
4) Exercise system integration by connecting the Retriever and Reader, enabling the Retriever to narrow down potential answers, and the Reader to pinpoint the correct response.

3.2 Query Development and Evaluation Criteria

To develop effective queries for the RAG, the Journal articles were read and evaluated by a licensed landscape architect. Drawing from both the research findings specific to the articles and professional expertise, we then formulated a series of targeted queries for the RAG. The queries for our model were crafted to reflect the type of knowledge and tasks landscape architects would encounter in their practice. The queries were designed based on the specific content, using terminology found in the Journal's articles. The queries ranged from identifying broad specific factors important in the design of public space to something more nuanced such as what landscape qualities might influence the economic vitality of retail chains.

The evaluation criteria of the model's performance included quantitative measures. While the RAG's performance can provide one data point it doesn't capture the efficacy and potential of its practical application in the field of landscape architecture. Can a RAG, trained on landscape architecture-specific data match that of a landscape architect? Would the responses generated fall short, equal, or surpass the capabilities of a landscape architect or other design professional? The intention of advancing the method is to determine if AI can become a trusted design ally and partner in the creative process with the potential to generate new directions in the field and practice.

The computational quantitative measures were based on the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) (LIN 2004) which is a set of metrics designed to assess the quality of text summaries. The evaluation process involves comparing the computer-generated summary with human-generated content. These measures include considerations for n-grams (sequences of words), and word pairs to assess this overlap. Recall, precision, and F1 score, are crucial components of ROUGE, with recall reflecting the proportion of information captured from the reference and precisions indicating the conciseness. The F1 score combines both aspects, providing a measure of overall summarization effectiveness. The evaluation approach, leveraging ROUGE metrics, ensures a balanced assessment of performance, capturing the depth and brevity of the model's summaries in comparison to the human-generated content.

4 Results

For the ten queries tested (Tab. 1), objective metric evaluation using ROUGE scoring provided a context in which to interpret the functionality of the CDQA paired with the RAG's performance. We also captured the subjective evaluation from a landscape architectural professional, highlighting key response themes to capture where the RAG either met expectations or needed improvement.

Table 1: Queries, Retrieved Responses and Human-Generated Responses

Query	Retrieved Responses	Human-Generated Responses
Acoustic grounds are formed landscapes intentionally designed to mitigate noise pollution, what are some limitations to using it as a noise mitigation strategy?	'Increasing in height in the direction of the noise dispersion provides less effective mitigation unless they are high enough to form a barrier adjacent to the noise source'	The position of the noise source (emitter) may impact or limit the location of the acoustic mound, the complexity of the parametric mound form, the level of noise absorption of nearby surfaces, and the way different ground covers may contribute to reduction
What blue-green infrastructure design approaches might a landscape architect use to achieve connectivity in an urban landscape to maximize ecological function?	'Renewable energy combined with blue and green infrastructure'	Maximizing consecutive bands of landscape patches

Query	Retrieved Responses	Human-Generated Responses
What strategies can be employed by landscape architects to achieve optimal leaf coverage and optimize ecosystem services of urban trees in cities?	'Semi-automated tree planning workflow'	Setting forbidden and allowable space for the canopy with target leaf voxel to achieve optimal spatial occupation of the tree canopy in a dense urban environment
What primary factors must a landscape architect consider when designing small-scale public spaces to optimize landscape performance?	'Correlative factors'	Traffic Capacity, Visual Openness, Greening Quality, and Service Capacity
Computational tree modelling is the process of creating dynamic, scientifically informed, and visually realistic digital representations of trees. What are the potential applications of this technology in urban planning and landscape architecture?	'User-specific research'	The use of digital tree models can be used for climate-based performance calculations
What landscape qualities influence economic vitality of retail chains and how might we measure this at both the micro and macro scale?	'scale and signage'	Access to transportation such as public transportation facilities, the presence of visitor-oriented urban functions, and the number of Points of Interest (POIs). To measure this, the authors used intensity of geo-tagged catering businesses from Points-of-Interest to measure economic aspects of vitality.
How can virtual reality (VR) technology be further developed and utilized to create more accessible and user-friendly urban and natural environments for individuals with sensory impairments or different sensory experiences?	'Representation and Design Studies'	Implementing during the design phase, a sonic data collection process, integrating data from multiple geographic points at a site and using the process outlined in the paper to convert the data into a headphone signal that can be integrated into a VR experience.
What role will Artificial Intelligence (AI) and Machine Learning (ML) play in shaping future landscape design and planning?	'The next frontier'	Accelerate the transformation of our landscape and urban systems towards a more equitable, resilient, and adaptive environment
How can landscape architects incorporate parametric design and other computational tools into their practice?	'Digital design practice'	Using the Grasshopper plugin for the Rhino 3D software program.
How can we use algorithms to design landscapes that maximize ecological performance?	'Performance-based design'	Using parametric computer applications and integrate data from multiple sources that can include interpretation of landscape data

The below table (Tab. 2) outlines the quantitative performance of the RAG as average values for each of the n-grams analysed (Rouge-1, Rouge-2, and Rouge-L). The *recall* ("r"), *precision* ("p"), and *F-1 score* ("f") variables presented in the table are related to the ROUGE metrics.

Table 2: Quantitative Performance Measures using ROUGE Scoring

Metrics	Rouge-1	Rouge-2	Rouge-L
r	0.221	0.052	0.194
p	0.310	0.082	0.269
f	0.252	0.059	0.215

The ROUGE score evaluates the similarity and quality of text in model-generated responses by comparing them to the training data. A surprising result of the ROUGE scores indicates that there is a general overlap of key concepts, however, the overlapping of two-word combinations (Rouge-2) presented relatively low values, ranging only 0-10% in all the metrics (Tab. 1). On the other hand, the Rouge-1 and Rouge-L showed better values, especially the precision values of 0.310 and 0.269, respectively. All Rouge variables presented higher values for precision, followed by F-1, a metric used to assess precision and recall.

In Figure (A) below, a comprehensive bar plot is presented, illustrating the performance metrics of Rouge-1, Rouge-2, and Rouge-L demonstrating the RAG's performance. The plot highlights the averages of *r*, *p*, and *f* values, showcasing notably higher values for Rouge-1 and Rouge-L, with the maximum value observed for the precision (p) parameter within the Rouge-1 metric. This emphasizes the efficacy of the RAG in capturing linguistic nuances and overall content coherence. Figures (B), (C), and (D) further delve into the distribution characteristics of *r*, *p*, and *f* values through box plots. In Figure (B), it becomes evident that only the f parameter exhibits a normal distribution, while both *r* and *p* display distinct distributions in the second and third quartiles, resulting in an asymmetrical structure. Conversely, Figures (C) and (D) demonstrate normal distribution exclusively for the precision (*p*) parameter. These visual representations provide valuable insights into the nuanced performance of the RAG across various linguistic and content evaluation metrics within the landscape architecture domain.

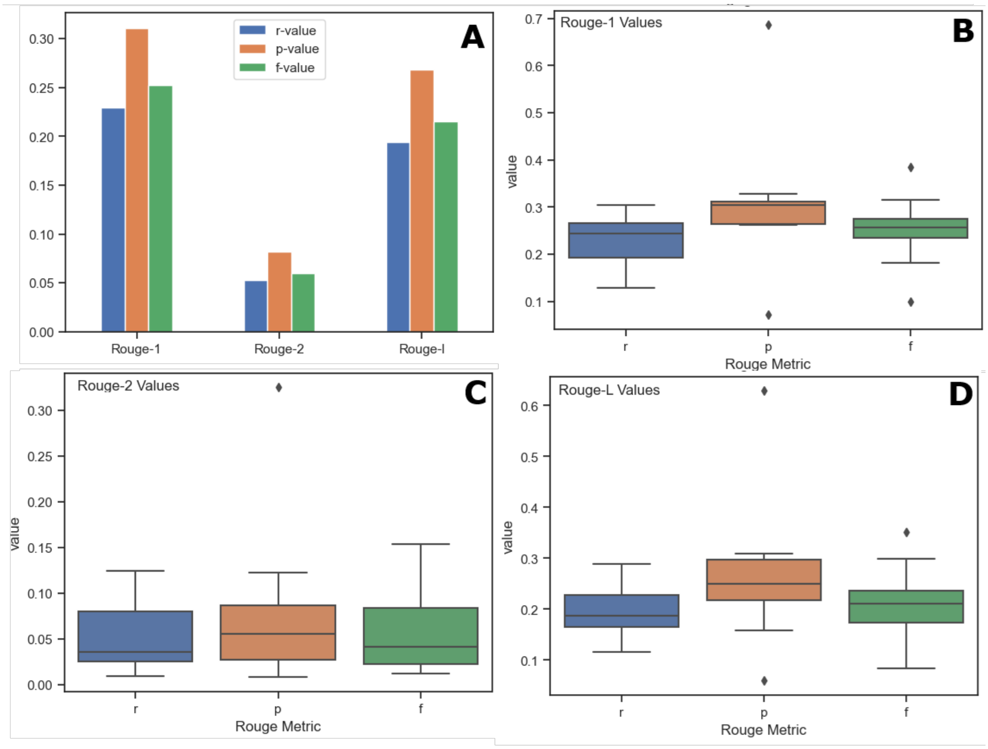

Fig. 1: Plots illustrating the performance of the ROUGE metrics. ROUGE-1 and ROUGE-2 assess the similarity by evaluating exact matches of single and double-word sequences between reference summaries and the systems-generated responses, whereas ROUGE-L measures the extent of longer shared text sequences within these responses compared to the reference summaries.

5 Discussion

The RAG's performance can be attributed to its ability to retrieve specific contextual information from the August 2023 issue allowing it to generate responses deeply rooted in the specific topical area of the Journal's August issue (Fig. 3). While large open-domain LLMs like ChatGPT and Google Gemini have a broader knowledge base, they lack the specificity and expertise that the RAG derived from its focused training data. Where the RAG met baseline expectations, we did identify that the model was not as effective in compiling multiple concepts or portions of text. To solve this, we would use improved retrievers while introducing more robust training data. This approach demonstrates a proof-of-concept in the potential of how a CDQA paired with a RAG could be deployed in a specific field, like landscape architecture or at a design firm.

The results demonstrate that certain landscape architectural tasks executed by specialized AI models can be performed but highlight the value of domain-specific, representative, and robust training data. Where the reader-retriever components of the RAG performed as designed by providing concise and relevant responses based on the training data, it did fall short on innovation and creativity. Training the model on a single issue of the Journal of Digital Landscape Architecture constrained the model's ability to imagine and curate more muscular and skillful responses. Future research should explore training the model on a more robust training dataset to increase the model's ability to generalize to other datasets and tasks.

Consideration of future AI/Human workflows is important as the technology advances and becomes integrated into production pathways for design work. Future integration would need to balance the human contributions with that of AI-driven approaches or solutions. The theoretical relationship building between AI tools and humans can advance our understanding of the universe-of-possibilities for domain-specific AI in design.

Practical application of neural networks in a design-discipline, like landscape architecture may become common-place. Employing a RAG approach on a design-firm's discrete knowledge base may enhance many aspects of a firm's practice, but it also may result in the stagnation of design ideas if training data are not continually updated or refreshed. We propose that further training methodologies be explored to reduce this possibility. As has been demonstrated in computing, engineering, and finance, the quality of the data sources used to train will significantly shape the ability for truly inspired and imaginative responses in LLMs.

The continued improvement and refinement of RAGs could have a revolutionary impact on the practice of Landscape Architecture and other design disciplines, especially as firms invest in the development of their own AI tools. This work underscores the role of a CDQA paired with a RAG in landscape architecture practice as partner, practioner, and associate.

6 Conclusion and Outlook

This study demonstrates that a Closed Domain Question Answering (CDQA) system integrated with a Retrieval-Augmented Generation (RAG) model and trained on specialized datasets like The Journal of Digital Landscape Architecture has the potential to enhance a landscape architect's workflow. Successful task execution included generating outputs to design concepts, retrieving responses to theoretical problems, and providing context specific summarizations. Possible application of a CDQA system integrated with a RAG model could include analytical tasks including identification of patterns, relationships, and design features or elements native to the design firm's approach, generating content for project proposals or summarizing site data. The integration of CDQA with a RAG model underscores their potential as invaluable assets in landscape architecture, particularly in refining data-driven design methodologies or applied to large datasets of past projects.

However, the practical application of this approach in landscape architecture is not without challenges. The efficacy of CDQA integrated with a RAG model hinges on the availability of extensive, well-curated datasets. The development of domain-specific metrics is crucial for accurately assessing the quality and relevance of the outputs generated by these models. Beyond data availability and quality, issues remain with bias, the limitations of creativity, and the attribution of ideas and content. A current limitation is the inability to incorporate visual data into this specific type of model, however advancements may make this possible in the near future. Addressing these challenges is essential for maximizing the models' utility and success in professional practice.

Future research should not only delve into these limitations but also explore strategies for diversifying the training datasets. Incorporating knowledge from related fields such as architecture, ecology, and urban design could foster a more holistic and interdisciplinary approach. This cross-pollination is vital for the evolution of AI applications in design, ensuring that these tools remain adaptable and relevant across various domains.

References

CANTRELL, B. & ZHANG, Z. (2018), A Third Intelligence. Landscape Architecture Frontiers, 6 (2), 42-51. https://doi.org/10.15302/J-LAF-20180205.

DAS, R. DHULIAWALA, S., ZAHEER, M. & MCCALLUM, A. (2019), Multi-step Retriever-Reader Interaction for Scalable Open-domain Question Answering. https://doi.org/10.48550/ARXIV.1905.05733.

DEVLIN, J., MING-WEI, C., LEE, K. & TOUTANOVA, K. (2019), BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.Org. https://www.proquest.com/docview/2118630252?parentSessionId=0tmGCY%2BQzvaL8%2F%2Bfn3I9GB9ddSDreF9VM5NexAdR9hQ%3D&pq-origsite=primo&sourcetype=Working%20Papers.

FERNBERG, P. & CHAMBERLAIN, B. (2023), Artificial Intelligence in Landscape Architecture: A Literature Review. Landscape Journal, 42 (1), 13-35. https://doi.org/10.3368/lj.42.1.13.

FU, J. XU, J. & JIA, K. (2009), Domain Ontology Based Automatic Question Answering. 2009 International Conference on Computer Engineering and Technology, Singapore, 346-349. https://doi.org/10.1109/ICCET.2009.132.

LEWIS, P., PEREZ, E., PIKTUS, A., PETRONI, F., KARPUKHIN, V., GOYAL, N., KÜTTLER, H., LEWIS, M., YIH, W., ROCKTÄSCHEL, T., RIEDEL, S. & KIELA, D. (2021), Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401. https://doi.org/10.48550/arXiv.2005.11401.

LIN, C.-Y. (2004), ROUGE: A Package for Automatic Evaluation of Summaries. Text Summarization Branches Out, 74-81. https://aclanthology.org/W04-1013.

QIU, Y., CONG, Z., OPINIANO, K. N., QIAO, X. & CHEN Z. (2023), Landscape Architecture Professional Knowledge Abstraction: Accessing, Applying and Disseminating. Land (Basel), 12 (11). https://doi.org/10.3390/land12112061.

SIRIWARDHANA, S., WEERASEKERA, R., ELLIOTT, W., KALUARACHCHI, T., RANA, R. & NANAYAKKARA, S. (2023), Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering. 1-17. https://doi.org/10.1162/tacl_a_00530.

TAWARE, U. & SHAIKH, N. (2018), Heterogeneous Database System for Faster Data Querying Using Elasticsearch. 1-4. https://doi.org/10.1109/ICCUBEA.2018.8697437.

VILA, K. & FERRÁNDEZ, A. (2011a), Model-driven restricted-domain adaptation of question answering systems for business intelligence. In: Proceedings of the 2nd International Workshop on Business Intelligence and the WEB, 36-43. https://doi.org/10.1145/1966883.1966893.

ZHANG, Z. & BOWES, B. (2023), The Future of Artificial Intelligence (AI) and Machine Learning (ML) in Landscape Design: A Case Study in Coastal Virginia, USA. arXiv.org. https://doi.org/10.48550/arxiv.2305.02327.

ZHU, F., LEI, W., WANG, C., ZHENG, J., PORIA, S. & CHUA, T.-S. (2021), Retrieving and Reading: A Comprehensive Survey on Open-domain Question Answering. https://doi.org/10.48550/ARXIV.2101.00774.